Hearing-aid fitting in interaction: on optimal combinations of multiple acoustic

Franck, B.A.M.

Publication date
2004
Document Version
Final published version

Hearing-aid fitting in interaction

On optimal combinations of multiple acoustic signal-processing strategies

Bastien Alphons Marie Franck

Cover & illustrations: Frederieke Baars – FPW Rotterdam

Printing: Febodruk B.V. – Enschede

The research described in this thesis was supported by the Heinsius Houbolt Fund, that funded part of the equipment and the testing facilities.

Printing of this thesis was supported by the Atze Spoor fonds, Beltone Netherlands B.V., Oticon Nederland B.V., GN Resounds B.V., Schoonenberg Hoorcomfort, Siemens Nederland B.V., Veenhuis Medical Audio B.V.

Hearing-aid fitting in interaction

On optimal combinations of multiple acoustic signal-processing strategies

Academisch Proefschrift

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van Rector Magnificus

prof. mr. P.F. van der Heijden

ten overstaan van een door het college voor promoties ingestelde

commissie, in het openbaar te verdedigen in de Aula der Universiteit

op dinsdag 29 juni 2004, te 10.00 uur

door Bastien Alphons Marie Franck

geboren te Maastricht

Promotiecommissie:

Promotor: prof. dr. ir. W.A. Dreschler

Overige leden:

dr. ir. J.M. Festen
prof. dr. W.J. Fokkens
prof. dr. J. Kiessling
dr. ir. J.A.P.M. de Laat
prof. dr. ir. J.A.E. Spaan

Faculteit Geneeskunde

‘Hoe langer het duurde voor ik een blikje vond, hoe meer ik me verheugde op de pure eenvoud van wat ik ging doen. Een schop tegen een leeg blikje. Het beste moment lag, zo vond ik, zo rond een uur of vijf, en dan moest het in een stille, stoffige straat zijn. Het rinkelen van het vooruitstuiterende blikje moest goed weerkaatst worden door de huizen langs de straat [...] Blikje midscheeps op de wreef. Plokkend, hol geluid, half metertje de lucht in, de grond raken, rollen. Daarna – stilte [...] De uithaal, de rake trap, het geluid, de stilte, de enorme en tegelijk minuscule voldoening die volgt – even volledig synchroon te lopen met de wereld [...]
Ik wist mij onbegrijpelijk gelukkig.’

– Martin Bril
Hollandse luchten, Amsterdamse miniaturen (2000)

PREFACE

The motivation of my thesis is based on combination of three audiological themes. First, cochlear hearing impairment is generally more complex than middle or outer ear hearing loss. Second, the world around us consists of more complex signals than signals used in audiometric tests. And third, a digital hearing aid is an acoustic signal processor rather than an amplifier. In order to fuse these elements properly an optimisation method was developed.

Audiological theses often contain a series of articles, accompanied by an *Introduction* and a *Final discussion*. In my view it would be more elegant and more clear to write this thesis in a more coherent way. The advantage of such an approach is that problems that arose answering my research questions can be fitted better in the whole thesis.

The rationale of the optimisation method is described in the *Introduction* (chapter 1). This chapter outlines the research questions, the consequences of algorithm interaction, the importance of optimisation in interaction with the subject, and the scientific approach to tackle the questions.

In chapter 2 the test methods that were used are given. The details of the multi-directional pattern search (optimisation method) and the round-robin tournament are presented in appendices 1 to 5. The relation that was found between the two methods can be found in appendix 6.

The signal-processing algorithms are gathered in chapter 3. These strategies either reduce the undesired effects of background noises and reverberant room acoustics, or enhance the desired speech signals.

Chapter 4 contains the results of six studies that relate to the effects of combination of algorithms and the effects of simultaneous optimisation of multiple algorithms. The core of the experimental work that was relevant for this thesis is incorporated in this chapter.

In chapter 5 the effects of simulation of cochlear hearing loss on the optimal combination of three algorithms was studied. The results of two simulation models were compared to the results of normal hearing and hearing-impaired subjects.

Chapters 6 and 7 include overall discussions with respect to the methodological aspects and the audiological aspects of the studies described in chapters 4 and 5. The similarities and differences found between the studies are discussed in depth. This leads to a general discussion in which reasons are proposed to explain the results that are found, and alternative approaches are suggested.

In the *Final Discussion* (chapter 8) the problems that relate to subjective evaluation are discussed. Finally recommendations are given to cope with the dynamical nature of subjective judgement in the process of optimising hearing-aid algorithms.

In order to assist the reader this thesis includes a list of illustrations, a list of tables, a list of abbreviations, and a glossary. The glossary is given at the end and the lists are given at the beginning of this thesis.

ABSTRACT

Hearing-aid fitting in interaction

On optimal combinations of multiple acoustic signal-processing strategies

by

Bastien Alphons Marie Franck

The focus of this thesis is to facilitate the fitting of complex digital hearing aids. A fitting strategy is studied that tries to find the optimal combination of hearing-aid techniques in interaction with the hearing-impaired individual. In several studies, we examined the applicability of an optimisation procedure for different signal-processing strategies. These signal-processing techniques combine noise reduction and speech enhancing algorithms that are expected to improve either the signal-to-noise ratio or certain speech features in the temporal, spectral, or amplitude domain. Hence, we investigated the audiological consequences of combinations of signal-processing strategies and their interactions.

Since many objective methods cannot distinguish differences in performance, we use a paired-comparison format in order to determine subjective preferences for combinations of algorithms, based on either listening comfort or subjective speech intelligibility. In order to study the interaction between the different algorithms we performed a round-robin tournament for combinations that have been varied systematically. Moreover, the tournament gives information about the response distributions, i.e. contours of equally perceived comfort or intelligibility. We also performed experiments to measure the difference limens of the signal-processing techniques. The outcomes can predict useful step-sizes for the optimisation strategy. Finally, we studied the reliability and consistency of the optimisation strategy. The multi-directional pattern search that was developed includes an adaptive step size: the step size for each dimension was determined by the perceptual differences between settings for each listener. The abovementioned methods might be essential for future fitting of hearing aids, because interaction between several non-linear algorithms makes fitting based on prescriptive rules alone inadequate.

The results showed that the addition of a signal-processing strategy does not necessarily yield higher speech intelligibility. This seemed to be due to non-linear interactions between algorithms. That non-linear interactions exist was proved in a study with three signal-processing strategies. Though the interactions were found for the listening comfort criterion only.

Generally, application of signal-processing strategies was found not very successful, neither to achieve higher speech intelligibility nor to achieve higher listening comfort. This appeared to be due to trade-off effects between individual phonemes, between signal-processing strategies, and between perceptual strategies. Furthermore, additional detrimental distortions and acclimatisation effects appeared to be possible explanations.

The reliability of the multi-directional adaptive pattern search appeared to be reasonably high. Especially the presence of flat or multi-modal response distributions affected the reliability negatively. The reliability was affected also by other procedural, cognitive, auditory and acoustical factors. A change in the step-size configuration could improve the reliability dramatically.

The optimisation procedure can be valuable for clinical purposes, but some methodological problems remain to be solved first. Two directions are possible. First, the implementation of a more objective evaluation paradigm can be studied. Second, the subjective pair-wise comparison paradigm can be enhanced to circumvent problems originating from the dynamic nature of subjective evaluation.

SAMENVATTING

Hoortoestelaanpassing in interactie

Optimale combinaties van meerdere akoestische signaalverwerkingsstrategieën

door

Bastien Alphons Marie Franck

Het doel van dit proefschrift is om het aanpassen van complexe digitale hoortoestellen te vergemakkelijken. Daartoe is een aanpasprocedure bestudeerd die de optimale combinatie van hoortoesteltechnieken probeert te vinden in interactie met de slechthorende. In verschillende studies hebben we de toepasbaarheid van de optimalisatieprocedure voor verschillende typen signaalbewerking onderzocht. Deze signaalbewerkingstechnieken behelzen ruisonderdrukking en de opscherping van spraak die worden geacht om respectievelijk de signaal-ruisverhouding te verbeteren en specifieke kenmerken van spraak in het temporele, spectrale of amplitude domein. We hebben dus ook de audiologische consequenties van combinaties en interacties van signaalbewerkingsstrategieën onderzocht.

Aangezien veel objectieve methoden niet geschikt zijn om onderscheid te maken tussen verschillen in prestatie, maken we gebruik van paarsgewijze vergelijkingen om subjectieve voorkeuren te bepalen van verschillende combinaties van algoritmen. De keuzes zijn daarbij gebaseerd op luistercomfort of subjectieve spraakverstaanbaarheid. We hebben een round-robintoernooi uitgevoerd voor systematisch gevarieerde combinaties van algoritmen om interacties tussen deze algoritmen te bestuderen. Daarnaast geeft het toernooi informatie over de respons patronen die contouren van gelijk comfort of spraak verstaan vertegenwoordigen. Ook hebben we experimenten uitgevoerd om 'net waarneembare verschillen' van signaalbewerkingstechnieken te meten. De uitkomsten kunnen geschikte stapgroottes voor de optimalisatiestrategie voorspellen Tot slot hebben we de betrouwbaarheid en consistentie van de optimalisatiestrategie onderzocht. De ontwikkelde *multi-directional pattern search* gebruikt een adaptieve stapgrootte. Op basis van de perceptuele verschillen tussen instellingen werd voor iedere luisteraar de stapgrootte per dimensie bepaald. De genoemde methoden kunnen essentieel zijn voor het aanpassen van hoortoestellen in de toekomst, omdat interactie tussen verschillende niet-lineaire algoritmen het aanpassen op basis van alleen prescriptieve regels minder geschikt maakt.

De resultaten laten zien dat het toevoegen van een signaalbewerkingsstrategie niet per sé hoeft te leiden tot betere spraakverstaanbaarheid. Het lijkt dat dit komt door niet-

lineaire interacties tussen algoritmen. Dat niet-lineaire interacties voorkomen is aangetoond in een studie met drie typen signaalbewerking, hoewel de interacties alleen werden gevonden voor het criterium luistercomfort.

In het algemeen was het gebruik van verschillende typen signaalbewerking niet erg succesvol, noch om betere spraakverstaanbaarheid te bewerkstelligen noch voor een beter luistercomfort. Dit bleek met name te kunnen worden toegeschreven aan nivelleringseffecten tussen afzonderlijke fonemen, tussen verschillende typen van signaalbewerking en tussen perceptuele strategieën. De verklaring kan bovendien worden gezocht in nadelige vervormingen en acclimatisatie-effecten.

De betrouwbaarheid van de *multi-direction adaptive pattern search* is redelijk hoog gebleken. Met name de aanwezigheid van vlakke responspatronen en patronen met meerdere pieken hadden een negatieve invloed op de betrouwbaarheid. De betrouwbaarheid werd ook beïnvloed door andere procedurele, cognitieve, auditieve en akoestische factoren. De betrouwbaarheid kon aanzienlijk worden verbeterd door de stapgrootteregels aan te passen.

De optimalisatiestrategie kan waardevol zijn voor klinische toepassing, maar hiervoor moet een aantal methodologische problemen worden opgelost. Er zijn twee alternatieven. Enerzijds kan het gebruik van meer objectieve methoden worden onderzocht, anderzijds kan het paradigma van subjectieve paarsgewijze vergelijkingen worden verbeterd om problemen te omzeilen die voortkomen uit het dynamische karakter van subjectieve evaluatie.

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS

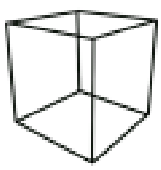
NR	Monaural noise reduction strategy
BNR	Binaural noise reduction strategy
DR	Dynamic range
DRV	Binaural de-reverberation strategy
CVC	Consonant-vowel-consonant word
DL	Difference limen; just noticeable difference
HI	Hearing-impaired subject group
IT	Information transfer; reliability measure
MCL	Most comfortable loudness level
NH	Normal hearing subject group
PhC	Phonemic compression strategy
SC	Syllabic compression strategy
SE	Spectral enhancement strategy designed
SIM1	Simulated hearing-impaired subject group for which spectral smearing was applied
SIM2	Simulated hearing-impaired subject group for which spectral smearing was applied and masking noise was added
SL	Spectral lift strategy
SNR	Signal-to-noise ratio
SRT	Speech reception threshold
STE	Spectro-temporal enhancement
u	Coefficient of agreement
ζ	Coefficient of consistency

Chapter 1

PRELUDIUM

'If we know the rules, we consider that we "understand" the world [...] The world is something like a great chess game being played by the gods, and we are observers of the game. We do not know what the rules of the game are; all we are allowed to do is to *watch* the playing. [...] *The rules of the game* are what we mean by *fundamental physics*. [...] However, we might not be able to understand why a particular move is made in the game, merely because it is too complicated and our minds are limited [...] Aside from not knowing all off the rules, what we really can explain in terms of those rules is very limited, because almost all situations are so enormously complicated that we cannot follow the plays of the game using the rules, much less tell what is going to happen next.'

– Richard Feynman
Six easy pieces – The fundamentals of physics explained (1998)



1.1 General introduction

What do hearing-impaired and deaf people lack? What do hearing-impaired people hear? What makes perception different from sensation? Can our senses give a solid representation of the external world? What tools does psychophysics provide to acquire knowledge of the perception processes? Does fundamental research of carefully assembled, artificial signals provide more valuable information about hearing (impairment) than real-life sounds? Or are complex signals appropriate to investigate, explain, describe or understand the normal hearing and the pathological? Should medical research be concerned with restoration of hearing defects only? What is the role of the audiological scientist at present? And is the hearing-impaired individual inevitably a victim of a scientific medical authority?[1]

Some sinister ideas have penetrated audiology. One of the most persistent ideas is our conviction that we can decide what is best for our unhappy hearing-impaired or deaf fellows.[2] More importantly, 'outsiders' think they know about the handicaps of hearing-impaired and deaf people. This conception culminates in preaching the never-ending benefits of hearing aids or compensations for cochlear deficits. 'Buts' are classified irresponsible or, even worse, immoral.[3] In more common practice this means that specialists try to convince people with sloping cochlear hearing losses that high-frequency gain provides optimal speech intelligibility, despite the uncomfortable shrieking sound. In other words: better speech intelligibility is the one and only way to go. But for whom? It seems that innovations with respect to hearing aids and hearing-aid strategies dominate the present audiological research. However, continuing in the directions that we followed thus far, it would be sensible to ponder what we have achieved so far and, as a consequence, ask ourselves what are the consequences for the hearing-impaired person.

It is remarkable that audiometric data that are translated into specific gain and frequency compensation still dictate hearing aid fitting practice. Prescription rules reign, because numbers yield true confidence. However, does this so-called 'objective' perspective really meet the demands of hearing-impaired individuals? Are impairments simply measurable and can they be characterized in one or more simple numbers?

Another difficulty of the present hearing-fitting practice is that tuning of individual hearing-aid techniques takes place in successive stages. This method presupposes no interaction between hearing-aid algorithms. This is a simplification of the actual situation, as will be shown in this thesis. Algorithms interact when the optimal setting of an algorithm depends on the setting of another algorithm. As an example a hypothetical

---

[1] This thesis is written as a *critique*, a study in line with what was usual for studies in the time of the Greeks. The question is always why something is as it is and why not different? The *critique* concerns primarily the optimisation method, signal-processing strategies and the role of the hearing-impaired subject in the process of fine-tuning of a hearing aid.

[2] Prosper Ménière, wrote in 1853: "The deaf believe that they are our equals in all respects. We would be generous and not destroy that illusion. But whatever they believe, deafness is an infirmity and we should repair it whether the person who has it is disturbed by it or not."

[3] Of course, the consequences of deafness or hearing-impairment are tremendous. Nevertheless, the consequences of hearing loss can be advantageous as well. Sound can be very annoying during sleep or when attention is needed for a certain intellectual task, i.e. reading, writing, etc. And as a result of acquired lip-reading, severely hearing-impaired people can 'listen' to conversations at large distances, although only a part of the information can be obtained in this way.

response curve is presented in figure 1.1. The optimal response on a factor x1 depends on the setting of factor x2. To meet interaction between signal-processing strategies, an adaptive multi-directional pattern search was developed. This optimisation procedure searches for the best combination of hearing-aid techniques, based on individual judgments. Details about the pattern search can be found in section 2.1.1.

In this thesis we advocate a symbiosis of prescription rules and feedback from the users' perceptual judgements to attain optimal fitting of state-of-the-art complex hearing aids. Based on the findings of several studies, we try to find out more about the perceptual world of hearing-impaired subjects. Therefore, we question the very concepts that define the subject. We consider the contents of perception research and related disciplines and try to disentangle concepts like perception and sensation. We also try to analyse the experimental consequences of the use of complex signals. As a conclusion this thesis will underline the importance of a shift of the dominant role of the expert to the hearing-impaired individual. The central theme is the problem that is experienced by these persons. It may turn out to be quite different what the perceptual world is for the hearing-impaired person himself or herself and what we – outsiders – think it sounds like.

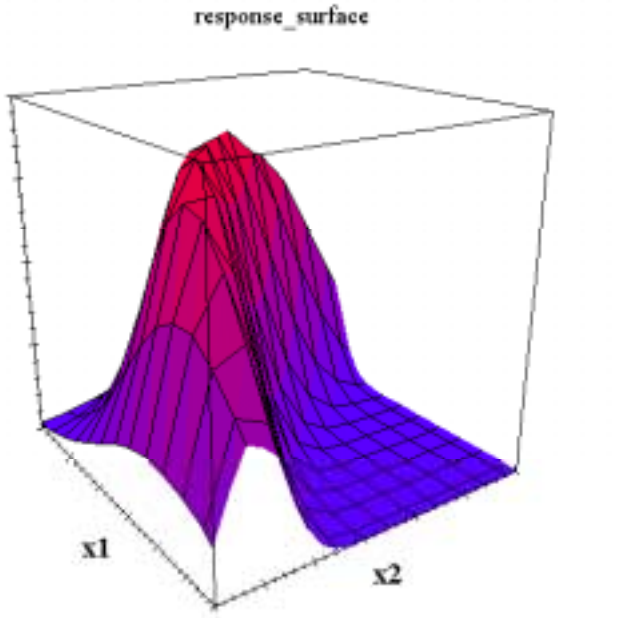


Figure 1.1 Hypothetical response curve as a function of factors x1 and x2. Interaction is present between factors x1 and x2.

1.2 Hearing impairment

To understand the importance of subjective interaction in fitting hearing aids, the perceptual consequences of hearing impairment will be discussed in this section.

In a normally functioning hearing system three principle components can be discriminated. First the eardrum transforms air vibrations into mechanical oscillations of acoustic signals reaching the ear. Secondly these oscillations are transmitted by three middle ear bones into the inner ear (cochlea) and transformed into pressure variations in

the cochlear fluids. According to hydrodynamic properties the basilar membrane is forced to move. As a result a travelling wave is generated. In this way the spectral content of the signal is transposed to specific places, which result in a tonotopic organization of the spectral information of the signal. After this first kind of spectral analysis receptor cells convert the basilar membrane movements into electrical nerve impulses. These firings are called spikes. Finally, the auditory nervous system transmits the neural representations of the signals to the brain. In this third stage many processes are active. Essentially the information is processed and in the end displayed as an image of auditory qualities in the cortex.

Hearing impairment is primarily defined as a defect. Somewhere in the hearing system the functionality is impaired. For the periphery of the system, i.e. before the signal is transformed into neural activity, a distinction is made between functional deficits of middle ear or inner ear origin. For this demarcation the terms 'conductive' and 'cochlear' hearing losses have been introduced.[4] Defining hearing impairment as a defect suggests that sooner or later the problem can be solved. This approach presupposes that in the end scientific knowledge can solve all problems.

The behaviour of individuals is not fully predictable as well.[5] A different approach is to accept that the world of the hearing-impaired subject is different and the impairment should be dealt with in interaction, i.e. between the physician or audiologist and the hearing-impaired individual.[6] Additionally, the effects of hearing aids are grossly overrated. The expectations of hearing-impaired subjects are high both as a result of commercial advertisement and as a result of underrating the complexity of hearing problems on the one hand and the time-consuming 'recovery' process on the other. The effects of a hearing aid also depend on individual hearing capacities. At present it appears difficult to predict the impaired functionality of the hearing system based on

---

[4] For clinical purposes the term 'sensorineural' hearing loss is often used. In principle the sensory and the neural part are meant, which would be represented by cochlear and retro-cochlear deficits. However, the terms suggest that sensation takes place in the cochlea, which is quite misleading. Therefore we prefer to avoid this terminology and instead try to be clear about the place where the functionality of the hearing system is impaired, by using the terms middle ear hearing loss, cochlear hearing loss and retro-cochlear hearing loss.

[5] For example, based on audiometric results, high-frequency amplification can be implemented in a hearing aid. There is however good reason to presume that the hearing-impaired individual will not be satisfied. One of the reasons is the new soundscape of the hearing-impaired subject, which at once sounds harsh and shrill. Due to missing links the auditory world cannot be associated to the world that used to be with less or different auditory information. Especially for elderly people, considerable effort is needed to grasp the meaning of the auditory world. The new source of information can even give rise to a lot of confusion, which might be distressing. Hence, hearing-impaired subjects need time to acclimatise to their new auditory world.

[6] Peculiarly, it appears very difficult to imagine the land of the hearing impaired or of the colour blind, or another world that is seemingly close to our own world of experiences. Similarly, it appears very difficult to imagine subatomic particles, the fourth dimension, the world of the deaf, or even a human being that has an additional sense organ. In practice, we consider our senses as quite common and everything we experience we call reality. We will never be (fully) able to imagine what the world sounds like for a hearing-impaired, unless we become hearing-impaired ourselves. Personal observation or experiencing is better than an exact description, especially when the subject of investigation is the analysis of perception. Therefore, an audiologist should best be hearing-impaired, or should at least have a considerable amount of empathy, compassion and understanding. In a novel Bernlef describes the process of becoming demented. In the same way it would be interesting to experience, for example in a novel, what it is to become more and more hearing impaired.

standard audiometric tests alone. Signal-processing strategies can be beneficial, but only in certain circumstances. Moreover, the use of a hearing aid in speech-containing background noise is often not appreciated. Hence, from this alternative perspective, emphasis is especially needed on the usage of hearing aids. Therefore, a hearing aid with only a limited number of facilities is probably most effective. Every way to regain functionality should be taken into consideration. Hearing proved to be important for spatial orientation, communication and has existential value (Espmark and Scherman, 2003). By losing part of the functionality of the hearing system, social contact and connection with the world around can be seriously impeded.

1.3 Cochlear hearing loss

Hearing impairments are very diverse. In order to constrain the broad range of hearing impairments our main interest applies to subjects that suffer from cochlear hearing loss.

Physiological processes in the cochlea are vulnerable. The functioning can be deteriorated as a result of many factors, for example by ageing, noise exposure, anoxia (deficiency in oxygen), ototoxic agents, disease, drugs, and mechanical disturbance. Problems that occur as a result of poor cochlear functioning have appeared to be extremely difficult to solve. The key reason is the incomplete knowledge of the operation and processing characteristics of the cochlea. What we do know is that due to the loss of outer and/or inner hair cells some mechanisms are impaired. More than 90% of the afferent nerves originate from inner hair cells (Forge, 2002). About 80% of the efferent nerves terminate at outer hair cells. Based on this difference, mechano-receptors and mechano-motors can be distinguished (e.g. Van Dijk, 2001). The mechano-receptors transpose the acoustic signal as a coded message to the brain via the afferent nerves, while the mechano-motors have a mechanical influence on the motion of the basilar membrane. The mechano-motors are guided primarily by the efferent innervation of nerves. The vibration pattern of the basilar membrane is affected by the motility of the outer hair cells, i.e. fast reversible changes in length and shape. Therefore the outer hair cells can be considered as cochlear amplifiers. In all, for the conversion of mechanical energy into electrical energy, the inner hair cells follow the motion of the basilar membrane that is partly shaped by the motility of the outer hair cells. As a result the signal that reaches the inner hair cell is fine-tuned and amplified.

Audiologically the presence of outer hair cells has many advantages. First the dynamic range of sound levels that can be handled by the auditory system is enlarged (more details can be found in the next section in combination with figure 1.2). Secondly, the spectral and temporal resolution is high. Good spectral resolution is necessary to resolve the spectral fine structure of signals (Moore, 1989). Good temporal resolution is assumed to be necessary to resolve temporal variations adequately. Signals that are preceded or followed by other sound samples are less likely to be masked when the temporal resolution is high (Zwicker and Schorn, 1982). As a consequence even in adverse acoustic situations listeners can understand speech. Subjects with a well-functioning hearing system, including a healthy cochlea, usually have good speech

intelligibility despite much reverberation or a background noise with a sound level that can be considerably higher than the sound level of the speech.

Dynamic, spectral, and temporal capacities can deteriorate as a result of cochlear damage. Damage to the inner hair cells results in a transduction loss of information to the brain. Dysfunction of the outer hair cells diminishes the amplification characteristics and affects the hearing sensitivity and the frequency resolution. Total loss of outer hair cells without accompanying loss of inner hair cells will cause hearing losses of around 60 dB, since the inner hair cells still can transpose information to the brain, though not amplified (Kemp, 2002). Only a loss of inner hair cells can yield any degree of hearing loss.

For the subjects that participated in all studies discussed in this thesis, we assume that the major part of their cochlear hearing losses stems from a loss of outer hair cells. In effect, cochlear hearing loss can lead to a reduced dynamic range, loudness recruitment, temporal and spectral smearing, and excessive masking.

1.4 Quantification of residual capacities

In the audiological practice compensation of cochlear hearing loss with signal-processing strategies is mainly based on estimation of the residual capacities of the individual. Hence, the primary goal is to assess the hearing abilities of subjects with cochlear hearing loss as good as possible.

Despite the potential help of an ever-increasing amount of hearing-aid strategies to cope with background noise and enhance speech characteristics, at the very start it is necessary to estimate the magnitude of the hearing loss. Until now psychophysical methods were used to assess the effects of the loss, based on detection, discrimination, identification or recognition tests. When the residual hearing capacities have been estimated, the most suitable hearing-aid strategies can be selected and fitted to the individual's specific needs.

Audiometric methods that are used in clinical practice, such as the pure-tone audiogram and the speech audiogram appear not sufficient to encompass the effects of cochlear hearing loss. At best, an impression of the elevated threshold and impaired speech perception can be obtained. Consequently it will be not clear to what extent compression and temporal and spectral filter characteristics will be deteriorated when additional audiometric tests are not included.

Methods that can give additional information about the residual capacities and are used in our studies are the following. To measure the frequency-specific loudness growth, the Würzburger Hörfeld Skalierung can be used. The outcomes can diagnose loudness recruitment. Difference limen tests can be used to determine the just noticeable difference between two signals that differ in a certain respect. This can be either frequency, intensity or, as is used in one of our studies, a compound of physical variables of the signal. Frequency resolution can be determined by measuring the width of the auditory filters by means of the notched-noise method. General hearing capacities cannot be measured by detection or discrimination tests alone. Tests with an identification or a recognition task are also of great value. One such test encompasses

the measurement of the speech intelligibility. Two methods were used in our studies: the speech-reception threshold (SRT) test and the consonant-vowel-consonant (CVC) test. SRT tests measure the ability to recognise and reproduce short sentences, whereas CVC tests measure the perception of the individual constituents of words, the phonemes.

1.5 Auditory signal-processing algorithms

Hearing aids can help to compensate for hearing impairments due to cochlear hearing loss. It seems however that signal-processing strategies cannot compensate for every possible situation. Hence, hearing-impaired subjects will not always be as happy with their new hearing aids as developers and providers in the audiological field expect them to be.

There are at least four hearing problems that need closer inspection. The first problem concerns spectral components that are below threshold. Some parts of the frequency spectrum cannot be heard while other parts are still audible, due to a change in frequency balance. As a consequence, identification of signals can be deteriorated. One of the main goals of hearing aids is the restoration of the spectral balance. That is, using such amplification that the inaudible and softer parts are amplified more than the louder parts of the frequency contents of a signal. The second problem is the combination of a reduced dynamic range and loudness recruitment. Due to the loss or reduction of compression in the hearing system of a hearing-impaired subject, the dynamics of level variations in signals like speech and music become exaggerated. The softer parts become inaudible, whereas the louder parts become uncomfortably loud. To compensate for the changed mapping of signal level to loudness, hearing aids often use a compression algorithm. The hearing aid compensates for the loss or reduction of compression so that inaudible sounds can be heard again, while the loudness of the very loud sounds does not become too high. The third problem for which hearing-aid algorithms should compensate is the reduction in frequency resolution. Due to a broadening of the auditory filters that determine the frequency dispersion in the cochlea, the signals are supposed to become spectrally smeared. To compensate for the spectral smearing the opposite can be performed by a hearing aid. Specifically the spectral peaks and valleys can be enhanced. The fourth problem relates to compensation for the deteriorated temporal resolution, which can result in a poorer performance to hear a signal in the presence of other signals just before or just after the target signal. Some hearing aid strategies compensate for the presumed deteriorated temporal resolution by enhancement of the temporal contrast, i.e. lengthening and shortening of the duration of specific parts of the signal.

Many compensation strategies are based on psychophysical and physiological knowledge about loss of compression and about temporal and spectral smearing. However, by transforming the experimental knowledge of the (impaired) hearing system into hearing aid technology major problems have been encountered. Most signal-processing strategies are not robust for a large range of signal-to-noise ratios. For high signal-to-noise ratios, i.e. for low background noise levels or the absence of noise, many of the aforementioned compensation strategies do their work properly. For low signal-

to-noise ratios however, i.e. when the background noise level is high relative to the target signal, most strategies do not result in an increased performance. In most experimental studies, the speech intelligibility is equal or worse relative to no processing at all. The main reason is that the amplification and enhancement influence both the target signal and the background noise. Only for stationary noise some signal-processing strategies seem to have positive effects. Therefore, psychophysically and physiologically oriented technologies that have been developed to enhance the signal need to be accompanied by more common-sense oriented noise-reduction schemes. Again it appears that many of the recently developed and mathematically driven noise reduction strategies can improve the signal-to-noise ratio only for stationary background noise. For a multi-talker babble or "cocktail-party" environments, many digital noise reduction techniques perform rather poorly. In such cases hearing-impaired subjects experience hearing aids as unpleasant, as distracting environmental sounds become equally well amplified or enhanced as the target signal (Bronkhorst, 2000). Sound emitters become threatening, exhausting, and hectic, probably comparable to the new auditory world for people that at once can hear again, e.g. after acquiring a cochlear implant. Then it seems reasonable that hearing-impaired people like to turn off their hearing aid and prefer the quietness.

1.6 Combination and interaction of algorithms

In addition to limitations of assessing the residual capacities of persons with cochlear hearing loss, the effects of combining several signal-processing strategies is not known in detail as well. The optimisation method studied here can cope with both needs.

Several studies were performed in order to investigate the effects of combining different experimental signal-processing strategies. First the combination of spectral enhancement and two types of phonemic compression was studied (see section 4.1). Speech intelligibility tests were conducted for different combinations. Surprisingly, one of the outcomes was that for the combination of these algorithms the speech intelligibility was not higher, but often lower than for spectral enhancement alone. The order of processing of the individual algorithms could be a determining factor. Another influencing factor is non-linear interaction between the algorithms.

To study the effects of combinations of algorithms in general and the interaction between algorithms in particular, we developed an optimisation strategy. For that reason we applied the multi-directional pattern search strategy. In the following studies two directions were chosen. The optimisation method was evaluated and the optima were evaluated audiologically.

The methodological aspects of the optimisation procedure were studied in different configurations, in order to enhance – among other things – the reliability. For that reason the step size was varied and different evaluation strategies were used. With respect to the audiological aspects we evaluated the end-results of the optimisation strategies for five different algorithm configurations. Binaural noise reduction was combined with phonemic compression and spectral enhancement, or with de-reverberation and phonemic compression. Monaural noise reduction was combined

with phonemic compression and spectro-temporal enhancement. Two monaural noise reduction parameters were combined with spectral enhancement. Finally, monaural noise reduction was combined with spectral lift and spectral enhancement. For these combinations some algorithms that share the same name have diverging processing strategies. These different versions of the algorithm were used primarily because we found that the success of the optimisation method relies heavily on perceptual distinctness between algorithm settings.

1.7 Hearing-aid fitting

Assessment of residual capacities is limited because our scientific methods are limited. On the one hand scientific methods are sometimes constrained, due to the fact that these methods need to simplify aspects of auditory perception. On the other hand scientific methods cannot assess the experienced needs fully. Therefore an interactive approach is proposed to circumvent some of these problems, i.e. our limited knowledge of the perceptual world of the hearing-impaired subject.

The introduction of digital techniques in hearing aids resulted in a vast potential of possible parameters that can be varied to improve speech intelligibility. The large amount of parameters can contribute to a much more refined fine-tuning method. However, it is unclear how some of the digital processing schemes should be used and, maybe even more important, what the individual fine-tuning should look like. A fine-tuning method should focus not on the effects of digital techniques per se, but on its perceptual consequences.

In order to improve speech intelligibility in noisy situations, different signal-processing schemes can be applied. Different fitting strategies are available in order to optimise algorithm parameters. When fitting is based on more objective rules, i.e. prescriptive formulae, only parameters in isolation can be accounted for. Due to the increasing number of different signal-processing algorithms, needed to cover a wide range of hearing problems and acoustical situations in a flexible way, hearing aids have become more and more complex. However, there is little knowledge about the effects of combining different algorithms. As a consequence fitting of multiple algorithms in practice is more based on intuition than on a systematic optimisation process.

In previous studies multi-directional strategies have been proposed to optimise hearing-aid parameters (e.g. Levitt, 1978; Neuman *et al.*, 1987; Kuk and Lau, 1996). In these studies the (modified) Simplex procedure was promoted (for more details see section 2.1.1). This optimisation strategy is used and adapted in this thesis, but is called 'pattern search' because it is fundamentally different from a 'simplex method'. The optimisation strategy resembles a pattern search most for reasons described in appendix 1. The optimisation procedure is especially advantageous compared to tournaments with respect to time efficiency. Moreover, the procedure can take into account interactions between parameters by considering multiple responses at the same time. However, many methodological consequences have not been studied systematically. Therefore we investigated the applicability of a three-directional pattern search procedure in the perceptual domain.

In order to apply the pattern search in the auditory domain, a paired-comparison paradigm was used frequently. The use of paired-comparisons has several reasons. It often appears very difficult to use objective scores like speech intelligibility thresholds, because different parameter settings do not always yield different scores (Studebaker, 1982). A paired-comparison paradigm however has high sensitivity (Eisenberg *et al.*, 1997). In addition there is a psychological advantage (Kuk, 1994a) that results from the active role of the subject. That is, optimisation of the hearing-aid algorithms is performed in interaction with the subject when a paired-comparison paradigm is used.

The use of paired-comparison techniques can be very promising for optimisation of hearing aids. In several studies the method of paired comparisons was used to optimise low- and high-frequency gain (Byrne, 1986; Neuman *et al.*, 1987; Dirks *et al.*, 1993; Kuk and Pape, 1992, 1993; Keidser, 1995; Keidser *et al.*, 1995; Kuk and Lau, 1996a, 1996b), compression ratio and release time of slow-acting compression systems (Neuman *et al.*, 1994, 1995), and compression ratio and threshold of fast-acting compression (Keidser, 1995). Basically, optimisation can be applied either within or between specific algorithms. Most straightforward is optimisation of parameter settings within one well-defined hearing-aid algorithm. However, the trend in complex hearing aids is to combine different algorithms and they need to be optimised in interaction. The optimisation procedure can be advantageous because interactions between algorithms are integrated in the optimisation process. In the pattern search studies of this thesis a broad scale of algorithm combinations is present, including settings with unpredictable trade-off and interaction effects. The listener should determine which combination is appreciated best.

Depending on the specific needs of individual subjects, evaluation should be based on specific evaluation criteria. Evaluation criteria that may be assumed to be important for optimum fitting are loudness, listening comfort or pleasantness, annoyance (e.g. to background noise, distortion, reverberance), speech intelligibility, and sound quality. In a study of Byrne (1986) subjective evaluation of intelligibility and pleasantness appeared to be significantly different. Especially pleasantness is strongly affected in noise. Paired-comparison tests of Byrne (1986) suggest that differences in frequency response had more effect for speech in quiet when judged on the criterion "speech intelligibility" than when judged on the criterion "pleasantness". For judgments on pleasantness frequency response differences had more effect for speech in noise. Because listening comfort (pleasantness) and speech intelligibility may give opposite trends, both evaluation criteria have been chosen for further investigation.

What is a good starting point for hearing-aid algorithm optimisation? The multi-directional pattern search is a possible tool, but in combination with the method of paired comparisons several obstacles are present. Is any algorithm suitable for pattern-search optimisation? Until now the pattern search optimisation strategy was used principally to fine-tune frequency shaping, the frequency-specific amplification of hearing aids. The relative success in previous studies of the multi-directional optimisation method has resulted in a major methodological gap. There is little known about issues that relate to the optimisation method like reliability and convergence. How sure can we be that the optimisation strategy 'works' in the audiological domain? In the

past only the effectiveness of the optimisation strategy was tested: can the procedure find the optimum hearing-aid setting? Previous studies especially aimed at finding the optimum, without investigating the optimisation process itself. Therefore, we also like to investigate what the possibilities and the constraints of the optimisation procedure are. This problem is approached in a distinctive way, so that a formal approach is avoided as much as possible and instead traded for a more intuitive style.

1.8 Simulation of hearing impairment

Simulation of hearing impairment is one way to come closer to the perceptual world of the hearing-impaired individuals and understand the problems they encounter. This section discusses the opportunities and constraints of simulation models.

The communication between the philosopher Martin Milligan, who happened to be blind, and the philosopher Bryan Magee (2001) has contributed a lot to attaining similarities and differences between experiences of the 'normally' seeing and the blind. According to Milligan knowledge can be the same, but the experiences are different. Similarly, the real differences between normal hearing and hearing-impaired subjects are the consequences of lacking acoustic information. Becoming hearing-impaired changes especially social intercourse and orientation rather drastically.

A different approach of listening into the world of the hearing-impaired people is to simulate certain hearing losses. Thus we can try to imagine what a hearing-impaired person hears and what the consequences are. Additionally, the effect of compensation or restoration by signal-processing strategies can be measured and compared to the results of the hearing-impaired subjects. The performance on auditory tasks by normal hearing persons with simulated hearing loss can provide insight into the world of experiences of the hearing-impaired person, and the effects of deteriorated auditory performance. It would be very interesting to compare a hearing impaired and a normal ear in one person. We measured only one subject for which this combination applied.

In order to be able to simulate a certain hearing loss, we need to model the hearing loss. Models can summarize general properties of the hearing system. Three kinds of models have been developed: transmission models, psycho-acoustic models and cognitive models. In the transmission model the physiological functionality, i.e. the successive stages of processing in the ear, can be modelled. Beyond the periphery, where neural information is processed, cognitive models can be used, like auditory scene analysis and streaming models. A different approach that combines both areas in an integrated way is the psycho-acoustic model, which only describes the relation between the acoustic input and the psychological output.

The advantage of using models is that individual psycho-acoustic factors can be changed and its effects measured. Models can support the design of hearing-aid strategies. Models can also help normal hearing, including people who live together or close to hearing-impaired subjects, to seize the effects of the hearing loss. Modelling the hearing-impaired world has two disadvantages. First, there are numerous kinds of hearing losses. The methods to determine the deteriorated functionality are not exhaustive. Therefore hearing losses that seem to be the same can yield different (kinds

of) hearing problems. So we have to generalize to some extent. Secondly there is the problem of perception in itself. We can never know what another person experiences. We only have access to our own experience. We cannot experience what the effect is of a certain kind of hearing loss for another person.

Although a real simulation of a certain kind of hearing impairment for a specific individual cannot be realized, we will use psychophysical models. These models can predict the response of the auditory system on physical stimuli at least to some extent. The models we used were based on several assumptions concerning the functioning of the normal and the cochlear impaired ear.

The mechanical response of the normally functioning cochlea is based on non-linear compression, in order to transform the large range of intensities to the small range of neural responses. This is illustrated in figure 1.2. For a low or a high input (parts I and III), the output is assumed to be proportional to the input. For intermediate input levels (part II) compression is active. In figure 1.2 the hatched area illustrates the compressive characteristic of the healthy cochlea.

Next, in a healthy ear the movement of the basilar membrane is non-linear, due to the behaviour of the outer hair cells. We assume that for a cochlear hearing loss the non-linear characteristic is reduced. As a consequence, a cochlear hearing loss yields a threshold shift, i.e. the intensity of a signal needs to be larger in order to be just noticeable. The broken line in figure 1.2 represents the input-output curve of a cochlear hearing loss with complete absence of compression. In effect, only a small part of the input levels can be mapped on the range of output levels. Hence the dynamic range (of noticeable input signals) is much smaller. Additionally the frequency resolution is expected to be poorer for an impaired cochlea.

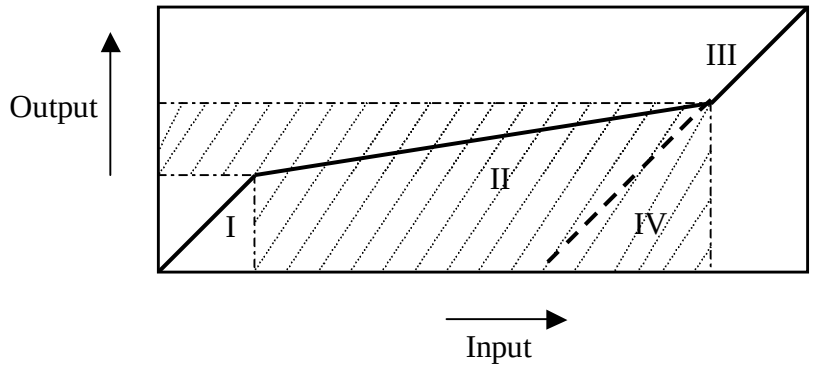


Figure 1.2 Input-output characteristics of a healthy cochlea (I-II-III) and an impaired cochlea (IV-III).

Finally, a cochlear hearing loss is assumed to result in loudness recruitment. That is, when the audibility threshold is higher than normal, but the uncomfortable loudness level remains the same, the loudness growth is assumed to be steeper than normal.

To simulate an elevated threshold and loudness recruitment we used masking noise.[7] Spectral smearing simulates the deterioration of frequency resolution, producing distortion among other effects.

1.9 Research questions

The main problem studied in this thesis applies to hearing aid fitting. At this moment, the hearing aid fitting practice is inadequate in at least two respects. First, experts decide what is best for hearing-impaired subjects, mainly based on technological-innovative grounds. Second, there is insufficient knowledge and empathy of the perceptual world of hearing-impaired subjects. A solution to these problems is the development of a fitting strategy that tries to find the optimal combination of hearing-aid techniques *in interaction* with the hearing-impaired individual.

We studied the applicability of a multi-directional optimisation procedure for different signal-processing strategies. Our goal was to develop an optimisation procedure that meets the following requirements:

a. it optimises subjective speech intelligibility or listening comfort of speech in background noise;
b. it optimises multiple signal-processing strategies simultaneously;
c. it is directed by subjective evaluation, i.e. by individual judgment of stimuli based on paired comparisons;
d. it takes (non-linear) interactions between signal-processing strategies into account;
e. it is time-efficient.

The multi-directional pattern search that was developed includes an adaptive step size.[8] The step size for each dimension was governed by the perceptual differences between settings for each listener. That is, the listeners should be able to distinguish the settings that need to be compared. The reliability of the optimisation procedure appears to be reasonably well.

We supposed that fine-tuning of a hearing aid that is based on subjective preferences and the multi-directional pattern search can yield an optimal combination of different signal-processing strategies. The reason that the 'true' optimum can be found relates to the fact that non-linear interactions between the signal-processing strategies will be taken into account. At the same time we expect that the multi-directional pattern search can be applied for almost any kind of signal-processing strategy.

---

[7] The model of masking noise keeps the signal intact. Other models, such as non-linear expansion, the mirrored version of compression, are based on manipulation of the signal itself.

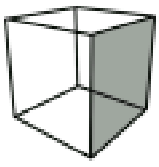
[8] In two studies described in sections 4.2 and 4.3 the step size was fixed. In order to increase the reliability of the multi-directional pattern search the step size was made adaptive.

Chapter 2

METHODS

'Each man will even end up by 'seeing' the world in ways organized according to his own *Gestalt*. For what he 'sees' when he looks down (say) a microscope will be governed not only by the structure of his eyes and his instruments, but also by his particular theoretical paradigm; this will determine what any particular specimen is 'seen as'.'

– Stephen Toulmin
Human understanding – The collective use and evolution of concepts (1972)



In this chapter the methods that are used in this thesis are described. Methods that are used generally in the audiological practice comprise objective and subjective tests. The objective tests are physical in nature. Tympanometry, Electrical Response Audiometry (ERA), and oto-acoustic emissions are typically objective tests, for these tests do not need cooperation of the subject. Objective tests were not used in our studies, because these measurements cannot be used to assess thresholds, loudness growth, and frequency selectivity with enough precision to be applicable in auditory rehabilitation. Subjective tests can be subdivided into psychoacoustic and almost purely subjective tests. Psychoacoustic measurements are subjective in nature due to the necessity of active participation of the subject. Though psychoacoustic tests can yield quantitative measures of the hearing system that can be related to systematically varied physical input signals. When this relation is absent the test might be called purely subjective. These methods lack the systematically varied physical input and sometimes the quantitative output measures. Psychoacoustic tests that are used in (some of) our studies are pure-tone audiometry, auditory filter measurements, loudness measurements, difference limens, and speech intelligibility measurements, i.e. phoneme and sentence tests. Purely subjective tests comprise rating tests (e.g. of listening comfort and sound quality), questionnaires, and two procedures that are based solely on subjective judgments: the multi-directional pattern search and the round robin procedure. These procedures are based on pair wise comparisons of different signals on a verbally described sound attribute.

2.1 Pure-tone audiogram

The pure-tone audiogram was determined using standard audiometric techniques in a sound-insulated test booth. For each subject the pure-tone threshold was measured at 0.25, 0.5, 1.0, 2.0, 4.0, and 8.0 kHz, with adequate masking of the contra-lateral ear.

As suggested by Moore (2002), a problem of the measurement of the pure-tone audiogram is that the hearing loss might be greater than suggested by the audiometric threshold. In the presence of so-called 'dead regions', regions in the cochlea where inner hair cells and neurons are totally absent, neighbouring regions with surviving inner hair cells and neurons will detect the stimulus and they are responsible for a "flattened" response.

2.2 Speech intelligibility measurements

Two types of speech intelligibility measurements were performed. Speech reception thresholds assess the intelligibility of sentences. CVC tests assess the intelligibility of words and, more in particular, phoneme intelligibility.

2.2.1 Speech reception thresholds

The speech reception threshold (SRT) is the signal-to-noise ratio (SNR) at which 50% of speech in noise is repeated correctly (Plomp and Mimpen, 1979). The procedure is adaptive and follows a one-up-one-down paradigm, i.e. after a sentence is repeated

correctly the SNR is reduced, after a sentence is repeated incorrectly the SNR is increased. After thirteen sentences the procedure stops. The first sentence was repeated until it was repeated correctly using speech level increments of 4 dB to converge quickly. Increments and decrements of the speech level were 2dB for the others sentences. The speech reception threshold was estimated by averaging the SNR of the last nine sentences and the SNR of the sentence that would have followed the thirteenth sentence. During the tests, the speech level was fixed.[9]

2.2.2 Consonant-vowel-consonant tests

In the consonant-vowel-consonant test the subjects were asked to listen to CVC words included in a short carrier sentence chosen from a set of five standard sentences. The target words are nonsense words, always consisting of three phonemes: consonant-vowel-consonant (CVC). The subjects were asked to repeat the CVC word. The male spoken Dutch sentences come from Hans Boer lists, developed by the TNO-IZF Institute in Soesterberg (Steeneken *et al.*, 1990). Each list consists of 51 sentences. In order to avoid fatigue the subjects listened to no more than four lists per session. Four categories of data were obtained: three phonemes scores and the overall scores.[10]

2.3 Loudness measurements

We measured loudness perception using the Würzburger Hörfeld Skalierung (Hellbrück and Moser, 1985). This method prescribes that subjects listen to 1/3-octave noises of different intensities and centred at four central frequencies: 0.5, 1.0, 2.0, and 4.0 kHz presented in random order. Their task was to scale the loudness of the noises. They could respond by giving any integer number between 0 and 50, corresponding with 'not heard' and 'too loud', respectively. The numbers 5, 15, 25, 35 and 45 corresponded to 'very soft', 'soft', 'middle loud', 'loud' and 'very loud', respectively. An automatic fitting procedure yields a curve relating subjective loudness to input level (given in sound-pressure level, dB) for each of the four frequency areas. Both the steepness and the horizontal position of the fitted curves carry information about the loudness perception of the subjects. The steepness is directly related to the subject's Dynamic Range (DR) per frequency. We calculated Dynamic Range values from the level-differences corresponding to the subjective responses of 40 and 10 (in Categorical Loudness Units), and the indications of loud-to-very-loud and soft-to-very-soft, respectively. For each subject the most comfortable level corresponds with a categorical loudness of 25 units.

---

[9] The speech level was fixed in order to avoid parts of the speech become below threshold. As a consequence the noise level was variable, introducing another matter of consideration in that the noise level can become uncomfortably loud. Especially for noises that tend to mask the speech rather poorly, uncomfortable loudness levels can occur. The background noises used in our studies did not yield any problems of this kind.

[10] The material of the SRT test is much more redundant relative to the material used in the CVC test. This does not necessarily imply that the test is less suitable, for the test simulates problems that occur in daily life more properly. The advantage of using a CVC test is the possibility to differentiate between problems that occur for individual phonemes. However, due to a shallow slope in the psychometric curve of the CVC test, the procedure is less sensitive for differences in signal-to-noise ratio.

2.4 Auditory filter measurements

To measure the frequency selectivity of the ear the so-called notched-noise method was used. The task for the listener is to detect a sinusoidal signal in noise that has a spectral notch.

Frequency selectivity was measured at one or more different frequencies, using the notched-noise technique and a Von Békésy procedure, described by Leeuw and Dreschler (1994). Parameter is the positioning of the notched noise bands, relative to the centre frequency f_c of the probe tone. As described in Leeuw and Dreschler (1994) this notched noise method for clinical purposes includes five different noises. The probe tones are, except for the noise without notch, asymmetrically located in the notch. The notched noise bands are either 600 or 1000 Hz wide, depending on the target frequency. For two noise conditions the higher frequency noise band is situated 400 Hz higher than the signal, so the lower band is either 200 or 600 Hz lower than the probe tone. For the other two conditions the opposite is true: the lower frequency noise band lies 400 Hz lower, while the higher noise band is either 200 or 600 Hz higher. With these five conditions it is possible to get a reliable estimate of the auditory filter (including the filter's asymmetry) in a reasonable time. Each condition was tested twice, unless the thresholds obtained differed more than 5 dB. In that case a third measurement was included.

The thresholds were fitted by means of the program FASTFIT (Glasberg and Moore, 1990), which yielded parameters for the lower and upper skirts of the triangular-shaped auditory filter, PL and PU, respectively. The Equivalent Rectangular Bandwidth (ERB) was calculated from PL and PU.

2.5 Difference Limens

In order to find out what differences in parameter settings are perceptually relevant, difference limen experiments were performed. The concept of threshold or limen can be explained by discerning three types of limens. The absolute limen (AL), the terminal limen (TL), and the difference limen (DL), which represent the lower boundary of detection, the upper boundary, and a relative boundary within the range of the other two boundaries, respectively. A threshold is actually a statistical value, due to the randomness inherent of the human hearing system.

The difference limen test we used uses a three-alternatives forced-choice paradigm. Two alternatives are equal, one is different. The task for the subject is to determine which alternative is different from the other two. The subject is forced to make a choice.

Difference limen tests were performed in one study that incorporates three parameters of two different signal-processing strategies. Each parameter range used is divided in 45 equidistant steps. The difference limens were determined at three parameter settings: the two extremes and the geometrical mean of the two extremes. For the geometrical mean parameter setting we performed two difference limen tests, i.e. directing towards the two extremes.

Apart from the fact that difference limen experiments can contribute to a well-considered choice of the step size used in the optimisation procedure, it is still

impossible to be sure that all steps in the optimisation procedure are perceptually distinguishable. That is, although these limens can be measured, e.g. for each algorithm separately without processing of the other two algorithms, there is still no knowledge of the limens for combinations of these non-linearly interacting algorithms. Equidistant perceptual steps within the parameter space where the optimisation takes place cannot be ensured, unless each individual parameter setting is compared with each neighbouring parameter setting.[11]

2.6 Paired comparisons

In the method of paired-comparison subjects indicate preferences for pairs of alternatives. A set of comparisons can result in indicating which combination of parameter settings is judged best (Zerlin, 1962; Punch and Howard, 1978; Punch *et al.*, 1980; Studebaker *et al.*, 1982). In earlier studies several advantages of a paired-comparison method relative to more objective word-recognition scores emerged. The method of paired-comparison is equally or more reliable, has greater sensitivity in differentiating between hearing-aid settings, and has a high predictive power of hearing aid performance when evaluation is based on speech intelligibility. Moreover, involvement of the end-user has proven to yield a psychological advantage (Kuk, 1994). This can result in better acceptance and helps the users to get accustomed to the renewed acoustical environment. Due to the fact that the paired-comparison strategy is based on subjective judgments, it is possible to evaluate signal-processing strategies on criteria that cannot easily be measured quantitatively.

We used the paired comparison strategy with a forced-choice paradigm (AB-paradigm). The listener had to decide which of the two samples A or B was preferred. When the two samples seemed equally good or bad, even after repeating the samples, the listener was forced to make a choice. Depending on the study we performed, the listener was instructed to judge the speech intelligibility, the listening comfort, or a combination of both, i.e. judge the speech intelligibility while taking into account that the algorithm setting has to be used all day long.

2.6.1 Multi-directional pattern search

In order to simultaneously optimise multiple hearing aid algorithms a multi-directional pattern search design was applied. The multi-directional pattern search looks for the steepest-hill to reach the optimum as quickly as possible. Levitt *et al.* (1978b) and Neuman *et al.* (1987) were the first that applied a two-dimensional Simplex procedure (actually a pattern search procedure, see appendix 1) in combination with the paired-comparison paradigm using digital master hearing aids. From the latter study it appears that the final estimate agrees remarkably well with the optimum selected by round-robin and double-elimination tournaments, whereas in general the Simplex (pattern search) procedure is much more time-efficient.

---

[11] We considered such an elaborate job not realistic, for when we optimise ten parameters in the future – which should be a realizable goal, or else the patterns search strategy would be too constrained in its utility – it is practically unrealisable to determine equidistant steps in all cases.

For the optimisation strategy, we use three dimensions that represent three (experimental) hearing aid algorithms. A flow diagram of the pattern search strategy is given in appendix 2. Several procedural features characterize the search strategy, which is aimed at finding the optimal combination of all algorithm settings. The search has a single initial setting. After the first comparison, the next setting is assumed to have improved performance. To search for improved performance, the setting under focus is compared pair wise with three neighbours (n paired comparisons for n dimensions). The settings included for paired comparisons are orthogonally related to the initial setting under focus and remain orthogonally related. So, three orthogonal search directions are used throughout the procedure for this pattern search design. The decision as to what will be the next setting under focus depends on the outcome of all three paired comparisons in the following way. When neither of the neighbours wins, the setting under focus remains the best setting. When one neighbour wins, this neighbour will be the new best setting. When two or three neighbours win, the new setting is estimated from the winning directions. That is, the assumed new best setting is found by addition of the vectors that represent the search directions, i.e. the directions of paired comparisons. As a consequence, the progress can be along a diagonal with respect to the orthogonal system of search directions (see appendix 1).

The new search directions for the novel best setting depend on the history of search directions. For each dimension the search direction remains the same when the neighbour yields higher listening comfort. The search direction is reversed when the setting at stake performs better than the neighbour in that particular dimension. The incorporation of the history of search directions is very important for this search strategy in order to accelerate the optimisation process. The settings can never escape a predetermined lattice structure. This feature is necessary for our study due to the fact that we had to process the speech and noise material off line.

For the enhanced multi-directional search used in the latest studies, the step size has been made adaptive[12], and can be characterized by perceptual distinctness. At the start of the search, the step size is largest, e.g. equal to four steps in the lattice structure of section 4.6. After one series of three paired comparisons the step size is halved. After another three rounds the step size is halved once more, at least when the minimum step size was not yet reached. The performance of the optimisation procedure is illustrated in figure 2.1. The halving process has been incorporated to yield convergence. The initial large step size was mainly chosen to ensure audible differences between settings and to make the listeners acquainted with the perceptual dimensions. At any time the samples to be compared can be repeated. When the listener cannot distinguish between the two settings, the step size for that particular dimension will be increased until it is equal to three. Again, after three series the step size will be halved and rounded, as long as the step size is not equal to one.

[12] In an adaptive test the conduct is determined by the data obtained during the test (Levitt, 1978). In the multi-directional search the search direction is adaptive. Additionally, for the multi-directional search with an adaptive step size the step size can vary, depending on perceptual distinctness of the samples and on convergence rules.

It is important to note that the 'no difference' button can be used only after the 'repeat' button has been used once. The instruction was that the listeners should be convinced that the samples are indistinguishable, instead of just rather difficult to judge in terms of listening comfort. Each pair of samples can be repeated once. The multi-directional pattern procedure terminates when one of the following three stop criteria was met. When the same setting is frequented three times or when a setting wins a paired comparison more than six times, the procedure will stop. These stop criteria are a compromise between good convergence and limited required average test time. The third stop criterion is the maximum number of paired comparisons in the procedure. This criterion has been included to limit the maximum test time of the procedure. In a pilot experiment a limit of 54 paired comparisons was found to be a reasonable choice, which corresponds to a test duration of fifteen minutes on average.

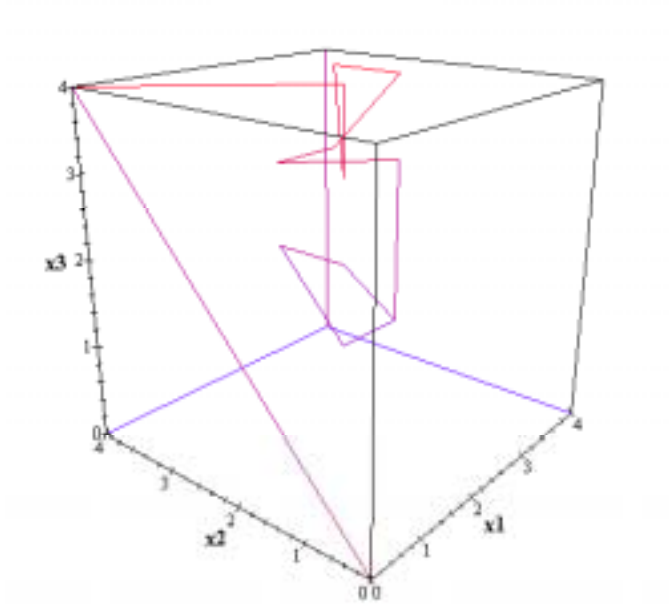


Figure 2.1 Multi-directional pattern search procedure with adaptive step size, which is variable for each dimension. Three dimensions are used, represented by the cubic space. The dimensions of the cube are represented by three algorithms, x1, x2, and x3. The step size is intended to be related to the minimum perceptual distance for a subject. The corner of the cube in the front below is the starting point in this example. The optimal setting, which is assumed to be the end point of the optimisation strategy, is not really found, but confined by the closed path in the middle of the cube. Note that the size of the step in this example does not follow the pattern search rules exactly for reasons of clarity.

