FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-6-4: Report on outcomes of research on automated fitting for compression hearing aids

Contractual Date of Delivery:	M32 + 4 month +45 days
Actual Date of Delivery:	10 October 2007
Editor:	Rolph Houben (NL-AMCA)
Sub-Project/Work-Package:	SP3/WP6
Version:	1.0
Total number of pages:	31

Dissemination Level		
PU	Public	X
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	
Project co-funded by the European Commission within the Sixth Framework Program (2002-2006)		
This information is confidential and may be used only for information purposes by Community Institutions to whom the Commission has supplied it		

Deliverable D-6-4

VERSION DETAILS	
Version:	1.0
Date:	10 October 2007
Status:	Final

CONTRIBUTOR(S) to DELIVERABLE	
Partner	**Name**
NL-AMC	Rolph Houben
NL-AMC	Wouter Dreschler

DOCUMENT HISTORY			
Version	**Date**	**Responsible**	**Description**
0.1	4.9.2007	Rolph Houben	First draft outline
0.2	8.9.2007	Wouter Dreschler	Second draft
1.0	10.10.2007	Rolph Houben	Edited after review

DELIVERABLE REVIEW			
Version	**Date**	**Reviewed by**	**Conclusion***
0.2	18-9	Rob Drullman	Accepted after corrections
0.2	23-9	Andy Faulkner	Accepted after corrections

* e.g. Accept, Develop, Modify, Rework, Update

Table of Contents

List of Figures

List of Tables

Acknowledgement

This work was supported by grants from the European Union FP6, Project 004171 HEARCOM. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Pre-Amble

This deliverable describes the investigation into a method to fit or fine-tune complex hearing-aid algorithms. It is part of work package 6: "Technical rehabilitation procedures and individualized fitting methods". The work package as a whole examines a range of current issues in rehabilitation with hearing aids and cochlear implants.

This work builds on the work described in the deliverable of work package 6 task 2, where the results of a multi-center questionnaire indicated that many professional hearing-aid providers (i.e., audiologists and hearing-id dispensers) were interested in interactive fitting procedures (D-6-2, stage 2).

The procedure under investigation is meant to improve the fine-tuning of hearing-aid fittings using generic methods or to fit newly developed algorithms such as the ones from SP3/WP5.

1 Executive Summary

This deliverable describes an experimental procedure for individualized fine-tuning of hearing-aid fittings. The novel procedure can also be applied to the fitting of new hearing-aid algorithms for which no generic fitting rules are available. As such, individualized fitting strategies might be applicable to state of the art hearing-aid signal-processing, such as under development in SP3/WP5.

The procedure is based on the mathematical simplex optimization procedure and uses direct comparisons of subjective sound quality. In other words, the method takes into account the preference of the hearing-aid user. The user is simply asked to compare two different hearing-aid settings and to indicate which one sounds better. This information is fed to the mathematical optimization system that efficiently chooses the next comparison, and that eventually converges to the optimal hearing-aid setting.

In order to improve the reliability of the input of the optimization algorithm, a two-step paradigm was developed. The subjects were presented with three sound samples. First, they had to determine which sample differed from the other two (odd-ball paradigm with a three-alternative forced choice, 3AFC). This step was included to increase the reliability of the second step in which the subjects had to judge which of the sound samples sounded better (two-alternative unforced choice, 2AUC). The new paradigm was tested for a simulated four-channel hearing aid.

The results indicate that for both normal hearing and hearing-impaired subjects, the minimally perceived differences between hearing-aid settings were large compared to the values used in clinical practice. This indicates that either the subjects did not hear the difference between the samples or that the discrimination task was too difficult. In view of previous results with optimization procedures, it is likely that the unfavorable results were largely caused by the length of the sound samples. Our paradigm might simply not be suitable for entire sentences. User comments agree with this assumption.

A possible improvement to the procedure is to abandon the current 3AFC setup with fixed sentences (with each a duration of about 3 s) in favor for one with continuously running speech for which the user (and not the computer) determines when one of the three alternatives becomes active. This will minimize memory requirements of the subjects, and hopefully this will improve the results. Or, alternatively the length of the stimuli might be shortened.

2 Introduction

This deliverable describes experiments with interactive fitting, i.e. a procedure that is intended to optimize hearing-aid fitting in interaction with listeners. The fitting procedure is built around an adaptive optimization strategy that selects sound samples for comparison, based on previous answers of the listeners.

The capabilities of current hearing-aids include sophisticated signal processing with many parameters whose effects on both objective and subjective measures of aid performance may interact. However, the available generic fitting-procedures are mostly based on the detection thresholds for pure tones (the pure-tone audiogram) and do not provide the information needed to set the parameters of the many new processing algorithms. An example of these new algorithms is active noise reduction. Unfortunately, it is common practice to keep proprietary processing algorithms secret, and this severely complicates the development of generic fitting-rules. And even if generic fitting-rules are available, in clinical practice the fitting is often subjectively optimized for an individual user. This optimization is purely based on the judgement of the professional. To improve upon this situation, a generally applicable procedure that can fine-tune a hearing-aid fitting can fill in this hiatus. Interactive fitting can bring an important innovation involving the end-user in hearing aid fitting and fine tuning.