2.6.2 Round-robin tournament

Basically, the multi-directional pattern procedure produces only one outcome; that is the optimal setting. Without the use of an extensive procedure like a round-robin procedure, or the more efficient double elimination tournament, there is no information about the overall distribution of preferences. The round-robin procedure is essentially a

tournament between algorithm settings, which can give detailed information because every set of settings is fully rank-ordered. The algorithm that is preferred most often is ranked highest. So, not only the optimal setting is known, but also information about serious competitors (suboptima) will become available.

The round-robin tournament and the multi-directional optimisation strategy have many procedural similarities. Again, the paired comparison paradigm is used to determine the preferred setting. Next, the presence of a 'repeat' and a 'no difference' button was included. When the 'no difference' button has been used each setting was awarded ½ points. A real winner of a paired comparison gets one point. The rank ordering is determined by counting the total number of points for each setting. Because five different settings are included in our studies, the tournament yields 25 different comparisons.

2.7 Rating

One way to investigate the perceptual landscape, the domain in which the optimal setting is located, and to determine a reference frame for the optimal settings found, is to judge different settings in a predetermined experimental design. For a restricted number of settings subjects were asked to rate listening comfort on a scale between 0 and 10 in steps of 0.1. The chosen combination of algorithm settings makes it possible to detect algorithm interactions. It is possible that, although each algorithm separately does not yield any improvement, the combination does yield at least a subjective improvement.

2.8 Questionnaire

A questionnaire was used in order to find out more about the listening strategies of the subjects. Therefore, we asked the listeners to indicate which sound feature they had been using. The features contain clarity, fuzziness, noisiness, loudness, sound quality, pleasantness, fluency, and speech intelligibility. The subjects could specify how much they used these features by choosing among four options: none or hardly any, some, reasonable, and much. Because we used speech-in noise stimuli, the subjects could also indicate if their answers are associated to the speech, the noise, or to both.

2.9 Simulations of cochlear hearing loss

The impaired hearing system of subjects with cochlear hearing loss suffers especially from a combination of elevated thresholds and a reduced frequency-specific dynamic range, and from poor frequency resolution. These three deficiencies were modelled by two different models in order to simulate the most important characteristics of the impaired hearing system.

Formerly, two models were generally used to model reduced frequency resolution. First, frequency smearing is simply achieved by multiplication of the original waveform and a low-pass filtered noise (e.g. Summers, 1991; Boothroyd, *et al.*, 1997). In effect a noise band replaces each tonal component of the speech spectrum. As a consequence

both the spectral fine structure and detailed waveforms, i.e. the temporal details, are manipulated. Second, spectral 'smoothing' of the signal in the frequency domain can be used (e.g. Ter Keurs *et al.*, 1992; Baer *et al.*, 1993). The effect of the technique of Ter Keurs *et al.* (1992) is that the spectral fine structure is kept intact. That is, the frequency-dependent product of a smeared spectral envelope and the original complex spectrum, effectively reduce the spectral contrasts while the harmonic structure and the phase structure remain intact. The smearing algorithm of Baer *et al.* (1993) also affect the spectral fine structure, simulating the reduced frequency selectivity as well, without disturbing the detailed waveform. They showed that spectral smoothing increases the disturbing effect of noise.

The elevated thresholds and reduced dynamic range are supposed to result from a loss or reduction of compression in the hearing system. The absence of the non-linear compression characteristic of the healthy hearing system was generally modelled in two ways. The signal can be dynamically expanded in multiple channels, such that when combined with the compression of the healthy ear, the transduction becomes, at least theoretically, linear for each frequency channel (e.g. Villchur, 1974; Lum and Braida, 1997). The signal can also be accompanied by a masking noise, modelling the elevated threshold such that many target signals become inaudible, (e.g. Fabry and Van Tassell, 1986; De Gennero & Braida, 1997).[13] The reduced dynamic range, with an upper limit similar for the impaired and the normal hearing system, is realised by both models.

Two types of simulation were used in one of our studies. In the first type of simulation we have chosen to solely use the spectral smearing. Lyzenga *et al.* (2002) developed the algorithm used to simulate reduced frequency discrimination, which is a modification of the smearing algorithms of Ter Keurs (1992) and of Bear *et al.* (1993). Essentially, the smearing method is based on the convolution between the envelope of the squared short-time FFT of the signal and a Gaussian-shaped filter. The degree of smearing is given by the bandwidth of the filter. The filter, therefore, represents the auditory filter of the (hearing-impaired) listener, i.e. the more serious the cochlear hearing loss, the broader the auditory filter. Lyzenga *et al.* (2002) added two alterations to this strategy, aiming at better sound quality characteristics. Random phase values are added for small excitation values to prevent introducing annoying artefacts[14] as much as possible. The random phase values produce, instead, a noise, which level is rather low, and therefore, will be masked by the much louder background noise used in our studies.

---

[13] It seems that for many studies that simulate certain aspects of cochlear hearing loss one problem is overlooked. Simulation of hearing impairment is principally a combination of alteration of the input signal - by manipulation of the signal or by additional signals – and processing of the healthy ear. However, the healthy ear is not the normally performing cochlea alone. Also the influence of the outer and middle ear should be accounted for. Unless these components operate linearly, the order of processes that need to be performed in order to obtain a correct simulation should be considered carefully. In fact, the alteration of the signal by a certain type of simulation strategy should take place after the signal has passed the outer and middle ear. Then two complementary processes like for example compression and expansion will result in a linear transmission, simulating the lack of compression in an impaired cochlea. In practice this solution is hardly realistic. Though for the linear addition of masking noise the intervening action of the outer and middle ear should not be too much of a problem. For non-linear strategies like dynamic expansion, however, this could be a serious problem.

[14] Avoiding the introduction of audible artefacts is especially important for our study, because the spectral enhancement and noise reduction strategies are susceptible to introducing artefacts as well.

Additionally, a more natural sound is preserved by using an overlap of 75% between successive smeared signal segments, due to the presence of relatively much energy around the fundamental frequency.

For the second type of simulation we incorporated both the spectral smearing algorithm of Lyzenga *et al.* (2002) and the masking noise concept from De Gennero and Braida (1997).[15] The order of processing was chosen spectral smearing first and masking noise second.[16] The masking noise was filtered according to the shape of audiogram of the subject that had the smallest high-frequency hearing loss (40 dB, averaged for 0.5, 1.0, 2.0 and 4.0 kHz). After that, the overall level of the filtered masking noise was determined based on pure-tone audiometry of three normal hearing subjects. That is, the level of the filtered masking noise is amplified by an amount, such that the thresholds that are found for the normal hearing subjects who listen to signals in the presence of the filtered masking noise are the same as the thresholds of the subject with the smallest audiometric hearing loss.[17]

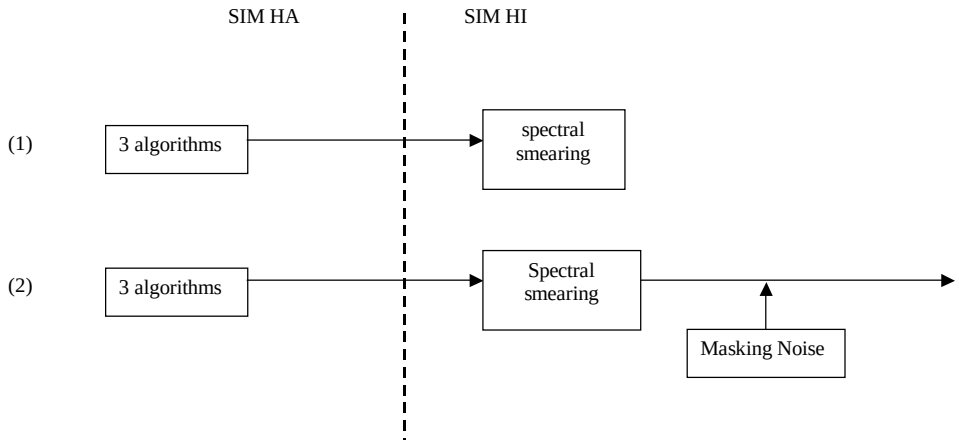


Figure 2.2 Combination of hearing-aid simulation (SIM HA) and two simulation types of cochlear hearing impairment (SIM HI).

---

[15] An advantage of simulation by expansion is that it works by attenuation instead of masking as for masking noise. Therefore, simulation of more severe hearing losses is still possible by expansion, not by masking noise. A suspected disadvantage relates to the problems that occur for multi-channel compression. For compressors we assume that, contrary to multi-channel compression, the single-channel compressor does not change the short-term spectra of the speech signals. Similarly, we assume that for a multi-channel expander the short-term spectra will be changed as well.

[16] The order of processing is not at all obvious. Oxenham asserts in a personal correspondence that compression and frequency resolution might well be two aspects of the same non-linear filter mechanism. The choice for the order just indicated, compression after filtering, is more practical in nature. That is, when the order is reversed the frequency channels would not be independent any more. However, the solution given is not at all elegant as well, because effects like suppression need a certain kind of interaction between filtering and compression. This would implicate that in order to model suppression, a specific kind of processing is needed that can simulate a poor frequency resolution and elevated thresholds (or more broadly stated, reduced compression) simultaneously.

[17] It appeared that the level of the masking noise should be chosen 86 dB SPL to achieve the same thresholds. This high level indicates that the use of masking noise to simulate elevated thresholds is useful for mild-to-moderate hearing losses only. This is even more so in combination with the speech-in-background-noise target signal. Because the presentation level is unacceptably loud, we decreased both the speech-in noise signal level and the masking noise level until the combination was equal to 85 dB SPL.

The simulated hearing impairment was tested in combination with the simulation of a hearing aid. Therefore, three signal-processing strategies were incorporated to reduce noise and to enhance the speech signal, developed by Lyzenga *et al.* (2002).[18] Two combinations of simulating a hearing aid and a simulated impaired cochlea we studied are presented in figure 2.2. The addition of experimental hearing-aid signal-processing strategies implies the incorporation of a compensation rule, or some other amplification strategy. Only then the results of simulated hearing-impaired subjects can be compared to the results of 'true' hearing-impaired subjects for samples of speech-in-background-noise, processed according to three signal-processing strategies. Because the hearing-impairment of subjects in our studies is compensated for by the National Acoustic Laboratories (NAL) prescription rule, also the simulated hearing-impaired subjects get amplification according to the NAL rule. It would be proper when the NAL compensation would take place immediately after the simulated hearing aid and before the simulated hearing impairment. However, due to the constraint that the signal processing had to take place off-line, i.e. both for the experimental hearing-aid strategies as for the spectral smearing, the NAL compensation could be applied only after both processing stages.

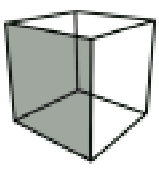
[18] The respective algorithms, spectral enhancement, spectral lift, and noise reduction that are developed by Lyzenga *et al.* (2002), are described in chapter three.

Chapter 3

ALGORITHMS

'The system is computer controlled, greatly simplifying the task of adjusting the master hearing aid. As a result, extremely efficient adaptive paired-comparison strategies can be used to converge on the estimated optimum. In the example given, only the frequency-gain characteristic was adjusted, this being the thrust of much recent research. Using the method of computer simulation, it is a relatively simple matter to manipulate other electroacoustic characteristics such as frequency-dependent amplitude limiting, compression amplification, and other forms of acoustic processing.'

– Harry Levitt
A digital master hearing aid (1986)



In order to help hearing-impaired people distinguish and comprehend speech in noisy environments, several algorithms have been proposed. At least two different approaches are possible: either by reduction of the undesired effects of background noises and reverberant room acoustics, or by enhancement of the desired (speech) signals.

3.1 Reduction of interfering sounds

Different types of reduction – or suppression – strategies were used. These comprise two different monaural noise reduction strategies, a binaural noise reduction strategy, and a de-reverberation algorithm.

3.1.1 Monaural noise reduction

Monaural noise reduction based on spectral subtraction

A first type of monaural noise reduction (NR_D) is an adapted version of the Ephraim and Malah algorithm and described by Dubbelboer *et al.* (1999). Basically, this strategy uses spectral subtraction to suppress noise, i.e. subtracting the estimated frequency spectrum of the noise from the original signal. For speech in noise signals, first the noise had to be separated from the speech. Therefore, a speech-pause detector was used (Van Beek *et al.*, 2000).

Two parameters of the speech-pause detector that control the monaural noise-reduction algorithm were varied in one experiment: the 'a priori signal-to-noise ratio update time constant' (N_{estim}) and the 'speech pause detection threshold' that estimates whether the signal is speech or noise (SN_{estim}). Parameter N_{estim} controls the estimation of the amount of noise present in a stimulus sample. N_{estim} is varied between 1 ms and 1001 ms. The shorter the time windows the more the algorithm suffers from artefacts. The artefacts are expected to occur due to the small internal fluctuations inherent for noise. Longer time constants inhere the problem that changes in the noise level cannot be followed accurately any more. Parameter SN_{estim} is determined based on the ratio between the minimum and maximum energy. Hence, the ratio determines the modulation depth of the signal and can vary between 0 and 1. In effect, when SN_{estim} is low the signal is expected to be noise, and maximum reduction will be applied. When SN_{estim} is high the signal is highly modulated. In this case the signal is assumed to be speech and therefore the attenuation factor is set minimal. The noise update time constant was set default at 132 ms.

Monaural noise reduction based on phase variations

The second type of monaural noise reduction (NR_I), developed by Lyzenga *et al.* (2002), attempts to suppress noise based on the temporal behaviour of the signal, i.e. the phase variations. When the temporal behaviour of spectral components is irregular, the signal is assumed to be noisy, and the linear suppression filter is set at small amplitudes. The amplitude of the filter is set high when spectral components behave temporally regular, and are assumed to be speech-like signals. The noise reduction multiplication factors varied by dimensionless values between -3 and 6 in steps of 1. A value of 0 represents

the unprocessed condition. The algorithm values have been chosen so that steps are approximately perceptually equidistant for normal hearing listeners.

3.1.2 Binaural noise reduction

The binaural noise-reduction algorithm was developed in Oldenburg (Kollmeier *et al.,* 1993; Marzinzik, 2000). In this strategy a directional filter reduces the background noise binaurally. The algorithm is based on the assumption that an incoming sound emerging from a frontal position is the target sound, whereas sounds that have incidence angles that are sufficiently different from the frontal position represent distracters, i.e. background noise. The signal processing of the binaural noise reduction strategy took place in the frequency domain. For the noise reduction scheme the amount of attenuation is variable.

Two parameters account for the attenuation of unwanted signals, the Level Difference Weighting (LDW) and the Phase Difference Weighting (PhDW). These weightings vary between 1 (no reduction: reference condition) and 0.1 (maximum reduction), and are related to the incidence angle of incoming signals. In the frequency domain the level and phase differences are used in order to estimate the spatial representation of sounds. When the phase and level differences are large between both ears, i.e. between two hearing aids, frequency bands will be attenuated. For incidence angles larger than 19^0, relative to frontal incidence (0^0), suppression becomes gradually active. For incidence angles larger than 33^0 the suppression is fully active. In our studies, only the amount of attenuation is variable. The averaging of the dynamic weights is performed in windows with a length of 60 ms. The level averaging time was set at 5 ms. Reconstruction of the signal was performed by overlapping additions of 20.8 ms segments. In order to relate the weighting functions to incidence angles of 0^0, 19^0, and 33^0, we trained the directional filter with 2 minutes white noise signals for these angles.

We measured the effect of noise reduction by processing both binaurally recorded frontal speech and diffuse car noise. The constant velocity car noise was recorded inside the car cabin at the driver's seat. The results are presented in figure 3.1. Parameter is the amount of attenuation, which varied between 0.01, 0.10, 0.55, and 1.00. The cumulative distributions clearly indicate that part of the car noise is attenuated because the incidence angles are partly larger than ±33 degrees (shown in lower panel), whereas the speech signals that are presented frontally are not attenuated (upper panel). Note the very small effect by reducing the noise suppression parameters from 0.10 to 0.01.

3.1.3 Binaural de-reverberation

The binaural de-reverberation algorithm that was developed by Kollmeier *et al.* (1993) and was slightly modified by Marzinzik *et al.* (2000) compensates for poor acoustics, i.e. rooms with much reverberation. A cross-correlation and an auto-correlation analysis were performed in the frequency domain to calculate the coherence between the signals retrieved from the binaural experimental hearing aids situated at both ears. The cross-correlations were used to determine the correlation between the two experimental hearing aids, whereas the auto-correlations were used to calculate the correlations between successive signal segments for each aid separately. Ultimately, the coherence is

estimated by the three values obtained by the different analyses. The coherence value is used to weigh each frequency component relative to the other components. In effect, the direct incident (highly correlated) signal remained the same, whereas the indirect reflected (poorly correlated) signal was attenuated.

In our experiments coherence values of 0%, 50%, and 100% were used that correspond to no processing, intermediate attenuation, and maximum attenuation, respectively.

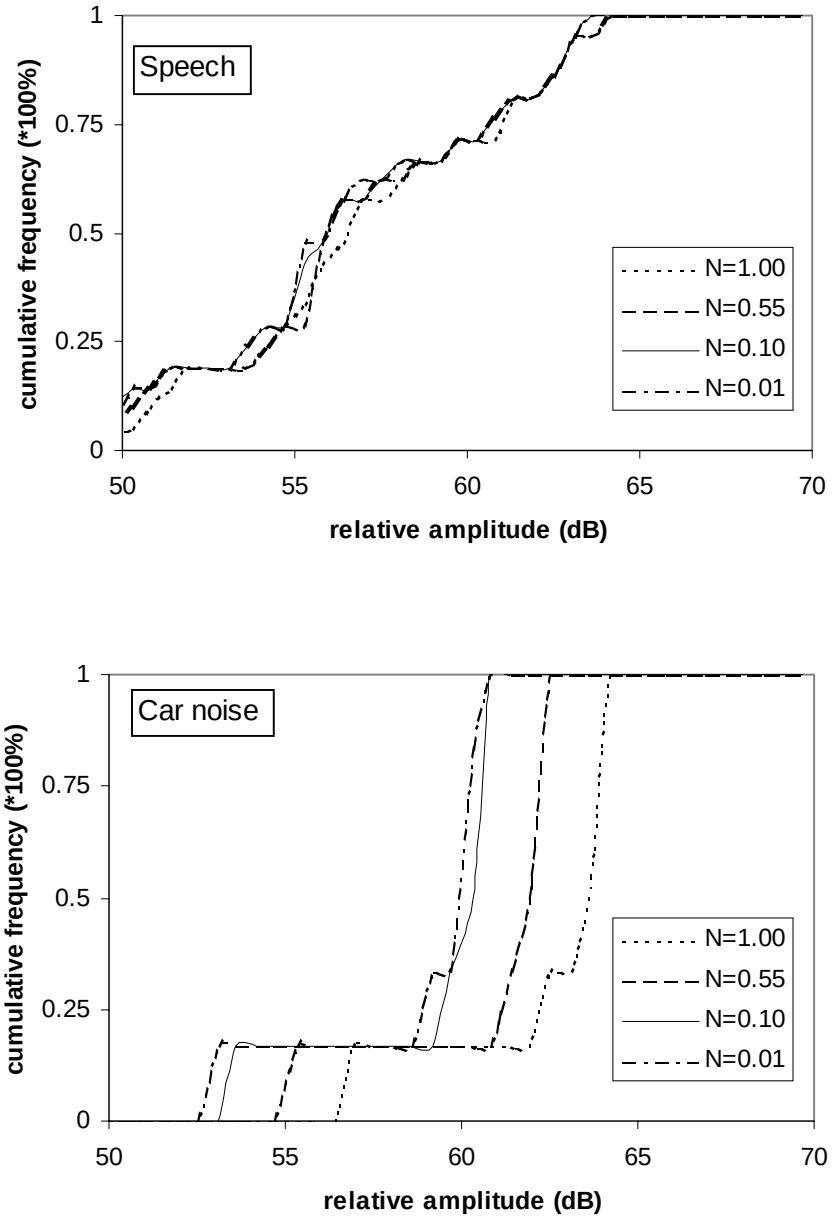


Figure 3.1 Cumulative distributions of binaurally recorded speech that comes from the front (upper panel) and diffuse car noise (lower panel). The Noise reduction parameters LDW and PhDW vary simultaneously between 1.00 (N=1.00: no reduction) and 0.01 (N=0.01: strong reduction). The intensity levels of both signals was 63 dB SPL.

3.2 Speech enhancement

Three types of experimental signal processing strategies were used in our experiments: spectral lift, spectral enhancement, and fast (phonemic) compression.

3.2.1 Spectral lift

The spectral lift algorithm was developed by Lyzenga *et al.* (2002). The algorithm is a linear filter that is assumed to reduce upward-spread-of-masking, i.e. trying to prevent low-frequency high-level formants from masking higher-frequency low-level formants. Therefore the algorithm is designed to progressively emphasize frequencies between 0.3 and 3 kHz, the region in which the second to fourth formants are dominant. For frequencies higher than 3 kHz the spectral envelope decreases again. The filter is sawtooth shaped, and the peak of the filter is centred on 3kHz. Emphasis of high frequencies is avoided, because sharp sounds may be expected to be uncomfortable. The spectral lift varied between -9 and 18, in nine 3-dB steps. A lift-value of 0 represents no spectral lift.

3.2.2 Spectral enhancement

According to Baer *et al.* (1993) the recognition of speech sounds is based on the use of spectral shape to detect spectral features, especially formants. Therefore, speech recognition may be facilitated by spectral enhancement.

Cambridge spectral enhancement

In most experiments the spectral enhancement algorithm designed by Baer *et al.* (1993) enhances both the spectral valleys and the spectral hills (SE_B). In one field study, however, only the spectral valleys were enhanced (deepened), because the experimental hearing aid needed a computationally less time consuming algorithm. In effect, it is expected that the SNR increases, i.e. that this algorithm type will act more like a noise suppressor.

The signals were sampled at a 16 kHz rate and low-pass filtered at 4 kHz for computational reasons. This may be expected to be broad enough for speech. In order to manipulate the signal in the frequency domain, the waveform of the input had to be translated to a spectral representation. Because the signal varied continuously, we needed an averaging process over a specified period of time. We chose the size of the Hamming window to be 1024 points, 32.0 ms long. This windowed segment was converted to the frequency domain, using the Fast Fourier Transform (FFT) process. After the spectrum was processed the spectral information is re-transformed into the time domain with an inverse FFT. All segments were added using an overlap-add procedure to get the ultimately processed waveform.

In the spectral enhancement process itself (which is in conformity with the method of processing of experiment 3 in the study of Baer *et al.*, 1993) only the magnitudes were manipulated. First the excitation pattern was calculated using the input spectrum. This pattern was used to filter out excessive details that would not be resolved by the ear anyway. It is a representation of the frequency contents after normal auditory filtering. Then the new pattern was subjected to a function that enhanced the peaks and valleys.

In order to create a gain function based on the read-out of these local extremes, a function was used that is composed of the difference between a positive and a negative rounded exponential function.

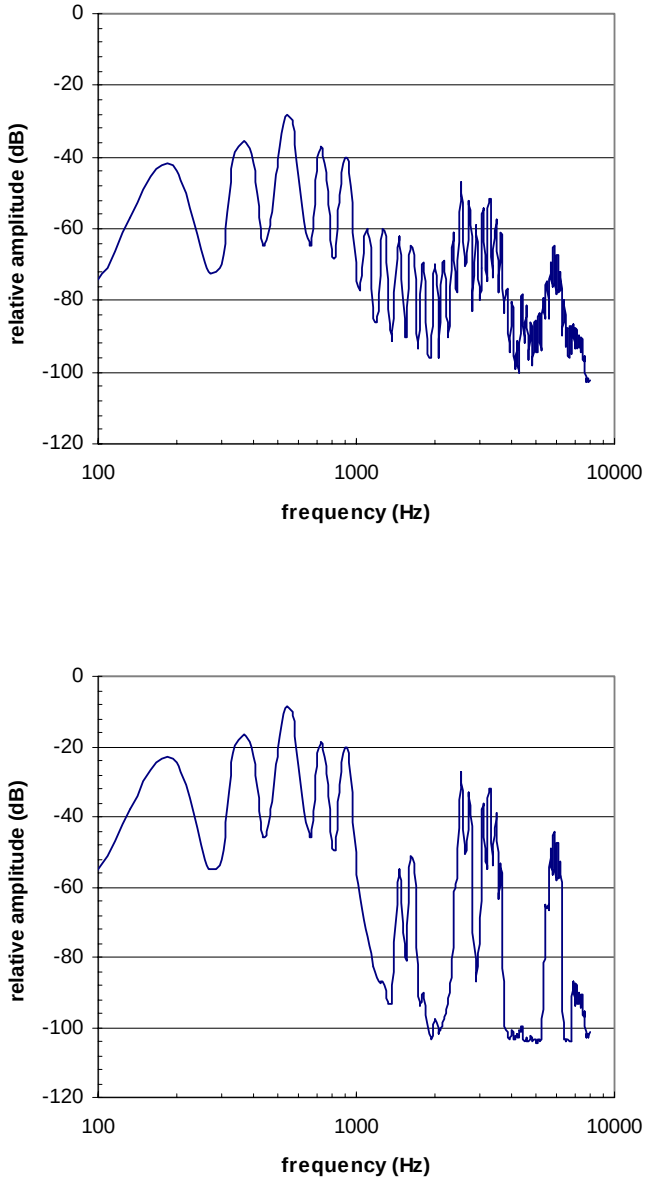


Figure 3.2 Upper panel: Spectral representation of the unprocessed vowel /o/ of the Dutch nonsense word 'woom'. Lower panel: spectrally enhanced vowel /o/, when the enhancement procedure developed by the Cambridge group Baer *et al.* (1993) has been used. For both figures a window size of 30 ms has been used, to calculate the average spectrum. For reasons of clarity, the amount of enhancement is chosen more extreme than actually used. The 4-kHz low-pass filter was not used in these examples.

For both functions the bandwidths are varied with centre frequency, based on ERB calculations of auditory filters.[19] To shape the output spectrum, the gain function was convolved[20] with the input spectrum. The minima and maxima were enhanced in this way and the fine structure of the original spectrum was preserved. In the inverted FFT the adapted magnitude data were combined with the original phase data and transformed into the time domain.

The effect of spectral enhancement is illustrated in figure 3.2. The spectral content of the spectrally enhanced Dutch vowel /o/ is compared with the unprocessed vowel /o/. To exaggerate the effects of spectral enhancement, the amount of enhancement is chosen larger than actually used. These pictures are mean Fast Fourier Transforms, using a Hanning window size of 30 ms, and sampled at 16 kHz. A logarithmic scale is used for the abscissa because of cochlear properties.

Modified spectral enhancement

The spectral enhancement algorithm of Lyzenga *et al.* (2002) is a modification of the algorithm of Baer *et al.* (1993) just described (SE_L). The algorithm expands the spectral peaks for frequencies between 200 and 5000 Hz. In that region slow spectral fluctuations are slightly compressed, intermediate fluctuations expanded, and fast fluctuations remain intact to avoid statistical variations of noise from being expanded. The spectral enhancement was realized using a power function. Enhancement factor *M* represents the exponent; spectral amplitude *a* of the signal represents the base of the power function. In our experiments *M* varied between 0.25 and 2.5 in steps of 0.25 dB/dB. For *M*=1 there is no spectral enhancement.

3.2.3 Phonemic compression

Phonemic compression is assumed to equalize level differences between successive phonemes. As a consequence, temporal masking of soft consonants preceding and following loud vowels will be reduced. To compress quickly, we used short attack and release times. In our studies we used two types of phonemic compression: (i) A single-channel phonemic compressor PhC_S developed by Verschuure *et al.* (1994), (ii) a multi-channel phonemic compression system PhC_M described by Van Harten-De Bruijn *et al.* (1997). The phonemic compressors contain a filter bank and a compression table. The filter bank that has been constructed contains FIR-filters. These have a phase characteristic that is assumed to be nearly linear.

Single-channel phonemic compression

The single-channel phonemic compression system is based on the hypothesis that the ears of hearing-impaired individuals who have a cochlear high-frequency hearing loss,

---

[19] The widths and centre frequencies of the auditory filters are related by the function $ERB = 6.23f^2 + 93.39f + 28.52$ designed earlier by Moore and Glasberg, 1983. Baer *et al.* (1993) performed informal listening experiments to determine specific multiplication factors for these functions, performed some scaling, and specified the gain function based on the enhancement function mentioned.

[20] The convolution process was preceded by a pre-emphasis on the enhancement filters, because the spectrum drops off at a rate of 6 dB/octave. In other words, the spectrum of the input signal was pre-whitened, so the peaks would tend to have equal weight across the spectrum.

mainly contain two different zones: a low frequency zone in which linear amplification can be applied, and a high frequency zone where a phonemic compressor is required. Therefore, the single-channel compressor contains only one FIR-filter with the shape of the mirrored audiogram of the hearing impaired individuals. An envelope detector controls the compressor. The result of this system is that for low input levels the response will look like the mirrored audiogram, whereas for high input levels the response will be almost flat. For intermediate levels the responses look more like flattened versions of the mirrored audiogram.

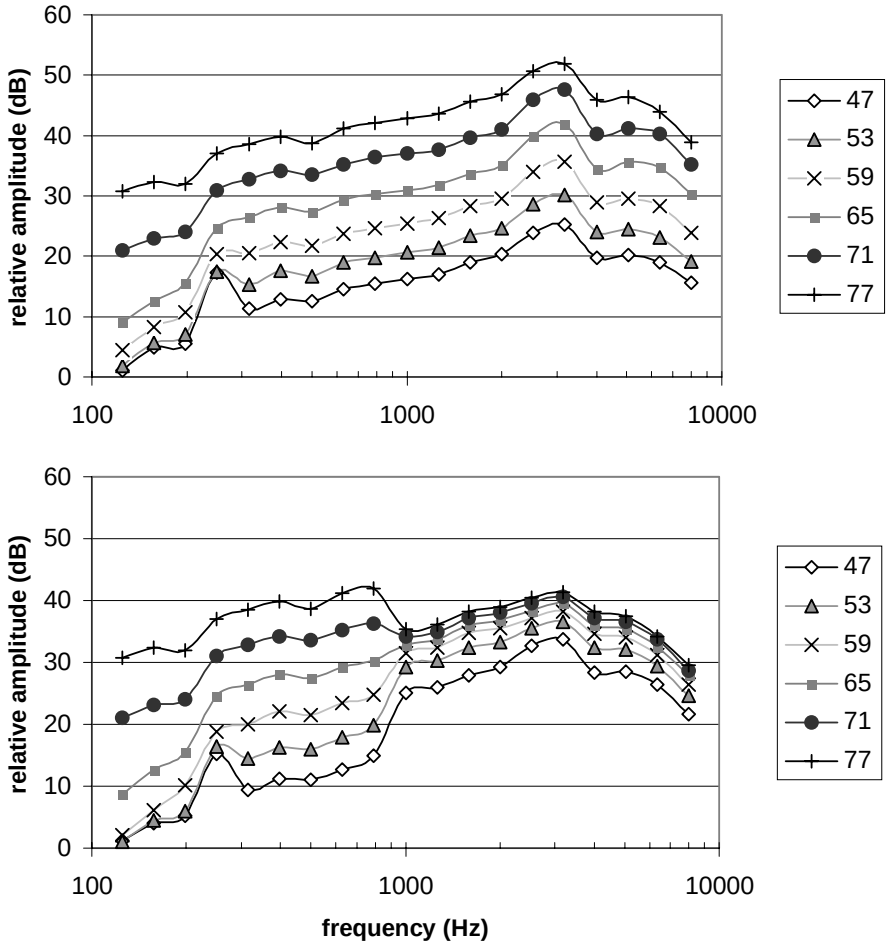


Figure 3.3 Output level measurements for the phonemic compressor as a function of frequency. Parameter is the input level of speech. The compression threshold of the compressor is 50 dB. The compression ratio is either 1.0 (upper panel) or 4.5 (lower panel). Frequencies higher than 861 Hz will be compressed.

The single-channel compressor reacts to the level of the broadband signal, whereas only high frequency components (above 891 Hz) are actually compressed. The compression knee-point is fixed at a certain sound pressure level. Signals with levels below this threshold are not compressed. Above this threshold, signals will be compressed by a factor corresponding to one of the compression ratios (CR) used in our studies.

We measured the effectiveness of the phonemic compressor. This is presented in figure 3.3. Two-minutes speech samples were recorded for several input levels from 47 to 77 dB SPL in 6 dB steps. Output levels were calculated in third-octave bands. In this example we used a knee-point of 50 dB SPL and CR of 4.5. The point at which a compression curve and a linear curve intersect is set at 65 dB SPL. This means that the speech segments below 65 dB SPL are amplified, and the segments that are above 65 dB SPL are attenuated relative to linear amplification. Effectively, this means that the level of the louder vowels and the softer consonants are indeed equalized. Note that we calculated the average speech spectra. Dynamically, the reduction in dynamic range is less dramatic (see Verschuure *et al.*, 1996). The single-channel phonemic compression system uses an attack time τ_a =5 ms and a release time τ_r =15 ms. The time constants were chosen high enough to compress modulations below 20 Hz effectively, because these are considered to contain important speech information (Goedegebure, 1999).

Multi-channel phonemic compression

The assumed advantage of the multi-channel phonemic compression system is the fact that the parameters in each filter band can be set to the needs of the hearing impaired person in that particular frequency region. The spectral energy within channels determines the compressive activity for each channel. This ought to be advantageous especially for hearing-impaired subjects with dynamic ranges that are highly frequency-dependent (Verschuure *et al.*, 1993). In the multi-channel set-up an envelope detector and a non-linearity table are embedded in each of the eight frequency bands with fixed centre frequencies around 543, 900, 1273, 1604, 2021, 2546, 3208, and 4041 Hz. These bands are 1/3 octave wide, except for the lowest frequency band. For computational reasons this band is 1 octave wide. The envelope detector together with the non-linearity table takes care of the multiplication factor of that specific frequency region in order to determine the level of each channel. Frequency bands need additional amplification because the dynamic ranges (DR) and most comfortable loudness levels (MCL) for every hearing impaired subject depend on frequency. Therefore, we determined the difference between the MCL level for each band and the MCL level for 500 Hz, for each subject.

Contrary to the single-channel compressor, the eight-channel compressor does not use a knee-point. All input-levels below -62 dB RMS were removed, for all frequency bands. To avoid saturation of the DSP the output level was limited to -20 dB by use of a peak clipper.

For high input levels a limiter is implemented to avoid signals becoming uncomfortably loud. The knee-point of the limiter is set at a threshold above the peak levels of speech. Therefore, not much distortion of the speech signals is expected. Like the single-channel compressor, the time constants and compression ratio used for the multi-channel compressor are chosen to give optimum performance on both speech intelligibility and comfort (Verschuure *et al.*, 1993; Harten-De Bruijn *et al.*, 1997). For each channel, the compression ratio of the multi-channel compressor is set at a CR = 2. The compressor uses attack and release times of τ_a=0.5 ms and τ_r =10 ms in all

channels. To reduce overshoots that can be interpreted as plosives, a digital delay τ_d was applied with a delay time of. [21]

3.3 Algorithms of a state-of-the-art hearing-aid

One experiment was performed using three state-of-the-art hearing-aid algorithms in a commercially available hearing aid (Siemens Signia). These comprise modulation-based filtering, syllabic compression, and temporal contrast enhancement. Although the details of these strategies, developed by Siemens, remain confidential, some general properties can be formulated. This hearing aid is a multi-channel digital behind-the-ear hearing aid. The eight channels range between 0 and 8000 Hz and the inter-channel boundary frequencies are 300, 565, 800, 1130, 2000, 3200, and 5000 Hz, respectively. Parameter settings used are preset test hearing aid settings.

Modulation-based filtering is a noise reduction strategy (NR_S) that is based on amplitude modulations. Signals are assumed to be noise when temporal modulations are low in amplitude. Speech is known to be typically highly modulated in amplitude. The modulation-based filtering is performed in eight frequency channels, distributed four by four in two frequency zones. The amount of filtering could vary between preset values off, min, and max for the which in combination yield five settings for our experiments: off-off, off-min, min-min, min-max, and max-max.

The syllabic compressor (SC) has attack and release times that act more upon whole syllables, than on the constituting phonemes. Fast compression can be included in four frequency zones that each comprise the lowest two, the third, the fourth and fifth, and the highest three frequency channels, respectively. Parameter in our experiments was the number of zones that use syllabic compression, which varied between 0 and 4. In the reference case only slow compression is active in all compression zones. For the other settings fast compression was included successively, starting in the highest frequency zone. For each zone, the compression ratio and compression knee-point were fixed at 2 and 54 dB, respectively.

Temporal contrast enhancement is a kind of spectro-temporal enhancement (STE). The algorithm is assumed to compensate for the poor temporal and frequency resolution of listeners with cochlear hearing loss by enhancing the temporal and spectral contrasts. The algorithm can be activated in six out of eight frequency channels, excluding the lowest two channels. In each of these channels the amount of contrast can be zero (off position), medium (med), or maximum (max). To achieve five different algorithm settings, two zones of frequency channels were created, containing channels three to five, and six to eight, respectively. The amount of contrast for both zones of five settings used in our experiment was off-off, off-med, med-med, med-max, and max-max.

---

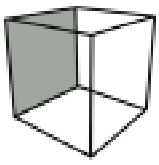
[21] Fast compression systems should suddenly react on fast level changes of transients. However, these higher and lower levels cannot be reached by the compressor that quick, so over- and undershoots are created, respectively. Overshoots are most easily perceived, and therefore should be reduced. The delay time used to suppress overshoots is τ_d =3 ms and τ_d =0.5 ms for the single-channel and multi-channel compressor, respectively. Verschuure and Dreschler (1996) reported on these technical considerations.

Chapter 4

EXPERIMENTS ON COMBINATIONS OF ALGORITHMS

'I was like a mountaineer who, not knowing his path, must climb slowly and laboriously, is forced to turn back frequently because his way is blocked but discovers, sometimes by deliberation and often by accident, new passages which lead him onward for a distance. Finally, when he reaches his goal, he finds to his embarrassment a royal road which would have permitted him easy access by vehicle if he had been clever enough to find the proper start.'

– Hermann von Helmholtz
On the sensations of tone as a physiological basis for the theory of music (1954/1887)



4.1 Spectral enhancement and phonemic compression

We investigated the separate and combined effects on speech perception of compensation of the reduced dynamic range by different forms of fast compression and compensation of the reduced frequency resolution by spectral enhancement. The study has been designed to compare the effects of signal processing on monosyllabic consonant-vowel-consonant words (CVC) for hearing-impaired listeners in conditions of quiet, fluctuating noise, and continuous noise. Eight subjects participated in this study.

4.1.1 Methods

Algorithms

The speech material (in the presence or absence of noise) was recorded on audio CD and was processed off-line by spectral enhancement in Cambridge by Baer, Stone and Moore (SE_B). For computational reasons the signals were low-pass filtered at 4 kHz and sampled at a 16 kHz rate. The remaining spectrum may be expected to be broad enough for speech perception. The enhancement gain *M* was chosen 2.

The signals were presented through an on-line compressor that was operating as a single-channel compressor (PhC_S) or as an 8-channel compressor (PhC_M). In case of the single-channel compressor (PhC_S) we used an attack time τ_a =5 ms, a release time τ_r =15 ms, a delay time τ_d =3 ms, compression ratio 2, and a knee-point 28 dB below the overall RMS value of the speech.[22] For input levels below this knee-point the action of the system is linear.

The multi-channel compressor (PhC_M) uses τ_a=0.5 ms, τ_r =10 ms, τ_d =0.5 ms, and a compression ratio of 2. The eight-channel compressor does not use a knee-point. All input levels below -62 dB RMS were removed for all frequency bands. To avoid saturation of the DSP the output level was limited to -20 dB by use of a peak clipper.

Experimental design

We selected four processing conditions from the combinations of spectral enhancement (Unprocessed and spectrally Enhanced) and the presence of a certain kind of compression Linear (no compression), Single channel compression, or Multi channel compression). This results in either a (ES) condition, in which the signal is spectrally enhanced and compressed in a single channel, or the (EM) condition achieved by spectral enhancement and a multi-channel compression algorithm. The reference condition (EL) for these two compression conditions is spectral enhancement without compression. The reference condition with respect to spectral enhancement is UL, which uses linear amplification and no spectral enhancement. We cannot compare one of the conditions including compression (ES and EM) with the unprocessed UL condition directly, because this would presuppose that we used two linear, independent processes (spectral enhancement and phonemic compression). The other possible conditions - US and UM - are not included in the test for reasons of comprehensiveness and time limitations. The effects of phonemic compressors alone have been studied extensively (e.g. Verschuure *et al.*, 1994).

[22] The maximum input level is was chosen 22 dB above the RMS level.

Two different kinds of tests were carried out: psychophysical tests with non-speech and with speech stimuli. The three tests with non-speech stimuli include pure-tone audiometry, loudness scaling, and measurement of auditory filters. These tests give information about the residual capacities of the ears of the individual subjects. The outcomes can be correlated with the individual results of the speech tests. Speech intelligibility was measured by nonsense consonant-vowel-consonant (CVC) words.

We used three acoustical conditions in which speech is presented: without noise (q), fluctuating noise (f), and continuous noise (c). In order to save time, the quiet condition (q) was the only one that was re-tested. The fluctuating noise was a male voice reading Dutch sentences, played backward. The continuous noise was a mixture of several speaking persons and was called restaurant noise, simulating that specific kind of background. The speech signals were presented at two signal-to-noise ratios relative to the noises: SNR= +4 dB and SNR= -1 dB. For the speech and both noise signals the average level (time constant = 1000 ms) was measured. In total 20 different conditions (5 acoustic conditions*4 processing schemes) were presented to the subjects.[23]

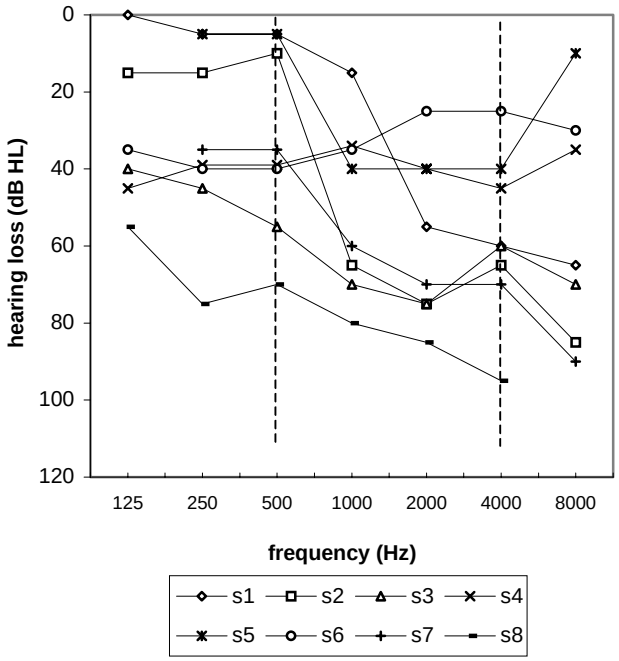


Figure 4.1 Hearing losses (dB HL) for eight subjects as a function of frequency (kHz). The subjects are indicated as s1 to s8.

The subjects listened to computer-generated stimuli through headphones. They used a keyboard to enter the CVC-words they heard. Because the spectral enhancement was done off-line the computer was only used for processing the compression algorithms. The computer output was converted to analogue, and fed into the audiometer. During this stage the desired level for each individual could be adjusted, so the subject could

[23] The order of the processing conditions and the order of the noise conditions for each day was randomized for each subject. For the processing conditions this was done by means of a Latin square.

hear the stimuli at the most comfortable level (MCL). It appeared necessary to implement a high-pass filter with a cut-off frequency of 0.3 kHz to avoid distortions.

4.1.2 Results

Pure-tone audiogram

Eight hearing-impaired subjects participated in this study. The subjects were 16 to 18 years old, all studying at the Alexander Graham Bell School in Amsterdam, a school for hearing impaired pupils. They were paid volunteers. With the exception of two subjects (s1 and s5) all subjects have linear hearing aids. Only one subject (s2) wears two hearing aids. For each subject the hearing loss (in dB) was measured at 0.125, 0.25, 0.5, 1.0, 2.0, 4.0, and 8.0 kHz with adequate masking of the contra-lateral ear. The hearing losses were predominantly cochlear. Only one subject (s6) had a mixed loss. All testing was done monaurally at the better ear. The better-ear audiograms of the eight participants are presented in figure 4.1. Most audiograms are sloping towards the higher frequencies, except for subjects s4 and s6. Subject s5 has a 'cookie-byte' audiogram between 1.0 and 4.0 kHz.

Psychophysical data

The Würzburger Hörfeld Skalierung method (see Section 2.8) measures loudness data in terms of Categorical Loudness Units (CLUs). From these data two important parameters can be extracted: the Dynamic Ranges (DR) and the Most Comfortable Levels (MCL) as a function of frequency. These parameters are given by $DR(f) = L_{CLU=40}(f) - L_{CLU=10}(f)$, and $MCL(f) = L_{CLU=25}(f)$.

	Parameter (dB)							
Subject	**DR.5**	**DR1**	**DR2**	**DR4**	**MCL.5**	**MCL1**	**MCL2**	**MCL4**
s1	26	33	13	20	97	103	109	109
s2	51	26	22	39	97	101	114	109
s3	36	26	27	27	109	113	113	107
s4	32	31	35	30	89	86	92	95
s5	47	34	37	30	67	84	82	76
s6	44	46	62	59	98	95	91	83
s7	21	23	21	28	99	108	115	102
s8	32	26	26	12	109	106	114	115
NH	55	54	59	58	72	71	75	69

Table 4.1 Dynamic Ranges (DR) and Most Comfortable Levels (MCL) for four different frequencies (0.5, 1, 2, and 4 kHz) for the subjects s1 to s8. These variables are derived from the loudness data (using the Würzburger Hörfeld Skalierung). DR values correspond to the difference between the subjective indications loud-to-very-loud and soft-to-very-soft. MCL values correspond to the indication medium loud. The normal hearing data are calculated from data proceeding from a study by Verschuure *et al.* (1993).

Table 4.1 presents for each subject the dynamic ranges and most comfortable levels for four different frequencies, 0.5, 1.0, 2.0 and 4.0 kHz. Reference data for normal hearing subjects measured with the same procedure are presented for purposes of comparison.

All subjects except s6 have a clearly smaller dynamic range, especially for higher frequencies. Next to this the levels at which the loudness of the speech is prescribed as most comfortable are shown for each subject and each frequency. The MCL levels are clearly higher than normal.

Using a Von Békésy notched-noise method for measuring the auditory filters centred at 0.5 and 3.0 kHz, 5 threshold values were obtained for each frequency (see Section 2.7). Three parameters were calculated and are presented for each subject in table 4.2. In the last row of the table we calculated the ERB, PL, and PU values for the average normal hearing person. A comparison between the data of the subjects and the reference data shows remarkable deviations, especially at 3 kHz (see ERB3, PU3, and PL3 for subjects 1,2 and 3). The shallower slopes of subject 3 reflect very broad filters (large ERB3 values). On the other hand, filter shapes proved to be close to normal at 500 Hz for s2 and s5 and at 3kHz for s6.

	Auditory filter parameter					
Subject	**ERB.5 (Hz)**	**PL.5**	**PU.5**	**ERB3 (Hz)**	**PL3**	**PU3**
s1	113	17	19	1035	6	50
s2	83	22	26	1038	49	7
s3	141	17	12	1884	3	50
s4	102	20	19	645	13	35
s5	74	21	37	984	13	12
s6	115	22	15	318	43	34
s7	96	20	23	591	25	17
s8	139	14	15	744	19	14
NH	79	25	25	348	34	34

Table 4.2 For all subjects three auditory filter parameters are given: the upper slope (PU), the lower slope (PL), and the rectangular equivalent bandwidth (ERB). Two centre frequencies are used to obtain a typically high frequency and low frequency filter: 0.5 and 3 kHz. The ERB values according to the average normal hearing subject are calculated by an equation suggested by Glasberg and Moore (1990): ERB = $24.7(4.37f_c+1)$. The filters are assumed to be symmetrical for normal hearing persons, so PL = PU and therefore PL = $4000f_c$/ERB. ERB is given in Hz, while f_c is given in kHz. PL and PU are without dimensions.

Speech perception data

A Word scores

A three-way analysis of variance was used for the three experimental variables: subjects (8), algorithms (4) and combinations of noise type and SNR (5). The results are presented in table 4.3. The main effects are all significant: $p<0.001$ for the parameter 'subject' and 'noise type', $p<0.01$ for 'algorithm'. The large variance between the noises is not surprising, because the noise type 'quiet' is very different from fluctuating noise and continuous noise. Post-hoc testing showed that the significant effect of 'algorithm' is due to the difference between the scores in the EM condition and the original condition UL ($p<0.05$). The parameters subject and noise type interact significantly ($p<0.01$). As indicated later, the scores of three subjects (s1, s3, and s8) are especially low in the fluctuating noise condition with the lowest signal-to-noise ratio.

To determine the significance of the differences, Wilcoxon tests (matched pairs signed ranks) were done. The average scores are highest in the quiet condition, for a signal-to-noise-ratio SNR = +4 the scores are moderate, and for SNR = -1 the scores are lowest. For equal signal-to-noise-ratios the scores in continuous and fluctuating noise are about equal. So it appears that the subjects hardly profit from the pauses in the fluctuating noise. In the original condition (UL) the scores are always best, although the differences are not always statistically significant. The largest differences appear between the EM condition and one of the other processing conditions. Condition EM scores significantly lower than the UL condition ($p<0.02$), and also lower than the EL condition ($p<0.02$). Except for continuous noise and an SNR of -1 dB, the scores for the EL condition and the ES condition are approximately equal.

Design	**Probability**
Subject	<0.001
Algorithm	<0.01
Noise type	<0.001
Subject x Algorithm	n.s.
Subject x Noise type	<0.01
Algorithm x Noise type	n.s.

Table 4.3 Three-dimensional analyses of variance, using subject (8), algorithm (4), and noise type (5) as parameters.

B Phoneme scores

The results can be analysed in more detail by looking at the scores of the different parts of the CVC words: the initial and final consonants and the vowel. These three phonemes are defined as c1, c2, and v, respectively. We expect confusions to be mainly determined by noise type, so the SNR will no longer be taken into account. Therefore, the two SNRs have been averaged both for the fluctuating and continuous noise conditions.

On top of the expected significance of the main effects, variance analyses of the different phonemes show more significant interactions than for the word scores. There is a weak effect of 'subject*algorithm' for c1. The significant interaction 'subject*noise-type' is found for all phonemes. An interaction between algorithms and noise types is only significant for the final consonants. In conclusion, these results suggest that the main trends are found for each phoneme, but some interaction effects may differ between phonemes.

In figure 4.2 we focus on the differences between continuous noise, fluctuating noise, and quiet, averaged across SNR (for the noises) and test-retest for each of the signal-processing conditions. The individual phoneme scores are presented in different panels of figure 4.2. In the original condition (UL) the c1, v, and c2 scores in quiet are very similar. In the conditions with background noise the identification of initial and final consonants is clearly more reduced than for vowels. This is especially true for the continuous noise condition ($p<0.01$ for both consonants). For fluctuating noise only initial consonant scores are significantly different from vowel scores ($p<0.05$).

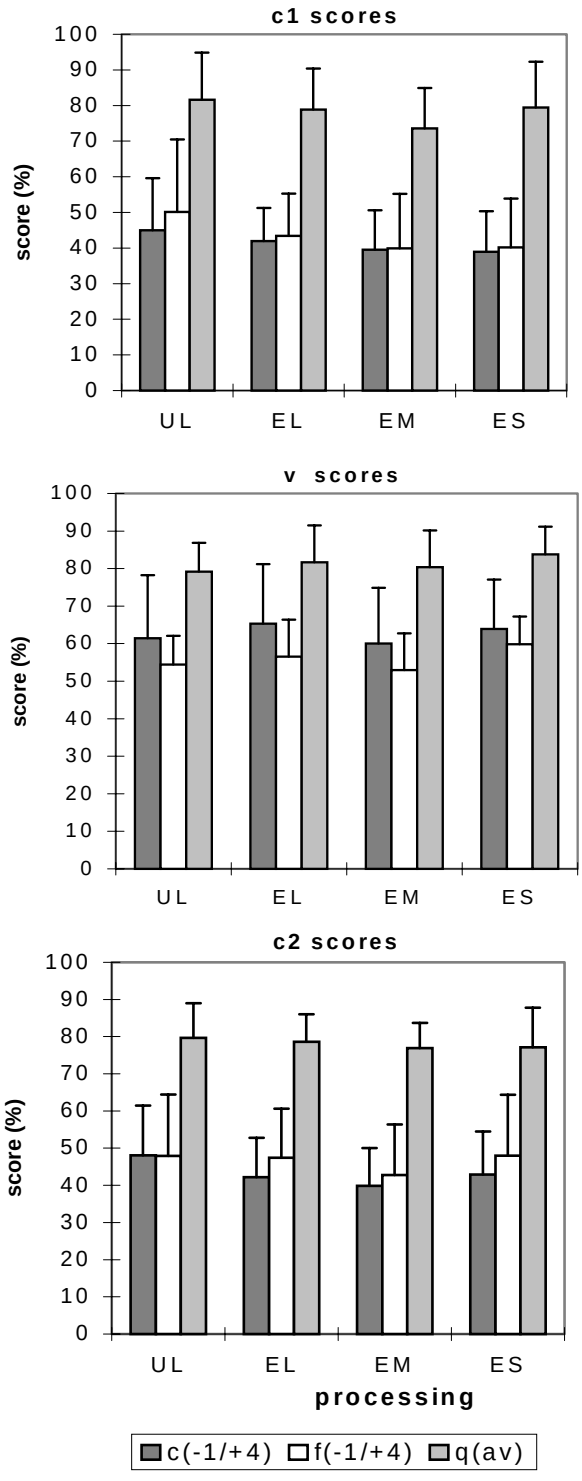


Figure 4.2 Three representations concerning the effects of noise type. The noise types are abbreviated: continuous noise (c), fluctuating noise (f), and quiet (q). The averaged SNR are indicated between brackets. On top the initial consonant scores (c1) are depicted, in the middle the vowel scores (v), and at the bottom the final consonant scores (c2).

In the quiet condition the consonant scores for UL tend to be higher than in other algorithms ($p<0.05$ for UL-EM in initial consonants). For vowels, however, there is a trend that higher scores are achieved in the conditions where spectral enhancement is used alone (EL), and where the spectral enhancement is combined with single-channel compression (ES), but these effects are not significant. These trends in the quiet condition are also present in continuous and fluctuating noise. Only for the final consonants there are significant differences in favour of the original condition (in fluctuating noise $p<0.05$ for UL-EM and EL-EM; in continuous noise $p<0.05$ for UL-EL and UL-EM).

For all three noise conditions there is a trend that the subjects can benefit from spectral enhancement only for vowels. For vowels we found that the scores for the spectrally enhanced condition alone (EL), and combined with the single-channel compression condition (ES) correspond closely. This might imply that single-channel compression could not add much to spectral enhancement. The combination gives no further improvement. On the other hand when the multi-channel compressor follows the spectral enhancement process the scores get lower relative to enhancement alone. The outcome is again the same as the original, implying that the improvements made by spectral enhancement are lost due to multi-channel compression.

For the vowels several significant differences in scores appear between conditions with continuous and fluctuating noise. Obviously, it is easier to perceive vowels in continuous noise than in fluctuating noise. For the perception of consonants an opposite trend is present in the data, especially for the c2 scores, but statistical significance was reached only for condition EL ($p<0.05$).

C Individual data on speech perception

In the group data some effects of spectral enhancement are present, but they are not very impressive. It is possible that the small effects found are due to averaging over the eight subjects. Individual data on speech perception show much larger differences. Subjects s3, s7 and s8 benefit from spectral enhancement in most conditions, but others do not. The extra benefits from single-channel compression could only be found for the easiest conditions c(+4), f(+4), and q(av) in subjects 1, 2, and 6. For multi-channel compression (EM) CVC scores were always worse than in the spectrally enhanced condition alone (EL), except for subject 1

Analysis of the results

We applied several methods of analysis to get a better picture of the advantages and disadvantages of using spectral enhancement and two types of dynamic compression. The outcomes can yield indications as to why new ways of signal processing are favourable for certain subjects and not for others.

A INDSCAL analysis

In order to get a qualitative impression of the perceptual mechanisms that cause the errors in speech intelligibility, we investigated the pattern of confusions for the different phonemes: c1, v, and c2. For that reason the confusion matrices were transformed into

symmetrical similarity matrices using the algorithm suggested by Houtgast in an article by Klein *et al.* (1970). Two- and three-dimensional INDSCAL analyses on the similarity matrices of the c1, v and c2 phoneme confusions were performed. The addition of a third dimension in the analyses resulted in all situations in a marginal increase in the variance accounted for. For that reason we present the results of the two-dimensional analyses. Three separate analyses per phoneme were carried out, each analysing a separate parameter: subject, noise condition, and processing type. We concentrated on the conditions with background noise. In a separate analysis it was found that the confusion patterns in the quiet condition are quite different from the confusion patterns in noise. For that reason, we did not combine quiet and noise conditions in the same analyses.

	Noise condition				Processing condition			
	c(-1)	c(+4)	f(-1)	f(+4)	UL	EL	EM	ES
c1 perception								
Plosiveness	0.63	0.68	0.73	0.56	0.77	0.62	0.65	0.57
Sibilance	0.49	0.62	0.44	0.49	0.38	0.59	0.62	0.60
v perception								
First formant	0.84	0.78	0.75	0.71	0.79	0.74	0.82	0.79
Second formant	0.37	0.51	0.46	0.62	0.51	0.54	0.45	0.50
c2 perception								
Sibilance	0.90	0.61	0.57	0.86	0.50	0.89	0.64	0.84
Voicing	0.24	0.69	0.69	0.35	0.80	0.32	0.68	0.37

Table 4.4 Results of a two-dimensional INDSCAL analysis on the c1 scores (a), the v scores (b), and the c2 scores (c). The dimensions correspond to specific acoustic characteristics. For each phoneme the weightings for each dimension are given as a function of the noise, and processing conditions.

B c1-perception

The accounted variance of the two dimensions of the c1 object space is relatively equally distributed between both dimensions: 43% and 31% respectively. The first dimension separates the plosives from the sonorants, the second dimension is mainly determined by the /s/ (sibilance). INDSCAL provides information about the perceptual weighting of these dimensions in different conditions (as shown in Table 4.4 for the noise conditions and for the processing conditions) and for the different subjects (to be used in the correlation analysis, see below). It is important to note that this information is related rather

to the relative perceptual importance than to the absolute perceptual importance. In most cases an increase in the perceptual weighting of dimension 1 will go along with a decrease in the weighting of dimension 2 and vice versa. Table 4.4a shows a high weight of plosiveness for the unprocessed condition and relatively better use of the sibilance information in conditions with spectral enhancement. The subjects that profited most from spectral enhancement (s3, s7, and s8) appear to be the subjects that rely on sibilance information the most.

C v-perception

In the INDSCAL-analysis of the vowel confusions, the first dimension corresponds to the frequency of the first formant, the second to the second formant. The variance accounted for in these two dimensions is 60% and 25%, respectively. The weighting factors of the formant information per condition are presented in Table 4.4b. The differences between conditions are only small, but some trends appear to be present. Higher signal-to-noise ratios yield relatively higher weightings of second-formant information, which can be due to improved audibility. In the individual data it is striking that subject 6, the only subject with a non-sloping hearing loss, also has high weighting of second-formant information.

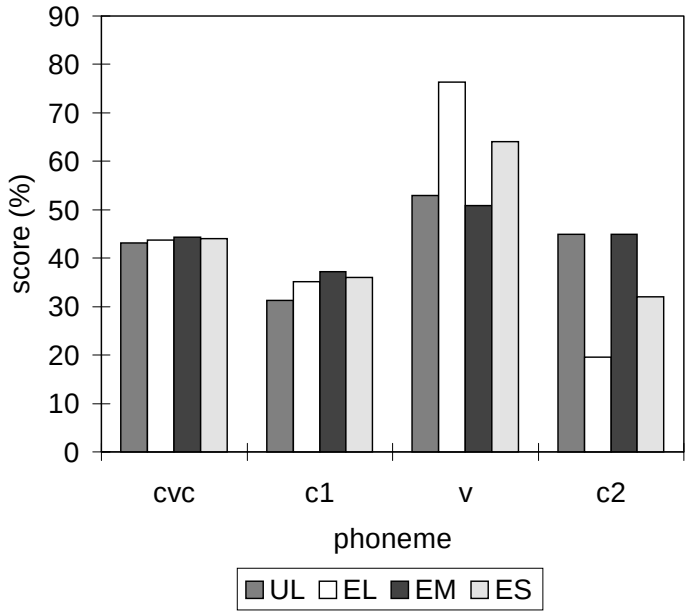


Figure 4.3 Scores of the phoneme parts of a CVC word for subject s7 measured in the c(-1) condition: the initial consonant (c1), the vowel (v), and the final consonant (c2). A trade-off effect between the vowel and the final consonant is clearly discernible for the spectrally enhanced condition alone (EL), and for the combination of spectral enhancement and single-channel compression condition (ES).

D c2-perception

The analysis of the final consonant confusions shows a distinction of the /s/ in the first dimension. In the second dimension the voiced and unvoiced consonants are separated (accounted variance is 84% and 11%, respectively). The weighting factors per condition

are given in Table 4.4c. The weighting factors appear to vary considerably for the different noise conditions in a way we cannot explain. However, for the processing conditions the sibilance dimension becomes even more important in the spectral enhancement conditions, although this effect is partly counteracted by the addition of multi-channel compression. It is questionable whether a further increase in sibilance weighting is beneficial for the phoneme scores of the final consonant, while the c2-perception is already dominated by sibilance. The subjects that profited the most from spectral enhancement (s3, s7, and s8) appear to be the subjects that are the least dependent on sibilance information in the final consonant, because they have relatively high weightings for voicing. This may be an explanation for the disappointing effect of spectral enhancement in final consonant perception (figure 4.3 illustrates this effect for s7).

E Analysis of correlations

We applied the Spearman rank-order correlation method to investigate the relations with psycho-acoustic parameters. These parameters encompass:

- Hearing Losses for four frequencies (HL.5, HL1, HL2, and HL4) and the average Hearing Loss of these frequencies (HLav)
- Most Comfortable Levels for four frequencies (MCL.5, MCL1, MCL2, MCL4) and the Dynamic Ranges for each frequency (DR.5, DR1, DR2, DR4)
- Slopes of the lower filter skirt for two frequencies (PL.5, PL3), and the Equivalent Rectangular Bandwidths for two frequencies (ERB.5, ERB3).

Note that slopes of the upper filter skirt (PH.5, PH3) are not included, because no strong correlations were found. The same holds for the threshold difference between 0.5 and 4 kHz, which is related to the slope of the hearing loss.

F Relation with the audiogram

There are only weak negative correlations between the speech scores and the mean audiometric loss. Please note that the speech was presented at the MCL level. From the analyses per frequency it can be concluded that the highest correlations are found for the higher frequencies. Detailed analyses on the phoneme level reveal that this can be attributed to the perception of consonants.

The results of the analysis of difference scores show that the effects of spectral enhancement (EL-UL) for the speech in fluctuating noise are significantly related to the average audiometric loss ($p<0.01$). There is no clear pattern of significant correlations for the additional effect of compression (EM-EL and ES-EL). For the combined effect of spectral enhancement and compression (EM-UL and ES-UL) the significant effects are again present for the fluctuating noise condition ($p< 0.05$ for hearing loss at 4 kHz). The significant correlations appear to stress the importance of the high frequency cues (sibilance) for large high frequency losses.

G Relation with the loudness scaling results

There are only few significant correlations with the MCL-values and they show roughly the same pattern as for the pure-tone thresholds: the highest correlations are found at 4kHz. For the difference scores the correlations are only mildly significant, without a

clearly interpretable pattern. There are strong correlations between the speech scores and the dynamic-range parameters at 2 kHz ($p<0.05$) and – even stronger – at 4 kHz ($p<0.01$). For UL, EL, and ES there are clear patterns of significant correlations, for the overall scores as well as for the separate phoneme scores (c1, v, and c2), indicating that high speech scores are found for subjects with a large dynamic range. Further analysis of the significant relationships shows a characteristic ceiling effect on the scores for dynamic ranges above 30 dB. The ceiling effects appear to be similar for both noise types and for both signal-to-noise ratios. For that reason we averaged all noise measurements per stimulus condition in order to compare the scores of UL, EL, EM, and ES. In figure 4.4 the relations between the scores and the magnitude of the dynamic range at 4 kHz show parallel patterns for the different processing conditions.

In the analysis of difference scores there are fewer significant correlations for the dynamic range data at higher frequencies, but there is a trend for negative correlations in continuous noise conditions and/or at low signal-to-noise ratios. These negative correlations appear to be related to the effects of spectral enhancement (El-UL) rather than to the additional effects of compression (EM-EL and ES-EL). This implies that the effects of enhancement are smaller for subjects with large dynamic ranges at 4 kHz for the most difficult noise conditions.

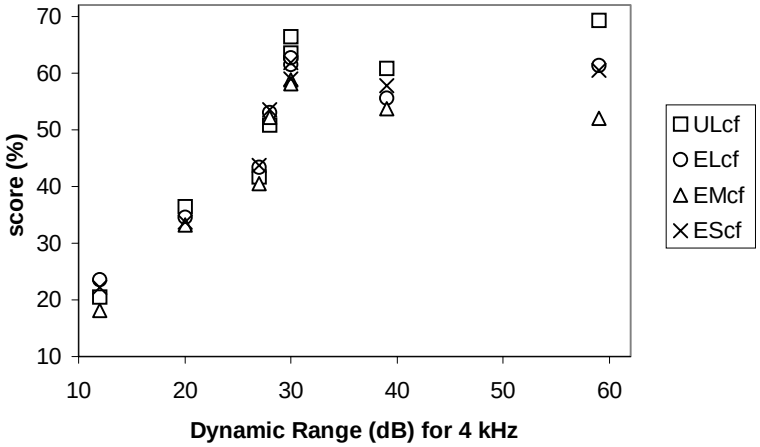


Figure 4.4 Speech scores as a function of the dynamic range at 4 kHz (in dB) for four different processing conditions, all averaged over the continuous(C) and fluctuating (F) noise conditions, hence the extension cf is used.

H Relation with the auditory-filter results

The correlations between the speech scores and the auditory-filter parameters (the filter width, ERB, and the steepness of the shallow filter skirt, PL) are only significant at 500 Hz ($p<0.05$). Narrow ERBs and steeper filter skirts go along with higher speech scores. This may be an indication for the involvement of 'upward spread of masking'. It is remarkable that the ERB-parameters at 500 Hz correlate mainly with the EM scores ($p<0.05$), while there are more significant correlations with the PL-parameters for UL, EL, and ES. The ERB-parameters appear not to be related to the effects of enhancement and the additional effects of compression, expressed in the difference

scores (EL-UL and EM-EL, ES-EL, respectively). The patterns of the speech scores in relation to PL values are more or less the same for all phoneme categories. There is a trend for negative correlations between PL at 500 Hz and the difference scores in continuous noise conditions and/or at low signal-to-noise ratios. Again, the negative correlations appear to be related to the effects of spectral enhancement (EL-UL) rather than to the additional effects of compression (ES-EL and EM-EL). This implies that the effects of enhancement (EL-UL) are smaller for subjects with steep filter slopes at 500 Hz for the most difficult noise conditions. Finally, the PL at 3 kHz appears to be significantly related ($p<0.05$) to the additional effects of compression in the single-channel-processing (ES-EL). Improvements in speech scores due to compression are higher for subjects with steep filter slopes at 3 kHz, especially in the relatively easy noise conditions.

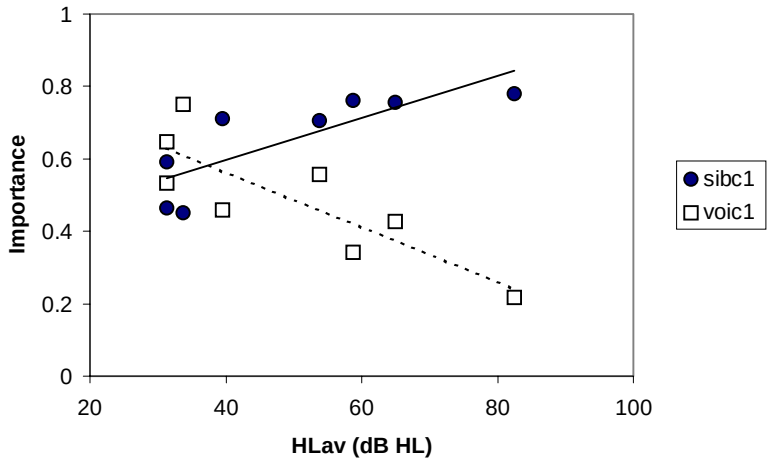


Figure 4.5 Importance of two different INDSCAL parameters, sibilance (sibc1) and voicing (voic1), as a function of the average hearing loss (HLav), for the initial consonants. The importance is expressed in a value ranging between zero and 1. The more important the INDSCAL parameter is for a certain subject, the closer the value is to the number 1.

I Relation with the INDSCAL-weightings

Correlations between the average hearing loss on the one hand and the INDSCAL parameters for plosiveness (Spearman's $r_s = -0.7904$) and sibilance ($r_s = 0.8743$) of the initial consonant on the other hand give opposite regression lines. The patterns can be seen in figure 4.5. Obviously, subjects with higher losses (s3, s7, and s8) tend to make use of sibilance more than of voicing/plosiveness in the perception of initial consonants. Note that these subjects benefit most from spectral enhancement. The opposite is true for average hearing losses lower than 40 dB.

In the correlations with vowel-scores it appears that the higher the PL-value at 3 kHz, the higher the weighing of the second formant.

With respect to the perception of final consonants correlations between the most comfortable level at 1.0 kHz on the one hand and the INDSCAL parameters for sibilance ($r_s = -0.8095$) and voicing ($r_s = 0.8571$) on the other hand give opposite regression lines. The patterns are shown in figure 4.6. Obviously subjects with lower

MCL1 levels tend to rely more on sibilance information than on voicing. For MCL1 sound pressure levels above approximately of about 110 dB (SPL) the importance of these two sources of information is about the same.

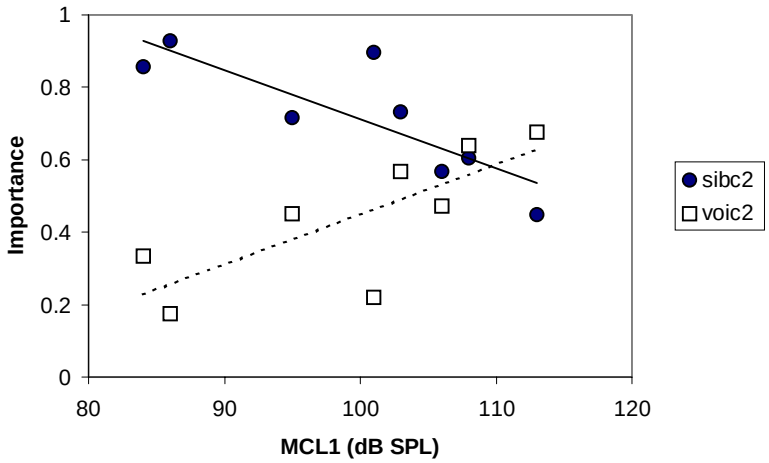


Figure 4.6 Importance of two different INDSCAL parameters, sibilance (sibc2) and voicing (voic2), as a function of the most comfortable level at 1000 Hz (MCL1), for the final consonants. The importance value ranges between zero and 1.

4.1.3 Discussion

In general, we see that the spectral enhancement especially improves vowel perception. It is disappointing that, in spite of better vowel scores, only few subjects can benefit from spectral enhancement for the CVC words. Obviously, the positive effect on the vowels is counteracted by the negative effect on the consonants. This trade-off effect might be the reason why almost no spectrally enhanced condition shows any improvement of the overall CVC scores.

The finding that spectral enhancement is particularly beneficial for vowel perception is not a real surprise because the key to perceive these phonemes is to distinguish the first three formants. Since this is a task that uses mainly information from the spectral domain, enhancement can easily lead to a better vowel recognition. But, apparently, information on consonants is lost for most subjects by using spectral enhancement. A possible reason can be the low-pass filtering at 4 kHz in the spectral enhancement process that was done off-line in Cambridge: parts of bursts of noise for voiceless stops and the random noise patterns in the higher frequency regions of fricatives get lost. From another point of view, improvement of consonant perception should possibly be sought primarily in the time domain, because temporal cues are important for the discrimination of consonants.

Improvements due to spectral enhancement show significant correlations with the measured dynamic range at 1 kHz; spectral enhancement gives some improvement only for subjects with small dynamic ranges at 1 kHz. Subject 2 is the only exception to the rule, maybe due to a completely different auditory filter at 3 kHz. It is the only subject with an auditory filter, which has a steep lower skirt and a flat higher skirt.

Considering the single-channel compressor, a beneficial effect was found for subjects s1, s2, and s6, but in the easiest conditions only (the noise conditions with an SNR = +4 and q(av)). In general the single-channel compressor does not show any negative effects, but does not add intelligibility to the spectrally enhanced signal. We assume that, contrary to multi-channel compression, the single-channel compressor does not change the short-term spectra of the speech signals. Given the fact that in an earlier study (Goedegebure *et al.*, 1998) the results of single-channel compression were more promising than the additional effects of phonemic compression shown in this study, it may be assumed that the single-channel compressor works best when it is not combined with spectral enhancement.

The absence of beneficial effects due to phonemic compression can possibly be explained by the specific choices made for the attack and release times. We have chosen those compressor parameters that came out best in pilot studies. In combination with spectral enhancement these choices are not self-evident. It is also possible that stronger effects of phonemic compression can be obtained if the compression takes place *before* spectral enhancement is applied. Hohmann (1983) underlined that specific characteristics in either the frequency or the time domain of the speech signal could not be changed without influencing the characteristics in the other domain. As long as the effects of certain kinds of processing are not well analysed both in the time and frequency domain it is hard to predict the results. The question is, to what extent can we compress and/or spectrally enhance the speech signal with a minimum of detrimental effects.

4.1.4 Conclusions

- Spectral enhancement alone gave no overall benefit in speech intelligibility, due to a trade off between improved vowel scores and deteriorated consonant scores
- Application of phonemic compression after spectral enhancement yielded equal or lower speech intelligibility
- Relations found between plosiveness/voicing and sibilance on the one hand and average hearing loss and most comfortable loudness level at 1 kHz on the other imply that hearing-impaired subjects with higher hearing losses seem to use different listening strategies to understand speech
- Subjects that profited most from spectral enhancement had smallest dynamic ranges around 1 kHz and relied least on sibilance information in the final consonant

4.2 Binaural noise reduction, phonemic compression, and spectral enhancement

In this section we describe the effects of three algorithms and their three-dimensional interactions on speech intelligibility and listening comfort. Two studies were performed. One in which the methodological and audiological implications of a multi-directional pattern search strategy were studied, the other to investigate effects of combinations of three algorithms used by means of a round-robin procedure. Background noise was attenuated by means of binaural noise reduction (BNR). Specific speech segments were

spectrally enhanced and equalised in intensity level by means of spectral enhancement (SE_B), and fast phonemic compression (PhC_S), respectively. Sentences were presented in six conditions, defined by three binaurally recorded noise types at two signal-to-noise ratios. Six hearing-impaired subjects (4 male and 2 female) with cochlear hearing losses and six normal hearing subjects (3 male and 3 female) participated in this study. All hearing-impaired subjects were paid volunteers and they were recruited from the Graham Bell School in Amsterdam. The average age of the normal hearing group was 25 and for the hearing-impaired group 18 years.

4.2.1 Methods

Algorithms

A binaural noise-reduction algorithm (BNR) was implemented to improve the signal-to-noise ratio. Parameters used to decide which signals will be attenuated and which will be unchanged are related to attenuation factors Level Difference Weighting (LDW) and Phase Difference Weighting (PhDW). The LDW and PhDW parameter values run parallel from 1.000 (no reduction), 0.775, 0.550, 0.225, to 0.100 (maximum reduction).

A single-channel phonemic compressor (PhC_s) was incorporated to reduce the masking effect of less intense phonemes by more intense phonemes. Parameters Compression Ratio (CR) and Release Time (t_r) have been co-varied in 'one' dimension. Parameter values were 1.0/not applicable (no compression), 1.5/80ms, 2.5/60ms, 3.5/40ms, and 4.5/20ms (maximum compression).

The spectral enhancement algorithm SE_B was implemented to enhance spectral peaks and valleys. The enhancement values range between 0.000 (no enhancement), 0.001, 0.010, 0.100, and 1.000 (maximum enhancement).

Stimulus material

For the multi-directional pattern search and round-robin tournament, the subjects were asked to compare differently processed sentences on aspects of speech intelligibility or listening comfort, in separately conducted tests. Both procedures are tested in six different acoustic conditions, i.e. three noise types, each at two signal-to-noise ratios.

A subset of 783 sentences, uttered by a Dutch female speaker, was used for the multi-directional pattern search and round-robin tournament (sentences VU98, Versfeld *et al.,* 2000).[24] The background noises feature continuous noise and fluctuating noise. We used uncorrelated real-life samples of car noise (car, 77 dBA), cocktail party noise (coc, 69 dBA), and electrical switchboards in a crane-cabin (cat, 73 dBA). The frequency characteristics of two noises (coc and cat) and the sentences (speech) are dominated by higher frequencies, mainly due to the use of Cosmea microphones that have high-pass

[24] To prevent subjects from losing their attention, a series of different sentences was used during the fitting procedure. To keep the acoustic conditions of the sentences the same, both the speaker and the background noise should be fixed in space. This requirement is very difficult to fulfill in realistic situations. Moreover, the background noise should be the same for each sentence to avoid mismatches in synchronicity. Therefore, we have chosen to record the speech and noise material separately. The speech material was recorded binaurally in a sound-attenuating booth, presented frontally (0 degrees) by a free-field speaker (Westra, LAB 1001) at a distance of 1 meter (L_{eq}=65 dBA). The noise material was recorded real-life, near to the sound source(s).

characteristics. To determine the filtering characteristics of the Cosmea microphones we recorded pink noise for five different presentation levels: 60 to 100 dB SPL. For intermediate levels, i.e. 80 dB SPL, the Cosmea microphones act to some extent high-pass with a slope of 7.5 dB/octave between 0.5 kHz and 4 kHz. The signal properties in the modulation spectrum show that the sentences and the crane-cabin noise (cat) both have a clearly modulating character, whereas the car noise and cocktail party noise are much more continuous.

The stimulus material was recorded and pre-processed offline. The off-line processing makes full automation of the pattern search possible and allows for fast and flexible switching. To simulate a realistic condition, binaural recordings have been derived from the microphones of in-the-ear hearing instruments (Siemens Cosmea), worn by the first author. We recorded the speech and noises on DAT tape, using a portable Sony PCM-M1 DAT recorder. The Cosmea devices were fixed into the ears with custom-made shells of soft (rubber-like) material. The microphone signals were directly fed into the DAT recorder. The speech-in-noise samples used in both procedures are 4 seconds long. Speech parts vary in length between 1.5 and 2.5 seconds and are preceded by a fixed period of 0.5 seconds of silence. Four-second noise samples were added to the speech samples in two different signal-to-noise ratios: 0 and 5 dB. In order to create these signal-to-noise ratios, the RMS-value of a speech sample without silence periods was compared with the RMS-value of a specific noise sample. After summing noise and speech in the appropriate SNR, the signals were converted into analogue signals (TDT-II system), and a DASi[25] system processed the signals by a binaural noise reduction and phonemic compression algorithm. The analogue output of the DASi unit was converted to digital again for further processing. To avoid distortion in the signals we applied an anti-aliasing filter. The speech-in-noise samples were sampled at a 16 kHz-rate to meet the requirements of the spectral enhancement processing (SE_B), which was the last processing step. The spectral enhancement algorithm was running on an IRIX workstation. Finally, the signals were scaled to equal RMS in order to prevent that the level would become an (unintended) evaluation criterion.

For the testing procedures the pre-processed speech and noise material was stored on a personal computer. The stereophonic output was attenuated by a Tucker Davis programmable attenuator (PA4) and processed subsequently via the DASi unit to compensate for the difference between recording and playback equipment. [26] Audiometric hearing losses have been corrected by use of the NAL formula (Byrne and Dillon, 1986).

The output of the DASi was transmitted to an AC5 audiometer for extra amplification. Subjects listened to the speech-in-noise samples via TDH-39P

[25] The DASi unit is a flexible experimental hearing aid platform developed by Rass and Steeger (2000).

[26] In our set-up the system delivered a signal by headphones to the ear canals, which had a spectrum very close to the realistic situation, i.e. when the in-the-ear hearing aid was actually worn. The mismatch between recording via hearing aids and playback via headphones has been compensated for by conversion rules that are described by Dillon (1997). The increase of 7.5 dB per octave for frequencies between 0.5 and 4 kHz for the in-the-ear Siemens Cosmea hearing aids has been compensated for as well.

headphones. Similar to the recording level of the sentences, the output was played back at a level of 65 dBA.

For the speech reception threshold measurements we used Plomp sentences (Plomp and Mimpen, 1979).

Experimental design

For both procedures the task of the subject was to indicate by paired comparisons which of the two sentences was more intelligible or sounded more comfortable (in separate experiments). The subjects were forced to make a choice between two differently processed realizations of the same sentence. However, by use of a repeat button the subjects could listen as often as they liked before they made their decision. For the paired-comparison procedure we used a slightly adapted version of the instruction of Purdy and Pavlovic (1992). A short trial period preceded the actual tests, in order to let the subject get accustomed to the task. During practice, another background noise was used (printing office noise). The test order has been randomised across subjects, evaluation criterion, SNR, and noise type.

Pattern search optimisation

The multi-directional pattern search was used as an optimisation method that accounts for complex algorithm configurations and their interactions. The step size was fixed during the search for the optimal combination of three experimental signal-processing algorithms. In this three-dimensional algorithm space 125 settings were included. That is, the dimensions each consist of five sub-settings, based on different weightings of one or more algorithm parameters. We have tested the pattern search in the six acoustical conditions for two evaluation criteria: the subjective evaluation of listening comfort (pleasantness) and the subjective evaluation of speech intelligibility. Due to the use of two different starting points (BNR_0,PhC_0,SE_0) and (BNR_4,PhC_4,SE_4), and a retest for of the starting point (BNR_0,PhC_0,SE_0), in total 36 pattern searches were available. Only normal hearing subjects performed all tests. Hearing-impaired subjects performed the pattern search for starting point (BNR_0,PhC_0,SE_0) in test and retest only.

Round-robin tournament

Also, we applied a modified round-robin tournament to study more details about the response distributions for each of the dimensions. For this procedure, the subjects performed the tests for different subsets of all conditions. The normal hearing subjects used low SNR only for the speech intelligibility criterion. Consequently, the normal hearing subjects performed the tournament for nine conditions: six conditions for the listening comfort and three conditions for the speech intelligibility criterion. The hearing-impaired subjects used the listening comfort criterion only for all six acoustic conditions. It seems most straightforward to compare all 125 different points. That would take a vast amount of time. Therefore, due to its many comparisons the round-robin procedure has a need for restriction of possible candidate settings. Therefore, round-robin measurements have been performed for the twelve edges that constitute the frame of the pattern search cube. The reason to choose the edges is that we wanted to

study algorithm interactions. By comparing extreme settings that lie on the edges it might be easier to show algorithm interactions. These measurements can be grouped into three sub-groups belonging to the three algorithms. Each sub-group consists of four parallel edges. So, each sub-group of four edges represents the variation of the parameter of one algorithm, whereas the other two algorithms are fixed at one of their extreme values. This parameter is called 'common edge'.

Better ear	**Hearing Loss f(kHz)**						**Speech reception threshold (dB SNR)**			
							Stat		**Fluct**	
Subject	**0.25**	**0.5**	**1.0**	**2.0**	**4.0**	**8.0**	**m**	**f**	**m**	**f**
s1	25	30	30	45	55	55	1.9	-1.3	-2.1	1.9
s2	35	40	65	70	70	95	3.3	1.3	0.7	2.5
s3	20	25	30	35	40	65	5.5	5.0	5.3	2.9
s4	35	40	40	50	45	40	-1.6	-1.8	-0.9	-2.6
s5	45	50	65	70	70	65	-1.0	-2.6	-1.6	-2.0
s6	5	5	15	60	60	65	4.8	3.7	3.4	3.9
HIavg	28	32	41	54	57	66	2.2	0.7	0.8	1.1
(stdev)	(13)	(14)	(14)	(13)	(11)	(14)	(3.0)	(3.1)	(3.0)	(2.7)
NHavg	8	7	2	4	9	16	-4.9	-5.1	-9.0	-10.1
(stdev)	(5)	(7)	(2)	(4)	(5)	(10)	(0.9)	(0.5)	(1.0)	(1.1)

Table 4.5 Hearing losses (dB HL) and speech reception thresholds (dB) of the better ear for six hearing-impaired subjects that are indicated as s1 to s6. Hearing losses are presented for the better ear for stimulus frequencies ranging from 0.25 to 8 kHz. SRT values are given for four different conditions, determined by noise type (fluctuating and stationary noise) and gender (male and female). Also average values and standard deviations (given in brackets) are given for the hearing-impaired (HI) and the normal hearing group (NH).

4.2.2 Results

Pure-tone audiogram and speech reception threshold

For each subject the hearing loss for both ears was measured at 0.5, 1.0, 2.0, 4.0, and 8.0 kHz, with adequate masking of the contra-lateral ear. Data of the better ear are presented in table 4.5. The hearing-impaired subjects have mild-to-moderate hearing losses that are rather symmetrical (differences between the ears are smaller than 15 dB). However, slight differences are compensated for during playback of the stimulus material. Only subject 2 and subject 5 actually wear their hearing aids.

Apart from pure-tone audiometric data, we also measured the critical signal-to-noise ratio for a sentence intelligibility of 50% for the better ear. Speech reception thresholds (dB) were measured for four conditions: male (m) and female speech (f) in stationary noise (stat) and two-band fluctuating noise (fluct). Both noises are spectrally matched to the target speaker. Average values for normal hearing subjects closely correspond to the results from Festen and Plomp (1990).

Pattern search optimisation

A Stop criterion

In our study it appeared that the pattern search only occasionally (14%) stopped on the first criterion, i.e. when the same point was frequented three times. This might be due to the low probability of passing the same point in a three-dimensional set-up. So, the second criterion, the number of wins, seems to be the most adequate stop criterion, at

least for a three-dimensional algorithm space containing 125 settings and a fixed step size.

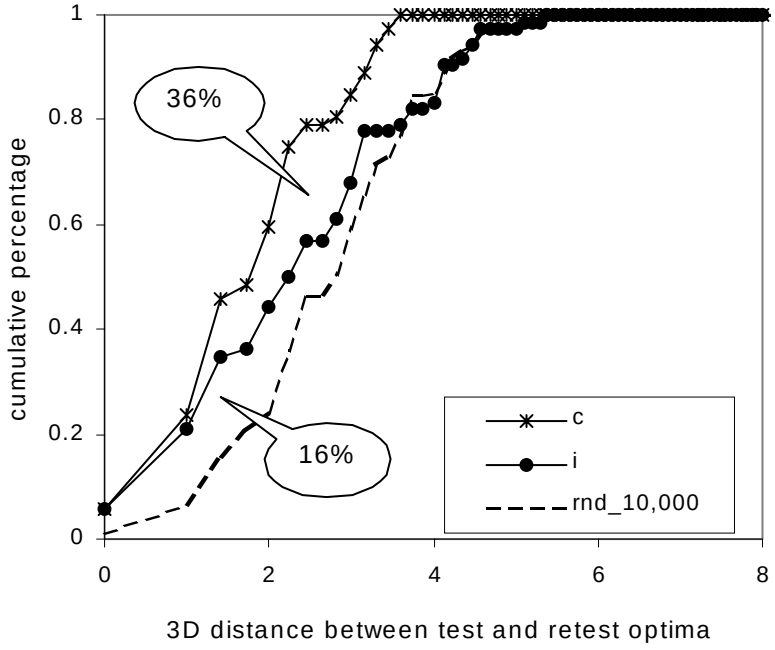


Figure 4.7 Cumulative distributions for the calculated three-dimensional distances between the optimal settings found for test and retest. The distributions are given for the comfort criterion (c) and for the intelligibility criterion (i). As a reference we also calculated the distribution when test and retest optima are randomly assigned in a procedure with 10,000 points (rnd_10,000). Numbers in call-outs indicate differences between the comfort and the reference distribution, and the intelligibility and the reference distribution, respectively.

B Reliability

The reliability of the multi-directional pattern search we used in this study appeared to be rather poor. This is shown in figure 4.7. Here the cumulative distributions are given for the three-dimensional distances between the test and the retest "optima". This measure is the distance calculated between the end points (assumed optima) for test and retest.[27] Random performance was determined by calculating the probability of the 3D-distance between randomly chosen end points of test and retest (n=10,000). From this probability function we can derive the cumulative distribution, which is given in figure 4.7 as a reference indicated by the dashed line. If we adopt 3D-distances smaller than $\sqrt{3}$ as an acceptable test-retest deviation, 49%, 36% and 23% of the data have 3D-distances below this criterion for comfort, speech intelligibility, and random evaluation respectively. To investigate the effects of acoustic condition, i.e. SNR and noise type, on reliability we also analysed different subsets of the data. In figure 4.8 the mean differences between test and retest have been calculated and they are shown for either noise type (upper panel) or SNR (lower panel). The differences are tested on significance

[27] Note that the distances are square roots of ascending integers. Also note that some distances cannot be a possible outcome: distances of square roots of 7, 14, 22, 27, etc. are not possible due to the discrete step sizes in the pattern search.

by use of the Wilcoxon matched pairs signed rank test. It appears that the differences between noise types are not significant for the comfort criterion, but for the intelligibility criterion test-retest differences are significantly smaller for crane cabin noise (cat) relative to car noise (Wilcoxon, $p<0.001$) and relative to restaurant noise (Wilcoxon, $p<0.05$). Apparently, the reliability of speech intelligibility judgments depends stronger on the noise condition than comfort judgements. For car noise, it also appears that evaluation on speech intelligibility is much less reliable than on listening comfort (Wilcoxon, $p<0.01$). Differences in SNR only affect the reliability of comfort judgments: for low SNR the results are significantly more reliable (Wilcoxon, $p<0.05$).

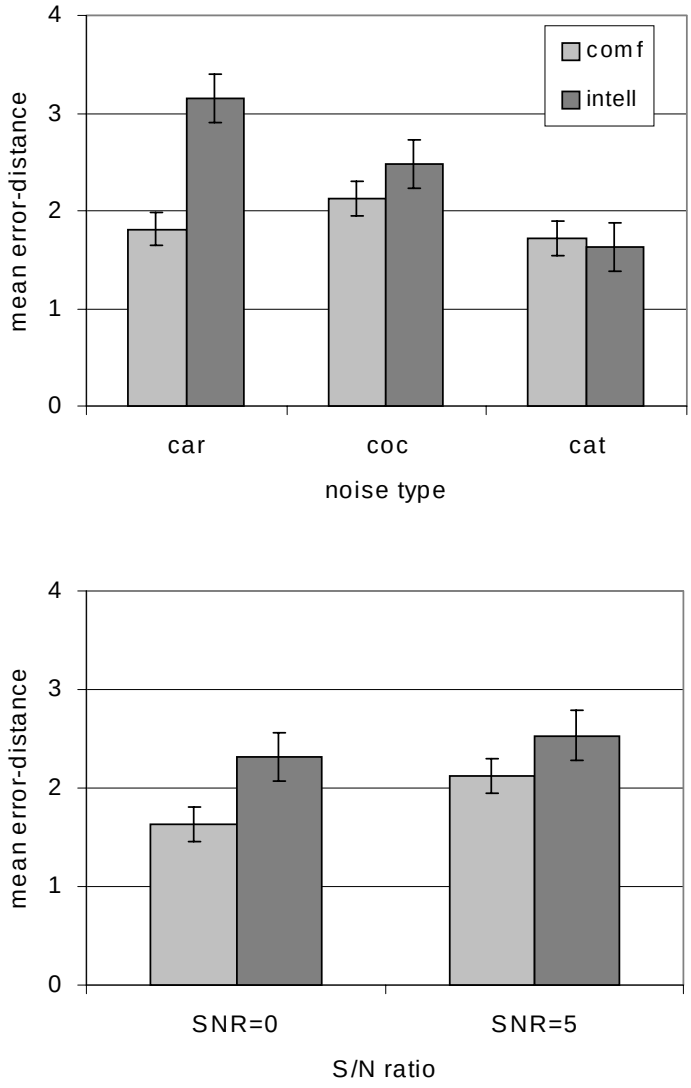


Figure 4.8 Mean difference (mean error distance) between test and retest as a function of either noise type (upper panel) or SNR (lower panel). Parameter is the evaluation criterion: speech intelligibility (intell) and listening comfort (comf).

Apart from the 3D-distance, which is a measure that encompasses all three dimensions, we can also look at the individual algorithms. In figure 4.9 we present cumulative distributions of the test-retest distances for separate dimensions.

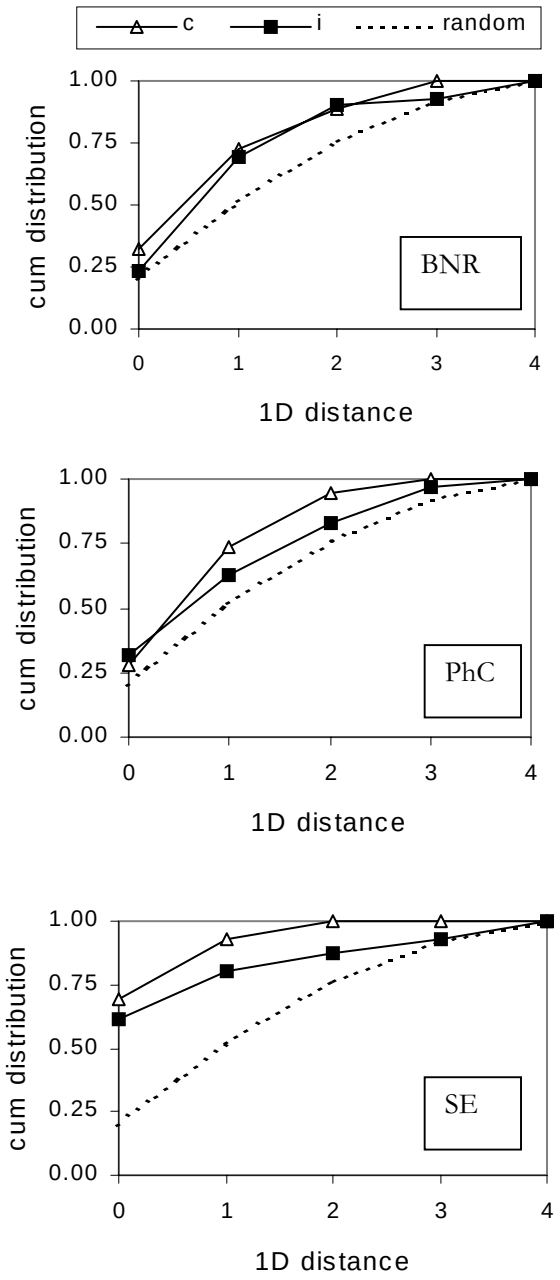


Figure 4.9 Cumulative distributions of test-retest differences given for each dimension as a function of the distance between test and retest optima. The individual dimensions represent Binaural Noise Reduction (BNR), Phonemic Compression (PhC_S), and Spectral Enhancement (SE_B). Parameter is the evaluation criterion: speech intelligibility (dotted line), and listening comfort (solid line), respectively. As a reference the dashed line represents the distribution retrieved from randomly assigned test and retest optima.

The (uni-dimensional) distances can vary between 0 and 4 for each dimension, due to the 5 parameter-settings. Distributions are given for the speech intelligibility criterion (dotted

lines), for the comfort criterion (solid lines), and for randomly assigned test and retest data (dashed lines). From top to bottom the results for binaural Noise reduction, Phonemic Compression, and Spectral enhancement are presented. Clearly, for spectral enhancement the test-retest reliability is highest. For the other two algorithms the reliability is clearly lower, though for the comfort criterion still approximately 75% of the data has distances smaller than 1. Again it appears that evaluation on comfort is more reliable for all three dimensions.

C Consistency

For the pattern search, variation of the initial setting influences the final estimate for all three dimensions considerably. For starting point (BNR_0,PhC-S_0,SE-B_0) on average the optimal setting ($BNR_{1.4}$,PhC-$S_{1.1}$,SE-$B_{0.4}$) is found. Starting point (BNR_4,PhC-S_4,SE-B_4) on average yielded a quite different optimal setting: ($BNR_{2.2}$,PhC-$S_{2.2}$,SE-$B_{2.0}$). This might indicate that the algorithms have an unfavourable response distribution. That is to say, a high consistency is obtained only for response distributions that are hill-shaped, containing only one clear peak. Apparently, the perceptual space contains either suboptimal settings or the perceptual differences between settings in some areas are rather small.

D Optimal settings

Pattern search outcomes have been analysed by analyses of variance. Outcomes are divided into three groups by differentiating between the three algorithms, i.e. the three dimensions of the pattern-search space. Six parameters are included: test, evaluation criterion, hearing, subject, noise type, and SNR. Because starting point (BNR_4,PhC-S_4,SE-B_4) was used only incidentally by hearing-impaired subjects test and retest data with starting point (BNR_0,PhC-S_0,SE-B_0) have been included only.

For *binaural noise reduction* there are no main effects and no interaction effects. It seemed that subjects like some noise reduction. On average, parameters LDW (Level Difference Weighing) and PhDW (Phase Difference Weighing) have a value 0.65 for optimal effects.

For *phonemic compression* there is a significant main effect between test and retest data (Wilcoxon, $p<0.05$). The retest optimum ends at more phonemic compression than the test optimum. Therefore, the significance of the other effects has been based on a separate analysis of the retest data only. There are no significant main effects, but significant interaction effects were found between evaluation criterion and noise type and between evaluation criterion and subject (nested in parameter hearing). For the speech intelligibility criterion the same optimal setting is found for all noise types. However, for listening comfort the optimal setting for phonemic compression depends on the type of background noise. That is, for the two fluctuating noise conditions (restaurant and crane-cabin noise) subjects prefer more phonemic compression than in continuous noise (car noise). This effect is significant for crane-cabin noise relative to car noise (Wilcoxon, $p<0.05$). The interaction effect between evaluation criterion and subject is difficult to interpret. It appears that opposite trends are present: subjects may

have similar, higher, or lower optimal settings for listening comfort with respect to speech intelligibility.

For *spectral enhancement* significant main effects were found for differences in hearing (Wilcoxon, $p<0.001$), for different individuals, and for different noise types (Wilcoxon, $p<0.01$ for car-coc, and car-cat). It is striking that there are no differences for different SNRs. On average the hearing-impaired group did prefer some spectral enhancement, although not much. Two hearing-impaired subjects prefer enhancement values between 0.001 and 0.01, whereas the other four subjects find optima for enhancement values between 0 and 0.001. Most normal hearing subjects did not like any spectral enhancement. It appeared that in general subjects like some spectral enhancement for car noise, but no enhancement for restaurant and crane-cabin noise. Significant interaction effects appeared between evaluation criterion and either subject or noise type. Parameter subject (nested in parameter hearing) interacted with noise type and SNR. Differences in optimal settings for different noises only appeared for judgment on speech intelligibility. Because trends seem to be different for the normal hearing and hearing-impaired group we also applied separate analyses of variance for each of the two sub-groups. These analyses might be indicative for trends that imply the need for individual fitting. Test and retest data have been pooled. For both groups main effects appear for parameter 'subject' and 'noise'. However, only for the hearing-impaired group we found significant interaction effects. These comprise significant interactions between evaluation criterion, subject, and SNR. More importantly, interaction appeared between parameter 'subject' on the one hand and parameter 'noise', 'SNR', or a combination of 'noise' and 'SNR' on the other hand. This strongly suggests the need for an individual fitting procedure.

Finally, for spectral enhancement no significant correlations (Pearson product correlation coefficient) were found with age, gender and audiometric data, i.e. hearing losses and speech reception thresholds in noise.

Round-robin tournament

A Consistency

The consistency of the round-robin procedure was determined, by calculating the number of circular triads. David (1963), who described the method of paired comparisons in general and in relation to the round-robin tournament in particular, illustrated problems that can occur and affect the reliability negatively. One of these problems is the occurrence of circular triads, for which $x1>x2$, $x2>x3$, and $x3>x1$. The consistency parameter ζ is derived from the number of circular triads c and the number of alternatives, which is 5 in our study. When $c=0$, $\zeta=1$, which indicates complete consistency. When $c=5$, $\zeta=0$, which indicates complete inconsistency. We calculated ζ for all round-robin tournaments and determined the percentage of outcomes that exceed $\zeta=0.6$. Based on these outcomes the difference in performance of the round-robin tournament was derived for different evaluation criteria, subject groups, and acoustic conditions. The results suggest that for the listening comfort criterion the performance for the normal hearing and hearing-impaired subjects was about equal: 65% of the data exceeded $\zeta=0.6$. For the different dimensions (algorithms that were varied) the

consistency was lowest for PhC_S (52%), intermediate for BNR (63%), and highest for SE_B (80%). For the normal hearing group there appeared to be no difference in consistency between the evaluation criteria. However, there was a trade-off between PhC_S and SE_B. That is, the consistency is higher for PhC_S but lower for SE_B when evaluated on listening comfort relative to the speech intelligibility criterion where the consistency for PhC_S and SE_B was equal. The noise type and SNR did not influence the consistency.

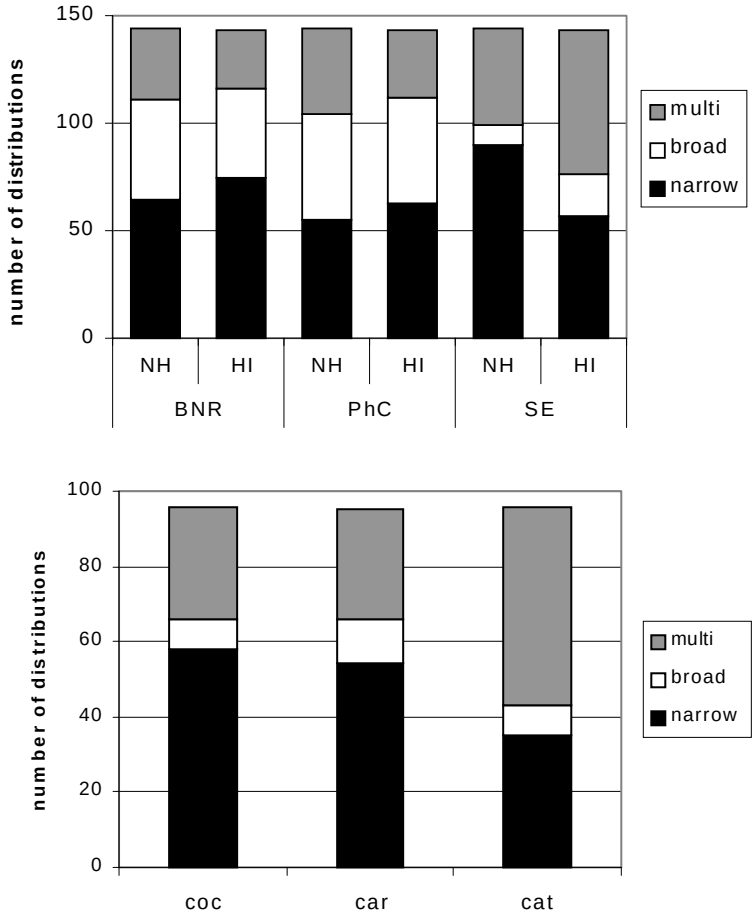


Figure 4.10 *Upper panel:* Number of response distributions for algorithms Binaural Noise Reduction (BNR), Phonemic Compression (PhC_S), and Spectral Enhancement (SE_B) and for Normal Hearing (NH) and Hearing-Impaired subjects (HI). *Lower panel:* Number of response distributions as a function of noise type. The types of response distributions are classified by peak shape: narrow, broad, and multimodal (multi).

B Response distributions

We analysed the round-robin results in two different respects. Both analyses are based on the response distributions measured, i.e. the rank ordering of the five settings of each of the edges. First, we estimated the winner for each tournament, i.e. the winner of each response distribution. This measure is used in the next section 'audiological aspects' to analyse the audiological implications of the round-robin results. Secondly, we determined the shape of the response distribution. The shape of the response distribution was

analysed to study the presence of algorithm interactions and to further explain the relatively low reliability and consistency of the pattern search.

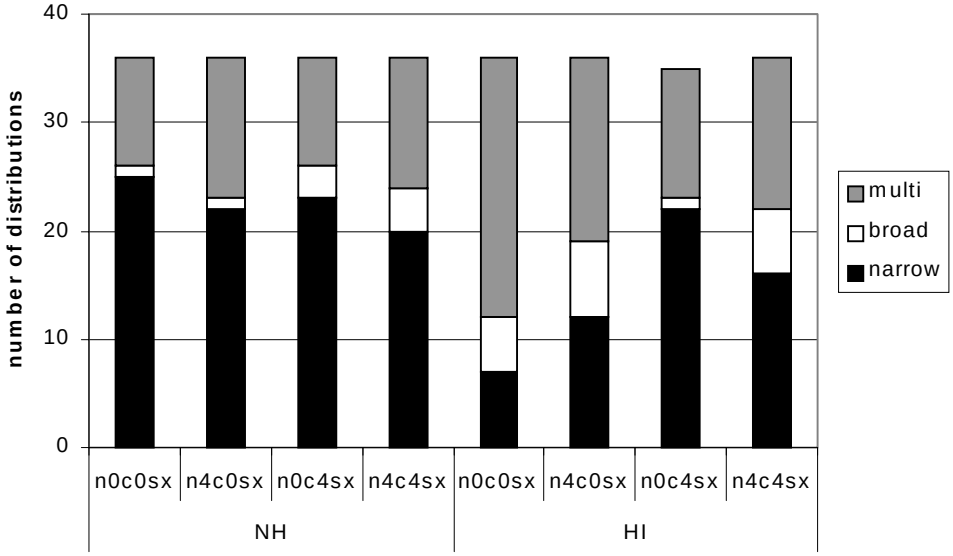


Figure 4.11 Number of response distributions for parameter 'common edge'. The setting of the fixed algorithms binaural noise reduction (n) and phonetic compression (c) is either minimum (0) or maximum (4). The variable algorithm spectral enhancement is denoted 'sx'. The results are given for normal hearing (NH) and hearing-impaired (HI) subjects. Response distributions are divided in distributions with a narrow peak, distributions with a broad (or flat) peak, and distributions with more than one peak (multi-modal). The number of distributions is calculated for the listening comfort criterion only. For condition n0c4sx there is one missing datum.

C Shape of response distribution

Use of the three-dimensional pattern search presupposes the presence of one optimum. To examine this prerequisite we analysed the round-robin data for each algorithm and we looked at the response distributions per subject and per algorithm. In order to identify differences between distribution patterns we classified these patterns after Kuk and Lau (1996a) into four categories: flat, broad, narrow, or multi-modal distribution. These patterns are defined having no peak, a very broad peak, a narrow peak, and more than one peak, respectively. Due to the fact that flat distributions hardly occur, these distributions were conceived as broad. So, three distribution types remain. We used the kurtosis of each distribution as an *a posteriori* measure to distinguish between the three distribution types. In appendix 4 the calculation method is discussed in more detail.

After classification we counted the number of distributions, differentiated for each algorithm and evaluation criterion. Roughly, the number of distributions does not depend on evaluation criterion. For listening comfort data only, we differentiated for algorithm, hearing, acoustic condition (noise type and SNR), and common edge. The number of narrow distributions depends on both hearing and algorithm (shown in figure 4.10). Most obviously, for spectral enhancement the difference between normal hearing and hearing-impaired subjects is largest. The number of multi-modal distributions for

hearing-impaired subjects is very high for SE_B. For binaural noise reduction and phonemic compression the number of broad distributions is considerably larger and, relative to spectral enhancement data, there are less multi-modal distributions. In the lower panel the number of distributions is given as a function of noise type. The number of narrow distributions is very low for the fluctuating crane-cabin (cat) noise. Further analysis showed that the number of narrow and multi-modal distributions depends on parameter 'common edge' for hearing-impaired subjects (see figure 4.11). That is, for different fixed combinations of binaural noise reduction (n) and phonemic compression (c), the number of narrow and multi-modal distributions for spectral enhancement varies considerably. This suggests that algorithm interactions are present.

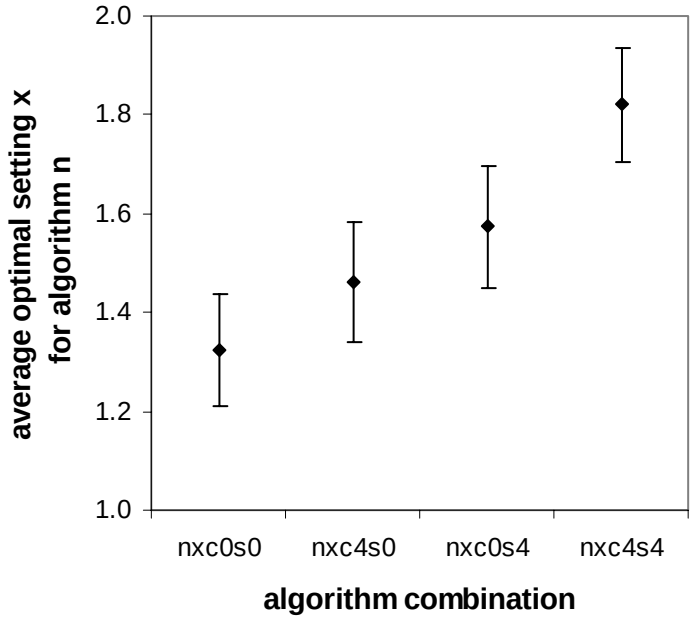


Figure 4.12 Average optimal binaural noise reduction for different combinations of algorithms (parameter 'common edge'). The phonetic compression and spectral enhancement algorithms are fixed at their minimum (0) or maximum value (4), and abbreviated by 'c' and 's', respectively. The binaural noise reduction algorithm is variable (x), and abbreviated by 'n'.

Audiological aspects

We analysed the estimated winners obtained by the round-robin procedure for different subgroups of subjects and different evaluation criteria, separately.

The round-robin tournament yields in fact too many data: for each axis the round-robin procedure gives a distribution of winners from 10 paired comparisons, because each axis consists of 5 points.[28] From the response distributions we like to extract single values that give the best estimate of the 'real' winner for each subject in a specific condition, using a certain evaluation criterion. The winner has been calculated as a kind of weighted average of the response distribution. Details of this estimation process are given in appendix 5. Note that we need the hypothesis that such a 'real' winner actually

[28] Each setting was compared with every other setting, such that only permutations are considered. That is, comparing setting A and setting B is assumed equal to comparing setting B with setting A.

exists. It is possible that this is not always true, e.g. when there are hardly any perceptual differences between the points, when the subject is not able to respond in a consistent manner, or when there is more than one optimal setting.

We performed an analysis of variance for normal hearing and hearing-impaired subjects, separately. Only the results for the listening comfort criterion were included, because hearing-impaired subjects did not evaluate speech intelligibility. The test parameter effects and algorithm interactions are described for subjects, for different acoustic conditions, and different (groups of) edges of the cube.

Most significant effects appear for the normal hearing population. For binaural noise reduction, there are significant effects for parameters noise type, SNR, and common edge. For normal-hearing subjects there is a significant difference (Wilcoxon, $p<0.01$) between car noise and crane-cabin noise: for car noise less noise reduction is preferred. Preferences of hearing-impaired subjects are independent of signal-to-noise ratio, whereas there is a significant difference (Wilcoxon, $p<0.01$) for normal-hearing subjects. That is, normal hearing subjects prefer more binaural noise reduction for a lower SNR. Different outcomes for different common edges are significant for hearing-impaired subjects only (see figure 4.12). When phonemic compression and spectral enhancement are maximal and noise reduction is variable (common edge nxc4s4) hearing-impaired subjects prefer more noise reduction than when one or both fixed parameters are put off (common edge nxc0s0, nxc4s0, or nxc0s4). Between common edge nxc0s0 and common edge nxc4s4 there is a significant difference (Wilcoxon, $p<0.01$).

For phonemic compression significant effects are present for normal hearing subjects in the interactions only. Interactions exist between subject and either noise or SNR.

For spectral enhancement there is a significant interaction effect between SNR and common edge in the hearing-impaired subgroup. It appears that for SNR=5 dB the winner for axis n4c4sx is on average significantly lower than for the other three axes, whereas for SNR=0 there is no difference between optima for different common edges. For normal hearing subjects and the listening comfort criterion there is a significant main effect for noise type and an interaction effect between subject and SNR.

We also performed an analysis of variance with evaluation criterion as main parameter, which is either speech intelligibility or listening comfort. Only results for SNR=0 are available for normal hearing subjects. No significant effect of evaluation criterion was found. That is, no significant difference occurred between preferences based on speech intelligibility and comfort, for any of the algorithms.

4.2.3 Discussion

Methodological aspects

We studied the performance, i.e. the reliability and consistency, of the multi-directional pattern search for combinations of binaural noise reduction, phonemic compression, and spectral enhancement. To determine the presence of suboptimal settings and interaction between algorithms we performed a round-robin tournament.

Analyses of the estimated winner and the types of response distributions found by the round-robin tournaments revealed considerable interactions between algorithms. For

hearing-impaired subjects, the estimated optimum for binaural noise reduction changes when the other two algorithms are varied (see figure 4.12). Response distributions of hearing-impaired subjects show considerable interactions between binaural noise reduction and phonemic compression (see figure 4.11). The analysis of distribution patterns reveals that the algorithms strongly suffer from either sub-optima or weak preferences.

The reliability of the multi-directional pattern search optimising combinations of binaural noise reduction, phonetic compression, and spectral enhancement is relatively low. Analysis of the results per algorithm shows poor test-retest reliability of the multi-directional pattern search for the phonemic compression dimension only. However, variation of the initial setting influences the final estimate for all three dimensions considerably. This might be due to unfavourable response distributions of the algorithms, or due to the fact that more than one optimal setting exist.

Most and largest test-retest deviations occur for the speech intelligibility criterion. But the evaluation criterion had no effect on the preferred setting. Therefore, listening comfort seems most suitable for optimisation of the fitting by means of a pattern search. It also appeared that we should use lower SNRs to evaluate both listening comfort and speech intelligibility. Moreover, for the speech intelligibility criterion, the pattern search appeared to be much more reliable for the fluctuating noises than for continuous noise. So, the stimulus material does not only influence the optimum, but also the reliability of the optimisation process.

Large deviations might partly be due to the fact that differences in speech intelligibility between samples were sometimes hardly detectable, even for normal hearing subjects. In addition, some hearing-impaired subjects reported that by hearing the samples twice, the second sample consistently appeared to be more intelligible. Therefore, they were biased to choose for the second sample. So, in future experiments the preference for the second stimulus should be prevented. We decided to show the sentence during the test on the computer screen in further testing in order to avoid poor reliability of the pattern search data.

Reliability and consistency possibly can be improved by means of a variable step size, the way the sentences are presented, and the application of preceding difference limen experiments. If reliability is improved a three-dimensional pattern search can become a powerful tool to take interactions between different (non-linear) signal processing algorithms into consideration.

Comparison of the pattern search data of this study with the round-robin data is only possible for the evaluation on listening comfort. Similarities between pattern search and round-robin results can be clearly found for all three dimensions. As a consequence, the large difference in test-retest reliability of the pattern search results can be explained by the consistency measure, based on the presence of circular triads. The subjects performed relatively poor for the PhC_S algorithm, intermediate for BNR, and good for SE_B. The consistency measure ζ calculated from the round-robin data can be used to predict the reliability of the pattern search. It should, however, be noted that the round-robin tournaments performed in this study only use a small subset of the pattern search

range of settings, i.e. the edges. Thus, the reliability of the pattern search can be predicted only for these edge-settings.

Audiological aspects

Analysis of variance shows hardly any differences between different conditions for *binaural noise reduction*. Moreover, it appeared that low values (high directionalities) are not preferred. Maybe this is due to the diffusiveness of the binaurally recorded noises: there are no isolated sound sources, typical for better performance in previous laboratory studies like Kollmeier *et al.* (1993) and Wittkop *et al.* (1997) found. It is possible that more directionality is preferred in configurations with isolated interfering noises, like one or two (separated) interfering talkers. Besides, gain from binaural configuration decreases rapidly with increasing reverberation, and the background noises were recorded in reverberant rooms. Although processing artefacts are reduced by time and frequency averaging of the weighing function, it is unavoidable that high directivity still introduces some artefacts as a result of rapid changes in the transfer characteristic of the adaptive filter. In combination with fast compression, these low-level artefacts can easily become audible. As a result, for this algorithm there is a clear trade-off between effective noise suppression in adverse listening situations and the speech quality. This could explain the overall choice of an intermediate noise suppression factor (LDW and PhDW), i.e. for different noise types and SNRs. Finally, SNRs of 0 and 5 dB are quite high for this algorithm to perform optimally, due to a ceiling effect (Wittkop *et al.,* 1997).

For *phonemic compression* there is a difference between test and retest, which might be due to acclimatization effects. For phonemic compression the retest optimum calls for systematically higher compression. At first, an explanation might be found in the time of acclimatization. From a study of Goedegebure *et al.*, (2001) it is known that subjects in general need time to get accustomed to compression. Moreover, interaction effects of parameters 'evaluation criterion', 'subject', 'SNR' and 'test' also suggest that this explanation is especially applicable for specific circumstances. In a relatively easy condition (SNR=5 dB), it is difficult to judge speech intelligibility. Both sentences can be understood relatively easily, so the subjects need time to get accustomed to higher compression ratios and shorter release times. On the other hand, it might be difficult to evaluate listening comfort for a low SNR. This might explain the higher values obtained in the retest. Previous studies showed that higher compression ratios (but not too high: CR<2 or 3) and lower release times might give slightly better speech intelligibility. However the opposite is true for comfort: long release times and small compression ratios show more comfort. Our results with fast high-frequency compression support results of Verschuure *et al.* (1998): especially in more fluctuating noises more compression is preferred. However, this interpretation should be treated with caution. Interaction effects show opposite trends: an increase in SNR yields preference for more compression for cocktail party noise only.

Hearing-impaired subjects prefer more *spectral enhancement* than normal hearing subjects do. For listening comfort no spectral enhancement was preferred. Our results suggest that speech intelligibility can be improved subjectively not only in quiet (Baer *et al.,* 1993) but also in stationary car noise. The fact that only very modest enhancement

values are preferred might be due to the counteracting effects of spectral enhancement. From the analyses of speech segments it appeared that for low enhancement values the algorithm acts primarily as a noise suppressor. For higher enhancement values the spectral peaks and values are enhanced considerably. However, these high values introduce artefacts that affect the listening comfort in a negative way. So, this might explain the difference in results for car noise for different evaluation criteria. It appears that especially in a stationary noise condition artefacts are very prominent, which act negatively on listening comfort. Apparently, preferences differ depending on noise type. Especially for hearing-impaired subjects preferences for spectral enhancement seem to depend on the particular condition, i.e. noise type and SNR.

4.2.4 Conclusions

- The results of the round-robin tournament showed that interaction was present between the three algorithms. This finding implies that optimisation of these three algorithms needs to be performed simultaneously, not successively
- The reliability and consistency of the multi-directional pattern search are relatively low and depended on the evaluation criterion. The reliability was highest for listening comfort and lowest for speech intelligibility
- The analysis of distribution patterns revealed that the algorithms strongly suffer either from sub-optimal or from weak preferences
- Phonemic compression is especially preferred for more fluctuating noises
- Hearing-impaired subjects preferred highest spectral-enhancement values to obtain better speech intelligibility in car noise

4.3 Binaural noise reduction, de-reverberation, and phonemic compression

In this study we investigated the additional effects of signal-processing algorithms for normal hearing and hearing-impaired subjects that work in noisy environments. This study was performed because for these subjects speech intelligibility is deteriorated due to the presence of loud background noise and reverberation. Two studies were performed. One in which the methodological implications of the pattern search optimisation strategy were studied for three algorithms, the other to investigate audiological effects of the optimal combination of the algorithms. The intended effects of the algorithms are as follows. Background noise was attenuated by means of binaural noise reduction (BNR). Reverberation was attenuated by means of a de-reverberation strategy (DRV). Specific speech segments were equalised in intensity level by means of phonemic compression (PhC_s). We investigated the effects of the three algorithms on speech intelligibility, by means of a sentence test (SRT) and word test (CVC). SRT and CVC measurements were performed in real-life noise recorded inside the cabin of a crane.[29] Nine subjects used three experimental hearing-aid algorithms[30] in combination

[29] Details of this study can be found in Koopman *et al.* (2000).

[30] These signal-processing strategies were incorporated in a flexible experimental hearing aid (DASi3-system; Rass and Steeger, 2000).

with an active hearing protector (Peltor Tactical 7)[31] for a period of six weeks. Eight subjects were accustomed to the use of hearing protectors; six subjects use custom-made in-the-ear protectors and two subjects use passive earmuffs. Subject s6 was not accustomed to the use of hearing protectors. The average age of the subjects was 42 years (s.d.=13).

4.3.1 Methods

Algorithms

The pattern search optimisation procedure comprised three algorithms that incorporated three settings for de-reverberation (DRV) and binaural noise reduction (BNR), and four settings for phonemic compression (PhC_S). Hence, there were 36 different settings. For the de-reverberation algorithm three coherence values were used comprising no processing (0%), intermediate de-reverberation (50%), and maximum de-reverberation (100%). For the binaural noise reduction algorithm two parameters were varied in parallel, i.e. the dimensionless factors Level Difference Weighting (LDW) and the Phase Difference Weighting (PhDW). The LDW and the PhDW factors varied between 0.1 (maximum reduction), 0.55 (medium reduction), and 1.0 (no reduction). Reduction is full on for angles larger than 33°. Reduction is gradually active between 19° and 33°. For angles smaller than 19° the signal is assumed to be the target, hence no reduction was applied. The single-channel phonemic compressor comprised parameters compression ratio (CR) and release time (t_r) combined in such a way that a broad range was created between intended high listening comfort and high speech intelligibility. Four settings were selected: no compression, CR = 1.5 & t_r = 80 ms, CR = 2.5 & t_r = 60 ms, and CR = 3.5 & t_r = 40 ms, respectively. The compression knee-point was chosen 50 dB SPL and the intersection point (input equals output) at 70 dB SPL.

Experimental design

The pattern search was conducted for speech in crane-cabin noise at a signal-to-noise ratio of 0 dB. Speech and crane-cabin noise were recorded by the earmuffs. Two pattern searches were performed, one for starting point (DRV_0,BNR_0,PhC-S_0), i.e. the reference setting and one for (DRV_2,BNR_2,PhC-S_3), which was the other extreme. The speech-in-noise material in the pattern search was evaluated on listening comfort.

Speech intelligibility was measured by a sentence test (SRT test, using sentences from VU98) and a word test (meaningful CVC words, Bosman, 1989). These tests yield a speech reception threshold (SRT) and a percentage correct score, respectively. The CVC-scores were measured for SNRs of -14, -7, 0, 7 dB. The speech tests were performed in real-life recorded background noise, i.e. inside a cabin of a crane. The noise was recorded with the microphones of the active hearing protector (Peltor Tactical 7). The frequency characteristic of the microphones was almost flat for frequencies between 0.2 and 3 kHz.

[31] The Peltor muff is used in its active mode only. The amount of amplification by the hearing protector is, in principle, adjustable. However, in order to make the results of the subjects comparable, the amplification was fixed and performed by the flexible hearing aid (DASi3). For input levels below 85A dB signals will be amplified by a maximum of 6 dB. The maximum output level of 120 dB SPL is achieved by a peak-clipper of the hearing protector and was implemented to avoid instantaneous loud sounds. The high-frequency hearing loss of subjects s6 and s8 was not compensated for.

The speech-in-noise material was processed by different combinations of the three algorithms, using an experimental hearing aid.[32] The stimuli were played back binaurally by means of the hearing protector via the soundcard of a laptop.

There were six different combinations of algorithms. Reference condition A represents no processing. Conditions B and C represent the winners of the pattern search optimisation procedure for the two different starting points. The other three conditions DRV-2, BNR-2, and PhC_S-3 were selected such that only one of the algorithms acted upon the signals. So, DRV, BNR, and PhC_S represent configurations (2,0,0), (0,2,0), and (0,0,3), respectively.

SRT tests were measured three times, i.e. before, halfway, and at the end of the six-week field trial. CVC tests were conducted only once for each condition during the six-week trial. Before the CVC tests the subjects performed a trial of three tests to get accustomed to the test using condition A, B, and C. After four weeks these conditions were actually tested.

Subject	*Optimum B*	*Optimum C*	*Optimum C**	*3D distance B-C*
s1	(2,1,1)	(1,1,1)	(2,2,3)	1
s2	(2,1,3)	(1,0,1)		√6
s3	(1,0,0)	(1,0,1)	(2,2,2)	1
s4	(1,1,3)	(1,2,3)	(2,2,1)	1
s5	(1,0,1)	(1,1,1)	(1,2,1)	1
s6	(1,1,2)	(1,1,1)	(1,1,0)	1
s7	(1,0,2)	(1,1,3)	(2,2,3)	√2
s8	(1,0,1)	(2,1,2)		√3
s9	(1,2,0)	(1,1,1)	(2,2,3)	√2

Table 4.6 Individual optimal settings B and C for starting point (DRV_0,BNR_0,PhC-S_0) and (DRV_2,BNR_2,PhC-S_3), respectively. The settings successively comprise de-reverberation, binaural noise reduction, and phonemic compression). The three-dimensional difference between these optima is given in the last column. When the three-dimensional distance between B and C is smaller than √3, an alternative setting C* was chosen, representing a sub-optimal setting of the optimisation procedure.

4.3.2 Results

Pure-tone audiogram

Two hearing-impaired subjects (s6 and s8) with cochlear hearing losses and seven normal hearing subjects participated in this study. The subjects were annoyed by loud noises, reverberation (subjects s4 and s8), and localisation and detection problems (subjects s1, s3, s5, s6, and s8) during work. The subjects participated on a voluntary basis. One welder (s6) worked at a company situated near Amsterdam. The other eight subjects worked at a ship unload company in Amsterdam and comprise two electricians (s2 and s8), and six subjects crane workers. Subjects s4 and s7 have binaural hearing

[32] The experimental hearing aid used, a DASi-3 device, was developed by Rass and Steeger (2000).

losses (BHL)[33] lower than 5 dB and s9 has a BHL between 5 and 10 dB. Subjects s1, s2, and s3 have BHLs between 10 and 15 dB. Subject s5, has a BHL between 20 and 25 dB. Subjects s6 and s8, have a BHL larger than 25 dB, and are defined hearing impaired, considering the WHO standard. These subjects have high-frequency hearing losses.

Optimisation data

The results of the pattern search strategy yield optima for different starting points that agree well. Table 4.6 shows the optima for each starting point. The three-dimensional distance between the optima found for different starting points is given in the last column for each subject (s1 to s9). In general condition B and C are represented by the winner for starting point (DRV_0,BNR_0,PhC-S_0) and (DRV_2,BNR_2,PhC-S_3), respectively. When, however, the winners are very similar, i.e. when the three-dimensional distance is smaller than or equal to $\sqrt{2}$, for the evaluation of the speech intelligibility (by means of an SRT and a CVC test) an alternative reasonable setting was chosen (C*). This second choice was that setting that appeared to be suboptimal.

The cumulative distribution of the three-dimensional distance between the optima found for the two starting points is given in figure 4.13. The three subjects with the largest hearing losses tend to have larger 3D-distances (average 3D-distance is 1.7) relative to the other subjects with no hearing loss (average 3D-distances is 1.1).

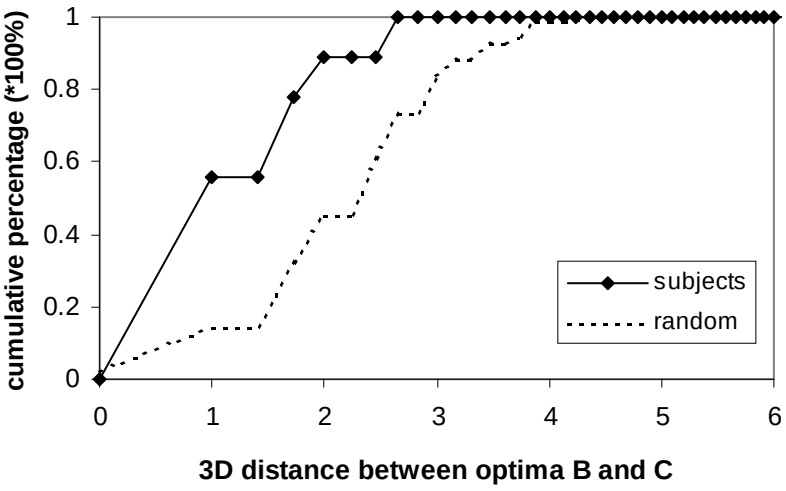


Figure 4.13 Cumulative distribution for the three-dimensional distance between optima found for starting point B=(DRV_0,BNR_0,PhC-S_0) and C=(DRV_2,BNR_2,PhC-S_3). As a reference, the random distribution is given, for which both optima were randomly assigned one thousand times.

From the cumulative distribution we also calculated the area between the cumulative distribution under study and the random distribution, relative to the 'perfect' distribution. This fraction was defined as the information transfer. By definition the information transfer of the random distribution is 0. In the latter case, When the

[33] The average binaural hearing loss (BHL) was calculated following the WHO standard, by weighing the hearing losses at the better ear and the worse ear by a factor of 5 and 1, respectively.

distance between test-and retest results is always 0, the information transfer is maximum, i.e. 1. The information transfer was calculated for the cumulative distribution both three-dimensionally, as well as for each dimension separately. The three-dimensional information transfer was 0.48, and for DRV, BNR, and PhC_S the information transfer was 0.50, 0.60, and 0.49, respectively.

Speech perception data

A Speech Reception Thresholds

No significant effect was found between the test results and two retests (Wilcoxon paired sign rank test).[34] Therefore we averaged the speech reception thresholds (SRT), given in figure 4.14. In general the lowest (best) speech reception thresholds (SRT) were found for the reference condition (A). The speech intelligibility for the reference condition (A) is significantly better relative to the optima B and $C^{(*)}$ (Wilcoxon, $p<0.01$). The SRTs for B and $C^{(*)}$ are not significantly different. Individually, the subjects do not always perform worse for configuration B and $C^{(*)}$. Within a 1-dB margin subjects s3, s4, s5, s7 perform equally well for configurations A and B, and subjects s1, s6, s7, s9 equally well for configurations A and $C^{(*)}$.

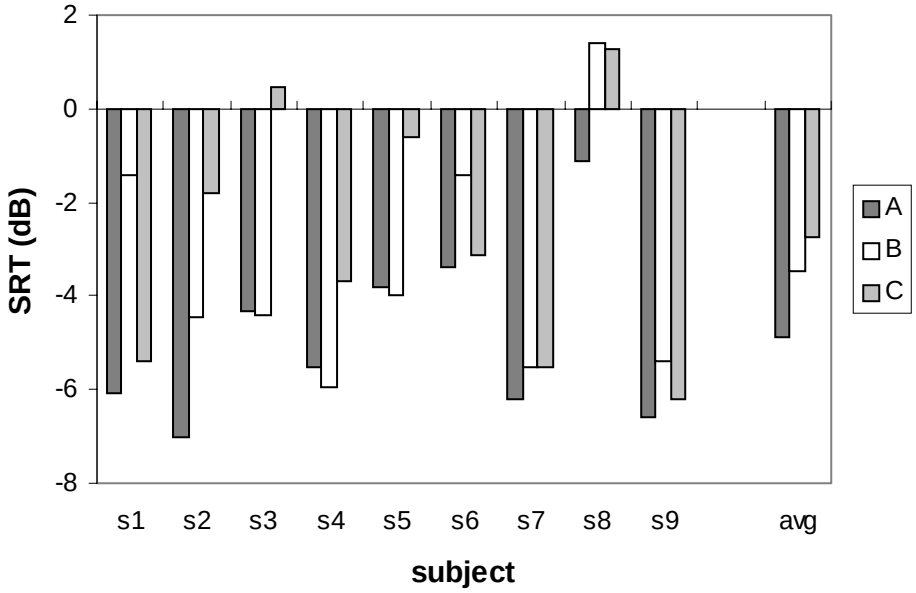


Figure 4.14 Individual speech reception thresholds (SRT) for the reference condition (A), and for the optima B and $C^{(*)}$ found for starting point (DRV_0,BNR_0,$PhC\text{-}S_0$) and (DRV_2,BNR_2,$PhC\text{-}S_3$), respectively. Also the average SRT (avg) for each condition A, B, and $C^{(*)}$ is given. The SRT values were averaged for three measurements.

B CVC tests

The CVC tests yield four types of results. The overall CVC scores and the amount of errors made for each individual phoneme (initial and final consonant c1 and c2, and vowel v). All subjects performed the CVC tests for SNR=0 and SNR=-7. Subjects s1, s4,

[34] Three SRT data that had a within-test standard deviation larger than 3 dB were discarded.

s5, and s9 also performed CVC tests for SNR=-14. The other five subjects s2, s3, s6, s7, and s8 performed the SRTs for SNR=7 in addition.

Several analyses of variance were conducted for the whole group, both subgroups, and groups based on the hearing loss. The main effects and interaction effects were determined for parameters subject, algorithm configuration, and SNR.

First the whole group was analysed, that is for SNR=0 and SNR=-7 dB only. CVC scores for these SNRs are on average 86% and 68%, respectively. For algorithm configurations A, B, and $C^{(*)}$ the percentage correct scores averaged for both SNRs and all subjects are 80%, 79%, and 73% respectively. The CVC results are significantly lower for algorithm configuration $C^{(*)}$ than for A ($p<0.01$, Wilcoxon). This result is mainly due to the significantly higher number of errors for the final consonant ($p<0.01$, Wilcoxon). There is also a significant interaction effect between the algorithm configuration and the SNR for the initial consonant. For SNR=0 dB the number of errors is not significantly different, whereas for SNR=-7 dB the number of errors is significantly higher for algorithm configuration $C^{(*)}$ relative to B ($p<0.05$, Wilcoxon). The analysis of different subgroups did not reveal any significant interaction effects. However, individually, some subjects did benefit from algorithm configurations found by the pattern search. For SNR=-7 subjects s6 and s8 have an average CVC score for the pattern search optima which is 14% higher than for the reference configuration. For SNR=0 the average CVC score for the pattern search optima of s4 and s9 is 11% higher than for the reference setting. These benefits cannot be related to a specific phoneme.

For algorithm configurations A, DRV, BNR, and PhC_S there is a main effect for the whole group. The CVC scores for BNR and PhC_s are significantly lower than for A ($p<0.01$, Wilcoxon). The CVC scores for A and DRV are highest and not significantly different. Subgroups reveal no interaction effects. Individually and relatively to A, DRV and BNR are sometimes advantageous. Subjects s8 and s9 have CVC scores for DRV that are 12% and 9% higher for SNR=-7 and SNR=0, respectively. The BNR scores for SNR=-7 are on average 12% higher relative to the A scores for s2, s4, and s8. The scores for phonemic compression (PhC_S) were always lower than for reference condition A.

4.3.3 Discussion

Pattern search optimisation

The optimisation method proved to be quite consistent. The differences between the optima found for two different starting points were much smaller than one would expect from random choice. That is, almost 60% of all three-dimensional distances between optima B and C were below 1. For each dimension separately the algorithms do behave quite similarly with respect to the consistency. For the signal-processing strategies specific parameter settings were chosen that ultimately determined the step size within the optimisation procedure. These choices appeared to result in a consistent behaviour of the patterns search. The three subjects with largest average hearing losses behaved relatively less consistent than subjects with small average hearing losses. This could be due to the fact that the settings were chosen by the normal hearing authors to be clearly distinguishable. The relative success of the pattern search is comparable to earlier studies (e.g. Neuman *et al.*, 1987; Kuk and Pape, 1992; Kuk and Lau, 1996) for which the

number of settings per dimension was comparably low. Reasons to include a small number of settings are on the one hand the auditory constraints of sound quality (introduction of artefacts for example) and on the other hand the perceptual distinctness between settings. However, a clear disadvantage of using only a few settings per dimension can be the lack of a real fine-tuning. By using a step size that is rather large the 'real' optimum might be missed, which could explain that most subjects have their optima very close to each other. Another reason could well be that the normal hearing subjects tend to like settings close to what they are accustomed to.

A cluster analysis (nearest neighbour, Euclidean distance) of the average optima for each individual yields three clusters when in each cluster the Euclidean distance between each average optimum and the cluster centre is smaller than 1. Seven subjects are clustered around setting ($DRV_{1.2}$,$BNR_{0.8}$,PhC-$S_{1.4}$). The other two clusters contain only one subject: (DRV_1,BNR_0,PhC-$S_{0.5}$) for s3 and (DRV_1,$BNR_{1.5}$,PhC-S_3) for s4. These results suggest that, except for two subjects, intermediate values for DRV and BNR were preferred, and some PhC_S. On average a de-reverberation factor of 50%, a noise reduction factor of 0.55, and a combination of compression ration of 1.5 and release time of 80 ms. So, the predominantly normal hearing subjects like some signal processing in the active hearing protector.

Speech intelligibility

In general, the SRT results showed that there were no improvements for the optimal algorithm configurations B and C (or C*) with respect to the reference condition. Also, no individual improvements were found for the optimal settings. This could be partly attributed to the subjects that are mainly normal hearing. They cannot benefit from digital signal-processing strategies that are meant at first for hearing-impaired subjects. However, we expected that at least the noise reduction and de-reverberation strategies could also be beneficial for normal hearing subjects, because these subjects work in noisy and highly reverberant environments. To test this assumption, we measured the SRT free field for the unoccluded ear and for the active muff without additional signal processing. The background noise used was standard continuous speech-shaped noise. The difference between the SRTs was on average only -1.3 dB, with lower (better) SRT values for the unoccluded condition. For subjects s3, s5, and s6 there was even no difference in performance at all.

Another reason might be found in the signal-to-noise ratio used. The CVC results were tested for four SNRs, ranging between -14 and 7 dB. In general, the CVC scores for algorithm configuration B did not differ significantly from reference configuration A. Only CVC scores for configuration $C^{(*)}$ were significantly poorer than reference configuration A. However, some subjects could benefit from one of the 'optimal' configurations for SNR=0 or SNR=-7 dB. The hearing-impaired subjects, s6 and s8, had higher CVC scores for SNR=-7, whereas normal hearing subjects s4 and s9 could benefit form the optimal configuration for SNR=0. Only for subject s4 this finding can be related to the fact that this subject has a clearly different optimal setting, given the information of the cluster analysis. A relation to a specific phoneme could not be found.

The influence of each of the algorithms was assessed by conducting CVC tests for each algorithm separately relative to reference setting A, i.e. DRV-2=(2,0,0), BNR-2=(0,2,0), and PhC_S-2=(0,0,3). The scores for configuration A and DRV-2 are highest and not significantly different. The CVC scores for BNR-2 and PhC_S-2 are significantly lower (worse) than for A ($p<0.01$, Wilcoxon). For some subjects DRV-2 and BNR-2 were advantageous relative to reference configuration A, especially for SNR=-7. No subject had higher scores for PhC_S-2 relative to A, including the subjects with hearing-impairment. The hearing-impaired subjects in another study (see section 4.5) could benefit from the level-equalisation between louder vowels and softer consonants provided by fast compression for speech in fluctuating noise, not in continuous noise. We also expect that the hearing-impaired subjects always have lower scores for PhC_S, because no frequency-specific compensation for the subjects' hearing loss was applied. Especially for subjects s6 and s8, who have high-frequency hearing losses, larger parts of the speech information could be lost.

Both the SRT results and the CVC results are on average poorer for the optima found by the pattern search relative to the reference condition (A). One major explanation for the relative poor scores of the optima found for each subject is the evaluation criteria that are different. The SRT and CVC tests focus on the speech intelligibility of the speech-in-noise material. However, for the pattern search the speech-in-noise material was evaluated on listening comfort, for reasons of higher reliability and consistency. This illustrates that there can be a discrepancy between good speech intelligibility and high listening comfort. This suggests that speech enhancing and noise or reverberation reduction strategies cannot achieve both goals simultaneously for this group of (near) normal hearing subjects using active hearing protectors.

4.3.4 Conclusions

- The consistency of the multi-directional pattern search was reasonably high in this study
- The optima with highest listening comfort found by the pattern search did not yield highest speech reception thresholds

4.4 Difference limens for monaural noise reduction and spectral enhancement

In this study we investigated the perceptual differences for three algorithm parameters, in order to determine distinguishable settings for a three-dimensional pattern search. Therefore, difference limen experiments were conducted. The outcomes were assumed to give a good indication of the possibilities within the perceptual constraints of a specific group of hearing-impaired subjects. Five hearing-impaired subjects (3 male, 2 female) participated in this study. The subjects were volunteers. This group comprised older subjects: the average age is 53 years. The hearing-impaired subjects were recruited from the audiological centre of the Academic Medical Centre in Amsterdam. All hearing-impaired subjects have a mild-to-moderate cochlear hearing loss for the test ear. Three subjects wear one hearing aid and two subjects (s2 and s4) wear two hearing aids.

4.4.1 Methods

Algorithms

Two algorithms were used for the Difference Limen tests (DL): monaural noise reduction (NR_D) and spectral enhancement (SE_B). From the monaural noise reduction algorithm we only varied two parameters: the a priori SNR update time constant (N_{estim}) and the speech pause detection threshold (SNR_{estim}). For the spectral enhancement algorithm the only parameter that was varied is the attenuation factor (SE_B).[35]

In each dimension 45 points were incorporated, counted from 0 to 44. The dimensions resemble the three parameters that were chosen to be optimised in a three-dimensional pattern search procedure. In each dimension the DL-experiments have two fixed parameters and one variable parameter. In this way the DL dimensions link to the pattern search dimensions, i.e. a connection exists between the three parameters. The parameter setting for the DL experiments were as follows:

1. Parameter: N_{estim} = variable between 1001 and 1ms
   SNR_{estim} = 0.27
   SE_B = 1 (maximum enhancement)
2. Parameter: N_{estim} = 183 ms
   SNR_{estim} = variable between 1 and 0
   SE_B = 1
3. Parameter: N_{estim} = 183 ms
   SNR_{estim} = 0.27
   SE_B = variable between 0 and 1

Experimental design

For each dimension, i.e. the variable N_{estim}, SNR_{estim}, or SE_B, we performed four DL-experiments. These comprise three different standard settings (settings 0, 22, and 44) and two different directions that are only applicable for the midpoint (setting 22). The standard settings and directions are presented in figure 4.14 and are represented by dots and arrows, respectively. In total 12 DL experiments were performed, four for each parameter. As an example a possible outcome is given in figure 4.15. From standard setting 0 for N_{estim} a difference of 22 scale units is needed to arrive at a just distinguishable signal. For SNR_{estim} again a difference of 22 scale units is needed for discrimination between the signals. For SE_B only a difference of 2 scale units is needed to distinguish parameter settings.

The difference limens for each parameter were determined by a 3-AFC procedure. The subjects listened to three sound samples and had to decide which sample is different from the other two. We used a 2-up 1-down method. Following this strategy, the subject

[35] The spectral enhancement algorithm only deepens the valleys and does NOT enhance the hills for reasons of computation time constraints. Factor E by which the valleys are deepened varies between no deepening (value 0) and full deepening (value 1). The shape of the enhancement filter is kept constant. This filter is composed of two auditory filters one with a bandwidth of 0.5 ERB and the other 2 ERB wide. It is expected that the SNR increases, i.e. that the algorithm will act more like a noise suppressor as well.

had to select the correct sound sample two times in row in order to decrease the difference. After one failure (wrong decision) the difference was increased.[36]

The initial difference between settings is chosen 8. In the very beginning of the experiment the differences are decreased already after 1 correct response, from 8 to 4 and from 4 to 2, or from 8 to 16 and from 16 to 24. The DL procedure ends after 8 reversals or 36 comparisons. The first three reversals are not used to calculate the average and standard error for each run.

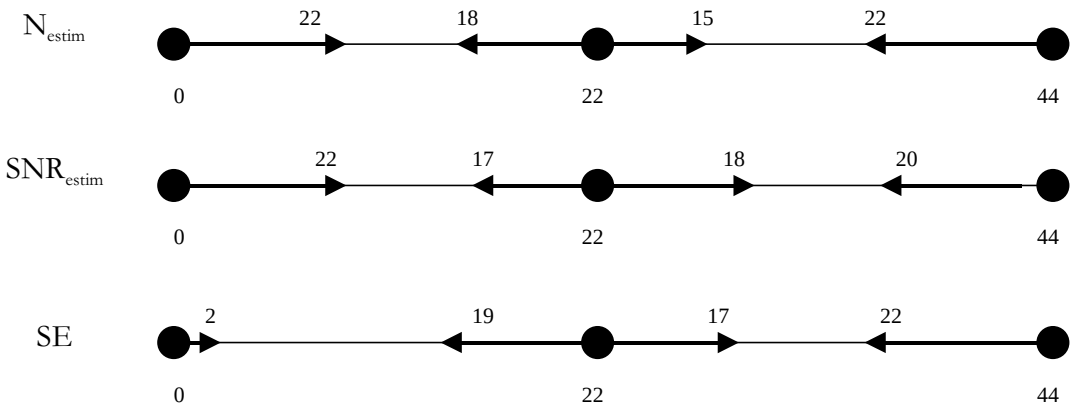


Figure 4.15 Example of DL results given for every initial value and for each algorithm. The settings (0, 22, 44) are indicated below each algorithm range. Arrows indicate the possible directions for which the DL experiments are performed. The perceptual differences found, are indicated as numbers above each arrow and by the arrow length as well.

Speech and noise were recorded separately, because it was desirable to control the signal-to-noise ratio (SNR). The sentences[37] and background noise were recorded monaurally by means of a Signia microphone attached behind the ear of one of the researchers and using a portable Sony PCM-M1 DAT recorder. The speech material, which was presented frontally (0 degrees) by a free-field speaker, was recorded in a sound-attenuating booth at a distance of 1 meter (Leq=65 dBA). The background noise comprised real-life samples of two speakers (1 male and 1 female) and Cuban folk music to create a typical complex realistic situation. The background speakers were reproduced time-reversed to avoid understandable speech segments.

The speech and noise samples were transmitted from the DAT recorder to a workstation that runs under UNIX. The speech samples, which vary in length between 1.5 and 3.5 seconds and are preceded by silence periods, are cut in four-second parts. The sentences are preceded by 0.5 seconds of silence and the length of the silence period

[36] The 2-up 1-down method converges when $(P[x])^2 = 1 - (P[x])^2$, with P[x] denoting the probability to give one correct response. So convergence takes place at P[x] = 0.707 (=√½).

[37] For the sentence material, a subset of 100 sentences, uttered by a Dutch female speaker, is selected from a database of more than 800 sentences (VU sentences98, described by Versfeld *et al.,* 1999). The database contains mainly back files of major Dutch newspapers, which gives a good representation of daily conversational speech.

after the sentence depends on the length of the sentence. Also, four-second noise samples were cut and added to the speech samples at a SNR of 0 dB.

The signals were down-sampled to 16 kHz and processed by the monaural noise reduction and spectral enhancement algorithm. After that, the signals were equalised for their RMS values. After the recording and processing of the stimuli, the signals were fed into the experimental Siemens Signia device, where the signals were linearly amplified according to NAL-RP prescription of the Siemens software. The output of the Signia loudspeaker was set to the individual MCL.

4.4.2 Results

Pure-tone audiogram

For each subject the hearing loss of the test ear is measured and presented in figure 4.16. The hearing-impaired subjects had a mild-to-moderate cochlear hearing loss for the test ear (better ear). Three subjects wear one hearing aid and two subjects (s2 and s4) wear two hearing aids.

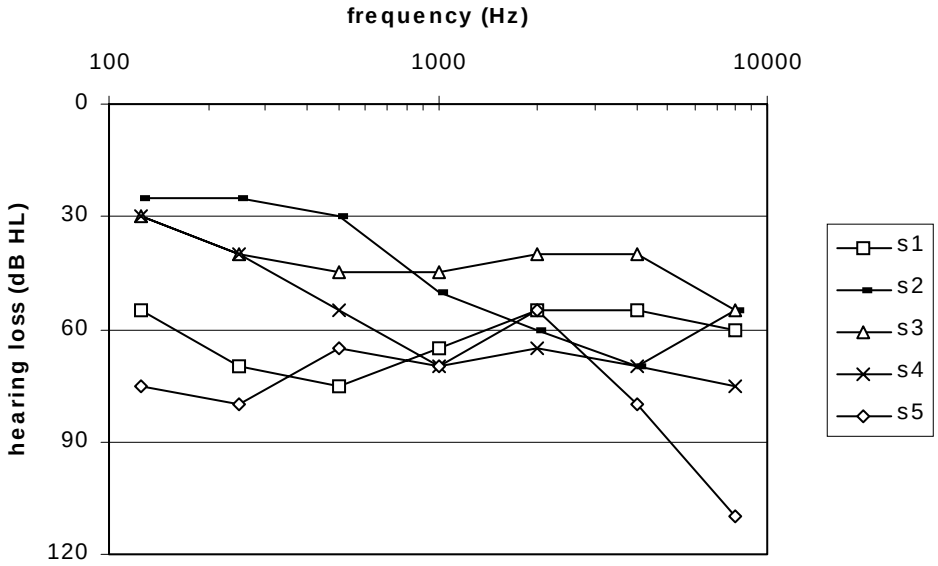


Figure 4.16 Pure-tone audiograms for five hearing-impaired subjects with cochlear hearing losses.

General findings

The difference limen results are presented in figure 4.17. The step-sizes found were in general very large. Difference limens were very high for standard setting '0r' for N_{estim} and SNR_{estim}, and for standard setting '44l' for SE_B. In these cases the algorithm parameters had largest parameter values. For all three algorithm parameters the difference limens strongly depended on the standard setting.

For the N_{estim} and SE_B parameter, difference limens for the mid-point did not depend much on the direction of comparison (left or right). For the SNR_{estim} parameter there appeared to be a significant difference between these two directions for all but one

subject.[38] For all three parameters the just noticeable difference for standard setting 22l is approximately equal, but still large: 18 scale units on average.

Parameter-specific findings

The difference limens found for each algorithm parameter depended on the hearing-impaired subject for some standard settings. For parameter N_{estim} and standard setting 44l, two subjects (s1 and s5) had a much larger difference limen than the other three (22 vs. 2). For parameter SNR_{estim}, and standard setting 0r and 22r, again subject s1 has a considerably larger step-size than the other three subjects, i.e. 42 vs. 27 and 18 vs. 6, respectively. For parameter N_{estim} and standard setting 22l, subject s1 had a difference limen of 20 scale units, s2 and s5 a DL of 15 scale units on average, and s3 and s4 a DL of 10 scale units on average. For SE_B and the four standard settings, the results were similar for all subjects, except for subject s5.

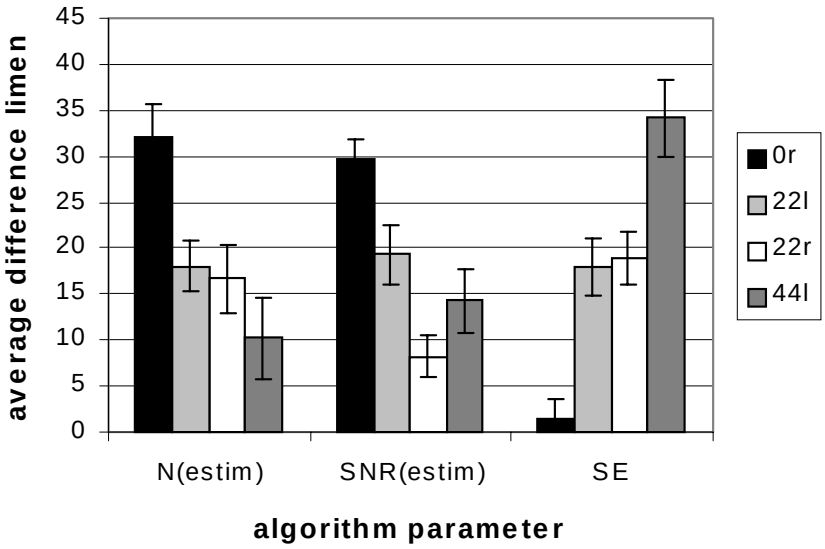


Figure 4.17 Average difference limen values and average standard deviations as a function of either axis. Parameter is a combination of standard setting and direction of comparison.

4.4.3 Discussion

One of the outcomes of this study is that the difference limens vary considerably for each parameter. In effect a fixed step-size in a multi-directional pattern search that optimises these three parameters will not be appropriate. This can be (partly) overcome by using an adaptive step size for the parameters.

From the DL experiment it appeared that difference limens were in general so large that only very few settings can be discriminated and thus can be used for each dimension in a pattern search procedure.

---

[38] Difference limens of most subjects show quite similar patterns within the indicated standard deviations, except for subject 5. For this subject a lot of amplification noise appeared. Using some adaptations in the experimental hearing-aid configuration, this unwanted noise has been removed largely. It seems that the change in configuration characteristics yield different results incidentally, especially for parameter SE_B.

For the combinations of algorithms in paragraph 4.2 and 4.3 the selection of good candidate parameter values was done by informal listening. Although this is not a highly systematic method, it can hint in what region of parameter values we should perform the DL experiments. Moreover, it seems of great importance that we search for algorithms that do have parameters that meet our requirements. That is, it is not at all obvious that for every algorithm enough settings can be found in order to properly use the pattern search optimisation strategy.[39] It is still possible that the setting range of the algorithm parameters need to be even larger. However, the parameters used in this study will yield more and more artefacts the more extreme the values are chosen. In conclusion, the parameters needed by the pattern search should be selected cautiously and the step sizes used in the pattern search should be chosen with great care in order to prevent the optimisation procedure become a random process.

4.4.4 Conclusions

- Difference limens for hearing-impaired subjects depended very much on the standard setting chosen for each algorithm
- In order to reliably optimise the algorithms applied in this study the step size should be made adaptive

4.5 Algorithms in a state-of-the-art hearing aid

Three state-of the-art hearing-aid algorithms were studied. The optimal combination was determined by a multi-directional pattern search algorithm with an adaptive step size. A round-robin tournament was used to compare different optima found with the two starting points, i.e. the reference setting and the setting with all three algorithms settings maximal. Speech reception thresholds (SRT) were determined for the winner of this tournament, the reference setting, and three settings for which only one algorithm was set maximally. Three hearing-impaired subjects s0, s1, and s2 participated in a pilot experiment. Ten hearing-impaired (s3 to s12) and four normal hearing subjects (s13 to s16) participated in the actual experiments. All experiments were conducted monaurally to the subject's better ear. The purpose of this study is to obtain knowledge of (1) the perceptual effects of advanced processing in a commercially available digital hearing aid, and (2) the applicability of a three-dimensional optimisation strategy to fine-tune the test hearing aid.

4.5.1 Methods

Algorithms

Three different processing techniques were included: syllabic compression (SC), spectro-temporal enhancement (STE), and noise reduction (NR_S). For each algorithm five settings were chosen to create a three-dimensional optimisation space of 125 settings. The settings are presented in table 4.7. For syllabic compression the parameter that was varied was the number of frequency bands that use syllabic compression, which varied

[39] In this perspective the remarks of Levitt *et al.* (1986), quoted in the beginning of chapter 3, seems to be a gross oversimplification of the success of an optimisation strategy based on paired comparisons.

between 0, 2, 4, and 6 channels (settings 0 to 4, respectively). In the reference case only slow compression is active in all frequency bands. For the other settings fast compression was included successively, starting in the highest frequency band. For each frequency band, the compression ratio and compression knee-point were fixed at 2 and 54 dB, respectively. The amount of spectro-temporal enhancement used could vary between preset values off, med, and max for the frequency bands from 565 to 2000 Hz and from 2000 Hz and higher, which in combination yields five settings: off-off, off-med, med-med, med-max, and max-max. For the noise reduction algorithm two frequency bands were created for frequencies below 1130 Hz and frequencies above 1130 Hz, respectively. The amount of contrast for both frequency bands was combined such that five settings were created ranging from off-off, off-min, min-min, min-max, to max-max.

	Ch1-2	**Ch3**	**Ch4-5**	**Ch6-8**		**Ch1-2**	**Ch3-5**	**Ch6-8**		**Ch1-4**	**Ch5-8**
SC_0	-	-	-	-	STE_0	N.a	Off	Off	NR_0	Off	Off
SC_1	-	-	-	SC	STE_1	N.a	Off	Med	NR_1	Off	Min
SC_2	-	-	SC	SC	STE_2	N.a	Med	Med	NR_2	Min	Min
SC_3	-	SC	SC	SC	STE_3	N.a	Med	Max	NR_3	Min	Max
SC_4	SC	SC	SC	SC	STE_4	N.a	Max	Max	NR_4	Max	Max

Table 4.7 Algorithm configurations of the test hearing aid for the multi-directional optimisation procedure. The dimensions correspond to three algorithms, Syllabic Compression (SC), Spectro-Temporal Enhancement (STE), and Noise Reduction (NR_S). 'Off' represents no processing. 'Min', 'Med', and 'Max' correspond to minimum, medium and maximum processing. 'N.a.' represents not applicable for those channels. For the compression algorithm '-' designates that only slow compression was active. Different channels are separated by boundary frequencies 300, 565, 800, 1130, 2000, 3200, and 5000 Hz, respectively.

Experimental design

The experiments were conducted in a laboratory setting. In order to use carefully selected and well-controlled test signals, the processing by the test hearing aid was performed off-line. These processed signals were used for the multi-directional optimisation procedures and round-robin experiments.

The speech material (VU98) and noises were recorded monaurally using the microphone of the test hearing aid, worn by the author. We used a continuous car noise (cont) and a fluctuating competing speech noise (fluct), a two-speaker babble consisting of a male and a female speaker. In order to avoid understanding of the competing speech noise, the competing speech was presented time-reversed. The signals were recorded from a frontal position. The sentences and noises were mixed at a signal-to-noise ratio of +5 dB. The output of the microphone was directly fed into a Sony portable DAT-recorder (PCM-M1). Next we presented these stimuli off-line to the experimental test hearing aid, for each experimental combination of signal, noise, and hearing aid setting. The resulting electrical output to the receiver was recorded and stored in a computer.

In the actual testing the recordings were presented electrically to the hearing aid receiver of the experimental test hearing aid, coupled by an individual ear mould to the subject's better ear. Signals were amplified by the test hearing aids, compensating the audiometric hearing loss. The linear target gain corresponds to the gain prescription by DSL-I/O for a 65 dB input signal, i.e. speech-shaped noise.

For the SRT measurements no off-line pre-processing of the signals was necessary and the test hearing aid could be used for real-time processing in the selected settings.[40] In this case, the noises were presented without interruptions in order to allow the NR_S-algorithm to be constantly active. Also for these experiments, both speech and noise were presented from a frontal position. SRT measurements were performed in free field. That is, the speech and noise were presented acoustically to the hearing aid. Therefore, the other ear was obstructed by a soft self-shaping foam earplug to avoid cross talk to the other ear. For the multi-directional optimisation and round-robin experiments, however, the signals were presented directly from the computer to the hearing-aid loudspeaker. Hence, an earplug was not required.

For the hearing-impaired subjects, all experiments – multi-directional optimisation, round-robin tournament, and SRT test – were performed for the two different background noises. For the multi-directional optimisation procedure tests were performed for two initial settings, T0=(SC_0,STE_0,NR-S_0) and T4=(SC_4,STE_4,NR-S_4). The different starting points were used to study the presence of possible sub-optimal settings. A retest R4=(SC_4,STE_4,NR-S_4) was performed to investigate the reliability of the multi-directional optimisation procedure. The retest was conducted for initial setting (SC_4,STE_4,NR-S_4) only, because a pilot experiment showed that for this setting subjects were able to make their decision most easily, at least in the beginning of the experiment. Test T0 was accomplished for continuous car noise only, because the effect of starting point is not likely to depend much on noise type. Normal hearing subjects performed the multi-directional optimisation experiments for continuous car noise only.

4.5.2 Results

Pure-tone audiograms

The pure-tone audiograms for the best ear of ten hearing-impaired subjects (s3 to s12) are given in figure 4.18. Three hearing-impaired subjects s0, s1, and s2 participated in a pilot experiment. Pilot experiments were carried out in separate experiments in order to determine whether differences between settings were distinguishable. Perceivable differences occurred most often for step sizes of 2 or larger. Therefore, we implemented a no-difference button to increase the step size when necessary. The subjects participated on a voluntary basis. The hearing-impaired subjects have at least one year of experience with one or two hearing aids. Two groups of hearing-impaired subjects were selected, based on the average hearing loss for frequencies of 0.5, 1, 2, and 4 kHz of the better ear. The first group comprised five subjects with an average hearing loss less than

[40] For the speech reception threshold measurements we used Plomp sentences (Plomp and Mimpen, 1979).

45 dB; the second group comprised five subjects with a hearing loss higher than 45 dB. The average hearing loss of the groups are 38 dB (s.d=4) and 54 dB (s.d=6), respectively.

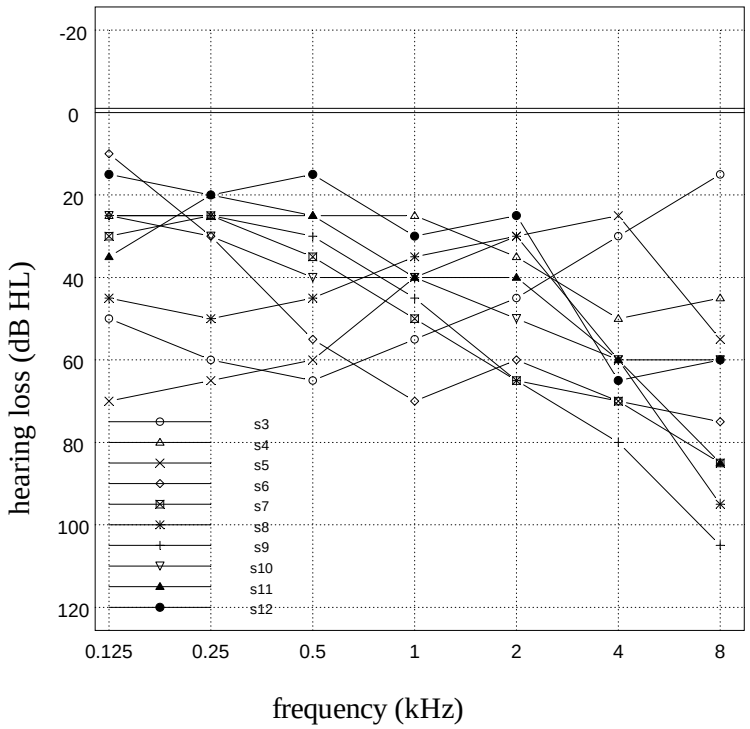


Figure 4.18 Pure-tone audiograms of ten hearing-impaired subjects, indicated as s3 to s12.

Multi-directional optimisation

To allow comparisons between normal and hearing-impaired subjects, we averaged data for continuous noise and condition T0, T4, and R4. On average, the hearing-impaired subjects needed more paired comparisons to end the multi-directional optimisation procedure than the normal hearing subjects. Moreover, the hearing-impaired subjects used the 'repeat' and the 'no difference' button more frequently than the normal hearing subjects did (see table 4.8). In the set-up of the multi-directional optimisation procedure, we decided to visualize the sentences on the computer-screen to avoid that the responses to the second sound sample would be favoured by the fact that the subjects heard the sentence for the second time. By comparing how often subjects chose the first or second sentence we found that this unwanted effect (favouring the second sentence) did not occur.

For the hearing-impaired subjects, an analysis of variance of the optimisation results did not yield any main or interaction effect for parameters subject, noise-type, and test, and for parameters subject and start. That is, the optima of the multi-directional search are not different for different subjects, noise types, test and retest, different starting points, or combinations of these parameters. Therefore we averaged the optima of the multi-directional search for the two noise types, test and retest, and different initial settings. On average the hearing-impaired subjects prefer some fast compression (SC=2). Hence, only in the highest 4 channels SC is preferred which represent the higher

frequencies. For STE the hearing-impaired subjects chose intermediate values (STE between 2 to 3). Thus, the amount of spectral and temporal contrast is medium for the frequency band from 565 to 2000 Hz and medium-to-maximum for the higher frequencies. The optimal NR_S setting is on average "minimum" (NR_S =2). That is, on the range between off, minimum and maximum, the minimum NR_S setting is preferred for all eight frequency channels.

Group	Step	#PCs	Repeat	ND	Fragm1	Fragm2
HI	***Adapt***	41.2	15.8	9.2	22.6	18.6
NH	***Adapt***	32.4	12.1	4.5	14.9	17.5
	Fixed	32.2	9.6	2.2	15.8	16.4

Table 4.8 Average number of paired comparisons needed to complete the pattern search (#PCs) and the average number the 'repeat', 'no difference', 'sample1', or 'sample2' (ND) button has been used during the procedure. The numbers have been calculated for the hearing-impaired group (HI) and the normal hearing group (NH). For NH the step size was either fixed (step = 2) or adaptive (step = 1 or 2). For HI the step size was always adaptive. The average values are determined for car noise and condition T0, T4, and R4 only.

We compared the optimisation strategy with fixed step size of 2 and the optimisation strategy with variable step size (size is 1 or 2) for the normal hearing subjects. Therefore, the results of the optimisation strategy were analysed by calculating the cumulative distribution of 3-dimensional distances between test and retest results labelled "reliability", and between the results for different initial settings labelled "consistency". The cumulative distributions for the reliability are illustrated in figure 4.19, in the upper panel.

The consistency patterns are not shown, but are comparable. In general, the cumulative distributions for the reliability and for the consistency of the normal hearing subjects are well above the random cumulative distribution.[41] The cumulative distributions also indicate that the reliability increases for the multi-directional optimisation procedure with a fixed large step size (NHs2), relative to the adaptive step size (NHa). For a fixed step size the consistency does not improve. Also the starting point appears to affect the reliability. There is a clear trend that the 3-dimensional distance between test and retest is lower for initial setting (SC_4,STE_4,$NR\text{-}S_4$) relative to (SC_0,STE_0,$NR\text{-}S_0$). It seems that differences between settings are more obvious in the vicinity of (SC_4,STE_4,$NR\text{-}S_4$). Apparently, the step sizes are perceptually larger when the algorithms settings are maximal.

We also determined the cumulative distributions for the hearing-impaired subjects. In general, both the reliability and consistency appear low, near to the random distribution for which the optimal settings are randomly assigned (figure 4.19, lower panel). As already indicated by the hearing-impaired subjects during the experiments, the results for fluctuating noise are worst. For hearing-impaired subjects the evaluation of

[41] The random distribution was determined by calculating the 3D distance for 1000 sets of randomly chosen test and retest optima.

the sentences in noise is based on differences in speech intelligibility. However, the subjects indicated afterwards that when the speech intelligibility was the same, they evaluated the clarity and quality of the background noise. This was only possible for continuous noise. This re-cueing may have affected the reliability and consistency.

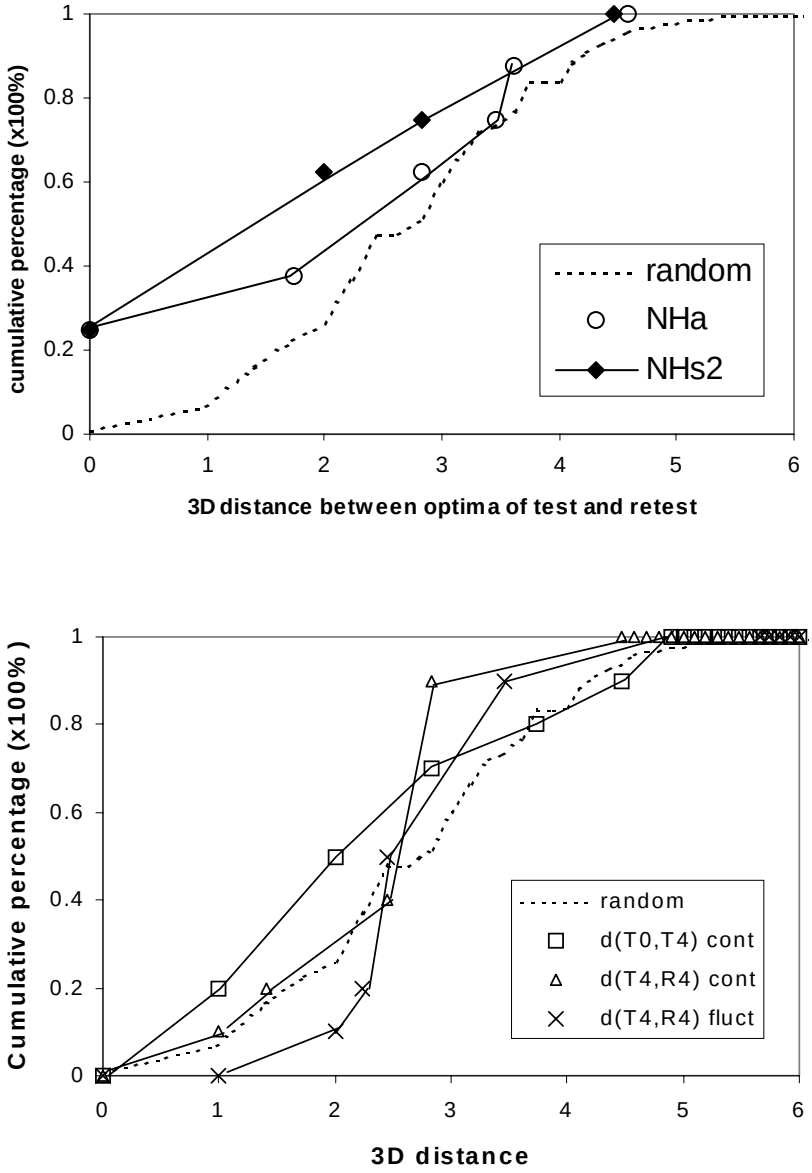


Figure 4.19 Cumulative percentage as a function of the 3-dimensional distance between (a) optima found for test and retest for normal hearing subjects (upper panel), and (b) between optima found for test and retest or for different starting points for hearing-impaired subjects (lower panel). Parameter for the normal hearing subjects is the step size configuration, which is either fixed at step size=2 (NHs2), or is adaptive, i.e. the step size is 1 or 2 (NHa). Parameter for the hearing-impaired subjects is the noise type, which is either continuous car noise, or fluctuating noise (fluct). For the random distribution the test and retest optima are randomly assigned.

In order to measure reliability and consistency from these distributions we calculated the information transfer. That is, we calculated the area between the cumulative distribution under study and the random distribution, relative to the area between the 'perfect' distribution with a 3-dimensional distance of 0 for all measured data and the random distribution. The results show that only for STE there is considerable information transfer with respect to the consistency. We also averaged the 3-dimensional distances to compare the two subgroups of hearing-impaired subjects. For subjects with lower hearing losses (hearing loss <45 dB) the reliability d(T4,R4) and consistency d(T0,T4) were in general higher, due to the lower average 3-dimensional distance. For fluctuating noise the results are relatively poor for both groups of subjects. Apart from the a priori division based on hearing loss, the hearing-impaired subjects were divided in subgroups a posteriori as well, based on age and on pure-tone hearing loss slope. On average the younger five subjects are 50.4 years old (s.d.=4.6). The average age of the older subjects is 63.0 years (s.d.=4.7). The audiometric slope was calculated by a balance equation:

$$balance = \frac{\frac{HL(2k) + (HL(4k)}{2} - \frac{HL(k5) + HL(1k)}{2}}{2}$$

The slope of the hearing loss was divided in either negative/flat (average=-4.7 and s.d.=7.6 dB/oct, for four subjects) or positive (average=11.0 and s.d.=3.7 dB/oct, for six subjects). The cumulative distributions for these subgroups diverge considerably. The consistency is high when subjects are relatively young and when subjects have a flat or negative slope for their pure-tone audiogram. Note that age and hearing-loss slope do not covariate in our test subjects. The other groups, the relatively older group and the group with a high frequency loss, perform very badly, i.e. comparable to the random distribution.

Cumulative distributions can also be used to determine the performance for each algorithm. The information transfer per dimension is presented in table 4.9. The numbers (*100%) are differentiated for subgroups of the hearing-impaired group, based on hearing loss, age, and hearing loss slope. Information transfer is highest for the consistency measure d(T0,T4). The conditions for which the transfer is larger than 50% are marked. The cells for which the transfer is much poorer than the random distribution are outlined. Especially for the STE and NR_S the consistency is high. Subjects that are relatively young or have a moderate hearing loss respond more consistently for the STE algorithm than older subjects or subjects with a mild hearing loss. For NR_S the younger subjects and the subjects with mild losses and flat or negative hearing losses performed most consistently.

The high negative transmission for SC for the older group can be interpreted as the presence of a sub-optimal setting for that algorithm. The distance between T0 and T4 appears consistently large, which means that each subject has found a different optimum when using a different initial setting for the multi-directional optimisation procedure.

% Information transfer			Subgroups HI					
			Mild(5)	Moderate(5)	Young(5)	Old(5)	Flat-neg(4)	Pos(6)
	Noise	Algorithm	38(4)	54(6)	50(5)	63(5)	-5(6)	11(4)
T0-T4	Cont	SC	0.00	-0.33	0.42	-0.75	0.38	-0.53
		STE	0.33	0.75	0.67	0.42	0.58	0.51
		NR	0.75	-0.08	0.75	-0.08	0.69	0.10
T4-R4	Cont	SC	0.33	0.33	0.67	0.00	0.58	0.17
		STE	0.33	0.25	0.17	0.42	0.06	0.44
		NR	0.33	0.00	0.08	0.25	0.27	0.10
T4-R4	Fluct	SC	-0.25	0.08	-0.25	0.08	-0.04	-0.11
		STE	-0.42	0.00	-0.42	0.00	0.06	-0.39
		NR	0.17	0.42	0.25	0.33	0.38	0.24

Table 4.9 Information transfer (IT) for the consistency (T0-T4) and reliability (T4-R4) of the multi-directional pattern search, differentiated for specific subgroups of the hearing-impaired subjects: mild vs. moderate hearing loss, young vs. old, and negative/flat vs. positive pure-tone hearing loss slope. The information transfer is also differentiated for noise type, continuous noise (cont) and fluctuating noise (fluct), and differentiated for algorithm SC, STE, and NR_S. Values for IT higher than 0.5 are presented in grey. It values lower than -0.5 are outlined.

The reliability (T4-R4) for continuous noise appears better than random for most subgroups and individual algorithms. Though, only young subjects or subjects with a flat/negative hearing loss respond very reliable for the SC algorithm. For the fluctuating noise the information transfer values are quite dramatic. It seems that subjects have responded almost randomly, irrespective of age or hearing loss.

Based on the information that certain groups performed the experiments more reliably and more consistently, the optima found were analysed for smaller subgroups. The results were studied for subgroups of hearing-impaired subjects (a) with a low average hearing loss, (b) with a negative or flat hearing loss slope, (c) with a low (negative) speech reception threshold, and (d) for relatively young age. Subgroups (a), (c), and (d) contain 5 subjects, whereas subgroup (b) contains 4 subjects. For subgroups (a) and (c) no effects were found. However, for subgroup (b) there is a significant difference (Wilcoxon, $p<0.05$) between the optimal SC setting found for continuous noise and found for fluctuating noise. For fluctuating noise more syllabic compression is preferred. The effect found for subgroup (d) relates to individual differences between optima found for the STE algorithm. For NR_S the optimum of s12 is significantly higher (Wilcoxon, $p<0.05$) than the optimum of s4 and s5.

Round-robin tournament

The round-robin tournament yields ratings across the settings (SC_0,STE_0,$NR\text{-}S_0$), (SC_4,STE_4,$NR\text{-}S_4$) and three winners of the multi-directional search (w1, w2, w3) for continuous noise, or two winners and one mirrored winner (w1, w2, m1) for fluctuating noise. In order to measure the appropriateness of the winners of the multi-directional optimisation we calculated the percentage of setting (SC_0,STE_0,$NR\text{-}S_0$), (SC_4,STE_4,$NR\text{-}S_4$), or one of the optimum settings found by the multi-directional search that has won the round-robin tournament. When the winner is randomly assigned, the percentages are 20%, 20% and 60%, respectively. The actual percentages obtained by the hearing-impaired subjects are 15%, 5%, and 85% for the continuous noise and

fluctuating noise results together. For fluctuating noise reference setting (SC_0,STE_0,$NR\text{-}S_0$) was the best setting for three subjects. These three subjects do not share a common SRT value, average hearing loss, hearing loss slope, or age. The average ratings (number of wins) for all subjects of setting (SC_0,STE_0,$NR\text{-}S_0$), (SC_4,STE_4,$NR\text{-}S_4$), and the three (or two) winners of the multi-directional search, is 3.7, 3.5, and 4.3 for continuous noise, and 4.7, 3.4, and 4.1 for fluctuating noise, respectively. When the round-robin ratings are averaged for settings 0, 2, and 4 of each dimension, the general optimal settings of each algorithm can be estimated. The results indicate that setting (SC_0,STE_2,$NR\text{-}S_2$) was preferred. That is, the hearing-impaired subjects judge the speech intelligibility highest, when there is no compression present, and when the enhancement (STE) and noise reduction (NR_S) have intermediate values, at least on the preset test hearing aid scale.

			Algorithm setting			3D distance	
Noise	**Cluster**	**#subj**	**SC**	**STE**	**NR**	**avg**	**stdev**
Cont	***cl1***	3	0	2	2	0	0
	cl2	3	1	4	2	0.8	0.7
	cl3	2	1	2	4	1	0
	cl4	1	2	2	0	0	
	cl5	1	4	4	4	0	
Fluct	***cl1***	5	3	3	3	1.6	0.2
	cl2	3	0	0	0	0	0
	cl3	1	0	4	4	0	
	cl4	1	1	0	4	0	

Table 4.10 Cluster analysis of pattern search results for continuous noise (cont) and fluctuating speech noise (fluct). The number of subjects (# subj) that form a cluster (cl1-cl5), the average 3-dimensional distance for the settings in each cluster (avg), and the standard deviation relative to the cluster centre (stdev) are also given.

We also performed a cluster analysis (Nearest neighbour, Euclidean distance) for the round-robin winners. The results are presented in table 4.10. The optimal settings in each cluster differ no more than √3 (1 step in each direction) relative to the rounded cluster centre. The average 3-dimensional distance and standard deviation are also given to indicate the compactness of the cluster. Clearly, (groups of) hearing-impaired subjects select different settings. The optimal settings selected for continuous noise are different from the settings selected for fluctuating noise. It is striking that most subjects do not prefer SC for continuous noise, whereas half of the subjects do prefer SC in many channels for fluctuating noise. Except for s6, the subjects that fall outside the first cluster for fluctuating noise (s7, s9, s11, s12) have the highest positive hearing loss slopes (high frequency losses). For fluctuating noise, subject s6 has used the 'no difference' button for more than 70% of the comparisons, which makes the winning setting (SC_0,STE_0,$NR\text{-}S_0$) less reliable. For the continuous noise condition, the subjects within the clusters showed no clear similarities with respect to average hearing loss, audiometric slope, age, and SRT for continuous noise.

For the round-robin results also the reliability and consistency were considered. Therefore, the coefficient of consistence "ζ" and the coefficient of agreement "u" were

calculated (David, 1963). The consistency measure is based on the number of circular triads, whereas the reliability measure is based on test-retest results. Only for four subjects the calculated coefficient of consistency and the coefficient of agreement is reasonably high for continuous noise ($\zeta>0.45$, $u>0$).[42] For fluctuating noise the consistency and reliability are very low: for all subjects ζ is lower than 0.45 and u is 0. For the four subjects s3, s4, s6, and s8 that have high consistency and reliability, SC is significantly higher (Wilcoxon, $p<0.05$) and STE is significantly lower (Wilcoxon, $p<0,05$) for fluctuating noise than for continuous noise.

Speech reception thresholds

We measured speech reception thresholds in continuous and fluctuating speech-shaped noise. Five subjects with speech reception thresholds higher than 3 dB have thresholds that are on average equivalent for both noises. This is found often in hearing-impaired subjects with cochlear hearing losses. The other subgroup of five subjects has lower thresholds for fluctuating noise than for continuous noise. The difference is on average 4 dB, which is slightly less than for normal hearing listeners. The SRT values of subject s5 are comparable with normal SRT values, both in absolute as in relative sense (-10 and -4 dB, respectively). Though, this subject has a very negative hearing loss slope of -11 dB/oct and an average hearing loss of 39 dB.

We compared the SRTs of the reference setting (SC_0,STE_0,$NR\text{-}S_0$) with setting (SC_4,STE_0,$NR\text{-}S_0$), (SC_0,STE_4,$NR\text{-}S_0$), (SC_0,STE_0,$NR\text{-}S_4$) and the best optimum of the multi-directional search, determined by the round-robin tournament. To decide which setting had the best (lowest) SRT, we compared the SRT values for each dimension. For each dimension we averaged the SRTs of three settings for which the setting was 0 for that particular dimension. For example, for SC the SRTs of setting (SC_4,STE_0,$NR\text{-}S_0$), (SC_0,STE_4,$NR\text{-}S_0$), (SC_0,STE_0,$NR\text{-}S_4$) were averaged. This value was compared with the SRT of setting (SC_4,STE_0,$NR\text{-}S_0$) and the best optimum of the multi-directional search.

When the SRT of the best optimum is at least 1 dB lower than the other two SRTs, this setting was defined as 'good'. When the SRT of the best optimum was within a 1 dB range of the other two SRTs, the setting was defined 'equal'. In the other case, i.e. when the SRT of the best optimum was worse (higher) than the other two SRTs, this setting was defined as 'bad'. These labels (good, equal, bad) are determined for each dimension.

Finally, the labels of the individual dimensions were compared to determine the optimum of the multi-directional search behaving optimal, suboptimal, equally well, or bad. The multi-directional search behaves optimal when for all three dimensions the labels are 'good'. The search behave suboptimal when at least on of the labels is 'good' and the other label(s) is (are) 'equal'. When all labels are 'equal' the search behaves equally well, and when at least one of the labels is bad, the search is defined to behave bad. The results of the analysis are presented in table 4.11. These results imply that for both noise conditions four out of ten hearing-impaired subjects have higher (i.e., poorer) SRTs for the best optimum of the multi-directional search than for the other settings.

[42] There appears to be a relation between ζ and u based on calculations for all ten subjects ($R^2 = 0.93$), namely $\zeta = 0.5u+0.5$.

Subjects s3 and s6 have chosen a best optimum that also has the best (lowest) SRT for both noise conditions.

Continuous noise				Fluctuating noise			
Optimal	**Suboptimal**	**Equal**	**Bad**	**Optimal**	**Suboptimal**	**Equal**	**Bad**
s3		s8	s4	s3	s8	s5	s4
s6		s10	s5	s6	s9		s7
s12		s11	s7	s12	s11		s10
			s9				s12

Table 4.11 Comparison of the SRT of the best optimum of the multi-directional search and SRTs of settings (SC_0,STE_0,NR-S_0), (SC_4,STE_0,NR-S_0), (SC_0,STE_4,NR-S_0), and (SC_0,STE_0,NR-S_4) for subjects s3 to s12. Comparisons were made for two noise types: continuous noise and fluctuating noise. Labels 'optimal', 'suboptimal', 'equal', and 'bad' are used to indicate that the SRT of the optimum is really best, best for at least one dimension and equal for the other dimensions, equal for all dimensions, or worse relative to the other settings.

4.5.3 Discussion

Methodological aspects

Hearing-impaired subjects have more problems to perform the multi-directional optimisation procedure than normal hearing subjects. They need more paired comparisons and they use the 'repeat' and 'no difference' button clearly more frequently. This suggests either that it is more difficult to hear differences between the samples, or that it is more difficult to judge which of the two samples is more intelligible. However, the subjects were instructed to only use the 'no difference' button when they really could not distinguish the two samples, i.e. the two test hearing aid settings. Therefore, it seems most plausible that the hearing-impaired subjects have problems distinguishing the samples. That is, the settings of the SC, STE, and NR_S algorithms of the test hearing aid that were chosen in this study are difficult to distinguish for a step size of one, and sometimes also for a step size of two. Note that even for the normal hearing subjects in the fixed step-size optimisation procedure (step size =2), the 'no difference' button was used on average 1.8 times. Two of these normal hearing subjects can easily distinguish the samples for the fixed step-size procedure, whereas one subject does hear differences, but has a lot of difficulties to determine which sample is more intelligible. Again, this suggests that the 'no difference' button has also been used when the samples were judged equally intelligible.

With respect to the reliability, the cumulative distribution patterns show that the percentages of transmitted information are higher for the fixed step size procedure. However, the consistency (the 3-dimensional distance between the optima for different initial settings) appeared equally low for the fixed and variable step size, near to the random distribution. This might imply that sub-optimal settings are present. However, on average, the same optimum is found for the two initial setting. Thus far, we cannot explain the discrepancy between the reliability and consistency data. Though, due to the fact that the cumulative distributions are close to the random distribution, it seems that – even for relatively large step sizes – it is difficult to decide which sample is more

intelligible. Either the samples are equally worse or equally acceptable. Due to the fact that the reliability for initial setting (SC_4,STE_4,$NR\text{-}S_4$) is better than for (SC_0,STE_0,$NR\text{-}S_0$), it appears that judgments can be made more easily for settings that are situated at that side of the multi-directional optimisation space. This implies that the algorithms do not have just a non-linear scale, but that the just noticeable differences between samples depend on the interaction between the algorithms of the test hearing aid. Because the algorithms behave differently for different combinations, fitting with an adaptive step size independently for each dimension becomes very valuable. Our results suggest that the multi-directional optimisation strategy is only valuable when the minimum step size is sufficiently large.

When the multi-directional search results are analysed for each dimension as well, it appeared that especially STE has a high information transfer. That is, the STE values are most consistent in transferring the ascribed optimal speech intelligibility. When the hearing-impaired subjects are divided into subgroups, it appears that the NR_S algorithm is only consistent for relatively young subjects (avg.=50 years) or for subjects with mild hearing losses (avg.=38 dB). The SC algorithm is most reliable for estimation of the best setting for young subjects or flat or negatively sloping hearing losses. For older subjects (avg.=63 years) the multi-directional optimisation strategy seems to suffer from sub-optimal settings. These implications only hold for continuous noise. For the fluctuating noise the information transfer is always low. The implication of our findings is that for these algorithms of the test hearing aid the success of the multi-directional optimisation strategy depends heavily on the degree of hearing impairment and of the subjects and on their age.

Audiological aspects

When the winners of the multi-directional search are averaged for all hearing-impaired subjects, the optimal setting of the test hearing aid is near to (SC_2,STE_2,$NR\text{-}S_2$). That is, the hearing-impaired subjects like, on average, intermediate values of algorithms SC, STE, and NR_S. When the round-robin results are analysed for each algorithm individually, it appears that SC, STE, and NR_S, have best ratings for setting 0, 2, and 2 respectively. This is slightly different from the average multi-directional search results.

The optimal setting depends on the hearing-impaired individual. There are two reasons for this conclusion. First, for subgroups that perform the pattern search reliably, individual differences were found, especially for STE. Second, this result is confirmed by the cluster analysis of the round-robin winners. For STE and both noise types the largest clusters differ especially for STE. The cluster analysis has the additional advantage that the results can be studied three-dimensionally, whereas the pattern search results could be analysed only one-dimensionally. The three-dimensional results of the cluster analysis appeared more diverse than the one-dimensional results of the pattern search analysis. As a consequence individual fine-tuning seems justified.

However, how optimal are the winners of the multi-directional optimisation procedure? When the SRTs of the best winners of the multi-directional optimisation procedure are compared with the SRTs of setting (SC_0,STE_0,$NR\text{-}S_0$), (SC_4,STE_0,$NR\text{-}S_0$), (SC_0,STE_4,$NR\text{-}S_0$), and (SC_0,STE_0,$NR\text{-}S_4$), six out of ten subjects score equally well or

worse with the best winner of the multi-directional optimisation procedure. This suggests that the SRT is not very discriminative for the optimal behaviour of the winners of the multi-directional optimisation procedure. However, the round-robin results suggest that the winners of the multi-directional optimisation procedure in general have higher speech intelligibility ratings than expected from chance. The reason for the occasionally occurring mismatch between the setting with the best SRT and the SRT of the subjectively acquired best setting, might be due to the fact that the subjectively found best setting was guided not only by the speech intelligibility criterion, but also by listening comfort. This might be ascribed to the instruction for the subjects, which states that the subject has to compare the speech intelligibility of the two samples, but also has to keep in mind that he has to use that setting all day long.

The cluster analysis revealed that there are two main clusters for the continuous noise and for the fluctuating noise. These clusters represent settings of the test hearing aid that are different for both noise types; (SC_0,STE_2,NR-S_2) and (SC_1,STE_4,NR-S_2) for continuous noise and (SC_3,STE_3,NR-S_3) and (SC_0,STE_0,NR-S_0) for fluctuating noise. The outcomes for each noise type are not trivial, because the subjects that form a cluster could not be characterized by a priori information regarding hearing loss, slope, age, or speech reception threshold.

For the fluctuating noise condition the subjects of the first cluster (SC_3,STE_3,NR-S_3) have low frequency hearing losses or flat hearing losses. Relative to the subjects in the other three clusters that have high frequency losses, the five subjects of the first cluster like much syllabic compression. For the continuous noise condition there are smaller clusters that could not be related to hearing loss, slope, age, or SRT of the hearing-impaired subjects. It is striking that for this condition eight out of ten subjects do not like syllabic compression. All subjects prefer intermediate or high STE values, and except for one subject they also prefer intermediate or high NR_S values.

From the cluster analysis it also appeared that different individuals prefer different combinations of test hearing aid algorithms for different background noises. Algorithm SC appeared to be mainly valuable for the fluctuating noise condition. This noise-dependency is also found when a subgroup of subjects is created based on relatively good reliability and consistency of the tournament results. For fluctuating noise SC is preferred for the higher frequency bands, whereas for continuous noise almost no SC is preferred. This outcome is in agreement with results of Verschuure *et al.* (1998). Though in the present study the preference for SC is only found for subjects with flat or negatively sloping hearing losses. For the STE and the NR_S algorithms many hearing-impaired individuals prefer intermediate or high values. Again, this depends on the subject and the background noise type. It is especially noteworthy that the contrast enhancement in the STE algorithm is appreciated for continuous noise. This might be due to the fact that spectral and temporal contrasts in the speech signal can be found more easily for a steady background noise.

Because of the fact that the best algorithm configuration could not be predicted from a priori information – the preferred configuration was not related to hearing loss, slope, age, and speech reception threshold – the present multi-directional optimisation

strategy, with some modifications in the procedural set-up, might be a valuable tool for individual fine-tuning taking the interactions between algorithms into account.

4.5.4 Conclusions

- The reliability of the adaptive pattern search procedure depended on the step size
- The reliability and consistency of the adaptive pattern search were low for hearing-impaired subjects
- The consistency of the adaptive pattern search seemed to depend on age, average hearing loss, hearing loss slope and noise type
- The subjectively obtained most intelligible settings did not always correspond with the objectively obtained best (lowest) speech reception threshold

4.6 Noise reduction, spectral lift, and spectral enhancement

Two studies were performed. In experiment 1 the multi-directional pattern search was evaluated for the inclusion of an adaptive step size. In experiment 2 different step size configurations were compared. The optimisation procedure was performed for three experimental auditory signal-processing algorithms. The individual dimensions of the optimisation procedure represent spectral enhancement (SE_L), spectral lift (SL), and noise reduction (NR_L), each comprising ten different settings. Nine normal hearing subjects (s1 to s9) participated in this study. The average age of the subjects was 29 (7) years. In the experiments, the subjects listened to speech in background noise. Two background noises have been used: continuous car noise and fluctuating speech-shaped noise. The speech and noise were presented at a signal-to-noise ratio of 0 dB. Seven normal hearing subjects conducted experiment 1: s1 to s7. Only part of these subjects were available for experiment 2: s1, s3, and s4. Therefore, two other normal hearing subjects were added, s8 and s9, thus the data of experiment 2 rely on the results of five subjects. We verified that the performance of s8 and s9 was in agreement with the other subjects for a subset of conditions in experiment 1.

4.6.1 Methods

Algorithms

Three experimental auditory signal-processing algorithms were used in the optimisation procedure. The individual dimensions were spectral enhancement (SE_L), spectral lift (SL), and noise reduction (NR_L), developed and described in detail by Lyzenga *et al.* (2002). All algorithm values have been chosen such that steps are perceptually approximately equidistant. It was found to be difficult to assemble a series of ten easily discernable settings for each algorithm, without introducing intolerable distortions. To overcome this difficulty, we included settings that, at first inspection, seemed counterintuitive. Not only spectral expansion is allowed but also spectral compression, not only spectral lift but also spectral suppression, and finally, not only noise reduction but also noise amplification. The settings are labelled from -3 to 6, such that 0 is the reference setting without processing for the specific algorithm.

The spectral enhancement algorithm expands the spectral peaks for frequencies between 200 and 5000 Hz. The spectral enhancement was realized using a power function. Enhancement factor *M* represents the exponent, the spectral amplitude of the signal *a* represents the base of the power function. The spectral enhancement factor *M* ranges between 0.25 and 2.5 in steps of 0.25 dB/dB. For setting 0, *M*=1 and there is no spectral enhancement.

The spectral lift algorithm is assumed linear filter that is assumed to prevent low-frequency high-level formants from masking higher-frequency low-level formants. Therefore the filter was designed sawtooth shaped, with the peak of the filter centred on 3kHz. The spectral lift settings have values between -9 and 18, in nine 3-dB steps. A lift-value of 0 represents no spectral lift.

The noise reduction algorithm attempts to suppress noise based on phase variations. The amplitude of the filter is set high when spectral components behave temporally regular, and are assumed to be speech-like signals. The noise reduction multiplication factors are represented by dimensionless values between -3 and 6 in steps of 1, and setting 0 corresponds to a value of 0.

Experimental design

For experiment 1 test (T) and retest (R) measurements were performed for four different starting points. In the multi-dimensional set-up we used combinations T(-3)=R(-3)= (SE-L_{-3},SL_{-3},NR-L_{-3}), T(1)=R(1)=(SE-L_1,SL_1,NR-L_1), T(3)=R(3)=(SE-L_3,SL_3,NR-L_3), and the centre of gravity of the winners of T(-3), R(-3), T(1), R(1), T(3), and R(3). This starting point, which could be a suboptimal setting instead of an optimal setting is labelled as T(S) or R(S) depending on the test or retest status.

In experiment 2 the experimental procedure was the same as for experiment 1, except for the step size. Using the step size as a parameter, the dependence of the reliability on the step size can be estimated. In this experiment, the step size is either fixed or variable. The fixed step size is either 3 or 4 lattice grid steps. The variable step size has a lower limit of 2 and an upper limit of 3 steps. Therefore, only the lower limit is different from the lower limit of experiment 1, which was 1 step. Hence, in combination with experiment 1, there are four step size configurations. The labelling of the configurations follows the possible step sizes in the procedure. So, the procedure of the experiment 1 is 'step123', the variable step size with elevated minimum step size is labelled 'step23'. The fixed step size configurations are labelled 'step3' and 'step4'. The first 'orienting' step of the multi-directional pattern procedure is twice as large as the fixed step size throughout the procedure, i.e. 6 and 8 respectively. For the variable step size, the first step is equal to 4, comparable to experiment 1. Experiment 2 has been completed for starting point (0,0,0) and continuous noise only.

In the experiments, the subjects listened to speech in background noise. The duration of the samples in the optimisation procedure range between 1.5 and 2.5 seconds. We used standard Dutch sentences (VU98; see Versfeld *et al.*, 2000), uttered by a Dutch female speaker and equated for their RMS-values. Two background noises have been used: continuous car noise (cont) and fluctuating speech-shaped noise (fluct, Festen and Plomp, 1990). The speech and noise were presented at a signal-to-noise ratio (SNR)

of 0 dB. This SNR was assumed to be appropriate for judging the listening comfort of the speech-in-noise samples used for the multi-directional search.[43]

The stimulus material was processed off-line. The speech samples were preceded by a fixed period of 0.5 seconds of silence and the length of the silence period after the sentence depended on the length of the sentence, creating a sample of four seconds. Likewise, four-second noise samples were cut and added to the speech samples at a signal-to-noise ratio of 0 dB. To achieve this, levels of the speech samples were compared to the levels of the specific noise samples in dBA. The A-weighting was used to correct for differences in spectral contents of the two noise types. After the speech and noise were mixed, the signals were processed by the three algorithms.

The order of processing is quite straightforward. The noise reduction was used first, because spectral peaks are more pronounced for higher signal-to-noise ratios and may be processed more successfully after noise reduction. The spectral lift algorithm is applied last, after the enhancement stage, because the lift will be defined better once the signals are enhanced. When the order of processing is reversed, the amount of lift will depend on the amount of expansion. We carefully avoided clipping of the signals in the recording. To avoid loudness becoming a cue when samples are compared pair wise in the experiments, the samples were equalized after processing with respect to their root-mean-square (RMS) values.

Before running the optimisation procedure the 44100-Hz sampled signals were stored on disk. During the experiments these monophonic signals were transferred to a signal-processing hardware system (stage 1), filtered using an experimental hearing aid (stage 2), transferred back to the signal-processing hardware system (stage 3), and finally presented through headphones (stage 4).

In the first stage the monophonic signal was fed into channel 1 of the signal-processing hardware system (Tucker Davis Technology TDT system). An appropriate DC-voltage was fed into channel 2 of the signal-processing hardware system in order to create positive voltages needed for the experimental hearing aid in stage 2. After that, the signals of both channels were transferred via the hardware modules in different ways. That is, the AC-part was converted to analogue (using a TDT DA3-unit), filtered by an anti-aliasing filter (TDT FT5) with a cut-off frequency of 16 kHz and a slope of 48 dB/oct, and attenuated 18 dB by a programmable attenuator (TDT PA4). The DC-component was also converted to analogue (TDT DA3) and summed with the attenuated AC-component (using the TDT SM3 signal mixer).

In stage 2, the signal with the appropriate DC-offset was transferred to a flexible experimental hearing aid (DASi-system; Rass and Steeger, 2000) to filter the signal

[43] In order to check this assumption, the Speech Reception Thresholds (SRT) of six subjects were measured. This threshold represents the signal-to-noise ratio at which 50% of sentences in noise are repeated entirely correctly (Plomp and Mimpen, 1979). On average, the subjects' SRT is -3.6 (1.6) and -9.1 (2.5) dB for two standard noises, continuous speech-shaped noise and fluctuating speech-shaped noise, respectively. This corresponds well with results of Festen and Plomp (1990) and is substantially lower than the SNR of 0 dB for the speech-in noise samples used in this study. The SNR for continuous car noise, which was not measured here, can be expected to be lower than the SNR for continuous car noise due to the fact that continuous speech-shaped noise will mask speech more efficiently than car noise. So, the paired comparisons were performed for speech-in noise presented at levels for which the percentage correct score was well above 50%.

following a predetermined frequency shape. We applied the correction factors described by Dillon (1986) to compensate for the presentation of signals by headphones instead of by a behind-the-ear hearing aid. In order to use this study as an appropriate reference for future studies with hearing-impaired subjects the normal hearing subjects in this study receive slight amplification for certain frequencies as well. The NAL-prescription rule was used for this purpose (Byrne and Dillon, 1986). In effect some frequency shaping was applied.[44]

After this filtering, the signal was transferred back to the signal-processing hardware system in stage 3. In this stage the signal is amplified with a microphone amplifier (TDT MA2) and buffered by a headphone-buffer module (TDT HB6). Finally, in stage 4 the signal was presented monophonically by TDH-39P headphones to the subjects seated in a sound-attenuating booth. The subjects were using their 'better' ear; three subjects used their left, six subjects their right ear.

For experiment 1 the reliability of the multi-directional pattern search was determined. The reliability is the test-retest reproducibility for the pattern search. The reliability was calculated by the three-dimensional distance between the optima found in the test and in the retest condition. For experiment 2 two measures were determined. First the reliability was calculated for different pattern search designs that comprised different step size configurations. Second a convergence measure was determined. Convergence relates to the ability of the pattern search to find the optimum. The degree of convergence was studied by comparing the optima found for each step size configuration. That is, for each subject the average optimum setting (avg2) was determined for each step size configuration. After that the results were compared with the average optimum setting for the continuous noise condition of experiment 1 (avg1). Finally, the three-dimensional distance between avg1 and avg2 was calculated.

4.6.2 Results

Experiment 1. Pattern search optimisation with adaptive step size

A Statistics

The optimisation procedure ended most often on meeting two stop criteria simultaneously. These were that a setting was frequented three times and that a setting has won a paired comparison more than six times. This might indicate that there is a close relation between these stop criteria. The total number of paired comparisons depended on subject and on noise type. To determine the significance of these differences, Wilcoxon (matched pairs signed rank sum) tests were performed. For continuous noise the number of paired comparisons was significantly lower (Wilcoxon, $p<0.05$) than for fluctuating speech-shaped noise, on average they were 30 and 38, respectively. In the paired comparisons the second alternative is used not significantly more frequently than the first alternative, although the second alternative tends to be used more often.

[44] The effect of the filter was very small. For five subjects (s1 to s5) we also performed six pattern searches for continuous noise with no correction. The results showed that the optima found differed only slightly for the spectral lift algorithm, as might be expected. Without correction the preferred amount of spectral lift was on average one scale unit smaller.

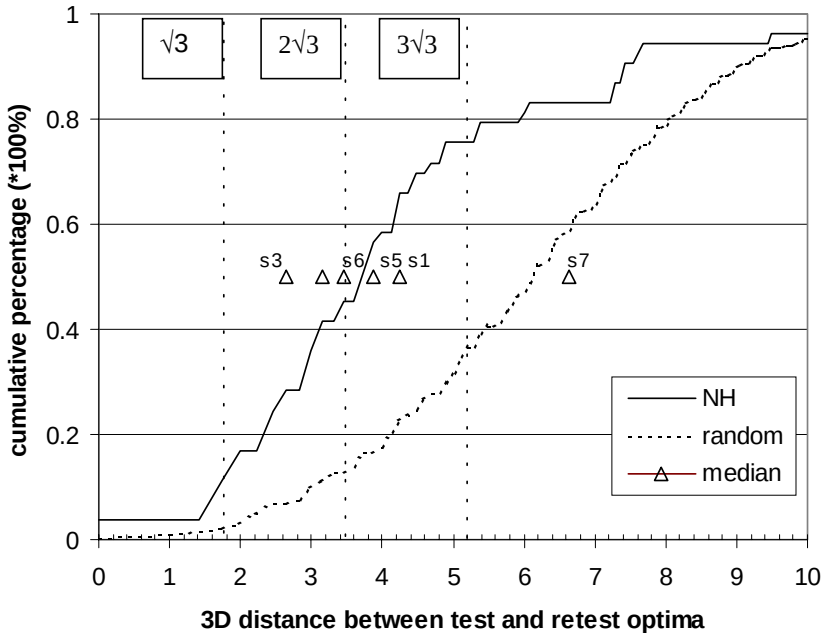


Figure 4.20 Cumulative distributions for the calculated 3D distances between the test and retest optima, averaged for two noise types and seven normal hearing subjects. As a reference, the random distribution is also given as calculated based on randomly assigned test and retest optima. The triangles represent the median 3D distance values for each individual (s1 to s7).

The relation between use of the 'repeat' and 'no difference' button was also subject dependent. For most subjects the 'repeat' button was used much more frequently than the 'no difference' button. So, these subjects did not seem to have frequent problems in distinguishing the samples, irrespective the step size. Two subjects used the 'no difference' button relatively often, probably because they sometimes had difficulties hearing differences between the samples. One subject almost always used the repeat button when there was no perceived difference. For the fluctuating noise relative to the continuous noise, the 'repeat' button was used twice as often and the 'no difference' button a factor four more often.

After a start-up round with a step size of 2, the step size was halved to 2. From this point on the step size was recorded to determine the minimum step size and maximum step size used in the search. The minimum step size could either be 1 or 2. The maximum step size could be either 2 or 3. The results showed that the minimum step size per dimension was comparable for both noise types and had a value of 1 for approximately 80% of all paired comparisons. However, the maximum step size was noise dependent. A maximum step size of 3 was found in 7% and 27% of the cases, for continuous noise and fluctuating noise, respectively. This large difference was found to be mainly caused in the spectral enhancement dimension. Obviously, spectral enhancement settings may be harder to discriminate in fluctuating noise.

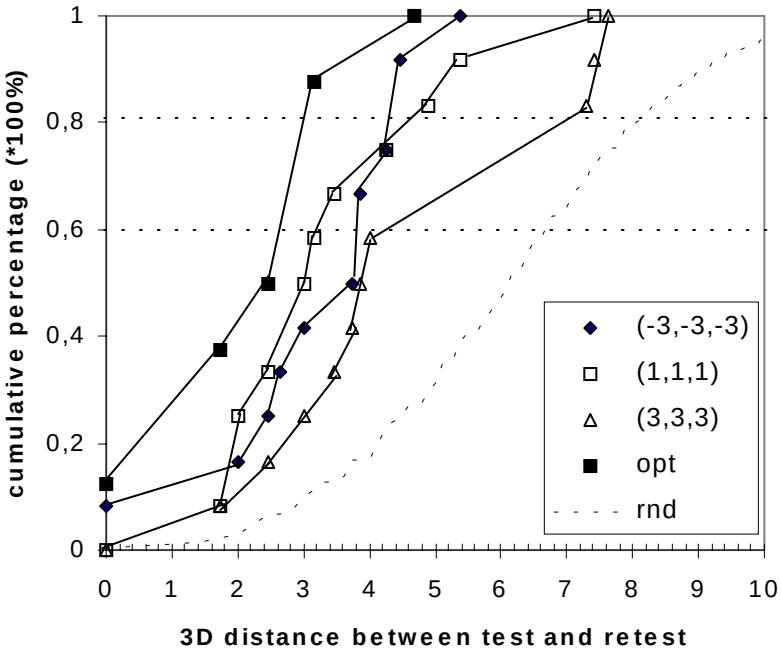


Figure 4.21 Cumulative distributions for the calculated 3D distances between the test and retest optima, averaged for six subjects and differentiated by initial setting. The initial setting was (SE-L_{-3},SL_{-3}, NR-L_{-3}), (SE-L_1,SL_1,NR-L_1), (SE-L_3,SL_3,NR-L_3), or the average optimal setting (opt). As a reference, the random distribution is also given as calculated based on randomly assigned test and retest optima.

B Cumulative distributions

The three-dimensional (3D) distance between the optima found for the test and retest conditions has been calculated in order to determine the reliability of the optimisation procedure. Basically, the procedure has maximum reliability when the test and retest results always yield the same optimum settings for all conditions. Cumulative distributions were used in order to analyse the spread of possible test-retest differences. The cumulative percentages were calculated as a function of the maximum distance between test and retest optima. These percentages are given in figure 4.20. From this graph it appears that 11%, 45%, and 76% of the data have a test-retest distance smaller than $\sqrt{3}$, $2\sqrt{3}$, and $3\sqrt{3}$, respectively. These distances represent a maximum step size of 1, 2 or 3, in each dimension. As a reference the random distribution is also given, presented as a dashed line. For this random distribution 1000 test and retest results have been chosen randomly, after which the 3D distances have been calculated. The cumulative distribution of the normal hearing subjects deviates clearly from the random cumulative distribution. The cumulative distribution of the normal hearing subjects contains a few distances that are very large. This is mainly due to the results of subject 7. To illustrate the large test-retest differences, we have plotted the median 3D distance for each individual (see figure 4.20), which shows that the performance of s7 was very close to random.

To study the dependence of the reliability on noise type and initial setting the results of s7 have been discarded. For different noise types, the cumulative distributions were found to be approximately equal. For different initial settings, however, differences in the distributions appeared, as shown in figure 4.21. Especially for percentages around 80% differences occur. In general, the test-retest reliability was highest when the average optimal setting (opt) was chosen as the initial setting of the procedure. Initial setting ($SE\text{-}L_3$,SL_3,$NR\text{-}L_3$) produced the lowest reliability.

The one-dimensional (1D) distances were calculated for each separate dimension. These results are presented in figure 4.22. It appears that the test-retest reliability depends strongly on the algorithm. The reliability was highest for the noise reduction algorithm and poorest for spectral lift. Additionally, the one-dimensional reliability depended on noise type. The reliability for SE_L and NR_L was highest for continuous noise, whereas for SL the reliability was highest for fluctuating speech-shaped noise (not shown in figure 4.22).

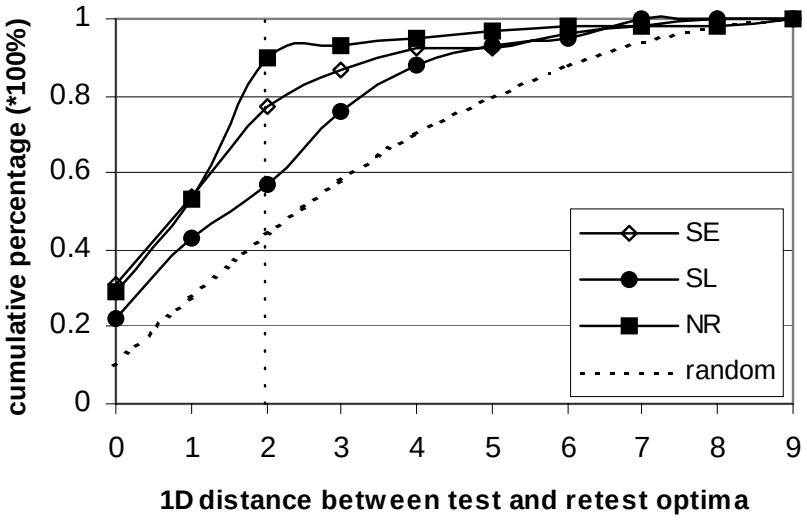


Figure 4.22 Cumulative distributions for the calculated 1D distances, i.e. the difference between test and retest optima for each dimension separately. The dimensions of the optimisation procedure represent the algorithms Spectral Enhancement (SE_L), Spectral Lift (SL), and Noise Reduction (NR_L). For comparison, the random distribution is also given.

C Cluster analysis

A cluster analysis has been performed to investigate the spread of optima found for different conditions in the three-dimensional space of algorithm settings. The method of hierarchical clustering has been used, based on nearest neighbours and Euclidean distances. The results of the cluster analysis are presented in table 4.12. The numbers of final settings that are included in the cluster analysis are indicated in the second column (total). The criterion to assign different optima to the same cluster is the average distance between the cluster centre and the other points, which was be lower than $\sqrt{3}$. The largest cluster is assumed to be the optimum. The one but largest cluster is considered a suboptimum. The algorithm settings that represent the optimum or suboptimum were

rounded values of the cluster centre. The number of winners (# winners) that form the optimum and suboptimum are also included in table 4.12. Clusters that contain only one or two final settings are considered to be outliers and are given in the last column (# outliers).

For subjects s1 to s6 at least 50% of the final settings is grouped in one cluster. So, for the majority of the subjects the optimisation strategy can find an optimal combination of algorithm settings. For s1, s4, s5, and s7 there is also a second cluster of reasonable size. For s1 and s4 this is a true suboptimum, i.e. a serious competitor for being most comfortable. The presence of two large clusters for subjects s5 and s7 appear to be optima that relate specifically to the two noise conditions.[45] The percentage of outliers is 25% or less for all but one subject. Subject s7 has a large scatter among the final settings found by the optimisation procedure.

There proved to be a clear relation between the number of paired-comparisons needed and the size of the clusters. For the large and medium optimum and suboptimal clusters the procedure stopped on average after 25 and 30 paired comparisons, respectively. For the outliers, 54% of the procedures that were performed took 45 to 54 paired comparisons, which is close to, or at, the upper limit of paired comparisons needed.

			Cluster 1: optimum				**Cluster 2: suboptimum**				**Outliers**
Subject	**#searches**	**#clusters**	**SE**	**SL**	**NR**	**#winners**	**SE**	**SL**	**NR**	**#winners**	**#winners**
s1	**16**	**4**	-1	5	0	11	0	1	0	3	2
s2	**16**	**4**	1	4	0	12					4
s3	**16**	**3**	-1	-1	0	14					2
s4	**16**	**6**	0	1	0	8	0	-3	1	4	4
s5	**14**	**4**	0	3	2	7	-1	-2	1	4	3
s6	**14**	**4**	-1	1	1	10					4
s7	**16**	**8**	5	-3	5	5	-1	-3	6	4	7

Table 4.12 Cluster analysis of pattern search results for seven normal hearing subjects (s1 to s7). The total number of points (total) included in the analysis and the number of clusters (#clusters) that result from the analysis are given in the second and third column. The two largest clusters are assumed to form the optimum and suboptimum. The algorithm settings of the optimum and suboptimum are indicated. SE_L, SL, and NR_L represent algorithms Spectral Enhancement, Spectral Lift, and Noise Reduction. Clusters that contain only one or two final settings are called outliers. The number of points (#points) is given for the optimum cluster, the suboptimum cluster and the clusters of outliers.

D Direction analysis

To verify whether the executed movements in the optimisation procedure were performed correctly, i.e. whether a step was in the right or wrong direction, we designed a direction analysis. By means of this analysis we tried to determine the consistency of

[45] When for each noise type cluster analyses were carried out for each individual, the same conclusion could be drawn. Moreover, it appeared that the number of optima that form the largest cluster and the number of outliers were comparable for subjects s8 and s9 and the other seven subjects. Hence, it is assumed that the results of the five subjects in experiment 2 can be compared safely amongst each other and with the results of experiment 1.

the movements in the pattern search optimisation strategy. In this analysis we compared all executed steps with the expected directions based on the assumed optimal setting. That is, for each dimension the percentage of matches between the executed and the expected steps were calculated. The analysis took two features of the optimisation strategy into account. Both the 'direction' of paired comparisons and the direction of movement need to be considered.[46] For each dimension the assumed optimal setting is chosen systematically from settings -3 to 6, for each dimension separately. Each setting is used as a reference. That is to say, for all settings that are visited in the procedure the expected direction of movement is determined based on the position of the reference setting. And for all settings that were visited, the percentage of matches was calculated.

The hypothesis is that the percentage of match is highest for either the average optimal setting, or for the cluster centre. This direction analysis was performed for each subject and each noise type. Hence, the percentage of successful movements was calculated for eight different procedures, i.e. for test and retest and for four starting points.

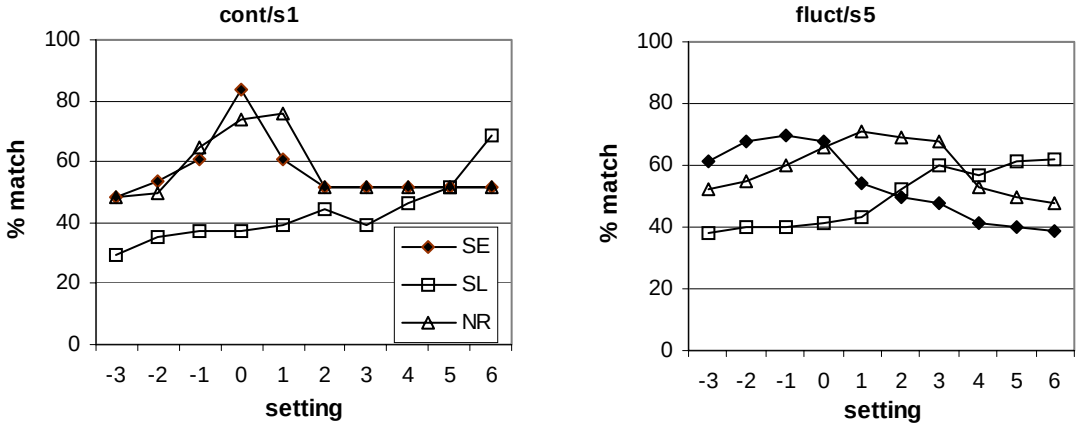


Figure 4.23 Percentage of matches between the executed and the expected steps in the pattern search as a function of algorithm setting, based on the direction analysis. The curves represent the response pattern for each algorithm separately, which are Spectral Enhancement (SE_L), Spectral Lift (SL), and Noise Reduction (NR_L). The response patterns are given for subject 1 in the continuous noise condition (cont/s1, left panel), and for subject 5 in the fluctuating noise condition (fluct/s5, right panel).

In figure 4.23, representative examples are given of the outcomes of the direction analysis. The shifting reference settings yield a response distribution for each dimension.

[46] When a neighboring setting is judged less comfortable than the setting that was judged best so far, the direction of comparison is called unfavorable. In that case the optimisation strategy prescribes that for the dimension under consideration there should be no step. As a consequence, the direction of comparison is reversed for the next paired comparison. In the other situation, i.e. when the neighbor is more comfortable than the best setting so far, the direction is called favorable. This information is included in the estimation of the quickest route towards the optimum, along with the outcomes of the paired-comparisons for the two other dimensions. There is a match when the measured direction of movement is equal to the expected direction of movement. Note that the expected and measured step can be both equal to 0 as well.

Taken together the peaks yield a combination of algorithm settings for which the percentage of match is maximal. For subject 1 in the continuous car noise condition the estimated optimum is ($SE\text{-}L_0$,SL_6,$NR\text{-}L_1$). For subject s5 in the fluctuating speech-shaped noise the estimated optimum is ($SE\text{-}L_{-1}$,SL_6,$NR\text{-}L_1$). For these two examples, there is a marked difference between the reliability of the estimations of the optimum based on the movements in the procedure. For s1, the peaks in the distributions are narrow, whereas for s5 the peaks are very broad.

The direction analysis results for subjects s1 to s7 are presented in table 4.13. On average 75 and 90 movements were included in the direction analysis for continuous and fluctuating noise, respectively. This reflects the higher average number of paired comparisons that was needed to end the optimisation procedure for fluctuating noise.

			setting			match		
		count	SE	SL	NR	m(SE)	m(SL)	m(NR)
Cont.	***s1***	54	0	6	1	83	69	76
	s2	58	1	5	-1	76	68	71
	s3	80	-1.5	-3	0	69	63	81
	s4	87	-1	-3	1	62	62	74
	s5	90	-1	-3	1	67	67	77
	s6	79	0	0	1	72	65	77
	s7	78	6	-3	6	64	74	77
	avg	75.1				70.4	66.9	76.1
Fluct.	***s1***	64	-3	6	-1	75	67	64
	s2	74	3	5	-1	68	60	72
	s3	98	-2	-1	1	67	64	66
	s4	92	-3	1	1	61	60	70
	s5	100	-1	6	1	70	62	71
	s6	90	-1	2	2	69	69	69
	s7	111	-3	-3	6	61	70	64
	avg	89.9				67.3	64.6	68.0

Table 4.13 Direction analysis of pattern search results for seven normal hearing subjects (s1 to s7) for continuous noise (cont) and fluctuating noise (fluct). In the second column the number of performed movements (count) is indicated that is used for the analysis. In the third column the setting is indicated for which the percentage of matches is highest. SE_L, SL, and NR_L represent algorithms Spectral Enhancement, Spectral Lift, and Noise Reduction. In the fourth column the maximum percentages of matches are shown for each dimension. Average counts and average matches (avg) are given in the third and fifth row.

On average, the maximum percentage of matches is highest for the noise reduction algorithm. The average minimum percentage of the distribution for each dimension is around 40%, which corresponds to the probability of 0.4 when movements are performed randomly in the procedure.[47]

To determine whether the movements in the procedure have been executed consistently or rather randomly, we assumed that the more peaked the distribution is, the more reliably the movements resulted in finding the optimum. As a consequence a

[47] Due to a reduction in possible movements at the boundaries, the random probability is 0.5 at the extreme settings 0 and 9 of each dimension.

perfectly flat distribution, i.e. a distribution for which the percentage of matches is equal for all settings, was expected to be the result of random movements. Therefore, the second derivative near the maximum of the distribution was calculated by taking the difference between the slopes around the maximum. Additionally, the range of percentages of matches has been expected to influence the randomness of the movements. For a small range of percentages, i.e. when the difference between the maximum and minimum percentage is small, we expect that the optimum setting will not be found easily. The combination of these two measures has been used to determine the consistency of the movements in the optimisation procedure. Hence, to distinguish peaky from flat distributions, the second derivative and the range have been multiplied. These peakyness values, calculated for all subjects and different noise types, are presented in table 4.14. For subject s1 and the continuous background noise, the peakyness measure is equal to 7.8, 3.2, and 3.6 for the SE_L, SL and NR_L dimension, respectively (see figure 4.23 and table 4.14). For subject s5 and the fluctuating background noise, the peakyness measure is equal to 0.6, 0.1, and 0.8 for the SE_L, SL and NR_L dimension, respectively (see figure 4.23 and table 4.14). The distributions of subject s1 in continuous background can be defined peaky, whereas the distributions of subject s5 in fluctuating noise can be defined flat.

Noise	Subject	SE	SL	NR
Cont	***s1***	7.8	3.2	3.6
	s2	6.8	4.8	4.0
	s3	1.4	0.6	5.9
	s4	0.7	1.4	3.2
	s5	4.8	5.2	10.8
	s6	1.7	0.7	5.0
	s7	2.9	10.0	7.4
	avg	3.7	3.7	5.7
Fluct	***s1***	8.6	2.1	1.3
	s2	1.6	0.5	3.3
	s3	2.3	0.6	1.2
	s4	0.2	0.1	0.8
	s5	0.6	0.1	0.8
	s6	1.9	2.2	0.8
	s7	0.8	5.8	1.7
	avg	2.3	1.6	1.4

Table 4.14 Measure of peakyness as a function of subject (s1 to s7) and noise type for each dimension separately. SE_L, SL, and NR_L represent algorithms Spectral Enhancement, Spectral Lift, and Noise Reduction. The peakyness values were calculated by multiplication of the range of percentages and the second derivative near the maximum percentage of matches. Average peakyness values (avg) are given in the third and fifth row for continuous (cont) and fluctuating noise (fluct), respectively. Values lower than 1.5 are presented in grey.

We can distinguish flat distributions from peaky distributions by defining a boundary value. For a peakyness value larger than 1.5 the distributions appear to be rather peaky, for values lower than 1.5 the distributions appear to be rather flat. Flat distributions have

been marked in table 4.14, based on values lower than 1.5. The peakyness of the distribution is on average highest for NR_L in continuous noise and SE_L in fluctuating noise. In continuous noise the peakyness is lowest for SE_L and SL, whereas in fluctuating noise the peakyness is lowest for NR_L. However, the values depend on the individual. The results indicate that almost twice as many of the distributions are peaked for continuous noise than for fluctuating noise.

Experiment 2. Pattern search with step size as parameter

The results of all four step-size configurations (1≤step≤3 is step123, 2≤step≤3 is step23, 3≤step≤3 is step3, 4≤step≤4 is step4) are plotted as cumulative distributions in figure 4.24. For comparison, the random distribution, based on randomly assigned test and retest optima, is plotted in the dashed line. The random distribution appeared approximately the same for all step size configurations (for more details see appendix 5).

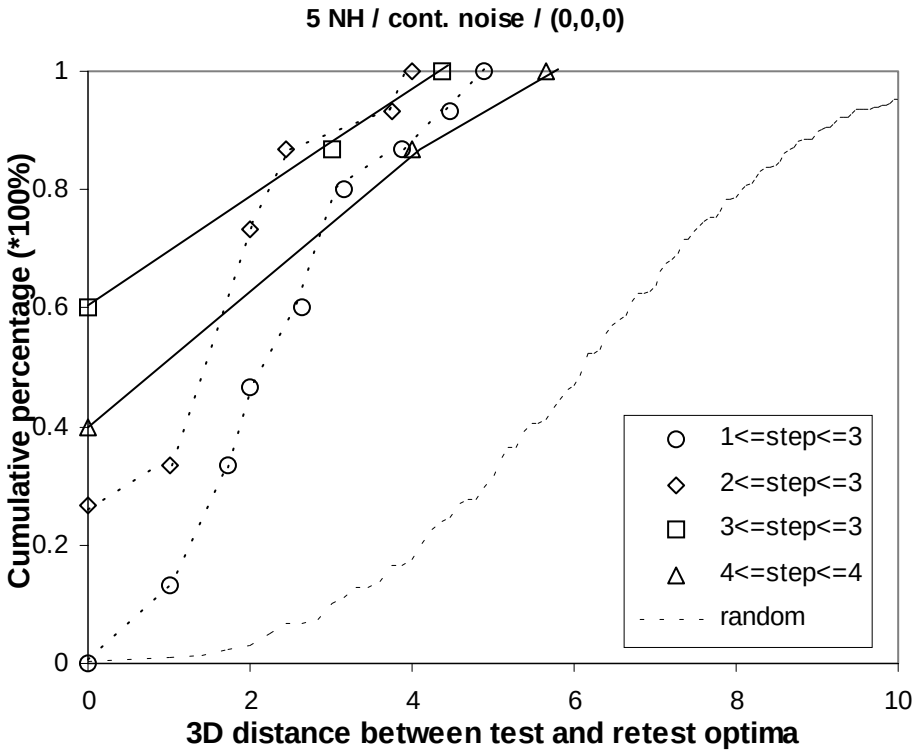


Figure 4.24 Cumulative distributions for the calculated 3D distances between the test and retest optima, averaged for five normal hearing subjects (5NH) and differentiated by step-size configuration. The pattern searches were performed for continuous noise (cont. noise) and initial setting ($SE\text{-}L_0$,SL_0,$NR\text{-}L_0$) only. The step size can be either adaptive and vary between 1 and 3 (1≤step≤3), or between 2 and 3 (2≤step≤3), or the step size can be fixed at 3 (3≤step≤3) or fixed at 4 (4≤step≤4). As a reference, the random distribution is also given as calculated based on randomly assigned test and retest optima.

The results show that the reliability of the procedure tends to increase when the minimum step size increases from one to three (for step123, step23, and step3, respectively). For a fixed step size of four the reliability decreases relative to a step size

of 3. Both for the fixed (solid lines) and for the variable step size configurations (dotted lines) there is a parallel shift of the distributions. This shift indicates an increase in reliability for the variable configurations and a decrease for the fixed configurations when the step size is enlarged.

The degree of convergence was studied by comparing the optima found for each step size configuration. That is, for each subject the average optimum setting (avg2 of experiment 2) was determined for each step size configuration. After that the results were compared with the average optimum setting for the continuous noise condition of experiment 1 (avg1). Finally, the three-dimensional distance between avg1 and avg2 was calculated. The average results indicate that the 3D distance increases with minimum step size, at least for minimum step sizes between 1 and 3. The 3D-distance is 1.0, 1.8, 2.0, and 1.7 for configurations step123, step23, step3, and step4, respectively. There is a significant difference (Wilcoxon matched pairs signed rank sum) between step123 and the step3 configuration, which indicates that the optimisation procedure converges better for the variable step size configuration.

4.6.3 Discussion

Adaptive pattern search

An advantage of the pattern search method is that no a priori information is required. For perceptually driven optimisation that depends on the listener's capacities and experience this method shows promising capabilities. However, the perceptual capacities of the listeners also constrain the methodological possibilities of the optimisation method. The adaptive step size that is used in this study helps to create perceptually distinguishable samples to be used in the paired comparisons. The results of the pattern search procedure suggest that settings that differ 2 or 3 steps can be judged reliably for our set of stimuli. That is, the reliability of the optimisation procedure is high for a variable step size that ranges between 2 and 3, and for a fixed step size of 3.

The optimal settings have been estimated in three different ways considering all pattern searches for each listener. The estimators comprise the average optimum setting (O_{avg}), the centre of the largest cluster (Cl), and that combination of algorithm settings that has the highest percentage of matches between the estimated and recorded movements in each direction (Dir). The three-4dimensional distance between the optimum setting and the other two estimators can be determined. The individual results, average results, and standard deviations are presented in table 4.15 for each background noise type. The results of this study suggest that for both noise types the correspondence between the cluster centre and the average optimum setting is close for all but one subject. This can be seen from the small average 3D-distances that are smaller than $\sqrt{3}$ for all but one subject (distances larger than $\sqrt{3}$ are presented in grey in table 4.15). Although the results presented in table 4.15 cannot simply yield information about the quality of the estimators, the cluster centre of the largest cluster is assumed to be the best estimator of the optimum. For in some pattern searches (for different tests and starting points) the 'real' optimum can be missed when there are grave discontinuities in the three-dimensional perceptual landscape or due to procedural inaccurateness as a result of step sizes used.

Noise	Subject	3D-distance d(Cl,Dir)	d(CL,Oavg)	d(Oavg,Dir)
Cont	***s1***	1.0	1.3	0.5
	s2	1.7	0.2	1.6
	s3	2.0	0.2	1.8
	s4	1.0	3.2	3.2
	s6	1.0	0.3	0.8
	s7	1.0	0.4	0.8
	avg	1.3	0.9	1.5
	stdev	0.5	1.2	1.0
Fluct	***s1***	3.7	3.1	0.7
	s2	2.8	0.3	2.5
	s3	2.2	0.3	2.3
	s4	3.0	0.6	2.8
	s6	1.0	0.5	0.7
	s7	2.0	0.5	1.5
	avg	2.5	0.9	1.8
	stdev	0.9	1.1	0.9

Table 4.15 Three-dimensional distances between the cluster optimum (Cl) and the winner of the direction analysis (Dir), three-dimensional distances between the cluster optimum and the average optimum (Oavg), and three-dimensional distances between the average optimum (Oavg) and the winner of the direction analysis (Dir) as a function of subject (s1 to s7) and noise type. The noise type was either continuous car noise (cont) or fluctuating speech-shaped noise (fluct). Additionally, the average three-dimensional distances (avg) and standard deviations (stdev) are indicated. Values larger than √3 are presented in grey.

The correspondence between Cl and Dir for fluctuating speech-shaped noise is consistently poorer for all subjects. The large differences that occur between the best match (Dir) and the cluster centre (Cl) are probably caused by the mutual independency of the dimensions for the direction analysis, where we expect that the perceptual decisions are influenced by effects of interactions between algorithms.

Reliability & convergence

The cumulative distributions of all subjects except for s7 indicate that the three-dimensional distance between test and retest is lower than 2√3 for about 50% of the results. The cluster analysis illustrates the presence of large and intermediate clusters around the optimum, small clusters around a suboptimum, and some outliers.

In studies of Neuman *et al.* (1987), Kuk and Pape (1992), and Kuk and Lau (1996) the reliability of the modified 'pattern search' procedure was higher than in this study. The most important explanation for that may well be connected to the configuration and the stimulus material of the optimisation procedure. In our set-up the subjects listened to sentences that were between 1.5 and 2.5 seconds in duration. Moreover, in our experimental paradigm speech-in-noise samples were presented in succession and could be repeated only once. A problem of this design is that the auditory memory is not sufficient to compare the samples: at the end of sample 2, a part of sample 1 may be

already forgotten. In the studies mentioned above the paradigm is very different in that continuous discourse is used and listeners are allowed to switch freely between two parameter settings. By switching freely, the effects of auditory memory are less acute, because of the possibility to return to the alternative. In a small parameter or algorithm space such a paradigm appears to be convenient, but for large multidimensional set-ups, like in this study, such a slow approach is unpractical.

Another, maybe more influential difference is the level equalization of the stimulus material in our study. For the study of Neuman *et al.* (1987) the loudness of low frequencies and high frequencies of the stimulus material was one of the optimisation parameters under study. For the studies of Kuk and Pape (1992) and Kuk and Lau (1996), the parameter settings of the two-dimensional optimisation procedure were the degree of amplification used for high and low frequencies. However, because the subjects had to adjust the level of the test material to their most comfortable level (MCL), the effective levels can be quite different for different settings. The advantage of these approaches is that the reliability is high because level differences are relevant cues for converging on the optimum setting. An additional advantage is that (parts of) the stimulus material can never be too loud or too soft. Our optimisation procedure was intended to focus on the effects of the algorithms themselves and not on level differences. The reason for this is that loudness differences can act as a separate and different evaluation criterion. The procedure used in this study avoids that the optimum setting found is the result of only a loudness judgment.

The choice of the background noise type in combination with the selected signal processing algorithms also affects the reliability. The direction analysis shows that the performed movement towards the optimum is more consistent for car noise. For fluctuating speech-shaped noise the performed movements appear to be much more random, probably due to the flat distributions of the direction analysis. An explanation could be that the differences between settings, in the paired comparisons in the optimisation process, are much more difficult to detect for the fluctuating noise than for the continuous noise. Subjects indicated that processing artefacts are much more prominent for the continuous car noise condition relative to the fluctuating shaped noise condition.

When the maximum number of paired comparisons was reached the end point is not likely to be the optimum, and the setting that has won most paired comparisons is in most cases not the optimal setting either. Conversely, when the end point is reached rather quickly, within nine to eighteen paired comparisons, this point is much more likely to be the true optimum. The reliability found for each noise type can be directly related to the stop criterion, if addressed. For fluctuating noise, that had a relatively low reliability, the procedure stopped most often when the maximum number of paired comparisons was reached. For continuous noise, the procedure ended both on three visits and more than six wins. This might imply that combining these both stop criteria yields highest reliability.

Auditory constraints

In this study we did not record the difference limens for each particular dimension. Although these limens can be measured for each algorithm separately, that would still not produce complete knowledge of the limens for combinations of these non-linearly acting algorithms. The present study shows that the adaptive step size can partly compensate for any unfortunate choices of the step size. The step size configuration appears to affect the reliability. Apart from the fixed step size configurations, we found that the highest reliability was achieved for a minimum step size of two and a maximum step size of three (step23). A variable step size helps to create perceptually distinguishable samples. In the same line, the elevated minimum step size has an advantageous effect in avoiding the presentation of indistinguishable samples, which is not only difficult for the subjects, but may also reduce their motivation and alertness.

The discrepancy between the peaked distributions in the direction analysis and the close to random cumulative distribution for s7 is difficult to explain. The consistent movements in the searches may possibly be ascribed to different evaluation strategies used by this subject. That is, when different strategies are consistently used in different runs of the search it is possible that the reliability across the searches is low, while the pattern of the direction analysis is nicely peaked.

In the studies of Neuman *et al.* (1987) and Kuk and Pape (1992), it seems that the possible existence of multiple optimum settings played a minor role. That is, the combination of the parameters under study yielded a response surface that curved smoothly around a single optimum. Yet, in the study of Kuk and Lau (1996) category rating results showed a multimodal response pattern for three out of seven subjects, when evaluated for clarity for SNR=0 and SNR=5 dB. Also the results of a previous pilot study indicated that the response surface could sometimes be multimodal. In this study, the cluster analyses imply the presence of multiple optima. The direction analysis is unsuitable for this purpose, because it presupposes one optimum. Additionally, it does not take small irregularities and flat areas into account. As a consequence the direction analysis does not reckon with the possibility that a longer route can be necessary – due to the peculiarities in the perceptual landscape – to reach the optimum setting. An even more fundamental factor is the inability of the direction analysis to take algorithm interactions into account. For subjects s4 and s6 such interactions seem to have played an important role.

4.6.4 Conclusions

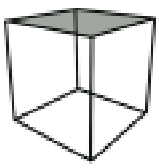
- The inclusion of an adaptive step size has the advantage that the perceptibility of differences between samples can be controlled
- The reliability is reasonably high for the adaptive multi-directional pattern search
- The reliability of the adaptive step size depends on many factors like the initial value, stop criterion, step size constraints, background noise, algorithms, and the presence of suboptimal settings
- The reliability improves for larger step sizes, whereas the convergence deteriorates

Chapter 6

METHODOLOGICAL ISSUES

'So the bottom line here is while we present in listeners the set of physically identical stimuli, stimuli that were equally spaced in a grid, neither of the two groups perceive sound invariably, neither of them show perceptual maps that correspond to the real physical world. In both cases perception is warped: American space is warped to highlight the difference between /r/ and /l/, very good for perception. The Japanese world is structured so as to unify /r/'s, and not show magnet effects nor boundaries because this is one integrated category. We don't see a clustering because we have not, in this set, found an 'ideal' for the Japanese point of view.'

– Patricia Kuhl
Speech perception (JASA video)



In this chapter a review is given of the methodological aspects of the studies that are described in chapters 4 and 5. This leads to a general discussion in which reasons are proposed to explain the results that are found.

In our first study, described in section 4.1, CVC word scores for two combinations of spectral enhancement and phonemic compression were found to be not better than for spectral enhancement alone. It was assumed that combining the best configurations of the two algorithms led to even better CVC scores. There are at least two reasons to question this hypothesis. The order of processing could have been chosen not optimally, i.e. by reversing the order the CVC scores do improve. Additionally, and maybe even more importantly, it is possible that there is interaction between the two non-linearly operating algorithms. This would imply that combining the optimal settings for each algorithm separately does not necessarily result in the best scores for combination of algorithms. If this is the case, this finding will have important implications for the current practice of hearing-aid fitting in which each individual signal-processing strategy is usually tuned successively. When we assume that many signal-processing strategies act non-linearly, it would be more appropriate to fine-tune multiple algorithms simultaneously. Therefore, the methodological target of this work is to find an optimisation strategy that acts multi-directionally, i.e. the multi-directional pattern search.

The results of our studies suggest that finding the optimal combination of multiple signal-processing strategies is not an easy and unproblematic task. Additionally, it appeared that the pattern search strategy used in our studies is not simply adaptable for every signal-processing strategy, or for every combination of signal-processing strategies.

In this chapter, we will evaluate a number of methodological issues. These are the presence of algorithm interactions, the reliability of the pattern search, and underlying procedural, cognitive, auditory, and acoustic reasons to explain the moderate reliability. Additionally, alternative approaches are discussed that might be more appropriate to optimise multiple signal-processing strategies simultaneously.

6.1 Algorithm interaction

The use of the multi-directional pattern search has been designed to take algorithm interaction into account (see figure 1.1). When interaction is absent, the algorithms can be fine-tuned successively as well. This would be much less complicated and demanding. The main problem is that the presence of interactions is rather unpredictable. That is, the interactions that matter are the ones that appear in the perceptual domain. Therefore, it is relatively hard to predict a priori whether interactions will appear by a specific combination of signal-processing strategies.

However, based on a-posteriori data, obtained in this study, we can try to trace algorithm interactions. A straightforward approach would have been to both use the pattern search three-dimensionally and for each algorithm separately (one-dimensionally) and in succession. Instead, we used a round-robin strategy to study the response distributions of systematically varied algorithm settings. To verify whether the optimal setting of algorithm x1 depends on the combination of algorithms x2 and x3 we used the edges of the cube studied in section 4.2. Analyses of the estimated winner and the types of response distributions found by the round-robin tournaments did reveal interactions

between algorithms for hearing-impaired subjects. So, interactions seem to exist for the combination of binaural noise reduction, phonemic compression and spectral enhancement. For the other signal-processing strategies the presence of interactions was not verified. However, interactions also appeared to be present between spectral enhancement and spectral lift, as used in section 4.6 (Lyzenga *et al.,* 2002).

The direction analysis was used in section 4.6 in order to verify whether the stop criterion of the pattern search did not yield premature termination. However, the direction analysis could have difficulty in finding the global optimum. One of the reasons is that strange interactions between algorithms are not considered by the direction analysis. That is, small irregularities in the perceptual landscape and the presence of multiple valleys and peaks cannot be accounted for. However, for a smoothly curved response distribution, like the one presented in figure 1.1, the global optimum can be found by the direction analysis of the pattern search results.

Study		NH	HI
4.2		0.37 (n=6)	0.37 (n=6)
4.5	adaptive	0.25 (n=4)	0.10 (n=10)
	fixed=2	0.36 (n=4)	-
5		0.48 (n=5)	0.28 (n=9)
	SIM1	0.41 (n=5)	-
	SIM2	0.40 (n=5)	

Table 6.1 Information transfer values (IT) given for different studies, and differentiated for different subject groups. The subject groups comprise normal hearing subjects (NH), hearing-impaired subjects (HI), and two types of simulated hearing-impaired subjects (SIM1 and SIM2). The number of subjects (n) included in each subject group is also presented. IT values were only derived for continuous noise, the listening comfort criterion, and for test-retest differences in three dimensions. The step size configuration in study 4.5 was either adaptive (step size is 1 or 2) or fixed (2).

6.2 Evaluation of multi-directional pattern search

One way to investigate the appropriateness of the multi-directional search is to determine the reliability of the results. We compared the test and retest results of the studies described in the study of section 4.2, the study of section 4.5, and the study of chapter 5. From these results the information transfer (IT) was derived from the cumulative distributions of the common condition: speech in continuous car noise, evaluated for the listening comfort criterion. The information transfer was derived from the cumulative distributions of three-dimensional distances between test and retest results of the pattern search. It is the ratio between two areas enclosed by cumulative distributions. That is, the area between the cumulative distribution under study and the random distribution, relative to the area between the 'perfect' distribution with a

3-dimensional distance of 0 for all measured data and the random distribution. It should be kept in mind that the recording of the car noise was not identical for the studies. The results are given in table 6.1.

The reliability varied considerably, depending on the study. The normal hearing subjects performed most reliably for the study of chapter 5 (IT=48%), intermediate for the study of section 4.2 (IT=37%), and poorest for the study of section 4.5 (IT=25%). The hearing-impaired subjects performed less reliably; the information was on average 12% lower. The information transfer of the simulated hearing-impaired groups SIM1 and SIM2 was close to the normal hearing group, but slightly lower. The low reliability for the study of section 4.5 probably reflects the constraint that we could select only preset algorithm settings of the hearing aid. Hence, it was not possible to guarantee perceptual differences between algorithm settings. The increase in performance of the study of chapter 5 relative to the study of section 4.2 can be due to the choice of different algorithms. However, it is more likely that the addition of an adaptive step size is the major cause. This revision facilitates the use of comparisons that are distinguishable. The results of studies of section 4.6 and chapter 5 both indicate that the step size configuration affects the reliability. An elevated minimum step size resulted in an increase in information transfer for the studies of section 4.5 and chapter 5 of 11% and 14%, respectively (not shown in table 6.1). The reliability also depended on the starting point. The difference in information transfer could be up to 17% (not shown in table 6.1).

Reliability did not depend much on the noise type used, at least for the tests on listening comfort. However, when the speech intelligibility criterion was used (the study of section 4.2), the information transfer not only decreased (a drop of 21%), but also a strong dependence on noise type was found. Reliability proved to be higher for noises with more fluctuations (see figure 4.8, upper panel). The reliability of the pattern search for the listening comfort criterion did depend on the SNR. For SNR=0 the reliability was higher than for SNR=5.

The reliability depended on each algorithm separately. An information transfer higher than 50% was found for SE_B in the study of section 4.2 for NH and HI, and in the study of chapter 5 for NH only. For the study of chapter 5 the reliability was also very high for NR_L for NH, HI, and SIM1, but not for SIM2. For the study described in section 4.3, which contained predominantly normal hearing subjects, the reliability was high for DRV and BNR. The difference in reliability for BNR found between the study of section 4.2 and 4.3 could be ascribed to the number of settings that were incorporated. Due to the use of only three setting in the study of section 4.3 relative to five settings used for the study of section 4.2, the step size was effectively doubled. Therefore, the increase in information transfer of 17% could be ascribed to the increased step size. The same could apply to the DRV algorithm and less to the PhC algorithm for which 3 and 4 settings were chosen in the pattern search configuration.

Generally, the reliability of the multi-directional pattern search was only moderate. There are many factors to explain this result. Although these factors do not operate independently, we discern four main groups of factors: procedural factors, cognitive

factors, auditory factors[56], and acoustical factors. These are described in separate sections hereafter.

6.3 Procedural factors

Procedural factors that might influence the reliability of the multi-directional pattern search involve the stop criterion, starting point, step size, direction of comparison, movement rules, and boundary rules. We will now discuss these topics in further detail.

Stop criterion

In our studies the pattern search terminated when (a) a setting is frequented three times (b) a setting has won a paired comparison more than six times, or (c) 54 paired comparisons were performed. Given the result that in general the pattern search terminated before the maximum number of paired comparisons was reached, it seems that the other two criteria are appropriate. An alternative approach is to base the stop criterion on the results of a direction analysis (see section 4.6.2, part A). The two approaches essentially differ in designating which setting is optimal. The pattern-search stop criteria presuppose that the final setting is the optimum. This assumption is not made for the direction analysis. For the direction analysis the optimum is determined by considering all movements. If the stop criterion would have been determined by the direction analysis, the information transfer would have been significantly higher (Wilcoxon, $p<0.05$); the increase was 11% for the direction-analysis approach relative to the stop criterion of the pattern search. However, we found that the direction analysis can not cope well with an optimum that shifts during the procedure, due to some sort of reconsiderations of the subject. To date, we have not been able to address this problem. An optimisation strategy that starts from the principle that the end point resembles the optimum can deal with shifting optima in any case.

Starting point

A Reliability

The reliability is the test-retest reproducibility for the pattern search. The results of sections 4.5 and 4.6 indicated that the starting point influences the reliability of the pattern search for normal hearing subjects. In section 4.5 the reliability appeared much higher for starting point (4,4,4) than for (0,0,0), which is the reference setting. The information transfer found for the results in section 4.6 for starting points ($SE\text{-}L_{-3}$,SL_{-3},$NR\text{-}L_{-3}$), ($SE\text{-}L_1$,SL_1,$NR\text{-}L_1$), ($SE\text{-}L_3$,SL_3,$NR\text{-}L_3$) and the average optimum (Oavg) were 56%, 56%, 39%, and 72%, respectively (see figure 4.20). So, when the starting point was chosen near the optimum the pattern search behaves very reliable. A reason that the other starting points yield lower reliability could be the presence of broadly peaked response patterns). The response patterns near setting ($SE\text{-}L_3$,SL_3,$NR\text{-}L_3$) might be typically flat, given the low information transfer for this starting point. Flat response patterns could be due to relative large difference limens or due to the fact that

[56] Auditory factors comprise factors that relate to peripheral hearing system, i.e. the (impaired) cochlea, only.

settings are considered equally well or equally bad. From the study described in section 4.4 it appeared that the difference limen heavily depended on the starting point. Therefore, it would be possible that in one area the subject hardly perceives differences making the pattern search move rather randomly. In another area the differences can be easily detected and the pattern search converges towards the optimum without problems.

B Consistency

The consistency of the multi-directional pattern search was studied in sections 4.2, 4.3, 4.5 by determining the variation of optima found for different starting points. We found quite different consistency values in our experimental results. For the pattern search procedure of section 4.2, variation of the initial setting influences the final estimate for all three dimensions considerably. For the study described in section 4.3 the optimisation method proved to be quite consistent. The consistency of the study presented in section 4.5 was low. The consistency was higher for normal hearing subjects than for hearing-impaired subjects.

Low consistency of the pattern search can be explained by the presence of multi-modal and flat response distributions. That is to say, a high consistency is obtained only for response distributions that are hill-shaped, containing only one clear peak. Apparently, the perceptual space contains either suboptimal settings or the perceptual differences between settings in some areas are rather small. The high consistency found in the study of section 4.3 indicates that response distributions are uni-modal and sufficiently peaked. However, there is no direct evidence for this. In the study of section 4.2 the round-robin results show that multi-modal response distributions were often found. The presence of multi-modal response distributions seems to be the most important explanation of low consistency. This is evidenced indirectly in the study of 4.5. The consistency was found sometimes lower than for random performance, i.e. the information transfer was negative. This could be explained by assuming that, due to a multi-modal distribution, different starting points yielded consistently different (sub)optima. A reason that multiple optima were found could be that the subjects used multiple listening strategies. Some subjects indicated that when the speech intelligibility was the same, they evaluated the clarity and quality of the background noise. This re-cueing may have affected the reliability and consistency.

While the presence of multiple optima (multi-modal distributions) can explain low consistency, low reliability is more likely due to the presence of broad or flat distributions. There are three reasons for this assumption. First in the study of section 4.5 the reliability increased when the step size was enlarged, whereas the consistency did not increase Second, the results of the study presented in section 4.2 showed that the modest reliability of the multi-directional pattern search relates to the coefficient of consistency ζ of the round-robin tournament results. The coefficient of consistency was derived from the number of circular triads found for the round-robin tournaments. The presence of circular triads seems to indicate that the response patterns were broadly peaked or even flat. Third direction analyses were performed in section 4.6 to study the consistency of the movements in the pattern search. By means of this analysis we tried to

determine the consistency of the movements in the pattern search optimisation strategy. When the movements in the pattern search were made randomly the distribution was flat, i.e. a distribution for which the percentage of matches is approximately equal for all settings. The results of the direction analyses showed that twice as many of the distributions were peaked for continuous noise than for fluctuating noise. This is in agreement with the finding that the reliability of two out of three algorithms was higher for continuous noise than for fluctuating noise.

Step size

Levitt (1978) underlined that "when little is known about the spread or location of the psychometric function it is recommended to use a large step size at the start, which is gradually decreasing during the course of the experiment." By this means maximal rate of convergence is secured. However, this idea is based on the assumption that the response curve is approximately linear. This might not be too problematic, because most data were found near the optimum. For the normal hearing subjects in the study described in section 4.6, we calculated the percentages of settings that are near the optimum for that specific pattern search. For continuous noise 48% of the settings that were visited had a three-dimensional distance smaller than $\sqrt{3}$, and 81% smaller than $2\sqrt{3}$. For fluctuating noise these percentages are considerably lower, i.e. 29% and 66%, respectively. The standard deviation between subjects was around 12%.

The studies of sections 4.2 and 4.3 used a fixed step size. The studies of sections 4.5 and 4.6 used a variable step size. For the adaptive pattern search the step size was partly controlled and partly adaptive. That is, the step size was halved in run 1 and after three rounds of paired comparisons and was increased by one, when the subject heard no difference between the samples. The step size was increased by one, when the subject heard no difference between the samples.

It appeared that the reliability was highest when either the fixed step size was sufficiently large (see figures 4.12 and 4.18, upper panel), or when the adaptive step size was used (see figure 4.19). The study of section 4.5 did use an adaptive step size, but the minimum step size hardly yielded perceptual differences, which made the pattern search again unreliable. These outcomes emphasize that the settings to be compared should at least be distinguishable. In some sense, the size of the step gives us some measure of how much we "trust" the information we know about the judgments. The importance of the minimum step size was underlined in section 4.6. The results showed that for high reliability the minimum step size had to be chosen at a value of 2 (see figure 4.23). An increased fixed step size has the disadvantage that only very few settings are involved, resulting in poorer convergence at the 'true' optimum.

Search directions & movements

The lattice structure in the multi-directional pattern search defines the search directions. This implies that paired comparisons are performed always in the same directions. In the algorithms cube the direction of comparison is always left or right, up or down, forward or backward. The search direction, therefore, coincide with variation of one algorithm. A change in orientation of the search directions, e.g. of 45 degrees, can be beneficial to

improve the reliability. When the orientation can be varied in the pattern search peculiarities in the response patterns such as saddle points and ridges can be coped with. When the orientation is changed 45 degrees for all three dimensions permanently, the reliability could improve, because subjects do not focus solely on one aspect, but on the whole percept. All three algorithms are varied simultaneously. The opportunity to use this approach for further improvements is outside the scope of this thesis.

Movements from the starting point towards the optimum follow the lattice structure of the pattern search only in some cases. The movements are directed by the outcomes of three paired comparisons, and are therefore based on the search directions. When multiple neighbouring settings are preferred, the movement is diagonal. Hence, the multi-directional pattern search could, in fact, be viewed as a crude approximation of the method of steepest ascent with finite-difference approximations to the gradient. Subjects indicated that this approach could lead to confusions. For it is possible that subjects get the impression that one of the neighbours is the optimum, but this optimum is left because also other neighbours were preferred. Due to a diagonal movement the optimum will not be present in the next round of paired comparisons. And the optimal setting just heard cannot be retrieved easily. To circumvent this problem a set of three paired comparisons, each for one dimension, could be divided in three and a movement should be made after each paired comparison. However, the crude approximation of the steepest ascent will be lost then.

The results of the direction analysis described in section 4.6 and using the search directions, indicate that the reliability was higher for continuous noise than for fluctuating noise. The information transfer for continuous noise and fluctuating noise was 57% and 36%, respectively. The difference of more than 20% strongly suggests that the movements in the pattern search were in general less consistent for fluctuating noise.

Boundaries

In the studies described in sections 4.2, 4.3, 4.5, and 4.6 the pattern search was constrained to a finite range of algorithm settings for all dimensions. The reason to constrain the range of settings is dictated by the fact that the algorithms could not process the stimuli real time. The advantage of off-line processing was that full automation of the pattern search was possible. Due to the limitation of the number of settings boundaries were created. These boundaries represent the edges of the cube of algorithm settings. When a paired comparison was required with a setting that fell beyond the boundaries, that setting was put at the boundary. To keep the step size the same the other setting was shifted. The direction of comparison was reversed in order to avoid crossing the boundary again.

In the study described in section 4.6 we tried to create boundaries of settings that are not preferred. We succeeded only partly, because sometimes setting -3 and 6 appeared to be the optimal setting (e.g., see table 4.11 for NH, and table 5.2 for HI). For the normal hearing group 18% of all pattern searches yielded SL=-3. For the hearing-impaired group 19% and 44% of all searches yielded SE_L=-3 and SL=-3, respectively. Algorithm setting 6 was hardly ever selected to be optimal. Also for the studies of sections 4.2, 4.3, and 4.5 the boundaries were important, because these settings are serious competitors

for some subjects as well (e.g. see tables 4.5 and 4.9). As a consequence the boundary rules could have influenced the reliability. For a cube the boundaries contain a number of settings 'z' that can be calculated by the equation:

$$z = \frac{x^y - (x-z)^y}{x^y}$$

Parameter x represent the number of settings for each dimension and y represents the number of dimensions. In figure 6.1 the fraction of settings that lie on the boundary relative to the total amount of settings is given as a function of the number of settings selected for each dimension. Parameter is the number of dimensions used. For the set up of the studies described in section 4.2 and 4.5 a 5*5*5 settings space was used, whereas for the study described in section 4.6 a 10*10*10 settings space was used. For these configurations almost 80% and close to 50% of the settings lie on the boundaries. Considering the fact that the boundaries contain a considerable amount of all possible settings the boundary rules should be carefully selected. To date, we did not study the effect of boundary rules in more detail.

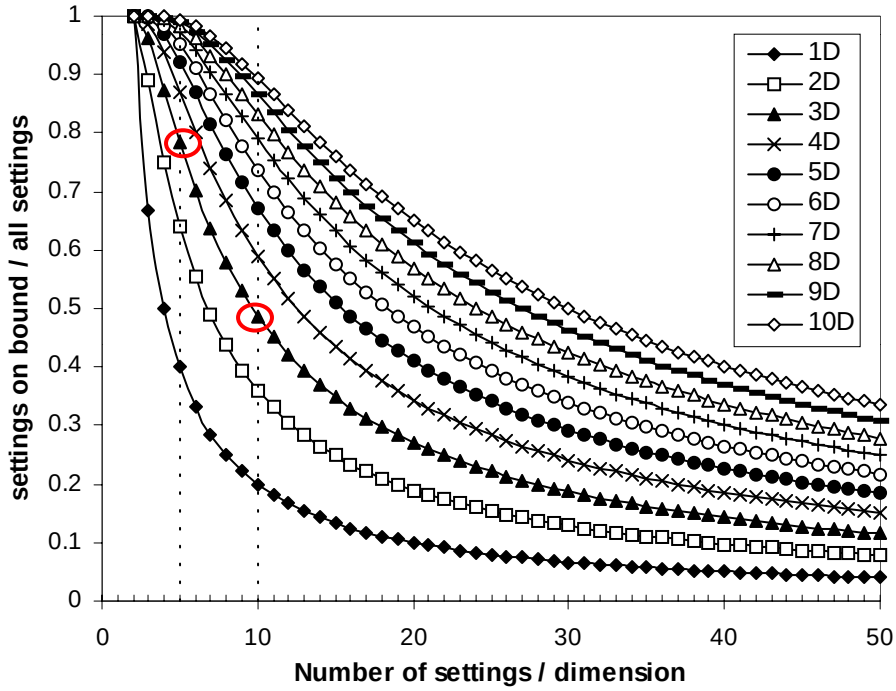


Figure 6.1 Fraction of settings on the boundary relative to the total number of settings as a function of the number of settings selected for each dimension. Parameter is the number of dimensions used. The calculated values apply to a cubic setting space only.

6.4 Cognitive factors

Cognitive factors that might influence the reliability of the multi-directional pattern search comprise the evaluation paradigm, the evaluation criterion, and the listening strategy. The last factor is subdivided to describe drifting cues, acclimatisation, and centre of attention.

Paradigm evaluation

The evaluation paradigm used throughout the studies was the method of paired comparisons. The main advantage of using this paradigm involves the possibility to use subjective evaluation criteria. Other paradigms comprise, among others, the method of rating, speech intelligibility measures such as SRT and CVC tests, and measurement of the reaction time. These methods yield perceptual scores that have a one-to-one relation to the acoustic algorithm settings. This is not the case for the results of the paired-comparison paradigm. This paradigm yields results that are relative, not absolute, in nature. An important advantage of the score-based paradigms is that these methods circumvent the problem of distinguishing difference limens from comfort limens.[57] The scores are either the same or different, whereas the comparisons can both yield judgments like 'not distinguishable' and 'equally comfortable' (or some other criterion). Especially in the study of section 4.6 it appeared that one subject sometimes mixed up these two judgments in selecting the 'no difference' button, used in the pattern search to guarantee perceptual distinctness of samples to be compared.

A Difference limens

In the study of section 4.4 we determined the difference limens (DL) for two parameters of monaural noise reduction and one parameter of spectral enhancement. It appeared that the DL varied considerably depending on the starting point chosen for each parameter (see figure 4.16). Due to this variation the use of a fixed, relatively small step size in the pattern search can cause settings to be indistinguishable during paired comparisons. There are at least two solutions to this problem. First, the step size can be made adaptive in order to guarantee perceptual distinctness between settings. This adaptation was used in the studies described in sections 4.5 and 4.6. The results of the latter study also showed that a minimum step size is required to get maximum reliability (see figure 4.23). This knowledge is hard to retrieve in advance, unless difference limen experiments are conducted in advance.

A second solution to avoid indistinguishable settings is to include settings such that the step size always is larger than the DL for each dimension. However, the results of DL experiment of three edges of the cubic set up only (comparable to the study of section 4.4) should be treated with care, because the DLs cannot predict the perceptual step size for combinations of algorithms other than the three edges that were used for the DL experiments. In a study not described in this thesis it was shown that parallel edges that vary in one dimension only, have very different DLs. This outcome implies that DL experiments should be conducted in principle for the whole algorithm space. But this theoretical requirement will be impossible to meet in practice. Therefore, the adaptive step size seems to be the best approach.

B Comfort limens and memory

The reliability of the pattern search procedure depends on the success of the paired comparisons that are performed. The uncertainty depends on the feasibility of the task

[57] A difference limen is the threshold for which two samples are just distinguishable, whereas a comfort limen is the threshold for which one of the samples is just preferred.

and the subjects' memory. The subjects had to compare speech-in-noise samples of four seconds. Due to the dynamic characteristics of the material it appeared sometimes hard to recall which of the two samples was preferred, although the sample could be repeated once.

The paired comparison method we used was forced choice. That is, the subjects had to choose between two samples A and B (AB paradigm). The use of the AB paradigm probably affected the reliability and consistency, due to relatively poor judgments or due to confusions between difference limens and comfort limens.

By counting the number of circular triads (a>b, b>c, c>a) that occurred in a round-robin tournament, a consistency measure ζ could be calculated. It appeared that for the study described in section 4.2 the consistency was lowest for PhC, intermediate for BNR, and highest for SE_B. This agrees with the reliability found for each algorithm based on the information transfer. This suggests that the reliability to a large extent could be ascribed to circular inconsistencies. To verify whether this assumption is justified we compared the results of sections 4.2, 4.5, and 4.6. Therefore, the relation was studied between the reliability of the pattern search and the coefficient of agreement and coefficient of consistency of the round-robin tournament results. The results are presented in appendix 6. In general, the reliability of the pattern search, indicated by the information transfer IT, can be derived largely from the measure of consistency ζ of the round-robin results.

There are many ways to explain the presence of circular triads. These are (a) guessing, (b) absence of a valid ordering, (c) there is more than one characteristic (aspect) that can be attended, (d) the samples have different features. Explanation (a) relates to perceptual differences. Parizet (2000) showed in his study that the forced-choice paradigm (AB) could lead to very high numbers of circular errors, especially when perceptual differences between samples are small. That is when loudness differences are equalized and this is exactly what we did for the studies described in sections 4.2, 4.3, 4.5, and 4.6. We wanted to exclude a loudness cue. Explanation (b) was not studied. However, it seems that ordering could be a problem for evaluation of listening comfort. Explanation (c) and (d) will be discussed in more detail in the next section called '*listening strategy*'.

The use of a 'no difference' alternative sometimes resulted in mixing up difference limens and comfort limens. This could be traced for subject s4 in the study described in section 4.6. An alternative approach that accounts for comfort limens has an additional 'no preference' alternative (ABN paradigm). The problem, of course, is how to use these two limens. A solution could be to use a rating paradigm in addition. Apart from the decision for one of the two samples, the subject could also rate how confident he or she is or how strong the preference is. The scale could range between very confident and very unsure or between much more preferred and equally preferred. The 'no difference' alternative can be left out then. Basically, the step size could be coupled to the confidence or preference rating. When the subject is, for example, very confident the step size could be decreased, whereas when the subject is not at all confident the step size could be increased. In this way the subject could follow equal-sensitivity contours throughout his or her perceptual space in order to find the optimum. Consequently,

problems that relate both to difference limens and comfort limens can be accounted for. That is, two signals that are either indistinguishable or are equally (un)comfortable can both be characterised as being equally preferred, or the decision could be rated 'very unsure' in both cases.

A disadvantage of using the ABN paradigm only is that ties introduce difficulties. That is, some judges may declare ties more readily than others. For the AB paradigm the area of possible disagreement between judges is reduced to a minimum. Though, from the study of Parizet (2000) it appeared that the coefficient of consistency ζ is only high when stimuli are presented at unequalised loudness levels. This, again, could explain that the reliability for the pattern searches of previous studies (e.g. Neuman *et al.*, 1987) was higher than in our studies.

Evaluation criterion

From the studies described in chapter 4 and chapter 5 several conclusions can be drawn. The test-retest reliability for listening comfort was highest. The low reliability found for the speech intelligibility criterion could be explained by small intelligibility differences and the tendency to select the second sample. The bias could be due to the fact that the second sample is more intelligible due to hearing it for the second time. The results presented in section 4.5 prove that this bias can be avoided by presenting the sentence material visually in addition. An alternative is to use different sentences in the paired comparisons. However, this could yield other biases. When, for example, spectral enhancement is used effects will be more pronounced for sentences that contain sibilants. It is likely that subjects will encounter tremendous difficulties in deciding what to compare. This, however, is only an assumption. We did not study this alternative approach.

In section 4.5 for hearing-impaired subjects there appeared to be a discrepancy between the speech reception thresholds (SRT) and the subjective preferences for different algorithm combinations. Thus it seems that the speech intelligibility was difficult to judge subjectively. Instead subjects maybe evaluate the difficulty (or ease) to understand speech, i.e. an effort measure. When hearing-impaired subjects evaluate listening comfort, speech intelligibility is also incorporated partly. This was shown by the questionnaire results of chapter 5. However, optimal listening comfort cannot generally predict higher speech intelligibility scores, i.e. lower SRTs. This applied to normal hearing subjects (see section 4.3) and hearing-impaired subjects (see section 4.5).

A problem of evaluating listening comfort is that the listening comfort criterion can be judged on many aspects, as was illustrated by the questionnaire results presented in section 4.6. This can explain the relative large spread in optima, but other explanations may be valid too. Poor reliability (differences between test and retest optima) and poor consistency (differences between the optima found for different starting points) may also be due to unfavourable response distributions, like multi-modal and flat distributions. This will be discussed in more detail in a subsequent section called '*acoustic factors*'.

We considered evaluation criteria covering a narrower range of aspects, such as clarity, continuity, and naturalness. Although these evaluation criteria are probably less ambiguous, the major disadvantage of using such 'one-dimensional' criteria is that the

complex behaviour of subjects in real-life situations cannot be accounted for. And one of the main goals for this study was to optimise multiple algorithms using real-life stimuli. However, we have to realise that the use of complex acoustical conditions in our experiments also introduced highly differentiated listening and judgment strategies. The experimental consequences are discussed in the next part.

Listening strategy

By using complex stimuli in our studies, i.e. speech in background noise, some problems were introduced that affected the reliability of the pattern search. The listening strategy of the subjects varied within or across different pattern searches. This can be due to selective attention of the subject, i.e. attention to one signal among others, made possible by the refinement of the filtering mechanism of the subjects' brains. The attention differed in several respects. The subject listened either analytically by attending either the speech or the background noise, or listened holistically attending the speech in noise as a whole. Additionally, the stimuli were judged based on several aspects, such as brightness, loudness, and fluency. That is, the subjects were cued to respond to specific spectro-temporal parts of the complex stimuli. The normal hearing subjects, for example, attended predominantly to brightness and fluency of the stimuli, probably related to the timbre of the signals and the presence of artefacts, respectively.

The listening strategies were studied in chapter 5 (see figures 5.5 and 5.6). It seemed that subjects indeed shift their attention from one configuration to another. For example, when it appears too difficult to judge the speech intelligibility, subjects shift their attention to the background noise. When this process of shifting is used consistently, this should not be too problematic. However, when the cues that are used for each configuration also shift, the reliability of the optimisation process might decrease. The drift in cues was not studied explicitly, but the use of multiple cues suggests that cues could have varied within or across different pattern searches. The poor reliability of subject s7 in section 4.6 combined with the fact that this subject used four cues frequently illustrates that drift in cues may affect the reliability of the results.

In general, the normal hearing subjects (NH) attended the brightness and fluency, whereas the hearing-impaired group (HI) attended the speech intelligibility and listening comfort mainly. The SIM1 and SIM2 group stuck to the NH cues, except for subject s18. This was the only subject that participated in all groups by using his right and left (impaired) ear. Contrary to the other subjects his listening strategies varied, depending on the ear used and depending on the stimulus material that was manipulated to simulate different aspects of cochlear hearing impairment. The reliability of the pattern search decreased considerably for subject s18 for the SIM2 condition (see figure 5.2), whereas for NH, HI, and SIM1 the reliability was high.

It seems that the hearing-impaired subjects cannot adapt easily to their new acoustic environment, when spectral enhancement is applied. To appreciate spectral enhancement the hearing-impaired subjects need to use spectral cues. The use of these cues might have been lost gradually over the years, due to systematic changes in the sensory input. This could be a reason why the information transfer was much lower for the HI group relative to the NH group i.e. 27% versus 73%, respectively (not shown in

table 6.1). The large difference in reliability calls for long-term studies in order to account for acclimatisation effects.

For an optimisation strategy acclimatisation effects cannot be avoided, unless different stimulus material is used. Buus (1997) showed that speech intelligibility scores of simulated hearing-impaired subjects, created by masking noise, were similar to the scores for hearing-impaired subjects when nonsense words were used. For meaningful sentences the scores for the SIM group and HI group varied considerably. This could be due to the fact that semantic and syntactic information cannot be used for nonsense words. Needleman and Crandell (1997) reasoned that the SIM group used a stricter criterion. It appeared that the subjects of the SIM group were less confident of their identification abilities. Therefore, they only responded when they were sure they got the message. Hearing-impaired subjects use a less strict criterion, and therefore are inclined to respond quicker, assuming they got the message. Hence, the reliability of the pattern search could be increased when using (series of) nonsense words.

6.5 Auditory factors

The reliability of the pattern search was lower for hearing-impaired subjects than for normal hearing subjects in the studies of section 4.3, section 4.5 and chapter 5. Apart from cognitive factors, also the impaired transmission in the cochlea could explain this decrease in reliability.

The study described in section 4.5 showed that the reliability was very poor for hearing-impaired subjects (see figure 4.18, lower panel and also table A5). Principally, this could be related to the small step sizes used (see section 6.3). Additionally, the average hearing loss and the hearing loss slope determined whether the optima were found consistently, i.e. independent of the starting point. It appeared that the consistency of spectro-temporal enhancement (STE) was higher for subjects with a moderate hearing loss than for subjects with a mild hearing loss. The hearing loss slope did not influence the consistency much. For the noise reduction algorithm (NR_S) the consistency depended heavily both on hearing loss and hearing loss slope. The consistency was poor for subjects with moderate hearing losses and positive hearing loss slopes. Hence, the consistency of the pattern search depends on a combination of the algorithms used and the amount and slope of the hearing loss.

The simulations of cochlear hearing loss described in chapter 5 also suggest that the reliability of the pattern search depends on auditory factors. We studied the broadening of auditory filters by comparing the NH group and the SIM1 group. The additional masking noise simulated the combined effect of shaping of the frequency spectrum due to a frequency dependent hearing loss and narrowing of the dynamic range and the consequences were studied by comparing the results of SIM1 and SIM2.[58] These results show that the reliability was much higher for subjects with relatively sharp auditory filters (the NH group) for the spectral enhancement algorithm (SE_L), and slightly higher for

[58] The masking noise was shaped, based on the smallest high-frequency hearing loss of one of the hearing-impaired subjects that participated as well. To compensate for the elevated thresholds of the SIM2 and HI group the NAL rule was used. This rule compensates for the hearing loss only partly; hence the signals still become spectrally tilted following a shallower slope than the high-frequency hearing loss.

noise reduction (NR_L). However, the reliability was much lower for the spectral lift algorithm (SL). A smaller dynamic range in combination with spectral tilt resulted in a decrease in reliability for SE_L and NR_L.

The reliability of the pattern search of the HI group was comparable to the reliability of SIM1 for SL and NR_L, and comparable to the reliability of SIM2 for SE_L. Hence, the large decrease in reliability of SE_L found for HI relative to NH could be explained by a combination of two factors. First, the decrease is due to the broadening of the auditory filters, and second due to the combined effect of the narrowing of the dynamic range and the shaping of the frequency spectrum due to the hearing loss. The increase in reliability found for SL for HI relative to NH, could be attributed to the broadening of auditory filters alone.

6.6 Acoustical factors

In the studies of section 4.1, section 4.6, and chapter 5 we used standard speech and standard background noises. In the studies described in sections 4.2, 4.3, 4.4, and 4.5 real-life stimuli were used. These comprise sentences and realistic background noises. The sentences were always recorded in a sound-attenuating booth to avoid differences in recordings of the sentences. The background noises were recorded real life. The speech and background noises were recorded and presented by the microphones and telephones of in-the-ear hearing aids (sections 4.2 and 4.5), of behind-the-ear hearing aids (section 4.4), or of earmuffs (section 4.3). The stimuli were processed off-line to make fast flexible switching between settings in the pattern search possible. The processing was conducted predominantly by experimental signal-processing strategies (sections 4.2, 4.3, and 4.4). In the study described in section 4.5 a state-of-the-art hearing aid, also used for the recordings, processed the signals.

A Hard ware effects

Real-life stimuli were used to make the fitting strategy, i.e. the pattern search, as realistic as possible. It is our conviction that subjects benefit most from fine-tuning of their hearing aids, when the laboratory conditions agree as much as possible with their acoustic environment. However, several problems occurred due to the use of hearing aids and real-life stimuli. Therefore, standard material was selected for the studies of section 4.6 and chapter 5. The in-the-ear hearing aids shaped the stimuli considerably due to the high-pass characteristics of the microphones. The frequency-shape characteristics of the behind-the-ear hearing aids and earmuffs were much less colouring. These results suggest that the hardware choice can influence the success of signal-processing strategies. Another factor that might influence the success of signal-processing strategies is the smearing effect of reverberation present in real-life recordings. In all, the reliability of the optimisation strategy might be influenced negatively by introducing real-life into the laboratory.

The interaction between spatial hearing abilities and speech intelligibility might be underrated as well. The true effects of signal-processing strategies become known only when tested in real life. In our laboratory studies the spaciousness is lost for the most

part. In effect spatial separation of sources is not possible which might affect the performance of the algorithms.

B Signal-processing-strategy effects

In the studies of chapter 4 and chapter 5 a considerable amount of different signal-processing strategies were used. The underlying concepts have been presented in chapter 3. In section 6.1 it was shown that these algorithms act non-linearly, hence they influence one another. As a consequence, the reliability of the pattern search could be affected.

Additionally, the separate effects of the algorithms on the reliability of the pattern search were studied. In section 6.2 it was shown that some algorithms yield higher reliability than others. The cumulative distributions are different for different algorithms (see for example figures 4.9). The information transfer derived from these cumulative distributions also shows differences in reliability between algorithms (see tables 4.8, A4, and A5). In addition, the reliability depended also on SNR for listening comfort, and on noise type for speech intelligibility (see section 4.2). In the studies of sections 4.5 and 4.6, that were conducted for listening comfort only, reliability did not depend on noise type. However, this appeared to be due to a trade off in the results for each algorithm separately. In section 4.5, reliability was highest for algorithms SC and STE when continuous noise was used, while reliability was highest for NR_S when fluctuating noise was used. A divergent pattern was found in section 4.6: for SE_L and SL the reliability was highest for fluctuating noise, whereas for NR_L the reliability was highest for continuous noise. Interestingly, the difference in reliability between both noise types is different for noise reduction strategies and speech enhancement strategies. It is difficult to explain this outcome based on differences in acoustical or signal-processing factors as such. Instead, the explanation should be sought in the perceptual domain, i.e. the auditory transforms of the processed signals.

Although the response distributions that describe the perceptual consequences of the processed (acoustical) signals were determined for the study of section 4.2 only, the results might be extrapolated to the results found for the studies of sections 4.5 and 4.6. A round-robin tournament was used to determine response distributions for all twelve edges of the algorithm cube of section 4.2. Edges that varied for the same dimension, representing a specific algorithm, were averaged. For each algorithm the shape of the response distribution was determined and classified being either unimodal or multi-modal. A unimodal distribution could, in addition, be narrowly peaked or broadly peaked (or nearly flat). The results showed (see figure 4.10) that unfavourable distributions, i.e. the multimodal and (nearly) flat distributions, occur relatively often. Considering the reliability of the pattern search, the binaural noise reduction and fast compression algorithms suffer especially from broad response distributions. The information transfer for both algorithms was relatively low. Hence it seems that samples used in the pattern search were difficult to distinguish, or equally preferred. The spectral enhancement algorithm suffers predominantly from multimodal distributions. This effect was most pronounced for hearing-impaired subjects and when noise reduction and fast compression were not applied (see condition n0c0sx in figure 4.11). Still, the information transfer for this algorithm was very high. Hence, it seems that the pattern search could in

general avoid terminating at a suboptimal setting. For section 4.6 this was certainly not the case, for the cluster analysis showed that the pattern search ended at suboptimal settings frequently. For five of the nine hearing-impaired subjects a clear suboptimum resulted from a cluster analysis, that differed mainly and most consistently for the spectral enhancement algorithm (see table 5.2). A possible explanation could be that the preferences differ much less between the optima and suboptima for spectral enhancement of section 4.6 relative to section 4.2.

The problem of the presence of unfavourable distributions is that the pattern search does not necessarily end at the optimum. The presence of multiple optima, discontinuities, saddle points, and ridges can affect the reliability and convergence of the procedure. In general, problems emerge when the direction of paired comparison is parallel to the equal-preference-contours in the perceptual response surface, or when the step size is too small in order to leave a suboptimal setting. These problems can be avoided in at least three ways. First, by conducting several pattern searches for different starting points and performing a tournament between the optima found. Second, by using the direction analysis of a fixed number of paired comparisons, instead of the stop criteria used in the pattern search. Third, by rotating the search direction. That is, when settings that are compared pair wise differ not in one direction only, but in three directions simultaneously, the procedure possibly could better avoid early termination.

The direction analysis solution is very (time) efficient in that it uses all information gathered during the search. Additionally, the use of the directions analysis eliminates the poor reliability of the pattern search found for subject s7. Although the exact reason to explain this discrepancy can be only guessed, it seems that for this subject the low reliability of the pattern search is due to early termination. An important advantage of rotation of the search directions might be that subjects cannot focus solely on the effects of single dimensions, i.e. separate algorithms, as discussed in section 6.4.

6.7 Alternative approaches

In the present pattern search strategy we always used a paired-comparison paradigm. A disadvantage of this approach is that the reliability of the optimisation might not be high enough. This is due to factors that were discussed in sections 6.3 to 6.6. Some, but not all of the problems can be accounted for. Therefore, in chapter 4 and chapter 5 alternative approaches were studied that could be used as alternative paradigms for the optimisation strategy. These approaches comprise the SRT, CVC, and rating experiment. These tests yield a value or score, which could be less prone to drift, change or ambiguity.

From the studies of sections 4.3 and 4.5 it appeared that the SRT is not very discriminative. For both studies the range of SRTs, determined for different combinations of algorithms, was on average only 3 dB. This value is relatively small compared to the standard deviation of 1.5 dB found for SRT measurements in the study of section 4.3. Therefore, it seems that the SRT is not a suitable paradigm to guide the pattern search, at least for the combination of algorithms used in these studies.

In the study of section 4.3 and chapter 5 it appeared that for CVC words the difference in test and retest scores was on average relatively large, i.e. around 12%. For

the study of section 4.3 the range of CVC results for four extreme settings was around 30% for SNR=-7. The results of chapter 5 suggest that especially for subjects that have broad auditory filters the range of scores found for ten different combinations of algorithms was highest, around 30% on average for ten extreme settings. Hence, the accuracy is too low to discriminate between the ten algorithm combinations. Therefore, based on these outcomes the use of a CVC paradigm to guide the pattern search is not desirable either.

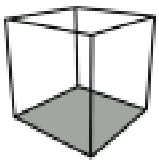
For the rating experiment in chapter 5, the difference between the test and retest results was on average 1.1 scale unit. Combined with the notion that the range of scores was on average 6.9 units, the rating procedure has a reasonable discriminative power. As such, the rating procedure could be used to guide the pattern search.

Chapter 7

AUDIOLOGICAL ISSUES

'But though our thought seems to possess this unbounded liberty, we shall find, upon a nearer examination, that it is really confined within very narrow limits, and that all this creative power of the mind amounts to no more than the faculty of compounding, transposing, augmenting, or diminishing the materials afforded us by the senses and experience. When we think of a golden mountain, we only join two consistent ideas, gold, and mountain, with which we were formerly acquainted.'

– David Hume
An enquiry concerning human understanding (1993/1748)



To date, the effects of combination of algorithms are hardly studied. In the studies described in chapter 4 and chapter 5 several combinations of speech-enhancement and noise- or reverberation- reduction algorithms were investigated. In the surveys of section 4.1 and section 4.2 combinations of algorithms could be analysed directly, due to the systematic approach used. From the other studies we only could analyse the effect of combinations of algorithms indirectly. This is due to the fact that a pattern search strategy was used. Additionally, we also analysed the effects of algorithms separately. For reasons of clarity the effects of combinations of algorithms will be discussed not before section 7.4.

7.1 Evaluation of spectral lift and spectral enhancement

Spectral lift

The results of chapter 5 indicated that hearing-impaired subjects did not prefer spectral lift around 3 kHz (see figure 5.3, middle panel, HI group). This could be explained by the fact that the elevated thresholds of the hearing-impaired subjects were already compensated for. This hypothesis was supported by the results of the SIM2 group, for which effects of poor spectral resolution and elevated thresholds were simulated. For this group the elevated thresholds, simulated by masking noise, were also compensated for. Spectral lift was only preferred for subjects with simulated poor spectral resolution (SIM1). Relative to the normal hearing group (NH), the SIM1 group preferred consistently more spectral lift. The increase was much higher than for spectral enhancement that was specifically designed to overcome problems that relate to broadened auditory filters (see figure 5.3, upper panel). This suggests that the effect of spectral smearing could better be compensated for by spectrally lifting the frequencies near 3 kHz. However, the isolated broadening of auditory filters is probably only existing in theory. In practice hearing-impaired subjects that have broadened auditory filters will also have elevated thresholds.

Spectral enhancement

The effects of spectral enhancement were studied in sections 4.1, 4.2, 4.5, and 4.6, and in chapter 5. The algorithm differed for these studies. The spectral enhancement algorithm (SE_B) of Baer and Moore (1997) was used in sections 4.1 and 4.2. For section 4.1 SE_B was fixed, and the enhancement gain M was chosen to be a value of 2. The spectral enhancement algorithm (SE_L) of Lyzenga *et al.* (2002) was used in section 4.6 and chapter 5 and the enhancement gain M ranged from 0.25 (setting '-3') to 2.5 (setting '6'). Enhancement factor M=2 equals setting '4' of the study described in section 4.6 and chapter 5. The spectro-temporal enhancement algorithm (STE) of section 4.5 could be compared to the other two (SE_B and SE_L) only qualitatively.

In general the speech intelligibility for spectral enhancement was not significantly different from the reference for any of the noise conditions (section 4.1). With respect to listening comfort and speech intelligibility the preferred setting was between M=1.0025

and M=1.0250[59] for the hearing-impaired subjects in study section 4.2. For these factors SE_B acts more like a noise reduction scheme. Highest spectral enhancement values are preferred for car noise. However, in the round-robin data a suboptimum was found for M=10.

As expected, normal hearing subjects did not prefer any spectral enhancement, i.e. M=1 for the studies of section 4.2 and chapter 5. For the study of chapter 5 hearing-impaired subjects preferred M-values around 0.75 (setting '-1'). This preference for spectral compression instead of spectral enhancement (or spectral expansion) is also found for subjects with simulated poor frequency resolution and elevated thresholds (SIM2). Some spectral enhancement was preferred (M=1.25, setting '1') by subjects with simulated poor frequency resolution only (SIM1).

Baer and Moore (1997) suggested three possible explanations. First, subjects maybe needed to get accustomed to the signal processing. Second, the broader the auditory filters, the more difficult to compensate for by means of spectral enhancement. For extremely broad filters an extreme compensation would result in a compound of sinusoidal components, reducing the information content of the stimulus substantially. Third, something other than restoration of the excitation pattern is required.

The first explanation was already discussed in chapter 6 with respect to the reliability of the pattern search. It is probably unrealistic to expect that hearing-impaired subjects at once could use spectral cues, i.e. appreciate spectral enhancement. This could be tested in a field study, where hearing-impaired subjects get time to adapt to their new perceptual environment, and learn to make use of spectral cues. However, since the SIM2 group and the HI group showed similar results in chapter 5, and since the SIM2 group did probably not loose spectral cues, this explanation cannot be the only answer.

The second explanation does not apply to the results of our studies. In our studies subjects had generally mild-to-moderate hearing losses, accompanied by a broadening of the auditory filters by a factor of about two. The asymmetry of the auditory filters was also grossly the same and compensated for, at least in paragraphs 4.1 and 4.2, and chapter 5.

The third, most loosely formulated explanation could explain our results. The results of chapter 5 indicate that spectral enhancement is only beneficial when subjects have broadened auditory filters. As already mentioned in the previous paragraph such effects do probably not occur without elevated thresholds. When masking noise was added SE_L was not preferred; simulated hearing-impaired subjects preferred mainly the opposite, i.e. spectral compression. Therefore, it seems that the making noise in some way destroys the potential benefits of spectral enhancement. The underlying explanation could be that subjects with elevated thresholds need higher signal levels, for which the auditory filters are broader (Van Tassell, 1993; Baer and Moore, 1997). Consequently, these broadened filters will pass more noise, which is detrimental for the temporal coding of the signals. The real problem is in fact that the spectral enhancement will not selectively process the speech. For SNR=+5, used for the HI and SIM2 group, both the target signal and the background noise have been manipulated.

---

[59] Note that in section 4.6 and chapter 5 logarithmic values were used, whereas in section 4.2 linear values were used.

An additional problem that appeared from our study described in section 4.1 is that there was a trade-off between improvements and deteriorations between individual phonemes. In general we see that the spectral enhancement improves vowel perception above all. The positive effect on the vowels is counteracted by a negative effect on the consonants. Hence, no overall benefit was generally found in our subjects and in our experimental conditions. Spectral enhancement can easily lead to a better vowel recognition, because this task requires mainly spectral information. For improvement of consonant perception we probably need also processing in the time domain, because temporal cues are important for the discrimination of consonants. Van Tassell (1997) suggested that short-term spectro-temporal characteristics, i.e. vocal tract gestures, could be relatively important for subjects that have mild-to-moderate hearing losses. The information carried in formant transitions characterise to a large extent the voicing, manner, and place of articulation.

In section 4.5 we studied a spectro-temporal enhancement strategy (STE) of a state-of-the-art hearing aid. It appeared that contrast enhancement in the STE algorithm is appreciated for continuous noise. The preferred STE setting was significantly lower for fluctuating noise than for continuous noise (Wilcoxon, $p<0.05$). This might be due to the fact that spectral and temporal contrasts in the speech signal can be found more easily for a steady background noise. For the group of younger subjects that performed the pattern search more reliably than the group of older subjects, individual differences were found between optima found for the STE algorithm.

Despite the poor results for spectral enhancement generally found, individual results suggest that spectral enhancement was beneficial for some individuals in specific conditions. In section 4.1 we found that for three out of eight subjects SE_B had higher CVC scores. The subjects with relatively large hearing losses (s3, s7 and s8) benefited from SE_B in most conditions, but others did not. These subjects appear to rely on sibilance information of the initial consonant mostly (implicated by the INDSCAL analysis). Improvements due to spectral enhancement show significant correlations with the measured dynamic range at 1 kHz; spectral enhancement gives some improvement only for subjects with small dynamic ranges at 1 kHz. Subject s2 is the only exception to the rule, maybe due to a completely different auditory filter at 3 kHz. It is the only filter, which has a steep lower skirt and a flat higher skirt.

These outcomes suggest that benefits from spectral enhancement do not relate necessarily to the broadening of the auditory filters (only), but (also) to the elevated thresholds and to the dynamic range for specific frequencies. It is possible that this factor relates directly to the listening strategies used by the subjects.

The effects of spectral enhancement were in general not preferred and generally did not improve speech intelligibility. To ultimately test the possibilities of spectral enhancement the important parameters should be varied systematically. In our studies we only varied the enhancement gain factor, not the bandwidth. Yang *et al.* (2003) studied the effects of both parameters, though only on simulations and not in interaction. The effects of spectral enhancement should be studied on its perceptual consequences. Additionally, the effects of bandwidth and gain factors should be studied

in interaction. The most optimal combination(s) of the parameters could be found by an optimisation method like the multi-directional pattern search.

7.2 Evaluation of phonemic compression

Syllabic and phonemic compression were studied for hearing-impaired subjects in paragraphs 4.1, 4.2, 4.3, and 4.5. The fast compression systems were single channel in paragraphs 4.1 and 4.3. Multi-channel fast compressors were used (also) in sections 4.1 and 4.5, comprising eight and four channels, respectively.

The study of section 4.1 showed that improvements in speech scores due to additional single-channel phonemic compression are higher for subjects with steep filter slopes at 3 kHz. In the study of section 4.5 the preference for syllabic multi-channel compression is only found for subjects with flat or negatively sloping hearing losses. These findings were contrary to our expectations. We assumed that subjects that had smaller dynamic ranges would benefit from phonemic compression. We also assumed that the fast compression systems used in the studies would be beneficial for subjects with high-frequency hearing losses.

No relation was found between the dynamic range and the effect of phonemic compression on speech intelligibility. This can be explained as follows. A relation can be found when the dynamic range of subjects is both larger and smaller than the dynamic range of speech. Speech intelligibility will be poorer when the dynamic range of a subject is smaller than that of speech, while speech intelligibility will be relatively good when the dynamic range is larger than that of speech. By determining the dynamic range the long-term average spectral characteristics are considered. For syllables with a time interval of 125 ms the dynamic range is around 30 dB. For individual phonemes the dynamic range is even higher, considering the fact that the dynamic range depends on the time window. Since the dynamic ranges of hearing-impaired subjects in our study were rather low a relation between the effects of phonemic compression on speech intelligibility is not likely to be found. Additionally, Van Tassell (1993) argues that for persons with mild-to-moderate cochlear hearing losses the information load of speech is carried predominantly by short-term spectro-temporal characteristics. This implies that the long-term average dynamic range is not a suitable predictor for measuring the effects of fast compression on speech intelligibility for persons with mild-to-moderate cochlear hearing losses.

The fast compression system was developed for persons with high-frequency hearing loss (Verschuure *et al.*, 1998). However, in the study described in section 4.5 we found that only subjects with flat or low-frequency hearing losses appreciated the multi-channel syllabic compressor with respect to listening comfort. An explanation could be that the long-term speech spectrum will be flattened by fast compression. According to Van Tassell (1993) this could be due to the fact that high-amplitude vowels are the most important contributors to the lower frequency components of the frequency spectrum.

In the study of section 4.1 improvements of the single channel compressor were found especially for the relatively easy noise conditions, i.e. fluctuating noise or SNR=5 dB. Analysis of the pattern search data of section 4.2 shows that phonemic compression is also preferred for more fluctuating noises. In our experimental set-up subjects prefer compression ratios around 2 in combination with a release time of 70 ms. The study of

section 4.5 showed that for fluctuating noise multi-channel syllabic compression is preferred for the higher frequency bands, whereas for continuous noise almost no syllabic compression is preferred. These outcomes do support results of Verschuure *et al.* (1998).

7.3 Evaluation of algorithms to reduce noise and reverberation

Binaural de-reverberation

In the study of section 4.3 all subjects, i.e. six normal hearing and two hearing-impaired subjects, consistently preferred an intermediate value for de-reverberation. Although this study was performed for subjects at work with earmuffs, this could imply that hearing-impaired subjects that also suffer from poor acoustic conditions like reverberation could benefit from de-reverberation.

Binaural noise reduction

The studies described in sections 4.2 and 4.3 showed that there was no clear preference for binaural noise reduction. Additionally, an analysis of variance of the results of section 4.2 showed hardly any differences between different conditions. As suggested by Kollmeier *et al.* (1993) and Wittkop *et al.* (1997), this could be due to the diffusiveness of the background noise, the dichotic presentation of the processed signals, and a ceiling effect present for higher SNRs.

Monaural noise reduction

The subjects that performed the pattern search results most reliably in study section 4.5 preferred intermediate values for modulation-based monaural noise reduction in continuous noise. For continuous noise, the SRT results were on average lowest (best) as well when noise reduction was maximal. The difference in SRT was on average more than 1 dB relative to the reference condition. This implies that the noise reduction scheme that was based on amplitude modulations was relatively successful. This was due to a slight improvement in SRT and higher listening comfort. However, more thorough studies are necessary to validate this trend.

Monaural noise reduction based on phase variations, developed by Lyzenga *et al.* (2002) appeared beneficial for improvement of listening comfort for continuous noise. In the study of chapter 5, hearing-impaired subjects preferred on average a noise reduction factor close to 3. The NH, SIM1, and SIM2 group did not prefer any noise reduction. The explanation for the discrepancy found between the simulated hearing-impaired subjects (SIM1&SIM2) and the hearing-impaired subjects (HI) could be that the HI group did not attend the sound quality and fluency, but rather the intelligibility of the sounds. An additional effect might be that the artefacts could be less detrimental in case of elevated thresholds.

7.4 Effects of combination of algorithms

For the combined effect of spectral enhancement and single-channel compression the scores are comparable to the scores for spectral enhancement alone, indicating that the

single-channel compressor gives no additional improvement to spectral enhancement (see section 4.1). The results of single-channel compression alone (Goedegebure, 1998) were more promising than the additional effects of phonemic compression this study shows. For the combination of spectral enhancement and multi-channel compression the scores are lowest, indicating that these two algorithms have opposite effects. We assume that, contrary to multi-channel compression, the single-channel compressor does not change the short-term spectra of the speech signals. Van Tassell (1993) reasoned that multi-channel compression could reduce spectro-temporal details. This could be harmful in the frequency domain for subjects with broadened auditory filters, because these filters already removed some spectral detail. This could explain the trade-off between spectral enhancement and multi-channel phonemic compression.

The question is, to what extent the speech signal can be compressed and spectrally enhanced with a minimum of detrimental effects. The pattern search is a means to determine the optimal combination. In the study described in section 4.2 the order of processing was reversed, in order to avoid the effect of reducing spectral detail due to compression. Hence, single-channel phonemic compression was used first and spectral enhancement second. The results showed that both normal hearing and hearing-impaired subjects preferred values 1.3 and 0.5 for phonemic compression and spectral enhancement, respectively. Again, there seems to be a trade-off. After fast compression no spectral enhancement was preferred. Influencing the amplitude of the signals by fast compression also influences the spectro-temporal details. In effect, spectral enhancement seems to enhance the blurring of important spectro-temporal details.

The order of processing is in fact a factor that was not studied systematically. Optimisation of combinations of algorithms requires also varying the order of processing. A pattern search should, for example, be performed for fast compression first and spectral enhancement second, and for the opposite order as well. The optima acquired for both orders can be compared afterwards.

The results of section 4.5 showed that most hearing-impaired subjects preferred some degree of spectro-temporal enhancement (STE) and modulation-based noise reduction (NR_S), both for continuous and for fluctuating noises. However, additional syllabic compression was only preferred by some subjects in the case of a fluctuating noise. The results of chapter 5 demonstrate that subjects in general prefer rather different combinations. Five out of nine hearing-impaired subjects preferred a setting near (SC_{-2},STE_{-3},$NR\text{-}S_4$). Hence, only noise reduction (NR_L) was preferred appreciably, whereas both for spectral enhancement and spectral lift mostly negative values were preferred. These results show that only NR_L was preferred and that negative values for either SE_L or SL was accompanied by a negative or zero value for the other algorithm. These preference results did agree to some extent with the CVC results for a subgroup of six hearing-impaired subjects.

7.5 Evaluation of individual needs

The results of section 4.2, section 4.5, and chapter 5 show that the optimal setting found by the pattern search is subject-dependent. In the previous section it appeared that the results of hearing-impaired subjects in the study of chapter 5 diverged. Five subjects

preferred a considerable amount of noise reduction, two subjects some noise reduction and one subject hardly any. The results of section 4.2 also showed that the amount of spectral enhancement and fast compression preferred is subject-dependent. The study of section 4.5 showed that different optima were found by two groups of hearing-impaired subjects. The optimal setting also depended heavily on the noise condition.

In all, the optimal settings depended on the combined characteristics of subject and acoustical background. The listener dependency can be differentiated by hearing-impairment factors like average hearing loss, hearing loss slope, auditory filter slopes, auditory filter bandwidths, and speech reception thresholds (SRTs). From the study of section 4.1 the most comfortable loudness level and dynamic range for different centre frequencies were additional factors that contribute to diverging effects of certain signal-processing strategies. The relations found underline the fact that subjects with higher hearing losses make use of different strategies to understand speech. A complicating factor is that the effects found diverged for initial and final consonants, see figures 4.5 and 4.6. Additionally, it appeared easier to perceive vowels in continuous noise than in fluctuating noise. For the perception of consonants an opposite trend is present in the data, especially for the c2 scores.

However, it appeared to be very difficult to relate these factors to the preferred settings for different signal-processing strategies, i.e. to find significant correlations. This could be due to other factors that are difficult to quantify, like evaluation criterion and listening strategy factors such as listening cues and attended stimulus material (see also chapter 6). The listening cues comprise aspects used to judge the evaluation criterion at issue. The stimulus material that was attended could be the speech signal, the background noise, or both speech and noise simultaneously. For example, preferences could be guided by the amount of 'mess' in the background noise. One of the subjects that participated in the study of chapter 5 remarked that two types of optima could be distinguished. When there was not much mess in the background, normal speech was preferred, whereas when there was much mess this subject preferred more emphasis in the higher frequencies. This suggests that when the background noise becomes distracting for some reason, the subject can benefit from high-frequency emphasis of the speech signal, despite the fact that the background noise will receive the same high-frequency emphasis.

Another complicating factor to explain the optimal results found is the presence of algorithm or parameter interactions. The systematic round-robin study performed in section 4.2 showed that there was interaction between non-linear algorithms that is difficult to predict. Algorithm interactions were found either between phonemic compression and spectral enhancement, or between noise reduction and phonemic compression. The evidence is based on the average winner and the shape of the response distributions, deduced from the round-robin results.

The abovementioned factors affect the search for the optimal combination of signal-processing strategies, but at the same time they call for an individually applicable multidimensional approach. Because the best algorithm configuration cannot be predicted easily due to the complex interplay between these factors, the present multi-directional optimisation strategy might be a valuable tool for individual fine-tuning.

Based on the results, specific hearing-aid settings can be selected for specific circumstances and interactions between algorithms can be taken into account.

7.6 Trade-off effects

Generally, the digital signal-processing strategies studied in this thesis did not yield higher speech intelligibility, i.e. higher monosyllabic word (CVC) scores or lower speech reception thresholds (SRT). The algorithms were, however, appreciated in certain conditions when evaluated on listening comfort. The studies described in chapters 4 and 5 showed that several trade-off effects could explain this discrepancy. Trade-off effects appeared between phonemes, algorithms, and evaluation criteria. These trade-off effects are discussed hereafter.

The study of section 4.1 demonstrated that there is a trade-off between the perceptual strategies of initial consonants and final consonants. For the perception of the initial consonant the higher the average hearing loss, the more important is the use of sibilance information relative to voicing/plosiveness. With respect to the perception of final consonants an opposite trend was found. The lower the most comfortable level at 1.0 kHz the more sibilance information was used relative to voicing information. A trade-off was also found between vowels and consonant scores. Consonant scores, especially final consonants scores, are higher in fluctuating noise than in continuous noise, while vowel scores are lower in fluctuating noise. Consonant and vowel scores showed a trade-off effect for spectral enhancement in particular. Only few subjects could benefit from spectral enhancement for the total scores for the CVC words. It appeared that the positive effects on the vowels are counteracted by the negative effects on the consonants. Finally the improvements made by spectral enhancement are lost due to multi-channel compression. These instances could be one explanation why spectral enhancement alone or in combination with phonemic compression was not beneficial in improving the CVC scores relative to no processing at all.

Finally, a trade-off appeared between listening comfort and speech intelligibility. The pattern search optima that have highest listening comfort did not show improved speech intelligibility. The optima found in section 4.3 showed no higher, often even lower speech reception thresholds (see figure 4.13). The results of chapter 5 showed higher CVC scores for one out of six subjects only, for SNR=5 dB (see table 5.4).[60] These outcomes suggest that listening comfort and speech intelligibility can be two separate diverging goals. Even when evaluating the speech intelligibility subjectively, the optima found in section 4.5 showed slightly higher speech reception thresholds for two or three out of ten subjects, for fluctuating and continuous noise, respectively (see table 4.10). This could be due to the fact that the hearing-impaired subjects also considered listening comfort to a certain degree.

These results show that for our studies there is a discrepancy between good speech intelligibility, measured by SRT or CVC tests, and high listening comfort. This implies

[60] Fortunately, for SNR=0 dB higher CVC scores were found for four subjects for different algorithm configurations. Though the largest improvement was found for additional spectral lift alone, which was not principally developed for that purpose.

that fine-tuning of hearing aids by a pattern search strategy could not achieve both goals. That is, although the multi-dimensional pattern search can handle multiple signal-processing strategies, it can handle only one response criterion at a time. A method to account for multiple response measures are therefore needed, e.g. speech intelligibility and listening comfort simultaneously. Otto (1998) described a method of fuzzy sets that provides flexible and efficient techniques for handling different and conflicting optimisation criteria.

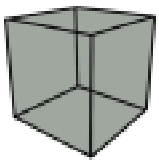
The questionnaire results presented in chapter 5 showed that hearing-impaired subjects consider the speech intelligibility as well, when they are asked to evaluate speech-in-noise samples on listening comfort. This implies that hearing-impaired subjects are predominantly focused on speech intelligibility. This is maybe not that surprising, since hearing-impaired subjects are accustomed to always solve puzzles of auditory messages. When this assumption is true, this explains that normal hearing subjects chose different algorithm settings in the study of section 4.2 for different conditions, whereas hearing-impaired subjects did not. As was emphasized by Bregman (1994) higher level organisation could be responsible for perceptual constancies. Due to difficulties in grasping the auditory message, hearing-impaired subjects could have some sort of speech intelligibility constancy. Hence, this constancy should be explained in terms of brain activity, instead of brain input. As such, variation of the physical input used in psychoacoustics should not be the ultimate goal. Additionally, variation in the perceptual domain could be a useful guide to study effects of signal-processing strategies. Similarly, Repp (1987) stated that relational properties are likely to be affected only slightly by transformations. Based on this assumption the robustness of speech perception could be explained. This would imply that in order to 'make' hearing-impaired subjects appreciate certain signal-processing strategies only long-lasting laboratory studies, or field studies can be used.

Chapter 8

FINAL DISCUSSION

'We can thus conclude that, structurally, perception depends on historically established human practices that can alter the system of codes used to process incoming information and can influence the decision assigning the perceived objects to appropriate categories. We can then treat the perceptual process as similar to graphic thinking: it possesses features that change along with historical development.'

– A.R. Luria
Cognitive development – its cultural and social foundations (1976)



8.1 A different perspective

The fundamental idea of this thesis is that hearing-impaired subjects should be able to fine-tune hearing aids based on their judgements. However, subjective preferences introduced problems that relate to perceptual differences, listening strategies, and drifting cues. By incorporating subjective factors the multi-directional pattern search appeared not immediately applicable for audiological purposes. In order to understand the underlying subjective mechanisms a different scientific approach might be required. The psychophysical approach might not be an appropriate tool. An alternative approach should be oriented not only at invariant characteristics, which is one of the primary goals of psychophysical research, but much more at individual dynamical changes in subjective judgments. Perceptions are ordered by constantly changing concepts and beliefs, responding to novel situations, thus adapting to specific individual demands. Instead of an acoustically oriented view, we propose to focus our way of thinking much more on perception and cognitive acts that are subject to individual preferences that can change over time. Therefore, we conclude this thesis with radically different views adapted from seemingly distant fields, i.e. cognitive psychology and philosophy.

The next section (8.2) will contain a more elaborate discussion of measuring the perceptual consequences of signal-processing strategies. The relatively poor performance of signal-processing strategies will be discussed in section 8.3. The methodological consequences of optimisation in the perceptual domain are discussed in section 8.4. Alternative scientific ideas and concepts in the domain of general perception are presented in section 8.5. In section 8.6 the scientific consequences of perception theories are discussed in relation to cognitive psychology and psychophysics in general, and speech perception in particular. This final discussion concludes with recommendations for the clinical practice and future research (section 8.7).

8.2 Understanding of auditory perception

Our studies on optimisation based on judgements of hearing-impaired subjects suggested that developing signal-processing strategies is useless without measuring the perceptual consequences. When even normal hearing subjects with high spectral resolution cannot distinguish between different hearing-aid algorithm settings, fine-tuning is pointless. Additionally, the multi-directional pattern search is only sensible when signal-processing strategies have a sufficient number of distinguishable settings.

A complicating factor is that it appears to be very difficult to determine in advance which signal-processing strategies are difficult to optimise. The perceptual constraints of hearing-impaired subjects are difficult to assess. Not only auditory factors play an important role, but also cognitive factors. It seems not possible to determine what a hearing-impaired subject perceives. Cochlear hearing loss not only affects the transmission process, but, indirectly, also neural processes. The fact that the brain receives less information, that might be distorted as well, influences neurophysiologic processes. Moore (2003) focused on the role experience plays in the development of auditory performance. Auditory performance reflects interaction between sensory and non-sensory factors (e.g. cognitive influences like attention). Consequently, experience

affects the way in which the brain processes future auditory stimuli that will be different for subjects with cochlear hearing losses relative to normal hearing subjects. This suggests that the outcomes of our optimisation strategy that is based on subjective preferences should be neatly controlled in order to avoid misleading conclusions. Auditory learning and procedural learning effects could be mixed up. These learning processes involve gaining auditory experience and mastering the rules of the optimisation strategy, respectively.

The complex interaction between cognitive and sensory processes brings us back to the role of the hearing-impaired subject. It could be questioned, first, to what extent hearing-impaired subjects actually want compensation for their cochlear hearing losses, and second to what extent compensation is advantageous or disadvantageous. For example, measuring speech perception scores or thresholds can assess optimal results with respect to speech intelligibility. However, the gain in speech intelligibility found for the speech enhancing or noise reducing techniques that were studied was often small or even negative. Additionally, when at the same time the listening comfort is reduced, this affects the benefit experienced by the hearing-impaired subject.[61] Moreover, Bronkhorst (2000) suggested that a closer examination of factors like spectral differences, fluctuations, voice segregation, spatial separation, reverberation, divided attention, and hearing impairment is needed to understand speech intelligibility in background noise. For example, there is clear evidence that performance increases for selective listening relative to divided attention.

Finally, it can be questioned to what extent scientists in audiological research have seriously tried to imagine how a subject with cochlear hearing loss perceives the acoustical world. And additionally, what are the consequences for the hearing-impaired subject? When the effort used to develop hearing-aid signal-processing strategies is to a large extent transferred to modelling the perceptual consequences of cochlear hearing loss, we might be better able to understand the behaviour of hearing-impaired subjects. Then, an explanation could be found why many digital signal-processing schemes studied here have relatively poor benefits.

8.3 Effects of signal processing

The previous section concluded by giving an explanation why many digital signal-processing strategies are not that successful as expected. Apart from trade-off effects that were discussed in section 7.6, there are at least three other explanations. These comprise additional detrimental distortions, acclimatisation effects, and wrong expectations due to a one-dimensional perspective. These explanations are discussed hereafter.

---

[61] In analogy, the subject should be able to choose among compensation by one or two hearing aid. A bilateral fitting can result in higher intelligibility, better localisation, better sound quality and a better stereophonic effect (Boymans, 2003; Espmark, 2003). Though, a unilateral fitting can be preferred due to higher pleasantness of the subjects' own voice and the use of a telephone by the unaided ear (Espmark, 2003). Hence, by weighing these factors, the subject should decide for a unilateral or a bilateral fitting of hearing aids.

Distortion

Cochlear hearing loss can distort speech, due to spectral smearing and loudness recruitment (e.g. Van Tassell, 1993; Needleman and Crandell, 1997). The effects are largest when speech is presented in background noise. In addition, digital signal-processing algorithms can also cause serious distortions, especially when rather severe degrees of processing are applied. In this study, this occurred specifically for high parameter values of the spectral enhancement and noise reduction schemes. Therefore, it has been shown that speech intelligibility can be reduced by distortions.

Transmission (audibility) problems that produce unbalance of spectral amplitudes can be compensated for reasonably well. It seems that even transmission problems of deaf or profoundly hearing-impaired subjects can, in some cases, be solved by the aid of cochlear implants; an advancement that Sachs considered impossible by whatever imaginable technology in 1989. Supra-threshold conversion problems that result from cochlear hearing loss can be compensated for as well, but compensation cannot always help regaining information that is partly lost or distorted.

Transmission problems due to conduction problems in the middle ear or inner ear seem to be better manageable than transduction problems like distortions and loudness recruitment. To illustrate this difference the effects of hearing aids and cochlear implants can be compared. It seems that the plasticity of the brain can manage dramatic changes in neural stimulation due to transmission problems more easily than distorted sensory input due to transduction problems. There are at least three possible explanations for this discrepancy. First, the two solutions, a hearing-aid strategy and a cochlear implant, differ in their places of activation. The impaired cochlea distorts the manipulated information presented by a hearing aid. This stage is skipped by the cochlear implant. In effect, neural manipulation seems more promising than sensory manipulation. Second, subjects might accept dramatic changes easier than small changes. This psychological effect is enhanced by the fact that an implantation is an irreversible process. Subjects have to make the best of it, which certainly is not the case for a hearing aid. Many subjects refuse to wear a badly fitted hearing aid. Third, profoundly hearing-impaired subjects that regain auditory information by a cochlear implant are, as it were, happy with any improvement. Mildly to moderately hearing-impaired subjects that still do have access to distorted auditory information are much more critical. Any change in their auditory world will be tested against what they are accustomed to. Therefore, it seems that a signal-processing strategy must be improving without any detrimental side effects.

Acclimatisation

A second explanation for rather poorly performing signal-processing strategies is acclimatisation. The effects of several signal-processing strategies were studied in chapters 4 and 5. The results of the multi-directional pattern search were analysed with respect to the reliability of the method and the optimal combinations of algorithms. The moderate reliability and the discrepancy between the results of hearing-impaired subjects and simulated hearing-impaired subjects were partly explained by acclimatisation. Subjects need time to get accustomed to their renewed auditory world (e.g. Gatehouse, 1992; Horwitz and Turner, 1997). Hearing-impaired subjects cannot benefit from

spectral enhancement, for example, when they are not trained to make use of spectral cues. Similarly subjects with cochlear implants need an extensive training program to learn using auditory information in order to (re)gain speech perception. It is strange that such a program is generally not used for hearing-aid use.

Acclimatisation has huge consequences for the fitting practice. In due time the optimal combination of algorithm settings can be expected to shift, when subjects learn to make use of additional cues. In this respect the results of chapter 5 are illustrative. Hearing-impaired subjects and simulated hearing- impaired do not share the same starting point. Yund and Crane (1997) state that when a hearing aid (in combination with the impaired cochlea) is able to retrieve signal patterns of normal-hearing subjects, it is not likely that a hearing impaired subject (HI) perceives speech similarly to normal hearing subjects (NH). The reverse is also true: when an impaired cochlea is simulated adequately it is still not likely that the simulated hearing-impaired subject (SIM) perceives speech similarly to a hearing-impaired subject. The reason for this is that the two groups, i.e. NH and HI or SIM and HI, do not share the same acclimatisation state. The normal hearing subjects with or without simulated cochlear hearing loss maintain using the same cues, whereas the hearing-impaired subjects stick to the acclimatised speech-cue utilisation, irrespective the use of a hearing aid. As a consequence there is a confounding effect when the optima found for the different groups are compared. The results found in chapter 5 might imply that this confounding effect is largest for the noise reduction strategy used there. In the optimisation process acclimatisation is not easy to control.[62]

There seems to be no direct solution to this problem. The acclimatised state can be changed only by longer experience, or by using a paradigm different from paired comparisons. The use of a discrimination paradigm, discriminating one phoneme among other phonemes in a VCV or CVC format, could help minimising this confounding effect. However, this discrimination paradigm most probably will not be discriminative enough for optimisation purposes, just like the sentence or word identification strategies discussed in chapter 6. Another drawback is that VCV or CVC words do not represent the real-life situation very well, as is more the case for the sentences used for our optimisation strategy.

Based on the assumptions described a training program can be taken into consideration. Repp (1987) stated that signal-processing strategies that try to enhance speech characteristics, i.e. improving the acoustic signal, cannot improve speech intelligibility, unless the 'prototypes' are rebuild by an extensive training. Fourcin (1989) showed that hearing-impaired subjects adopt different auditory processing strategies if they cannot use high-frequency cues to discriminate among consonantal plosives because of high-frequency hearing losses. In the same line it could be stated that speech-enhancing strategies like spectral enhancement and fast compression can be appreciated only when subjects with cochlear hearing losses develop different listening cues.

---

[62] It would be interesting to study the effects of acclimatisation and shifting cues by using hearing-impaired subjects and simulated hearing-impaired subjects that participate in different treatments over a longer period. The treatments would comprise either a training program or not for separate hearing-aid configurations. These configurations could be (a) 'no processing' only, (b) both 'no processing' and 'some signal-processing strategy' used in alternation, and (c) 'some signal-processing strategy' only.

One-dimensional approach

A third reason why specifically spectral enhancement and fast compression do not appear very promising could also be found in the one-dimensionality of the scientific perspective that is used in developing these speech-enhancing strategies. The manipulation of acoustic signals is performed, and thought of, in separate stadiums. That is, processing can be performed in the spectral domain, the amplitude domain, or the temporal domain. This perspective seems to dominate the development of signal-processing strategies used for hearing aids. An exception might be the state-of-the-art spectro-temporal enhancement algorithm, although the actual strategy is known by the manufacturer only. The concept of spectral enhancement might be problematic, because it presupposes that the broadening of auditory filters can be compensated for independently from other effects of cochlear hearing loss. This seems to be a gross oversimplification, given the results of chapter 5. Addition of masking noise (used to simulate elevated thresholds) after spectral smearing (used to simulate broadened auditory filters) cancels out the benefit of spectral enhancement. This result underlines that cochlear processes like spectral filtering and compression in the cochlea are probably performed simultaneously (Oxenham, 2003) and in interaction.

8.4 Optimisation in the perceptual domain

Today, there is a large collection of optimisation strategies. Methods that excel in simplicity are direct search strategies. Hooke and Jeeves (1961) defined direct search strategies as follows: 'We use the phrase "direct search" to describe sequential examination of trial solutions involving comparison of each trial solution with the "best" obtained up to that time together with a strategy for determining (as a function of earlier results) what the next trial solution will be.' The pattern search, which is a direct search method, was chosen for its simplicity, because optimisation in the multidimensional perceptual domain was expected to be very complex. Another advantage is that direct search methods do not use derivative information, which makes them not very susceptible to small irregularities in the response surface. Therefore, a carefully chosen, carefully implemented direct search method like the pattern search can be an effective tool for solving many non-linear optimisation problems. It should be noted, however, that there is no strategy that fits all optimisation problems.

Enhancement of the pattern search

The pattern search method has demonstrated to be quite reliable and converges towards the optimum setting reasonably well. However, there is still some scatter in the optima found, depending on procedural, acoustic, auditory, and cognitive factors. These factors can be fine-tuned to the specific circumstances in which the optimisation strategy is used. But in addition the procedure might be enhanced in several ways. A direction analysis as described in section 4.6 can be used as an alternative termination rule of the multi-directional pattern search. The optimal combination of algorithms might be found with more certainty when a sequential pattern search format can is used. Such a format prescribes a fixed large step size for the first search, and decreasing fixed step sizes for

succeeding searches that start at the optimum found in the previous search. The reliability of the movements can be enriched by the use of multiple paired comparisons for each round, by adding a rating paradigm, or by including a round-robin tournament. Also, the paradigm of subjective judgment can be exchanged by an objective identification format.

Another way to enhance the reliability is by using multiple paired comparisons for each pair of samples that need to be compared. An additional advantage is that the consistency of the procedure can be estimated by the use of multiple measurements. It is also possible that the number of paired comparisons for each round depends on the step size that is used. That is, for larger step sizes the need to have multiple paired comparisons is much smaller than for small step sizes. A disadvantage of the use of multiple paired comparisons is, of course, the time, needed to end the procedure.

To improve the sensitivity of the paired-comparison procedure, Fabry and Schum (1994) proposed to introduce that strength of preference rating can be elicited from the subject. The use of a rating procedure in order to find out to what extent one of the samples is preferred can help to get a kind of equalization between the separate dimensions on the one hand, and a better estimation of the steepest path, needed to find the optimum as quickly as possible, on the other. In practice, a higher rating indicates a definite step in that direction; a small rating indicates no step. In the last case either the step size should be increased, i.e. when there is really no difference between the samples, or the search direction should be reversed. The reversion acts as a solution of the problem to distinguish the difference limen from the comfort limen, i.e. there is really no difference, or the samples are preferred equally bad or equally well.

The stop criteria used in this study take information along the walk towards the optimum into account, because both the number of visits and the number of paired-comparison wins is recorded. Only the information about the direction of movement is not considered in the estimation of the optimum setting. The outcomes of the direction analysis suggest that the directions of movement can predict the optimum adequately only, when the response surface is smoothly curved. However, the analysis cannot take irregularities of the perceptual landscape into account. An alternative purpose of information gained from the response patterns is to use the degree of peakyness. The combination of the range of matches and the second derivative around the presumed optimum can yield a qualification of the consistency of moving in the pattern search. Low values of the measure of peakedness and small peaks in the response pattern can signify the presence of criterion drift. This is expected to be due to multimodal, broad or flat response distributions.

Other direct search strategies

Direct search strategies other than pattern search involve the simplex method and methods with adaptive sets of search directions.

The simplex method (for details see appendix 1) was designed to decrease the number of experiments needed to guide direct search optimisation strategies. That is, a full factorial design needs 2^k experiments, whereas the frugal simplex design needs only k+1, with k the number of dimensions (the pattern search needs 2k experiments). Full

factorial designs were developed to model interactions between factors. More elaborate designs like star designs and central-composite designs (mix of full factorial and star design) need even more experiments, i.e. 2^k+1 and 2^k+2^k+1 experiments, respectively. These designs give information about the local response surface and from this information elliptical contours of constant response can be drawn (see e.g. Öberg and Deming, 2000). The empirical nature of these designs makes them especially suitable to understand the underlying perceptual mechanisms. However, due to the large number of experiments that is required these elaborate designs are less appropriate for optimisation purposes in our type of applications.

Humphrey and Wilson (2000) proposed to revise the simplex search to avoid sensitivity to starting points and termination at a suboptimal setting. The simplex strategy was revised in several respects. The search is performed in multiple stages. In each stage the step size was fixed. The end point of each stage was the starting point of the next stage. It appeared important to step away from the current end point to avoid staying at a suboptimal setting. Finally, the final optimum was chosen to be the best end point. Such a revision could also be used for the pattern search we studied. For it appeared from section 4.6 that when the end point was used as a new starting point the search did not always end at the same optimum.

Due to the movement of the simplex by reflections for designs with more than two dimensions the reflected simplex does not necessarily lie on a predetermined lattice. That makes the simplex design difficult to implement when the algorithms do not operate real-time, as was the case for our studies. Additionally, when fast flexible switching is not possible, which could be the case for state-of-the-art hearing aids, the simplex design is also unattractive.

The abovementioned simplex strategies need response values in order to determine the best or worst vertex. The paired-comparison paradigm does not yield responses that can be used by the simplex strategy. To circumvent this problem, a round-robin tournament could be incorporated in order to retrieve values from several paired comparisons. In a three-dimensional set-up four vertices are required for the simplex strategy. Hence a round-robin tournament between these four vertices would require six paired comparisons.

The inclusion of a Round-Robin tournament in the optimisation strategy has two important advantages. First, the walk in the search procedure can be transformed into a simplex-based strategy. The advantage is that movements in the simplex-based method are simpler and include always one common setting. For clarity the pattern search movements and two types of simplex movements are described in appendix 5. Four settings are compared and the least preferred setting is reflected in the centre of gravity of the other three settings. It is even better to reflect the three least preferred settings in the most preferred setting for reasons of good convergence. For the pattern search it is possible that, due to a diagonal walk, all settings to be compared are new for the subject. Some listeners in our study complained that they heard the most preferred combination of algorithm setting, but that the procedure walks away from it and that it is hard to get back at the optimum. This is indeed possible when two or three settings are preferred, causing the procedure to walk diagonal. Second, reflections are based on the outcomes

of six paired comparisons (0.5n(n-1) for n vertices), which makes the movements trustworthier. Moreover, a consistency measure can be determined by calculating the number of circular triads, in order to know how much we can trust the movements (David, 1963).

The paired-comparison paradigm can also be exchanged by an identification paradigm. Just like the inclusion of the round-robin tournament the measurement of correct identifications makes it possible to use a simplex-based method. By designing a test that incorporates a limited number of phonemes, the speech intelligibility can be estimated and the advantage of several combinations of algorithms can be investigated. Maybe the most crucial advantage of such an optimisation strategy is the possibility to give feedback. A severe drawback of the subjectively oriented paired-comparison paradigm is the fact that it is difficult to give feedback. The three-dimensional search route could be presented visually to the subjects. However, this could affect the search towards the optimum. A major disadvantage of the identification paradigm is that the procedure will be time-consuming, and the sensitivity of the test could be rather poor, just like for the SRT or CVC paradigm.

A final adaptation, both suitable for the simplex method and the pattern search, is to rotate the set of search directions, e.g. by 45º. This adaptation could be incorporated to avoid termination at ridges or saddle points. In general, when the search directions are parallel to contours of constant response the simplex procedure or pattern search are deemed to terminate prematurely.

Strategies that avoid premature termination are methods with adaptive sets of search direction. These methods are designed to use information about the curvature of the response surface during the optimisation process.

Complex optimisation strategies

Direct search methods are simple and computationally efficient. However, these methods perform badly for non-smooth response surfaces. Among a large collection of complex optimisation strategies there are two methods that were designed especially to tackle the problem of a multimodal or discontinuous surface. These are the genetic algorithm (e.g. Durant *et al.*, 2002) and the simulating annealing global optimisation routine (e.g. Neptune, 1999).[63] An additional advantage is that these methods are not sensitive to the initial guess, i.e. the starting point, as is the case for the pattern and simplex search. However, the genetic and simulating annealing optimisation routines are time inefficient and are computationally burdensome. The simulating annealing routine has the additional disadvantage that it is not suitable for optimisation based on paired comparisons. Due to the relativity of the paired-comparison results it is not possible to explore the entire space and converge on the most promising region.

[63] In many studies that describe complex optimisation strategies it is often erroneously assumed that the global optimum can be found. Especially when only a few experiments can be performed, it is unrealistic to assume that the global optimum will be found. In fact, the global optimum can be found only when an infinitely small grid is used and all settings that lie on this grid are evaluated. For discontinuities can never be predicted.

8.5 Perceiving the world is changing it

Several approaches were used to understand why the reliability of the pattern search was poorer than expected. Additionally, possible explanations were suggested to understand why the signal-processing strategies were not very successful. These approaches and explanations evolved from ideas from the psychophysical and audiological field. The underlying assumptions were not discussed so far. Therefore, we conclude this thesis with a critique of these scientific assumptions. To understand the notions that were used, we explore ideas of how we perceive. In this way concepts like sensation, reception, and perception are addressed. These concepts structure the way we study perception in general and our optimisation strategy in particular.

Perception is both fascinating and intricate. Based on our perceptions we structure the world around us. Our perceptions are strictly individual but at the same time they are also communicable, i.e. universal to some extent. This illustrates the complexity and problematic character of the concept of perception.

The empirical tradition of science is founded on the idea that knowledge about the world around us can be derived from our experience. This questionable assumption has led to much confusion.[64] Magee (1997) complains repeatedly about mixing up the experienced world and the world in reality. Because we consider our senses for granted everything that we perceive we erroneously call reality. In the tradition of the philosopher Kant, Magee states that we are never able to know about the real world. This is due to the fact that we have to use our perceptual system to derive information about the real world. Therefore Kant discerns the 'Dinge als Erscheinungen' (Phenomena) and the 'Dinge an sich' (Noumena). Consequently, we can never escape from our perception or as Kant puts it: 'daß wir nähmlich von den Dingen nur das a priori erkennen, was wir selbst in sie legen'. As a consequence due to our limited perceptual abilities much of what really exists will be unnoticed.

This view has two major consequences. First, perception is characterised by the principle of ordering. Second, it is not possible for a normal hearing subject to experience what a hearing-impaired subject perceives. These consequences will be discussed in separate sections hereafter.

The ordering principle

The outside world is received by our sensory systems. Our mental system organises the material and delivers concepts needed to understand our experiences. The sensory or receptive process involves several stages in which mechanical energy is transformed into neural energy. The acoustical information is received by the sensory system, coded by

[64] Of course the rational qualities of empirical science are high in that predictions about future events can be made excellently. However, Hume and Kant make less optimistic and exaggerated claims on the powers of human reason. Kant emphasizes that empirical research cannot explain reason. Hume states that things outside human experience are unintelligible. For instance, it only seems as if there is a necessary connection between objects. Hume states that there is a connection only between the ideas of those objects, i.e. the perceived objects. Cause and effect exist in our minds only. Additionally, things within the range of experience cannot be known by rational inference. According to Russell (1961) this idea is already found around 500 by Protagoras who distinguishes percepts and inferences or implications. Man is inescapably confined to his own percepts, and can know about the percepts of others only by inference, i.e. by communication.

the hair cells into neural firing patterns, and transmitted to the brain where the information is processed by the central nervous system. But what, then, is perception?

Held and Richards (1972) state that the act of perception is characterised by internal processing of the sensory information, which is based on an internal code of the processor. This code is carried by the firing patterns and the spatial organisation of the neurons. The great advantage of this psychological information-processing approach[65] is that perception and other mental processes are treated in one single system. This system can be considered to be a representational system (Hofstadter, 1986; Hofstadter and Dennett, 1989). The heart of an intelligent system is that it is an active set of multi-layered structures that reorganises constantly. As a consequence, such a system needs to be highly flexible in order to be capable of representing the actual situation of the external world that is dynamic.[66] Hence, in line with Luria's beliefs (1976), reflection of reality by perception is not passive, but an active process that makes it possible to adapt to the environment and structure it. Or, as Cohen (1989) puts it, mental images of the real world are not static pictures, but plans for obtaining information.

We deliberately avoided the term sensation and avoided assigning a place in the brain where perception takes place. The reason not using the term sensation is that this term refers to immediate and basic experiencing.[67] The idea of immediate experience is also referred to as 'simple' or 'direct' perception, which is opposite to significant perception. This idea originates from Berkeley and is founded on the concept of sense data.[68] This concept assumes that there is something like immediate data that exist without intervention of any theory, judgment, or presumptions (Koningsveld, 1987). A lot of philosophers debated the question whether the idea of 'simple' perception is valuable (e.g., Brown, 1993; Dretske, 2000). They argued whether perception is always concept-laden; that we always 'perceive as'. Or is it possible that identification can be stripped off? Measuring instruments can be considered as sensing devices. However, these instruments cannot perceive. Hence, in perception there is always some degree of conceptualisation.[69] Consequently, the sensory system is a delivery device only.

---

[65] The information-processing approach of perception is one perspective among many others. There are, for instance, more structuralistic, functionalistic, Gestalt, and Behaviouristic approaches (e.g., see Kellogg, 1997; Matlin and Foley, 1997). A fundamental limitation of structuralism that is often used in scientific reasoning is that it presupposes an objective point of view. Mental processes like perception, cognition, and consciousness rather relate to subjective aspects of experiences. Therefore, it seems that these processes show the limits of scientific explanation. Psychophysics is typically a behaviouristic approach.

[66] Hofstadter (1986) digresses upon intelligent, representational systems by using an anthill as an example. The operation of an anthill is very illustrative for our understanding of the operation of the brain.

[67] This meaning of sensation was already used in the work of Kant (1781), but was also used in the work of Matlin and Foley (1997). When sensation is used in terms of sensory processes there are no problems.

[68] This term is defined by Russell and Ayer (Koningsveld, 1987) and is in the phenomenological tradition also called impressions (Hume), Empfindungen (Mach), and Elementarerlebnisse (Carnap).

[69] Perception is an active act; concepts represent the point of view. In his critical rationalistic view Popper states that perception and observation are theory-laden. He rejects the ideas of inductive confirmation and real facts acquired by scientific observation (Koningveld, 1987). On inductive confirmation, Kant (1981/1781&1787) says 'Erfahrung gibt niemals ihren Urteilen wahre oder strenge, sondern nur angenommene und komparative Allgemeinheit (durch Induktion), so daß es eigentlich heißen muß: so viel wir bisher wahrgenommen haben, findet sich von dieser oder jener Regel keine Ausnahme.' Also along with Kant, Popper denies the existence of 'naked' perception. Instead, Popper used the concept of a searchlight to indicate that (theoretical) expectations guide our perceptions and observations. Hence, conceptualisation implies learning to perceive in a specific way. That is, one has to know in advance the

Perception is connected with meaning (Merleau-Ponty, 1978), and is limited by a subject's collection of concepts. Knowledge and belief play a crucial role in what we perceive.[70] Magee (2001) asserts that it is impossible to have a representation of the outside world that is independent of experience (concepts) and mental processes. The sensory information is arranged and rearranged, guided by concepts. Hence, perception can be defined as becoming cognizant or being aware of entities in the outside world and is a state of mind.[71] Or as Raleigh (1903) puts it, 'a transaction between the outer powers that operate on the mind through the senses and the inner powers of the mind itself, which impose their own forms on the things submitted to it'.

This notion has major consequences on the experimental paradigms used for perception research. As for our studies the subjects' behaviour is probably seriously guided by their knowledge bases. It seems that listening without making associations is impossible. A direct consequence for our paired-comparison paradigm is that subjects have often made judgments on previous (pleasant and unpleasant) experiences. Presumptions are used, not a solid unbiased evaluation. This could explain some of the unpredictability of the subject's behaviour in the optimisation strategy. The listening strategies, cues, and foci of attention vary, which is in a sense typically human, and therefore is difficult to avoid. This is especially applicable to optimisation of multiple signal-processing strategies that change the perceptual world of the hearing-impaired subjects considerably and abruptly.

A common misunderstanding is that perception is considered as a second transducer.[72] Dennett (1995, 1996) indicates that there is no special place where consciousness 'happens'. Instead of a kind of special transduction he suggests that consciousness is somehow distributed in place and time throughout the brain. Similarly, perception could be considered as a state of mind and not as a medium. It seems difficult to consider neural activity as a resource of perception, instead of some central processor. The neural network itself contains perception, due to its complex multi-layered structure.

Concepts and percepts differ in that percepts directly relate to external entities, whereas concepts refer to relations between these percepts. Percepts apply to the present, concepts to the past. Concepts have to do with how we structure our experiences, how we categorise the entities, and how we use language. The ordering is not performed in formal structures or on logical relations, but more on associations. Therefore, concepts are also social, cultural, and historical in origin (Toulmin, 1972; Luria, 1976). This makes scientific psychology, as proposed by Wundt (1874), not adequate for all mental activities.

---

relevant factors in order to abstract, i.e. extract information. This is opposite to the empirical point of view that considers conceptualisation as a result of perception.

[70] We know, for example that a tick in a musical recording is an artefact and does not belong to the music. A young infant does not have this knowledge yet.

[71] For auditory perception the use of 'objects' or 'things' to designate percepts is problematic. Rée (1999) rightly claims that the objects that concern our hearing are the sounds produced by objects, not the objects themselves. We do not hear a car passing, but the sound it produces. Therefore, terms like 'entity' and 'event' are more appropriate.

[72] The first type is the transduction of mechanical energy into neural firings, e.g. in the cochlea.

An implication for subjects with cochlear hearing losses is that before neural processes take place information is lost or distorted. As a consequence perception will be different in some respects.[73] The concepts of hearing-impaired subjects might well be unchanged, because the internal representation of language or abstract thought is not. Milligan (Magee and Milligan, 2001) asserts that knowledge of deaf and hearing people can be the same, whereas their experiences are different. However, in chapter 5 the interpretation of the term listening comfort appeared different for normal hearing and hearing-impaired subjects. It seems that our normal hearing views have shaped our scientific premises wrongly. A typically normal hearing view is the idea that listening comfort might be an important evaluation criterion in hearing-aid optimisation. Normal hearing and hearing-impaired subjects have very different ideas of what listening comfort is. Normal hearing subjects especially attended brightness and fluency both of the speech and the background noise. Contrary, hearing-impaired subjects mainly attended the speech signal and experience this signal as comfortable mainly when the speech intelligibility improved.

It is possible that an evaluation criterion that avoids focusing at speech intelligibility improves the reliability of the pattern search. For instance, a paired-comparison strategy is also used in the visual domain by determining optimal glasses. Here the task is not to evaluate the readableness of characters, but the contrast of the black characters in the white background. Maybe our optimisation strategy should not be directed by evaluating the speech intelligibility/listening comfort, but by evaluating a certain type of contrast. Such a meaningless criterion is probably more suitable for optimisation.

Simulation of the perception in case of hearing-impairment

The experienced or empirical world is the world of all possible perceptions and therefore depends on the subject. Hence, we are never able to hear or perceive what another person hears or perceives, because we have to make an image of it ourselves. Even when it is possible to simulate cochlear hearing loss, it is still impossible to simulate the effects of cochlear hearing loss, i.e. what the hearing-impaired subject perceives. Although it might be possible to model the deteriorated processing in an impaired hearing system, whether the impairment is in the periphery or central, the perceptual consequences cannot be simulated.

However, when we realise what the limits are of simulating hearing impairment, models can help to systematically analyse the effects of cochlear impairment on peripheral or central processing. This can contribute to our understanding of the effects of signal-processing strategies developed for hearing aids. As an example, in chapter 5 an explanation was suggested why spectral enhancement did not yield higher listening comfort.

---

[73] Repp (1987) suggested that distortions might not only be present in the impaired cochlea, but also in the mental representations. Due to these distortions it might be more difficult for subjects with cochlear hearing losses to separate between speech and background noise. It seems that these subjects cannot filter the disturbing background noise. This problem can be enhanced by hearing aids that tend to deteriorate the spaciousness. The spaciousness is squeezed such that all sounds seem equidistant. As a consequence, subjects sometimes experience hearing aids in situations with multiple sources as annoying and tiring.

Additionally, hearing-impaired subjects with similar cochlear hearing losses at least share some perceptual consequences. For example, speech intelligibility deteriorates for subjects with elevated thresholds. But intelligibility is not equivalent to perception, or reception. This stresses the significance of using unambiguous terminology.

8.6 Consequences for scientific work

This thesis concludes with a discussion on the scientific consequences of the active processes in perception of complex stimuli. Psychophysical and cognitive psychological perspectives are compared to determine what models and ideas can enhance our understanding of the consequences of cochlear hearing loss. Our views about perception result in considering the concepts of relative perception, categorical perception, and speech perception.

Relative perception

Psychophysics, as traditionally defined, is the science of describing the relationship between objective (physical) and subjective (psychological) dimensions.[74] Psychophysical laws, founded for example by Fechner and Weber, generally apply to simple signals. For these sounds it is relatively easy to systematically manipulate physical parameters. Hence, detection, discrimination, identification, and judging were primarily studied in the psychoacoustical domain.

The classical psychophysical approach follows the idea of Descartes in that it assumes an isomorphic correspondence between entities in the external world and ideas in the human mind. Kant (1981/1781&1787) doubts this assumption, for we have knowledge of the ideas only. Heteromorphic primitives, as defined by Liberman (1989), have the ability to make adjustments to the internal mechanisms. These need to change for example when our heads grow bigger in order to be able to localise with two ears.

Norwich (1974) and Wong (2000) avoid the pitfalls of isomorphism by founding their work on Berkeley's concept of relative perception. Perception is relative, because it depends on the state of the perceiver. They promote the idea that perception is making a choice among alternatives. Everything is probability and thus perception involves uncertainty. The uncertainty is different for different perceivers either by assuming that the sets of perceivable alternatives differ (Norwich, 1974), or by assuming that the values of a priori probabilities differ (Wong, 2000). The approach of relative perception uses the idea that information is a reduction in uncertainty. Norwich adapted the entropy theory of Boltzmann to the information theory. The brain receives only some degree of uncertainty from the neurons, from which a representation of the external world needs to be apprehended. In this way perception reflects not what the world is, but what the neurones think it is.

The approach of Norwich might be limited, because it is founded on perceptual units, a structure that can perceive on its own. Contrary, a neural network uses the

[74] A scientific discipline can also be defined by the problems under consideration and the methods that are used. Toulmin (1972) discerns to aspects. A scientist learns to employ general scientific procedures. Additionally, he learns to recognise for what particular situations these procedures apply.

interplay of many simple units. However, neither of the two approaches uses a multi-layered structure that might better describe the process of perception.[75]

Categorical perception

Repp (1987) states that a definition of perception that includes categorisation and attentional filtering might agree better with the complex multi-layered structure of the brain. Cognitive psychology might be a valuable alternative approach in that it studies mental structures that are innately present and acquired by experience. From our studies it appeared that variation of acoustical parameters has limited explanatory power. The role of cognitive factors like context and exchanging relations between phonemes (Repp, 1987, Plomp, 2002) are missed when the data are explained by acoustical factors alone. Listeners use multi-dimensional criteria in perception that cannot be studied by psychophysics.[76] It seems that reasoning from a perceptual perspective, as used in chapter 6 to explain the moderate reliability, can reveal unexpected influencing factors. Perceptual knowledge can be used as a starting point that guides our anticipating behaviour and is based on the expectations of the perceiver (Cohen, 1989). Based on this knowledge base perception shapes gradually. Hence, listening is an active process that guides our way to what we intend to hear. Consequently, without a specific goal it is difficult to search for an optimum combination of signal-processing strategies. Searching for a combination that has highest listening comfort is probably difficult, because the subject does not know what to expect. However, when the evaluation criterion is tightly defined, e.g. a brilliance criterion, the subject possibly reaches the optimum much more reliably.

Speech perception

Speech is more than a collection of sounds. Psychophysics tends to study speech-sound perception more than speech or phoneme perception. In cognitive psychology, however, complex signals like speech and music can be studied using complex stimuli, although systematic variation will be hard or even impossible to accomplish. Hirsh and Watson (1996) stressed the significance of using complex stimuli. They argued that despite the use of a wide variety of non-speech sounds no strong correlations were found between psycho-acoustic abilities, like spectral and temporal discrimination, and speech perception. Hence, central factors may play an important role in speech processing, especially to explain individual differences in speech recognition in difficult listening situations. A time-consuming but important task would be to draw perceptual maps of effects of algorithm combinations. This could enhance our understanding of perceptual processes of (simulated) hearing-impaired subjects.[77]

---

[75] Both theories resemble the entropy theory of Boltzmann, for which only one step can be taken from molecules to a complete gas. The intelligent system as described by Hofstadter (1986) is much more differentiated; an anthill is organised in castes, groups, gangs, signals, and symbols.

[76] In 1886 Hyslop disputed the very optimistic expectations of the psychophysical approach. In his view the inquiries of this discipline exclude phenomena like circulation, digestion, assimilation, reasoning, memory and imagination.

[77] Kuhl presented perceptual maps to illustrate the concept of magnet effects and prototypes of phonemes in a video 'Speech perception,' released by the Acoustical Society of America.

Complex signals are especially useful to study speech perception.[78] It is suggested many times (e.g. Kuhl and Meltzoff, 1982; Liberman and Mattingly, 1989) that there could be something like a speech module or phonetic module. Such a module would be particularly apt to analyse speech, based on its own modes of analysis, and without using factors like pitch, loudness, or timbre. According to Kuhl and Meltzoff (1982) this mode can be attributed to a relation between audition and articulation. This relation is based on the ability of infants to detect correspondences between auditory and visual factors, due to the tendency to imitate. The existence of phoneme boundaries was explained by this motor theory of speech perception (e.g. Harnad, 1987; Hirsh and Watson, 1996). It is assumed that perception of speech is influenced by how it is produced.

Whether this specialised speech module is related to the production of speech or not, this module assumes that speech perception is a relational process. Speech is internally related to a phonetic knowledge base. The goodness of fit corresponds to the degree of uncertainty experienced (Repp, 1987). Therefore, speech perception is not only a complex activity, but also has a probabilistic structure, and involves analysis, synthesis, and decision-making. This process has been modelled by the concept of categorical perception (Harnad, 1987; Kuhl *et al.*, 1992). That is, incoming speech information is assigned to a familiar category that involves considerable abstraction. A common phenomenon that refers to categorical perception is the existence of prototypes. Prototypes are characteristic features that function as a kind of idealistic representatives of a given phonetic category. Kuhl *et al.* (1992) have shown that phonetic prototypes act like 'perceptual magnets'. As a consequence non-prototypic members of a phoneme category are perceived more similar to the prototype than to each other. Additionally, it seems that linguistic experience affects phonetic perception in that perceptual distances around prototypes get larger, relative to non-prototypes. Hence, this could explain the robustness of speech perception. Irrespective of the signal-processing strategy used, hearing-impaired subjects might tend to relate manipulated phonemes to their prototypes. This could explain that the beneficial effects of speech-enhancing strategies are generally smaller than expected.

8.7 Recommendations

The results of this thesis imply that a new kind of fine-tuning strategy needs to be implemented in the audiological practice. Future work should seek to attain an integrated experience-driven optimisation strategy. Such a strategy is confined to three main implications of the present work that need to be implemented: a synchronisation paradigm, a subject-interaction paradigm, and an active-relative perception paradigm. The principles of these paradigms are as follows.

First, the synchronisation paradigm asserts that signal-processing strategies used in hearing aids should be optimised simultaneously. The reason to use such an approach is

[78] In the literature many definitions are in use to describe speech perception. Of course these definitions are not always used to study the same aspects. However, there is a clear tendency to mix up definitions like speech perception, speech reception, speech intelligibility, speech understanding, and comprehension of speech. It seems that reception is used to describe measurements. We prefer to use the term perception, because this term illustrates that it is more than intelligibility or comprehension alone.

that signal-processing strategies interact. Thus the optimal setting cannot be determined by optimising the signal-processing strategies separately. Analyses of the estimated winner and the types of response distributions found by the round-robin tournaments in section 4.2 did reveal interactions between algorithms for hearing-impaired subjects.

Second, the subject-interaction paradigm asserts that optimisation of a hearing aid should be performed in interaction with the hearing-impaired person. The participation of the hearing-impaired person not only has a psychological advantage, but also is necessary to find the individual's optimum setting. The reason for this is that the audiologist will never know what the hearing-impaired person perceives. It appeared for example that the normal hearing subjects and the simulated hearing-impaired subjects in the study of chapter 5 seemed to use the same listening strategies to find the optimum. Therefore, the optima found for both groups of simulated hearing-impaired subjects were not always comparable to the optima found for hearing-impaired subjects. So, the hearing-impaired person him- or herself is the only one that determines the best combination of signal-processing strategies in the fine-tuning process.

Third, the active-relative perception paradigm asserts that an optimisation strategy should be based on a perception model proposed in this final chapter. This paradigm is founded on two questions posed earlier. To what extent hearing-impaired subjects actually want compensation for their cochlear hearing loss? And to what extent scientists in the audiological field have seriously tried to imagine how a subject with cochlear hearing loss perceives the acoustical world? We are never able to hear or perceive what another person hears or perceives, because we have to make an image of it ourselves. Therefore, not the physical dimensions of signals should guide the optimisation, but the perceptive and cognitive dimensions of the hearing-impaired individual. Two perception theories, active and relative perception, were discussed that could be used to improve the multi-directional pattern search. Active perception underlines that perception is an active process that makes it possible to adapt to the environment and structure it. Therefore, active perception inheres plans for obtaining information. Relative perception emphasizes that perception is relative, because it depends on the state of the perceiver. Hence, relative perception is making a choice among alternatives and involves uncertainty. In category perception the degree of uncertainty experienced corresponds to the goodness of fit in relation to the knowledge base of the perceiver.

An integrated experience-driven optimisation strategy requires all three aforementioned paradigms. The multi-directional pattern search combines simultaneous optimisation in interaction with the hearing-impaired subject. In addition, future studies should focus on perceptual measures instead of physical parameters in order to find the optimal setting of a hearing aid for hearing-impaired subjects. The optimisation strategy needs to be revised to meet auditory and cognitive aspects that caused the optimisation strategy studied in this thesis to be only modestly reliable. Aspects that deserve special attention are listening strategies and listening cues in relation to the evaluation criteria used. Active and relative perception can play a role in the optimisation process in two respects. First, instead of using a broadly defined listening strategy an active search goal might be more suitable. Second, uncertainty measures can be used to decide whether a movement is required and to determine the step size. As a result the problem of

responses like ‘equally comfortable’ or ‘equally uncomfortable’ and the existence of perceptual magnets (prototypes) can be managed.

This thesis showed the importance of subject interaction in fine-tuning of signal-processing strategies. In the near future this could lead to interesting applications in the audiological practice for the fitting of complex hearing aids. But at the same time this thesis showed that there are still some shortcomings that need to be studied. Therefore, additional refinements are necessary that follow the guidelines of a completely new perspective that, within the framework of this thesis, we called the integrated experience-driven optimisation strategy.

APPENDICES

Appendix 1

PATTERN SEARCH vs. SIMPLEX METHOD

Denotation

The optimisation method used in this thesis is called 'pattern search' instead of 'simplex method', as commonly used in the literature (e.g. Levitt, 1978; Neuman *et al.*, 1987; Kuk and Lau, 1996). The reasons are described below.

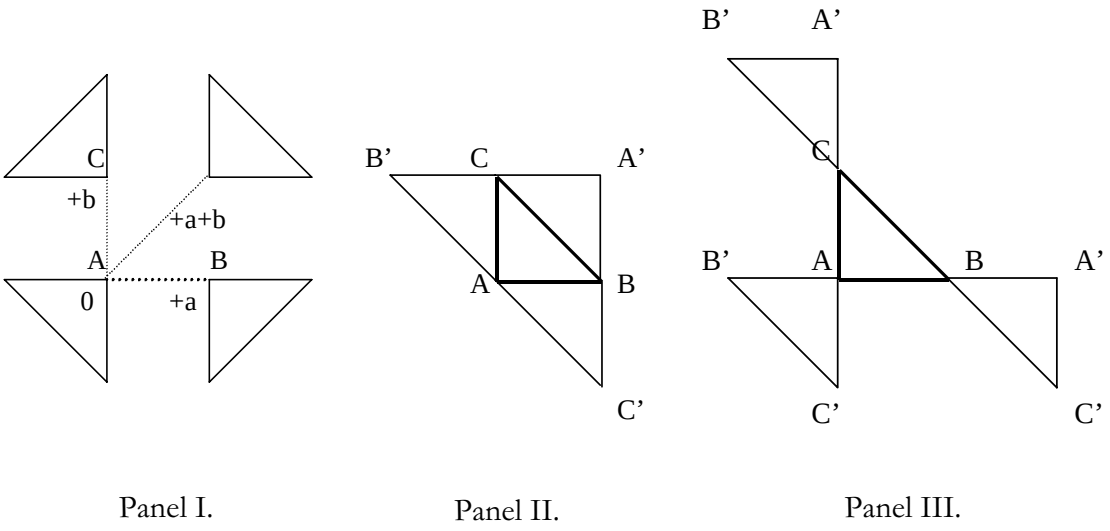
Using the paired-comparison paradigm the expression 'Simplex method' might be confusing. Spendley *et al.* (1962) have introduced the expression 'Simplex procedure' to track operating conditions by evaluation of the output of a set of points that form a Simplex. A 'Simplex' designates a Euclidean geometrical spatial element bounded by a minimum number of points. For two dimensions a Simplex is a triangle, in three dimensions a tetrahedron. So for n dimensions a Simplex contains n+1 vertices (boundary points). As such, the expression 'Simplex' can be used, because the settings that are compared pair wise correspond to vertices of a 'Simplex' configuration.

However, the composition 'Simplex method', defined by Spendley *et al.* (1962), was used for a procedure, a heuristic in fact, that is suitable to optimise unconstrained, multidimensional problems of independent dimensions by use of Simplices. In the procedure, reflection of one point in the centre of gravity (centroid) of the remaining points forms new Simplices. This so-called 'Simplicial' feature of the Simplex method is, among others, fundamentally different from the (modified) Simplex procedure of Neuman *et al.* (1987). When using that procedure in a paired-comparison paradigm, it is impossible to decide what vertex should be reflected. This means that different procedural rules had to be incorporated. Instead of reflection in one point, as is the case in the original procedure, the new 'Simplex' is created by a combination of translations and line-reflections. As a consequence, the new 'Simplex' can share at most 1 point with the old 'Simplex'. For Spendley's procedure the new 'Simplex' shares always n points with the old 'Simplex'. This procedural difference indicates that the original elegant movement of the 'Simplex' was abandoned in the modified procedure of Neuman. Hence, speaking in terms of 'Simplices' for Neuman's procedure is in a way unnatural. This is underlined by the fact that Neuman's procedure uses n responses instead of n+1 responses, as is the case for the original procedure.

To avoid ambiguous terminology, the most important features of our (and Neuman's) procedure have been selected and compared with other procedures that can be classified as direct search methods. Direct search methods involve comparisons of each trial with the best previous solution, without requiring any numerical function values. That is, relative ranks are sufficient, and that makes the use of the paired-comparison paradigm possible. Two features that characterize our pattern search procedure are the following. The use of a rational lattice, i.e., the underlying structure of the settings, comparisons, and movements, is shared with pattern search (Lewis, Torczon, and Trosset, 1998). The use of a history of information, needed for the adaptive set of search directions, is typical for the Rosenbrock method (Lewis, Torczon, and Trosset, 2000). However, the adaptive nature is restricted in that the search

directions remain always orthogonal, which makes it more pattern-search-like. That is, the search directions can be reversed only.

Basically, these features should be incorporated in the definition of our procedure. However, because it is likely that other features have been missed, and because the definition will become unnecessary complex it may be desirable to use a name that is relatively simple, without having some problematic connotation. 'Multi-directional search', as is used by Torczon (1989), might be a good alternative, although, again, the procedural rules are related to Spendley's Simplex method. Fortunately, this is not implicitly suggested by the term 'Multi-directional search'. To appreciate the lattice structure, the constrained search directions, and the factor interaction, it is most convenient to denote our procedure 'multi-directional pattern search'.



Panel I. Panel II. Panel III.

Figure A1.1 Different movements and search directions for (I) our pattern search, (II) the simplex procedure of Spendley *et al.* (1962), and (III) the multi-directional search of Torczon (1989).

Movements

Two different pattern search and two simplex procedures are described here to summarise their similarities and differences. The movements and search directions of each optimisation strategy are presented here. Our pattern search, the simplex procedure of Spendley *et al.* (1962), and the multi-directional search of Torczon (1989) are given in panel I, panel II, and panel III of figure A5.1, respectively. Symbols A, B, and C represent either settings to be compared (for panel I), or settings for which some sort of value can be measured (panel II and panel III). Symbols a, b, and 0 represent moves with sizes a, b, and 'no movement', respectively. Symbols A', B', and C' are reflections of settings A, B, and C.

In our pattern search the responses of the subject determine both the movement and the search direction. This is illustrated in figure A1.1, panel I. The movements in our pattern search are determined as follows. When setting B is preferred better than setting A, and setting A is preferred better than C, the pattern search moves by a step-size=+a

horizontally. When setting A is preferred better than B, and C is preferred better than A, the pattern search moves by a step-size=+b vertically. When both B and C are preferred better than A the pattern search moves by a step-size=+a+b diagonally. When A is preferred better than both B and C the pattern search does not move (step-size=0) Additionally, the search directions are determined as follows. When the step-size=+a the horizontal search direction remains the same, whereas the vertical search direction is reversed. When the step-size=+b the vertical search direction remains the same, whereas the horizontal search direction is reversed. When the step-size=+a+b the search direction remains the same both horizontally and vertically. When the step-size=0 the search direction is reversed both horizontally and vertically.

The original pattern search (e.g. Lewis, 1998, 2000; not shown here) is simpler in that a move is made after two paired comparisons in one dimension only (e.g. left and right). In the next round the other dimension is considered (e.g. up and down). Hence, the movements are never diagonal, but in one direction only. The step size was reduced *only* when no increase or decrease in any one direction further improved the fit. Hence the step size was ensured not to decrease prematurely.

The original simplex procedure of Spendley *et al.* (1962)[79] is given in panel II of figure A1.1. The setting with the lowest value is reflected in the centre of gravity (centroid) of the other two settings. When, for example the value of setting A is lowest this setting is reflected. A new simplex is, thus, created that comprises settings A', B, and C. Nelder and Mead (1965) revised the simplex procedure by allowing different step-sizes for different dimensions. Hence, the simplex geometry adapts itself to the local landscape, contracting in the vicinity of the optimum, and elongating along inclined planes. For that reason, the adaptive Simplex method has also been called an amoeba procedure (e.g. Press *et al.*, 1989, Hong and Vasaru, 1996), because of its perpetually changing shape. That is, the walls of the amoeba are elastic and adapt to the environment.

The multi-directional search of Torczon (1989) is an alternative simplex procedure that is based on the best setting and resizing the entire simplex (see panel III of figure A1.1). The only setting that is not replaced is the best setting. The other settings are reflected in the best setting. Contrary to the simplex methods of Spendley *et al.* (1962) and Nelder and Mead (1965) this alternative approach is robust. The multi-directional search of Torczon is the only simplex-based strategy that has a convergence theory.

[79] This procedure should not be confused with the simplex method for linear programming (e.g. see Press *et al.*, 1989). A linear program is solved by processing from one extreme to another. The only resemblance between the sequential Simplex of Spendley *et al.* (1962) and the simplex method for linear programming is the fundamental geometrical concept.

Appendix 2

MULTI-DIRECTIONAL ADAPTIVE PATTERN SEARCH

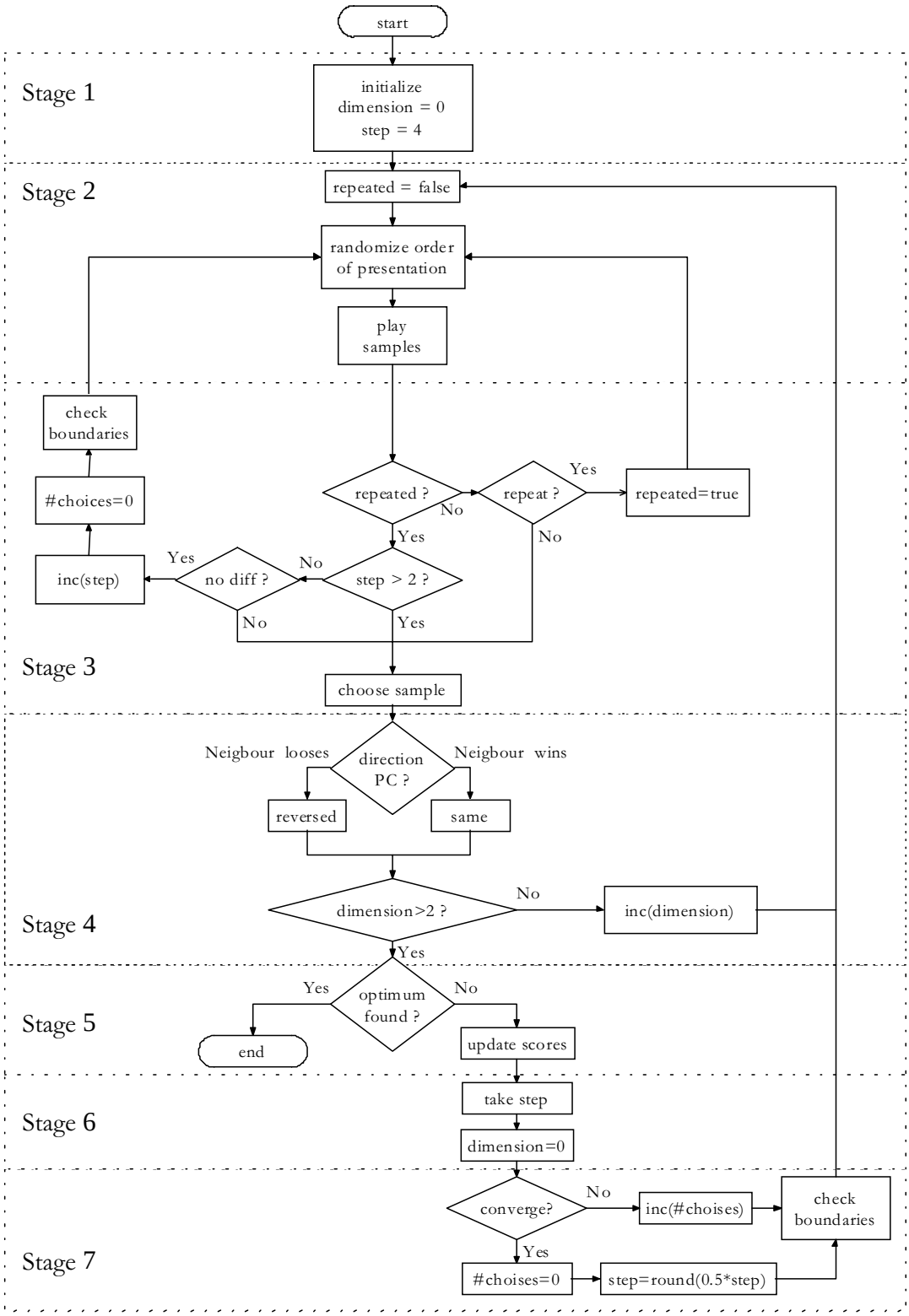


Figure A2.1 Flow diagram of the multi-directional pattern search with adaptive step size.

The multi-directional pattern search shown in figure A2.1 has basically seven stages. A more detailed description of these stages is given below.

1 Set-up stage
Initialisation of test parameters and the step size values.

The order of sentences and the order of dimensions that are applied successively are randomised and the initial step size value is chosen. In the studies that use ten settings for each of three dimensions (used in the flow diagram and in section 4.6 and chapter 5) the initial step size is four.[80]

2 Presentation stage
Signal presentation.

The 'order of presentation' of the samples is randomised. The fragments represent the algorithm settings. Two samples are presented ('play') that need to be compared pair wise by the listener.

3 Response stage
Response possibilities.

The listener can choose between sample 1 and sample2, or use the repeat button. When the repeat button is used the repeat button cannot be used again. The use of the repeat button can have two purposes. First, the repeat button can be used when the listener is not sure what decision to make. When the listener can make a choice after repeating the fragments the listener chooses one of the samples. Second, the repeat button can be used when the listener cannot distinguish the fragments. In that case, after repeating the fragments, a 'no difference' alterative (nodiff) can be used. When the step size is smaller than or equal to two, the step size will be increased by one in 'inc(step)'. The 'order of presentation' is randomised again to avoid bias.

By increasing the step size it is possible that next paired comparisons will take place with settings that lie across the boundaries of the settings cube. Details about the boundary check can be found in convergence stage 7. Also the initialisation of parameter '# choices' will be explained in convergence stage 7.

4 Search direction stage
Determination of the search direction for the dimension at issue.

---

[80] For a three-dimensional optimisation space with five settings for each dimension (section 4.5) the initial step size was two. In this case, variation of the step size was similar although much more constrained (only values 1 and 2 are allowed) than for the optimisation space of 10*10*10 settings illustrated here. In sections 4.2 and 4.3 a fixed step size of one was used. The rules to determine whether the optimum was found, described in decision stage 5, was also adapted slightly. That is, the optimum was found when the number of paired comparisons was larger than five. The other stop criterion, i.e. when a setting was frequented three times, remained the same. When the step size was fixed at one during the optimisation process, the flow diagram is similar as well, except for response stage 3 and convergence stage 7. In response stage 3, the 'No difference' button is no longer optional, so subjects can only choose between fragment1, fragment2, and repeat. Convergence stage 7 is not applicable for our pattern search with a fixed step size.

The search direction is 'reversed' when the setting at issue performs better than the neighbour in that particular dimension. The search direction remains the 'same' when the neighbouring setting performs better.

When parameter 'dimension' is not larger than two a new paired comparison should be performed in a different dimension. In that case the value of 'dimension' is increased by one in 'inc(dimension)'.

5 Decision stage

Decision rules used to end the pattern search.

The pattern search terminates when the optimum is assumed to be found. In that case the procedure ends. Three termination rules are used. First, when the same setting (algorithm combination) is frequented three times. Second, when a setting wins a paired comparison more than six times (or five, see footnote). Third, when the number of paired comparisons in the procedure was larger 54.

When the optimum is not yet found the scores are updated. We register the number of times a setting is frequented. Additionally, the number of times a setting has won a paired comparison is recorded. Due to the fact that the step size is adaptive it is possible that some settings are literally overstepped and therefore discarded. These settings are also credited by fractions of whole points (gained by one of the settings that are compared pair wise).

6 Movement stage

Take step towards a new, more optimal setting.

The number of dimensions that need to be compared is three in order to determine what step should be taken next. Consequently, the movement can also be diagonal as described in appendix 1. Parameter 'dimension' is initialised.

7 Convergence stage

Decision rules used to converge & next round of paired comparisons.

The step size is halved in two cases. First, the step size is halved when three rounds of three paired comparisons are performed with the same step size. In that case parameter '#choices' is three. Parameter '#choices' is set at zero either when the step size is halved, or when the step size is increased (see response stage 3). Second, the step size is halved when the step size is equal to four. A step size of four is only used at the start of the pattern search, for 'customisation' purposes. When parameter '#choices' is smaller than three and the step size (parameter 'step') is smaller than four the step size remains the same and parameter '#choices' is increased by one in 'inc(#choices)'.

In the next round of three paired comparisons it is possible that the boundaries of the settings space will be crossed. This is checked in 'check boundaries'. The boundary rules prescribe that when one of the two settings would be across one of the boundaries (of the cube) one of the settings is placed at the boundary itself, the direction of comparison is reversed, and the other setting is placed at a distance away of the boundary setting that is equal to the current step size.

Appendix 3

SHAPE OF THE RESPONSE DISTRIBUTION

We classified the distributions patterns, gained from the round-robin data, after Kuk and Lau (1996a) into four categories: flat, broad, narrow, or multi-modal distribution. In order to make a clear distinction between the four categories we, a priori, base the distinction between categories on the presence and distribution of settings with 3 and 4 wins only. This is for reasons of reliability. That is to say, settings that won 1 or 2 times out of 10 might have resulted from random choice only. We conceive 4 wins or two times value 3 next to each other as a clear optimum. In total we can distinguish nine different types of distributions, based on possible combinations of wins. The distribution patterns are constrained by the fact that a setting can maximally win 4 times, and there can only be one setting with four wins. Classification of the distributions patterns depends on the permutation of the numbers of these sets and is rather complex. Broadly outlined, the set that consists of values 2 only is defined as the only flat distribution. A distribution is defined as multi-modal when either value 3 and 4 or value 3 and 3 are separated by at least one 0 or one 1. A distribution is defined as broad when there are several adjacent values 2 and 3 in a row. Finally, distributions are defined as narrow when values 3 and 4 or 3 and 3 are not separated. Instead of this quite complex way of defining distribution types, we use a well-defined a posteriori measure to make a distinction between several kinds of distribution. A measure that represents the peakyness of a distribution is the kurtosis that is calculated by the following expression:

$$k = \frac{1}{\sigma^4} \cdot (\overline{x^4} - 4\bar{x}\overline{x^3} + 6\overline{x^2}\bar{x}^2 - 3\bar{x}^4) - 3 \qquad \text{(A1)}$$

This quite untransparent measure of the fourth central moment of x is dimensionless thanks to the fourth power of σ and is used with care here, because sometimes peculiarities appear. Basically, positive values represent a distribution with a relatively higher peak and broad tails. Negative values indicate a broader peak with small tails. Parallel to the calculation of the expected value (see appendix 1) we also make use of the third power of the frequency distributions for calculation of the kurtosis. On the one hand the values of the kurtosis lead to a clearer categorization. On the other hand for some distribution patterns awkward values appear for the kurtosis, which might relate to the use of the third power function for the expected value. In our data the kurtosis varies between –1.9 and 9.0. It appears that positive values indeed represent narrow peaked distributions when these distributions are considered by rules that are defined a priori. However, it would be too simple to conclude that negative values represent broad distributions. For k<-1.0 the distribution is in general multi-modal. When –1.0<k<-0.35 the distribution is in general broad. For –0.35<k<0.1 we encounter miscellaneous distributions, ranging between multi-modal, broad, and narrow. The kurtosis does not always predict the distribution category that is achieved by the a priori definitions. Therefore, there are some exceptions to the rules that relate to the kurtosis, which are accounted for in correction rules. One of these exceptions is the flat distribution that has a k-value of -1.3. Contrary to intuition, this distribution is not a boundary between the

multi-modal and broad distributions. Merely, there is quite a broad region between multi-modal and broad on one side and between broad and narrow on the other. Because flat distributions hardly occur, we conceived flat distributions as being broad.

Appendix 4

ESTIMATION OF WINNER OF RESPONSE DISTRIBUTION

The estimated optimal setting will be designated by the expected maximum value of the distribution. When this expected value E[x] is calculated from the linear average, some peculiarities appear. Depending on the distribution, the value deviates from values suggested by intuition. A clear example is given in figure A4.1, where the calculated expected value is larger than 1, which contradicts the idea that this value should be at least smaller than 1. The problem originates from the maximum amount of winners: each point can at most win four times. Intuitively, for a setting that wins four times it seems quite reasonable to assume that this point is the real winner. In the same way, we would classify a rate of three winnings also as a potential winner. However, when a configuration wins two times or even only once, we can assume that this configuration is not at all a potential winner.

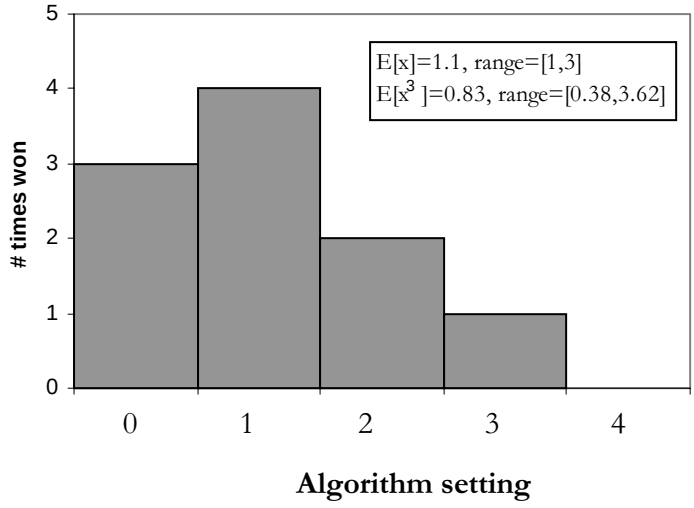


Figure A4.1 Example of a distribution of winners that result from the 10 paired comparisons. The 'real' winner is estimated by calculation of the expected value E[x] for two functions: x and x^3. The range of possible 'real' winners is also indicated.

Contrary to this reasoning, the weight of the configurations that are not likely to be 'real winners' can cause the expected value to be a poor estimate. Therefore, when we compare the calculated expected values with estimates made by common sense, deviations can occur: All possible distributions yield expected values that do not cover the actual range of winners. In order to cope with these problems we studied possibilities to extend the range of calculated winners and at the same time make the deviation of the estimated values that are obtained by common sense as small as possible.

One way is to give higher weightings to configurations that won more often. A way to increase the differences between the numbers of winning comparisons is to use a power function. In order to meet both the increase in range and a better estimate for the winner a power of three meets these requirements best. In that case the range of possible

winner values varies from 0.38 to 3.62, which is much better than that of the commonly used expected value with a power of 1 (range = [1,3]). Moreover, for a power of 3 it appears that deviations between the estimated values and calculated values (winners) are lowest. The estimated winners are based on an a priori prediction of 72 distributions, whereas the expected value E[x] is an a posteriori measure, because this value has been calculated, based on third power values.

Appendix 5

SIMULATION OF RANDOM PERFORMANCE

From simulations the random distribution appeared approximately the same for all step size configurations described in section 4.6. This was shown as follows. The random distribution that was given as a reference in figure 4.23 was determined by randomly assigning 1000 optima in test and retest and by deriving the cumulative distribution of three-dimensional distances between the test and retest optima. This result can also be obtained by randomly choosing between the two alternatives during the pattern search itself in 1000 pattern searches. This kind of simulation was performed for all four step-size configurations (sim123, sim23, sim3, sim4). The results are presented in figure A5.1. The results show that the random distributions are comparable. Sim0 is the random distribution that was used in figure 4.23, i.e. the distribution acquired by randomly assigning 1000 test and retest optima.

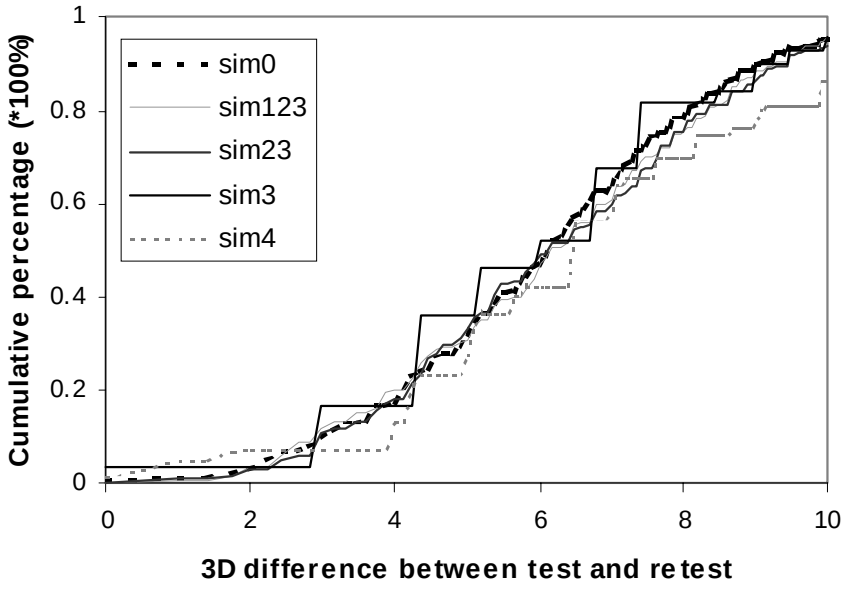


Figure A5.1 Cumulative random distributions acquired for the multi-directional pattern search. Parameter is the kind of simulation used to simulate the random behaviour. The random behaviour was simulated by randomly choosing between the two alternatives during the pattern search for test and retest for different step size configurations (sim123, sim23, sim3, sim4), or by randomly assigning the optima for test and retest (sim0).

Simulations sim0, sim123, and sim23 are very similar. The step-size-like distributions found for fixed step sizes of 3 and 4 (sim3 and sim4) can be explained by the fact that not all three-dimensional distances can be obtained. That is, some settings simply cannot be reached or are difficult to reach when a fixed step size is used.

The number of paired comparisons that were needed to end the pattern search when the winner of a paired comparison was randomly selected was also recorded. The results are presented in figure A5.2. For each number of paired comparisons the frequency is

given that was obtained when 1000 pattern searches were performed. The frequency of paired comparisons that were needed follows a beta-distribution showing that the variable is really random. The number of paired comparisons needed to end the pattern search varied between 9 and 435. The expected value E[x] of the frequency distribution was 89.4 paired comparisons.

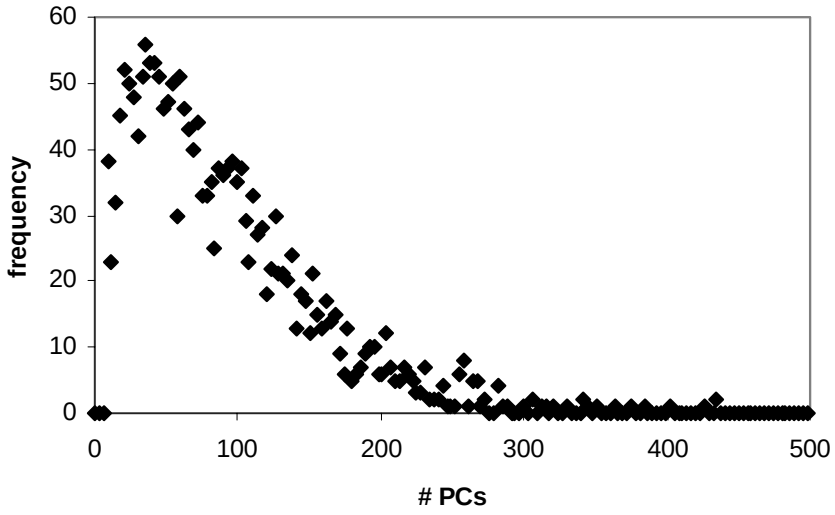


Figure A5.2 Number of paired comparisons that were needed to end the pattern search when the winner of a paired comparison was randomly selected was also recorded. For each number of paired comparisons the frequency is given that was obtained when 1000 pattern searches were performed.

Appendix 6

RELATION BETWEEN PATTERN SEARCH AND ROUND-ROBIN RESULTS

In chapter 4 a relation was found for two studies (4.5 and 4.6) between the coefficient of agreement 'u' and the coefficient of consistency 'ζ' for the round-robin results. It appeared that these two coefficient were linearly related following the equation ζ=0.5u+0.5. The results for the study described in section 4.6 showed that 'ζ' for the individual algorithms follow the same pattern as the reliability, given by the information transfer (IT), of these algorithms for the pattern search results. The question is whether the reliability of the round-robin tournament, given by the coefficient of agreement, agrees with the reliability of the pattern search.

The set up for the studies in sections 4.2, 4.5, and 4.6 differed in the settings that were included in the round-robin tournament. For the study of section 4.2 all edges of the algorithm-settings cube were included. In the study of sections 4.5 three winners of different pattern searches and two extreme settings were included. In the study of section 4.6 only the most diverse pattern search winners were included. The differences in set up could imply that the agreement between pattern search and round-robin results varied across the studies.

The results for the studies of section 4.5 and chapter 5 are presented in table A6.1. The information transfer results (IT) relate to the results presented in figures 4.18 and 5.2, respectively. The information transfer IT and the coefficient of consistency ζ have a one-to-one relationship and differ systematically around a value of 0.2.[81] The relation between IT and 'u' is also linear, but the relation is weaker. Hence it seems that the coefficient of consistency is a better indicator to assess the reliability of the pattern search. Coefficient 'u' indicates the reliability of the round-robin procedure itself.

			Pattern search	**Round-robin**	
Section	***Group***	***Noise***	**IT (%)**	**u**	**ζ**
4.5	***HI***	***Cont***	0.10	-0.28	0.34
	HI	***Fluct***	-0.03	-0.48	0.18
5	***NH***	***Cont***	0.48	0.47	0.74
	HI	***Cont***	0.28	0.37	0.54

Table A6.1 Information transfer (IT) of the pattern search results, and coefficient of agreement (u) and coefficient of consistency (ζ) of the round-robin tournament results. The results are differentiated for group and noise type. The group contained either normal hearing (NH) or hearing-impaired subjects (HI). The noise type was either continuous noise (cont) or fluctuating noise (fluct). The results refer either to the study of section 4.5 or refer to chapter 5.

[81] The systematic difference found between the values of ζ and the information transfer could be explained by two counteracting factors. On the one hand, the 3D distance between the settings selected in the round-robin tournament was large; on average a value of four. This is much larger than the step size used in the pattern search. On the other hand, the settings that were included comprise principally (sub)optima found by different pattern searches. This implies that the preferences would not differ much, although those optima were selected that were not found consistently.

The approach of the study of section 4.2 is much more systematic. For each edge only one algorithm varied. Therefore, the pattern search and round-robin tournament results are much more comparable, i.e. because the distances between settings that were compared were similar. The results of this study are averaged across normal hearing and hearing-impaired subjects, across different noise types and across parallel edges. The results are given in table A6.2, and differentiated for evaluation criterion and algorithm. The difference between ζ and IT is not consistently the same. It appears that the difference tends to be smaller for higher ζ values.

When the results of ζ and IT of all three studies are compared, a linear relation IT=1.28ζ-0.32 is found with high correlation (R^2=0.84). As a consequence IT can have negative values. This is in agreement with the results of some studies (e.g. see section 4.5). However, the IT and ζ values of tables A6.1 and A6.2 can be described much better by an integrated Gaussian-shaped curve. This is especially true when we assume that for ζ=1 also IT=1 and when we assume that for ζ=0 also IT=0.

Section 4.2, NH&HI		**Pattern search**	**Round-robin**
Evaluation criterion	**Algorithm**	**IT (%)**	**ζ**
Listening comfort	***BNR***	0.43	0.56
	PhC	0.43	0.58
	SE	0.85	0.79
Speech intelligibility	***BNR***	0.41	0.58
	PhC	0.27	0.57
	SE	0.61	0.66

Table A6.2 Information transfer (IT) of the pattern search results and coefficient of consistency (ζ) of the round-robin tournament results, of the study described in section 4.2. The results are averaged across normal hearing subjects (NH) and hearing-impaired subjects (HI), and across different noise types. For the round-robin tournament the results were also averaged across parallel edges. The results are differentiated for evaluation criterion and algorithm. The algorithms comprise binaural noise reduction (BNR), phonemic compression (PhC) and spectral enhancement (SE_B).

The integrated Gaussian-shaped relation between ζ and IT cannot describe negative IT values found in some cases. Additionally, the results of section 4.6 for the normal hearing group are not in line with the linear and the integrated Gaussian-shaped relation between ζ and IT. This can be explained only when we assume that the coefficient of consistency ζ describes only part of the reliability found for the pattern search. A second type of inconsistency that was described by David (1963) is the coefficient of separation 'S'. This type of inconsistency quantifies the presence of multiple peaks. Multiple peaks are present when for integers i, j, and k with i<j<k the round-robin response is R_i>R_j and R_k>R_j. In section 4.2 we described the shape of the response distribution differently by another measure of peakyness, i.e. the kurtosis (see appendix 3). It would be interesting to compare both analyses.[82] The results of section 4.2 clearly indicated that

[82] In our round-robin setup with five settings the coefficient of separation varies between 0 (no separation, e.g. distribution 0-1-2-3-4) and 6.25 (strong separation, e.g. distribution 4-1-1-1-3). These values were

multi-modal distributions could be relatively frequent (see figures 4.10 and 4.11). The cluster analyses performed in section 4.6 and chapter 5 clearly indicate the presence of multiple optima (see e.g. table 5.2, HI group, algorithm SE_I). This could explain the relative low IT value compared to the high value for ζ for the normal hearing group of section 4.6.

The combination of both coefficients of consistency might be able to describe the reliability of the pattern search, i.e. the information transfer. However, we did not calculate the average S value for the round-robin results of section 4.2. Hence we cannot specify to what extent this coefficient contributes to a better assessment of the information transfer of the pattern search.

---

calculated, following the equation $S=\Sigma C_i(t\text{-}i\text{-}R_i)$, with C_i is the column total, and R_i is the row total above the diagonal of the response matrix.

BIBLIOGRAPHY

Åberg, E.R., Gustavsson, A.G.T. (1982). "Design and evaluation of modified simplex methods", Analytica Chimica Acta 144: 39-53.

Arlinger, S. (1999). "Non-linear hearing instruments – clinical applications and experience", in *Auditory models and non-linear hearing instruments,* edited by A.N. Rasmussen, P.A. Osterhammel, T. Andersen, T. Poulsen, Scanticon: Kolding, Denmark, 309-323.

Baer, T., Moore, B.C.J. (1993). "Effects of spectral smearing on the intelligibility of sentences in noise", J. Acoust. Soc. Am. 94 (3): 1229-1241.

Baer, T., Moore, B.C.J. (1997). "Evaluation of a Scheme to Compensate for Reduced Frequency Selectivity in Hearing-Impaired Subjects", from: *Modeling sensorineural hearing loss*, edited by W. Jensteadt, Lawrence Erlbaum, Mahwah, New Yersey

Baer, T., Moore, B.C.J., and Gatehouse, S. (1993). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on the intelligibility, quality, and response times", J. Rehabil. Res. Dev. 30: 49-72.

Beek, J.H.M. van, Buitelaar, D., Festen, J.M., Houtgast, T. (2000). "Implementation of Ephraim-Malah noise-reduction algorithm in the DASi-3, Progress report V, SPACE project DE 3012.

Betteridge, D., Wade, A.P., Howard, A.G. (1985). "Reflections on the modified simplex – II", Talanta 32 (8B): 723-734.

Boothroyd, A., Mulhearn, B., Gong, J., Ostroff, J. (1997). "Simulation of Sensorineural Hearing Loss: Reducing Frequency Resolution by Uniform Spectral Smearing", from: *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum, Mahwah, New Yersey.

Bosman, A.J. (1989). "Speech Perception by the hearing impaired", Ph.D. thesis, University of Utrecht, The Netherlands.

Box, G. E. P. (1957). "Evolutionary operation: A method for increasing industrial productivity", Applied Statistics 6: 81-101.

Boymans, M. (2003) "Intelligent processing to optimize the benefit of hearing aids", Ph.D. thesis, University of Amsterdam, The Netherlands.

Bregman, A.S. (1994). "Auditory scene analysis – The perceptual organization of sound", MIT Press, Cambridge, Massachusetts.

Bril, M. (2000). "Hollandse luchten, Amsterdamse miniaturen", Prometheus Amsterdam.

Bronkhorst, A. (2000). "The Cocktail Party Phenomenon: review of Research on Speech Intelligibility in Multi-Talker Conditions", Acustica 86: 117-128.

Buus, S. (1997). Introduction to section "Spectral and Temporal processing in sensorineural hearing loss", part V of *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum, Mahwah, New Yersey.

Byrne, D. (1986). "Effects of frequency response characteristics on speech discrimination and perceived intelligibility and pleasantness of speech for hearing-impaired listeners", J. Acoust. Soc. Am. 80 (2): 494-504.

Byrne, D., and Dillon, H. (1986). "The National Acoustic Laboratories" (NAL) new procedure for selecting the gain and frequency response of a hearing aid", Ear Hear. 7 (4): 257-265.

Cohen, A. (1989). "Perception and Language", from: *Working Models of Human Perception*, edited by B.A.G. Elsendoorn & H. Bouma, Academic Press Limited, London.

Cox, R.M., Alexander, G.C. (1992). "Maturation of hearing aid benefit: objective and subjective measurements", Ear Hear. 13 (3): 131-41.

David, H.A. (1963). "The method of paired comparisons", Charles Griffin & Company Limited, London.

Davis, P.J., and Hersch, R. (1981). "The mathematical experience", Houghton Mifflin Company, Boston.

Dennett, D.C. (1995). "Is Perception the "Leading Edge" of Memory?", in Iride: Luoghi della memoria e dell"oblio, edited by A. Spadafora: 59-78.

Dijk, J.S.C. van. (2001). "Mechanical aspects of hearing", Ph.D. thesis, University of Amsterdam, The Netherlands.

Dillon, H. (1997). "Converting insertion gain to and from headphone coupler responses", Ear Hear. 18 (4): 346-348.

Dirks, D.D., Ahlstrom. J., Noffsinger, P.D. (1993). "Preferred frequency response for two- and three-channel amplification systems", J. Rehabil. Res. Dev. 30 (3): 305-317.

Dretske, F. (2000). "Perception, knowledge and belief – selected essays", Cambridge university press, Cambridge.

Dubbelboer, F., Buitelaar, D., Festen, J.M., Houtgast, T. (1999). "Implementation of Ephraim-Malah noise reduction algorithm", Progress report IV, SPACE project DE 3012.

Durant, E.A., Tassell, D. van, Rickert, M. (2002) "Hearing aid fitting with a genetic algorithm", IHCON conference.

Eisenberg, L., Levitt H. (1991). "Paired comparison judgments for hearing aid selection in children", Ear Hear. 12: 417-430.

Eisenberg, L.S., Dirks, D.D., Gornbein, J.A. (1997). "Subjective judgments of speech clarity measured by paired comparisons and category rating", Ear Hear. 18 (4): 294-306.

Ephraim, Y, and Malah, D. (1984). "Speech enhancement using a mnimum mean-square error short-time spectral amplitude estimator", IEEE Trans. on Acoust., Speech, Signal Processing 32 (6): 1109-1121.

Espmark, A.K., Scherman, M.H. (2003). "Hearing confirms existence and identity – experiences from persons with presbyacusis, Int. J. Audiol. 42: 106-115.

Evans, E.F. (1992). "Auditory Processing of Complex Sounds: An Overview", from: *Processing of Complex Sounds by the Auditory System*, Philosophical Transactions: Biological Sciences 336 (1278): 295-306.

Fabry, D.A., Schum, D.J. (1994). "The role of subjective measurement techniques in hearing aid fittings", from: *Strategies for Selecting and Verifying Hearing Aid Fittings*, by M. Valente, Theme Medical Publishers, New York: 136-155.

Fabry, D.A., Tassell, D.J. van. (1986). "Masked and filtered simulation of hearing loss: effects on consonant recognition", J. Speech Hear. Res. 29 (2):170-178.

Fearn, N. (2002). "De wereld zien is haar veranderen", Filosofie magazine 1 (11): 42-45.

Festen, J.M., and Plomp, R. (1990). "Effect of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing", J. Acoust. Soc. Am. 88 (4): 1725-1736.

Feynman, R.P. (1998). "Six easy pieces – The fundamentals of physics explained", Penguin Books Ltd, London.

Firth, R. (1949-50). "Sense Data and the Percept Theory", Part I: The phenomenological basis of the percept theory.

Forge, A., Wright, T. (2002). "The molecular architecture of the inner ear", British Medical Bulletin 63: 5-24.

Fourcin, A.J. (1989). "Links between voice pattern perception and production", from: *Working Models of Human Perception*, edited by B.A.G. Elsendoorn & H. Bouma, Academic Press Limited, London.

Franck, B.A.M. (2001). "Optimisation of listening comfort", from: *Verbetering vsn spraak-verstaan in lawaai*, Journaal van het Nederlands Akoestisch Genootschap 157: 23-34.

Franck, B.A.M., Boymans M., Dreschler, W.A. "Interactive fitting of three state-of-the-art hearing-aid algorithms", conditionally accepted in Int. J. Audiol.

Franck, B.A.M., Kreveld-Bos, C.S.G.M. van, Dreschler, W.A., Verschuure, J. (1999). "Evaluation of spectral enhancement in hearing aids, combined with phonemic compression", J. Acoust. Soc. Am. 106 (3): 1452-1464.

Franck, B.A.M., Lyzenga, J., Dreschler, W.A. "Methodological aspects of an adaptive multi-directional pattern search to optimize speech perception using three hearing-aid algorithms ", conditionally accepted in J. Acoust. Soc. Am.

Gatehouse, S. (1992). "The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural hearing aids", J. Acoust. Soc. Am. 92 (3): 1258-1268.

Gennaro, S.T. de, and Braida, L.D. (1997). "Effects of linear amplification and amplitude compression on sentence reception by listeners with hearing loss simulated by masking noise", from: *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum, Mahwah, New Yersey.

Gibson, J.J. (1968). "The senses considered as perceptual systems", Allen & Unwin, London.

Glasberg, B.R., and Moore, B.C.J. (1990). "Derivation of auditory filter shapes from notched-noise data", Hear Res. 47: 103-138.

Goedegebure, A., Hulshof, M., Maas, A.J.J., Dreschler, W.A., Verschuure, J. (2001). "Effects of single-channel phonemic compression schemes on the understanding of speech by hearing-impaired listeners", Audiology (40): 10-25.

Gulick, W.L., Gescheider, G.A., Frisina, R.D. (1989). "Hearing – physiological acoustics, neural coding, and psychoacoustics", Oxford University Press, New York.

Harnad, S. (1987). "Categorical Perception – The groundwork of cognition", Cambridge University Press, Cambridge.

Harten-De Bruijn, H.E. van, Kreveld-Bos, C.S.G.M. van, Dreschler, W.A., and Verschuure, J. (1997). "Design of two syllabic non-linear multichannel signal processors and the results of speech tests in noise", Ear Hear. 18: 26-33.

Held, R., and Richards, W. (1972). "Introduction", from *Perception: Mechanisms and Models, A Scientific American Reader*, W.H. Freeman, San Francisco, 1972.

Hellbrück, J., and Moser, L. M. (1985). "Hörgeräte Audiometrie: Ein computer-unterstütztes psychologisches Verfahren zur Hörgeräteanpassung", Psychologische Beiträge, 27: 494-509.

Helmholtz, H.L.F., von. (1954/1887). "On the sensations of tone as a physiological basis for the theory of music", Dover Publications, New York.

Hirsh, I.J., Watson, C.S. (1996). "Auditory psychophysics and perception", Annu. Rev. Psychol. 47: 461-484.

Hofstadter, D.R. (1986). "Gödel, Escher, Bach: een eeuwige gouden band", 4e druk, Uitgeverij Contact, Amsterdam.

Hoftadter, D.R., Dennett, D.C. (1989). "De spiegel van de ziel: fantasieën en reflecties over ego en geest", 2e druk, Uitgeverij Contact, Amsterdam.

Hohmann, V. (1989). "Störungen der Lautstärkewahrnehmung bei Innenohrschwer-hörigkeit und ihre Kompensation durch digitale Hörgeräte", Ph.D. thesis, University of Göttingen, Germany.

Hong, H., Vasaru, D. (1996). "Survey on nonlinear optimization", report 96-04, Research Institute for Symbolic Computation (RISC), Johannes Kepler University, Linz.

Horst, A.P.J.M. van den, Brouwer, J., Rensen, J., en Twilt, K. (1990). "Slecht horen: beschouwingen en ervaringen van slechthorenden", Boom, Meppel.

Horwitz, A.R., Turner, C.W. (1997). "The time course of hearing aid benefit", Ear Hear. 18 (1): 1-11.

Hume, D. (1993/1748). "An enquiry concerning human understanding", 2nd edition, edited by Eric Steinberg, Hackett Publishing Company, Indianapolis.

Humphrey, D.G., Wilson, J.R. (2000). "A revised simplex search procedure for stochastic simulation response surface optimization", J. Computing 12 (4): 272-283.

Hyslop, J.H. (1886). "Psychophysics", Science – supplement vol. III, no. 89: 259-262.

Jesteadt, W. (1997). "Modeling sensorineural hearing loss", Lawrence Erlbaum, Mahwah, New Jersey.

Kant, I. (1974/1798&1800). "Anthropology from a pragmatic point of view", translated, with an Introduction and Notes by M.J. Gregor, Martinus Nijhoff, The Hague.

Kant, I. (1981/1781&1787). "Kritik der reinen Vernunft", Wilhelm Weischedel, Suhrkamp, Frankfurt am Main.

Kant, I. (1983/1798&1800). "Anthropologie in pragmatischer Hinsicht", uit: Kant Werke-Band 10, Schriften zur anthropologie,Geschichtsphilosophie, Politik und Pädagogik, zweiter Teil, Wissenschaftliche Buchgesellschaft, Darmstadt.

Keidser, G. (1995). "The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids", Ear Hear. 16 (6): 575-586.

Keidser, G. (1996). "Selecting different amplification for different listening conditions", J. Am. Acad. Audiol. 7(2): 92-104.

Kellogg, R.T. (1997). "Cognitive psychology", Sage publications, London.

Kemp D.T. (2002). "Otoacoustic emissions, the origin in cochlear function, and use", British Medical Bulletin 63: 223-241.

Keurs, M. ter, Festen, J.M., Plomp, R. (1992b). "Effects of spectral envelope smearing on speech reception II", J. Acoust. Soc. Am. 93 (3): 1547-1552.

Klein, W., Plomp, R., and Pols, L.C.W. (1970). "Vowel spectra, vowel spaces and vowel identification", J. Acoust. Soc. Am. 48: 999-1009.

Kollmeier, B., Marzinzik, M., Tchorz, J., Wittkop, T. (2000). "Development and evaluation of advanced noise reduction algorithms", SPACE report DE 3012, RP-SP1c.

Kollmeier, B., Peissig, J., Hohmann, V. (1993). "Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain", Scand. Audiol. 38: 28-38.

Koningsveld, H. (1987). "Het verschijnsel wetenschap – een inleiding tot de wetenschaps-filosofie", Boom Meppel, Amsterdam.

Koopman, J., Franck, B.A.M., Pool, R., Dreschler, W.A. (2000). "Validation of communication devices with earmuffs in field tests at the working place with normal- and hearing-impaired listeners", Report SPACE project DE 3012.

Kuhl, P. "Speech perception", video produced by J. Acoust. Soc. Am.

Kuhl, P.K., and Meltzoff, A.N. (1982). "The bimodal perception of speech in infancy", Science 218: 1138-1141.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B. (1992). "Linguistic experience alters phonetic perception in infants by 6 months of age", Science 255: 606-608.

Kuk, F.K. (1994b). "A screening procedure for modified Simplex in frequency-gain response selection", Ear Hear. 15 (1): 62-70.

Kuk, F.K., and Lau, C. (1995). "Effect of initial setting on convergence to optimal hearing aid setting using a Simplex method", Br. J. Audiol. 29: 263-269.

Kuk, F.K., and Lau, C. (1996a). "Comparison of preferred frequency gain settings obtained with category rating and modified Simplex procedure", J. Am. Acad. Audiol. 7: 322-331.

Kuk, F.K., and Lau, C. (1996b). "Effect of hearing aid experience on preferred insertion gain selection", J. Am. Acad. Audiol. 7: 274-281.

Kuk, F.K., and Pape, N.M.C. (1992). "The reliability of a modified Simplex procedure in hearing aid frequency-response selection", J. Speech Hear. Res. 35: 418-429.

Kuk, F.K., and Pape, N.M.C. (1993). "Relative satisfaction for frequency responses selected with a Simplex procedure in different listening conditions", J. Speech Hear. Res., 36: 168-177.

Kuk, F.K.. (1994a). "Use of Paired Comparisons in Hearing Aid Fitting", from: *Strategies for Selecting and Verifying Hearing Aid Fittings*, edited by M. Valente, Theme Medical Publishers, New York: 108-135.

Leeuw, A.R., and Dreschler, W.A. (1994). "Frequency-resolution measurements with notched noises for clinical purposes", Ear Hear. 15: 240-255.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. 49 (2): 467-477.

Levitt, H. (1978). "Adaptive testing in Audiology", from: *Sensorineural hearing impairment and hearing aids*, edited by C. Ludvigsen and J. Barfod, Leif Sølborg, Copenhagen: 241-291.

Levitt, H., Neuman, A., Mills, R., Schwander, T. (1986). "A digital master hearing aid", J. Rehab. Res. 23 (1): 79-87.

Lewis, R.M., Torczon, V., and Trosset, M.W. (2000). "Direct search methods: then and now", J. Comp Appl Math 124: 191-207.

Lewis, R.M., Torczon, V., Trosset, M.W. (1998). "Why pattern search works", Optima (Mathematical Porgramming Society Newsletter): 1-7.

Liberman, A.M., Mattingly, I.G. (1989). "A specialization for speech perception", Science 243: 489-494.

Lum, D.S., Braida, L.D. (1997). "DSP implementation of a real-time hearing loss simulator based on dynamic expansion", from: *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum, Mahwah, New Yersey.

Luria, A.R. (1976). "Cognitive development – its cultural and social foundations", Harvard University Press, Cambridge, Massachusetts.

Lyzenga, J., Festen, J.M., and Houtgast, T. (2002). "Speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity", J. Acoust. Soc. Am. 112 (3): 1145-1157.

Magee, B. (1997). "Bekentenissen van een filosoof", Uitgeverij Anthos, Amsterdam.

Magee, B., & Milligan, M. (2001). "Beelden van een blinde filosoof", Athos uitgevers, Amsterdam.

Marzinzik, M. (2000). "Noise reduction schemes for digital hearing aids and their use for the hearing impaired", Ph.D. thesis, University of Oldenburg, Germany.

Matlin, M.W., Foley, H.J. (1997). "Sensation and perception", Allyn and Bacon, fourth edition, Boston.

Merleau-Ponty, M. (1978). "Phenomenology of perception", London and Henley, Routledge & Kegan Paul, The Humanities Press, New Yersey.

Mirzoeff, N. (1995). "Silent Poetry: deafness, sign, and visual culture in modern France", Princeton University Press, Princeton.

Moore, B.C.J. (1989). "An introduction to the psychology of hearing", 3rd edition, Academic Press, London.

Moore, B.C.J. (2002). "Psychoacoustics of normal and impaired hearing", British Medical Bulletin 36: 121-134.

Moore, B.C.J., Glasberg, B.R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", J. Acoust. Soc. Am. 74: 750-753.

Moore, D.R. (2002). "Auditory development and the role of experience", British Medical Bulletin 36: 171-181.

Nagel, T. (1997). "What is it to be a bat?", from *Mortal questions*, Cambridge University Press.

Needleman, A.R., Crandell, C.C. (1997). "Speech Perception in Noise by Listeners With Hearing Impairment and Simulated Sensorineural Hearing Loss", from: *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum, Mahwah, New Yersey.

Nelder, J.A., Mead, R. (1965). "A simplex method for function minimization", Comput. J. 7 (4): 308-311.

Neuman, A,C., Bakke, M.H., Hellman, S., Levitt, H. (1994). "Effect of compression ratio in a slow-acting compression hearing aid: Paired-comparison judgments of quality", J. Acoust. Soc. Am. 96 (3): 1471-1478.

Neuman, A.C., Bakke, M.H., Mackersie, C., Hellman S., Levitt, H. (1998). "The effect of compression ratio and release time on the categorical rating of sound quality", J. Acoust. Soc. Am. 103 (5): 2273-2281.

Neuman, A.C., Bakke, M.H., Mackersie, C., Hellman, S., Levitt, H. (1995). "Effect of release time in compression hearing aids: Paired-comparison judgments of quality", J. Acoust. Soc. Am. 98 (6): 3182-3187.

Neuman, A.C., Levitt H., Mills R., and Schwander T. (1987). "An evaluation of three adaptive hearing aid selection strategies", J. Acoust. Soc. Am. 82 (6): 1967-1976.

Norwich, K.H. (1974). "Information, Sensation, and Perception", Academic Press, San Diego, California.

Öberg, T. (1998). "Importance of the first design matrix in experimental simplex optimization", Chem. and Intell. Lab. Sys. 44: 147-151.

Öberg, T.G., Deming, S.N. (2000). "Find optimum operating conditions fast", Chem. Eng. Progress, 96 (4): 53-59.

Otto, M. (1988). "Fuzzy theory explained", Chem. Intell. Lab. Syst. 4: 101-120.

Oxenham, A.J., Bacon, S.P. (2003). "Cochlear Compression: Perceptual measures and implications for normal and impaired hearing", Ear Hear. 24 (5): 352-366.

Parizet, E. (2002). "Paired comparison listening tests and circular error rates", Acta Acustica 88: 594-598.

Plomp, P., and Mimpen, A.M. (1979). "Improving the reliability of testing the speech reception threshold for sentences", Audiology. 18: 43-52.

Plomp, R. (2002). "The intelligent ear – On the nature of sound perception", Lawrence Erlbaum, Mahwah, New Yersey.

Preminger, J.E., Neuman, A.C., Bakke, M.H., Walters, D., Levitt, H. (2000). "An examination of the practicality of the Simplex procedure", Ear Hear. 21: 177-193.

Press, W.H, Flannery, B.P., Teulolsky, S.A., Vetterling, W.T. (1989). "Minimization or maximization of functions", from: *Numerical recipes in Pascal: the art of scientific computing*, Cambridge University Press: 309-374.

Punch, J.L. (2001). "Paired-Comparison hearing aid preferences: Evaluation of an unforced-choice paradigm", J. Am. Acad. Audiol. 12: 190-201.

Punch, J.L., and Howard, M. (1978). "Listener-assessed intelligibility of hearing aid-processed speech", J. Am. Audiol. Soc. 4: 69-76.

Punch, J.L., Montgomery, A.A., Schwartz, D.M., Walden, B.E., Prosek, R.A., Howard, M.T. (1980). "Multidimensional scaling of quality judgments of speech signals processed by hearing aids", J. Acoust. Soc. Am. 68 (2): 458-467.

Punch, J.L., Robb, R. (1992). "Prescriptive Hearing Aid Fitting by Parameter Adjustment and Selection", J. Am. Acad. Audiol. 3: 94-100.

Purdy, S.C., Pavlovic, Ch.V. (1992). "Reliability, Sensitivity and Validity of Magnitude Estimation, Category Scaling and Paired-Comparison Judgments of Speech Intelligibility by Older Listeners", Audiology 31:254-271.

Putten, J. van. (2002). "Het gedachte-experiment (8)", Filosofie magazine 1 (11): 36-37.

Rayleigh, J.W. Strutt (1903). "Scientific Papers", Cambridge University Press.

Royal Institution Proceedings **14**, 524 (1895) [from *Rayleigh's Scientific Papers*, vol IV (Cambridge, UK: Cambridge Press, 1903)]

Rass, U., Steeger, G.H. (2000). "A high performance Pocket-size System for Evaluations in Acoustic Signal Processing", Acta Acustica 86: 374-375.

Rée, J. (1999). "I see a voice – A philosophical history of language, deafness, and the senses", Harper Collins Publishers, London.

Repp, B.H. (1987). "The role of psychophysics in understanding speech perception", from: *The psychophysics of speech perception*, edited by M.E.H. Schouten, Martinus Nijhoff, Dordrecht.

Roederer, J.G. (1979). "Introduction to the physics and psychophysics of music", 2nd edtion, Springer-Verlag, New York.

Rossing, T.D. (1990). "The science of sound", 2nd edition, Addison-Wesley, Reading, MA.

Rota, G-C. (1981). Introduction to Davis, P.J. en Hersch, R. "The mathematical experience", Houghton Mifflin Company, Boston.

Russell, B. (1961). "History of western philosophy – and its connection with political and social circumstances from the earliest times to the present day", George Allen & Unwin, London.

Sacks, O.W. (1989). "Seeing voices – a journey into the world of the deaf", University of California Press.

Smoorenburg, G.F. (1992). "Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram", J. Acoust. Soc. Am. 91: 421-437.

Spendley, W., Hext, G.R., and Himsworth, F.R. (1962). "Sequential Application of Simplex Designs in Optimisation and Evolutionary operation." Technometrics 4: 441-461.

Steeneken, H. J. M., Geurtzen, F. W. M., and Agterhuis, E. (1990). "Speech data-base for intelligibility and speech quality measurements", IZF-report A13, TNO Institute for Perception, Soesterberg, The Netherlands.

Störig, H.J. (2000). "Geschiedenis van de filosofie", Het Spectrum BV, vijfentwintigste herziene druk, Utrecht

Studebaker, G.A., Bisset J.D., Ort, D.M. van, Hoffnung, S. (1982). "Paired Comparison judgements of relative intelligibility in noise", J. Acoust. Soc. Am. 72 (1): 80-92.

Summers, I.R. (1991). "Electronically-simulated hearing loss of and the perception of degraded speech", from *Bioinstrumentation and biosensors*, edited by D.L. Wise, Marcel Dekker, New York.

Tassell, D. van. (1993). "Hearing Loss, Speech, and Hearing Aids", J. Speech Hear. Res. 36: 228-244.

Torczon, V.J. (1989). "Multi-directional search: a direct search algorithm for parallel machines", Ph.D. thesis, Rice University, Houston, Texas.

Toulmin, S. (1972). Human understanding, Volume I: General Introduction and Part I: "The collective use and evolution of concepts", Clarendon Press, Oxford.

Verschuure, J., and Dreschler, W.A. (1996). "Dynamic compression hearing aids", from *Psychoacoustics, speech and hearing aids*, edited by B. Kollmeier, World Scientific, Singapore: 153-164.

Verschuure, J., Benning, F.J., Van Cappellen, M., Dreschler, W.A., Boermans, P.P. (1998). "Speech intelligibility in noise with fast compression hearing aids", Audiology 37: 127-150.

Verschuure, J., Dreschler, W.A., Maré, M.J., Boermans, P.P., Capellen, M. van. (1993). "Compression and expansion in hearing aids especially in conditions with background noise", IOP-report 889cd14.

Verschuure, J., Maas, A.J.J., Stikvoort, E., De Jong, R.M., Goedegebure, A., Dreschler, W.A. (1996). "Technical assessment of fast compression in hearing aids", in *Psychophysics, speech and hearing aids,* edited by B. Kollmeier, World Scientific, Singapore: 171-182.

Verschuure, J., Prinsen, T. T., Dreschler, W. A. (1994). "The effects of syllabic compression and frequency shaping on speech intelligibility in hearing impaired people", Ear Hear. 15: 13-21.

Versfeld, N.J., Daalder, L., Festen, J.M., and Houtgast, T. (2000). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold", J. Acoust. Soc. Am. 107 (3): 1671-1684.

Walters, F.H.; Parker, L.R. Jr.; Morgan, S.L.; and Deming, S.N. (1991). "Sequential Simplex Optimization. A Technique for Improving Quality and Productivity in Research, Development, and Manufacturing", CRC Press, Boca-Raton, FL.

Weisskopf, V.F. (2003). "The privilege of being a physicist", Physics Today: 48-52.

Wittkop, T., Albani, S., Hohmann, V., Peissig, J., Woods, W.S., Kollmeier, B. (1997). "Speech processing for hearing aids: noise reduction motivated by models of binaural interaction", Acta Acustica 83: 684-699.

Wong, W. (2000). "Towards the physics of perception", IPO Annual Progress Report 35, Technical University Eindhoven.

Wundt, W. (1874). "Grundzüge der physiologischen Psychologie", Wilhelm Engelmann, Leipzig.

Yang, J., Luo, F-L., Nehorai, A. (2003). "Spectral contrast enhancement: Algorithms and comparisons", Speech communication 39: 33-46.

Yost, W.A. (1993). "Overview: Psychoacoustics", from: *Human psychophysics*, edited by W.A. Yost, A.N. Popper, and R.R. Fay, Springer-Verlag, New-York.

Yund, E.W., and Crain, T.R. (1997). "Voiced Stop Consonant Discrimination With Multichannel Expansion Hearing Loss Simulations", from: *Modeling sensorineural hearing loss*, edited by W. Jesteadt, Lawrence Erlbaum, Mahwah, New Yersey.

Zadeh, L.A. (1965). "Fuzzy sets", Information and Control 8: 338-363.

Zerlin, S. (1962). "A new approach to hearing aid selection", J. Speech Hear. Res. 5: 370-376.

Zwicker, E., Schorn, K. (1982). "Temporal resolution in hard-of-hearing patients", Audiology 21 (6): 474-492.

DANKWOORD

Vaak heb ik gedacht dat ik met iets bezig was dat alleen voor vakgenoten interessant kan zijn. Toch bleek dat de waarneming ook voor niet audiologisch georiënteerden zeer de moeite van het overpeinzen waard is. Zo schrijft Voskuil in 'Het Bureau' en in 'Bij nader inzien' zeer perceptief: hij neemt ons mee in de wereld zoals hij die ervaart, waarbij hij zijn waarnemingen zo puur mogelijk laat. En na een enthousiast ontdekking dat we de objectieve wereld volgens Immanuel Kant nooit kunnen kennen omdat we altijd onze zintuigen nodig hebben om haar te ervaren, concludeerde Kiek dat zij dat eigenlijk wel logisch vond.

Mijn inspiratiebronnen wil ik enorm bedanken en wel om verschillende redenen. Kiek wil ik bedanken voor haar relativerende humor en haar lessen in inlevingsvermogen. Roos en Gwen wil ik bedanken voor hun prachtige onbevangenheid en het feit dat ik tijdens mijn twee 'papadagen' weer even los kon komen en inspiratie kon opdoen.

Mijn vader en moeder hebben me doordrenkt met passie voor muziek en geluid. Jan en Lijn Jooren hebben me geholpen de perceptieve zenuwen van mijn proefschrift bloot te leggen. De waarneming van mijn vrienden Clemens van Gessel en Ernst Vermeulen heb ik uitvoerig mogen testen, mede omdat een eerder experiment verkeerd was opgezet waardoor ze nog twee volle dagen in de testcabine moesten afzien. Bovendien wilden ze hun oor lenen voor mijn vaak al te uitvoerige verhandelingen over mijn werk. Clemens van Gessel en Frederieke Baars hebben mijn proefschrift visueel prachtig vormgegeven. Dennis Römer, Koen Rhebergen en Maarten van Beurden hebben geweldig geholpen mijn proefschrift de juiste toon te geven. De mannen uit Limburg en mijn ex-collega's Niels Plasmans en Arjo Rozeboom hebben ervoor gezorgd dat ook de zintuigen de broodnodige ontspanning kregen.

De mensen zonder wie ik nooit mijn perceptieve gedachtespinsels had kunnen neerleggen in dit proefschrift zijn Jan Koopman en Wouter Dreschler. Hun aandeel was enorm. Wouter Dreschler was een geweldige promotor, die mijn waarnemingen en analyses altijd kritisch observeerde. Jan Koopman kan ik eigenlijk niet genoeg bedanken. Bij het opzetten van de experimenten was hij niet alleen een gewillig oor, maar was ook in woord en daad bijzonder loyaal. Wetenschappelijk hadden we weliswaar verschillende ideeën, maar dat heeft de keuze voor de te bewandelen weg naar mijn proefschrift alleen maar duidelijker gemaakt.

Mijn collega's uit het Europese SPACE project wil ik graag bedanken voor alle hulp, de algoritmes en de experimentele hoortoestellen. In het bijzonder ben ik André Goedegebure uitermate dankbaar voor zijn geweldige hulp in de beginfase en de bevlogen treinritjes van en naar het AMC. László Körössy is altijd mijn akoestische steun en toeverlaat geweest. Bovendien heeft hij een belangrijke rol gespeeld bij het schrijven

van de appendices en heeft me verschillende keren behoed voor valkuilen van de techniek. Johannes Lyzenga heeft na mijn SPACE tijd de kroon op mijn werk gezet door belangeloos zijn hoortoestelalgoritmen en zijn simulatiealgoritme uit te lenen.

De mensen zonder wie ik niets zou hebben kunnen opschrijven zijn de normaal- en slechthorenden die hun waarnemingen hebben willen afstaan. Hen ben ik zeer erkentelijk.

Tenslotte wil ik de leden van de promotiecommissie dr. ir. Festen, prof. dr. Fokkens, prof. dr. Kießling,[83] dr. ir. De Laat en prof. dr. ir. Spaan allerhartelijkst bedanken voor hun vriendelijke woorden, suggesties en bedenkingen. Bovendien is het een aantrekkelijke gedachte dat mijn proefschrift niet alleen door mijn promotor is gelezen.

[83] I like to thank prof. Dr. Kießling for his role in the examination board and his presence at my defence of this thesis.

GLOSSARY

Adaptive step size	variable step size that depends on perceptual distinctness between alternatives and on convergence rules of the pattern search
Auditory filter	filtering concept to describe the frequency selectivity of the ear
Cochlear hearing loss	hearing impairment in the inner ear (cochlea) subject to inner- and/or outer hair cell loss
Coefficient of consistence	measure of response stability derived from the presence of circular triads in a round-robin tournament
Compression	non-linear transformation in the ear that transform the large range of intensities to the small range of neural responses
Consistency	variation of optima found for different starting points in the multi-directional pattern search
Convergence	ability of the pattern search to find the 'true' optimum
Information transfer	area between the cumulative distribution of the distances between test-retest optima and the random distribution, relative to the 'perfect' distribution
Interaction I	interaction between algorithms: response on algorithm x1 depends on the setting of algorithm x2
Interaction II	interaction with listener: active participation of a person to optimise a combination of hearing-aid techniques
Loudness recruitment	larger than normal growth of loudness level for persons with a cochlear hearing loss
Paired comparison	evaluation paradigm based on pair wise comparison of two alternatives
Pattern search	optimisation procedure that is based on a predetermined lattice structure
Reliability	test-retest reproducibility for the pattern search
Round-robin tournament	a tournament in which every alternative is compared with every other alternative
Signal-processing strategy	(experimental) hearing-aid technique or algorithm
Simplex procedure	optimisation procedure that is based on a specific geometric configuration
Spectral resolution	measure for frequency selectivity of the ear
Temporal resolution	measure for temporal selectivity of the ear

CURRICULUM VITAE

Bas Franck werd geboren op 18 maart 1965 in Maastricht. Hij behaalde zijn VWO-diploma in 1984 in Venlo. Na een jaar Wiskunde aan de Technische Universiteit Eindhoven, besloot hij de Technische Lerarenopleiding Wiskunde en Natuurkunde in Eindhoven te gaan doen. In 1991 studeerde hij af en was hij inmiddels begonnen aan de studie Muziekwetenschap aan de Universiteit van Amsterdam. De doctoraalstudie heeft hij aan de systematische Muziekwetenschap gewijd, met als specialisatie Akoestiek. Een jaar na zijn afstuderen trad hij in 1996 in dienst van het Academisch Medisch Centrum en deed onderzoek in Europees verband (HEARDIP & SPACE project) naar het combineren en optimaal instellen van hoortoesteltechnieken. Aansluitend voerde hij zijn laatste studies uit voor zijn promotieonderzoek. Sinds 2003 is hij betrokken bij een project uitgevoerd in opdracht van het College voor Zorgverzekeringen (CvZ) over een Aangepast Zorgmodel voor Slechthorenden (AZOS) waarin mogelijkheden worden onderzocht om taken te verschuiven van voorschrijver naar audicien.

CURRICULUM VITAE

Bas Franck werd geboren op 18 maart 1965 in Maastricht. Hij behaalde zijn VWO-diploma in 1984 in Venlo. Na een jaar Wiskunde aan de Technische Universiteit Eindhoven, besloot hij de Technische Lerarenopleiding Wiskunde en Natuurkunde in Eindhoven te gaan doen. In 1991 studeerde hij af en was hij inmiddels begonnen aan de studie Muziekwetenschap aan de Universiteit van Amsterdam. De doctoraalstudie heeft hij aan de systematische Muziekwetenschap gewijd, met als specialisatie Akoestiek. Een jaar na zijn afstuderen trad hij in 1996 in dienst van het Academisch Medisch Centrum en deed onderzoek in Europees verband (HEARDIP & SPACE project) naar het combineren en optimaal instellen van hoortoesteltechnieken. Aansluitend voerde hij zijn laatste studies uit voor zijn promotieonderzoek. Sinds 2003 is hij betrokken bij een project uitgevoerd in opdracht van het College voor Ziekenhuisvoorzieningen (CvZ) over een Aangepast Zorgmodel voor Slechthorenden (AZOS) waarin mogelijkheden worden onderzocht om taken te verschuiven van voorschrijver naar audicien.