Another complication with hearing-aid fitting is that the fitting parameters can interact. For instance, the required amount of low-frequency gain can depend on the amount of noise reduction. This in itself would not pose a problem. However, current hearing aids have many adjustable parameters us as the amount of gain, compression ratio, compression speed (attack and release time), compression threshold (kneepoint), limiter threshold, noise reduction, feedback reduction, directionality, and speech “enhancement”. Moreover, these parameters can often be adjusted separately for up to 15 independent frequency-channels. In total, this leads to billions of possible combinations and it is impossible to perform a full parametric comparison. The proposed solution for this fitting problem is to use an interactive procedure that incorporates an adaptive procedure for the selection of the best settings. Such a procedure uses the answers of the subjects to determine the best next listening test, so that the fitting is optimized within a short amount of time.

2.1 Literature overview

Personalized fitting strategies have been investigated before. Levitt and coworkers (Levitt et al., 1986; Neuman et al., 1987) were the first to use a mathematical optimization algorithm for the fitting of hearing aids. They adapted the simplex algorithm (Spenley et al., 1962) for use with paired comparisons of subjective quality judgments. This work was followed by numerous other studies, notably from Kuk and colleagues (Kuk and Pape, 1992, Kuk and Lau, 1996), from Dirks et al. (1993) and from the group of Levitt (Preminger et al., 2000).

Table 1 shows an overview of research into the applicability of the (modified) simplex procedure.

Table 1: Overview of research into the applicability of the (modified) simplex procedure. The research used a fixed grid in which only the grid points could be used as parameter values. "Start point" refers to the value of the initial fitting. "Stop criterion" indicates which criterion is used to end the optimization procedure: "rev/dim" stands for reversals per dimension, which indicates the number of reversals that are minimally required for each optimization variable before the procedures stops. SNR means signal to noise ratio, NH means normal hearing subjects and HI indicates hearing impaired subjects.

Paper	Grid size	Start point	Stop criterion	Outcome measures	Noise type	SNR (dB)	Subjects
Levitt et al. 1986	5x5	point (3,3)	3 rev/dim	subjective intelligibility	cafeteria noise	unknow n	2 NH
Neuman et al. 1987	5x5	point (3,3)	3 rev/dim	subjective intelligibility	cafeteria noise	+5	8HI
Kuk and Pape 1992	5x5	NAL-R	3 rev/dim	subjective quality, syllable recognition	quiet multi-talker babble	+5	20 HI
Dirks et al. 1993	5x5 5x5x5	NAL-R	5 rev. in total	subjective quality, subjective intelligibility	party noise	+3	9 HI
Kuk and Pape 1993	5x5	NAL-R	3 rev/dim	degree of difficulty	quiet and babble	0, +5	19 HI
Kuk 1994	5x5	NAL-R	3 rev/dim	subjective clarity	multi-talker babble	0, +5	10HI
Kuk and Lau 1995	5x3	NAL-R	3 rev/dim	subjective clarity	multi-talker babble	0	7HI
Kuk and Lau 1996	5x3	NAL-R	3 rev/dim	subjective clarity	multi-talker babble	-5, 0, +5	7HI
Kuk and Lau 1996	5x3	NAL-R	3 rev/dim	subjective clarity	multi-talker babble	-5, +5	13HI
Preminger et al. 2000	5x5	NAL-R	3 rev/dim	subjective clarity	office ventilation system noise, cafeteria noise	+8. +20	11HI
Franck et al. 2004	9x9x9	point (-3,-3,-3)	6 wins or 3 times at	listening	car noise, fluctuating	0	7NH

		(1,1,1) (3,3,3) winner	same place	comfort	speech noise		
Franck et al. 2007	9x9x9	point (0,0,0) (4,4,4)	6 wins or 3 times at same place	subjective intelligibility	car noise, two-talker babble	+5	10HI 4NH

Most previous research into the modified simplex procedure was two-dimensional. The gain in the low and high-frequency channels was varied independently (cross-over frequency was mostly around 1 kHz). The exceptions were Kuk (1994) who used low-frequency gain and overall gain, and Franck et al. (2004) who used non-linear signal processing (noise reduction, spectral enhancement, and spectral lift) as optimization variables as opposed to the linear processing (gain) of the other papers. Franck et al. (2007) used noise reduction, temporal signal enhancement and amplitude compression (in which the optimization parameter was the number of frequency channels).

The original simplex procedure was modified in such a way that it could be used with paired comparisons, and not with ratings only (Neuman et al. 1987). The modification resulted in a fixed grid of possible parameter values. The research from Table 1 used the modified simplex procedure with subjective judgments of speech (in quiet or embedded in a background noise). The starting point for the optimization procedure was either the commonly applied NAL-R fitting rule (Dillon, 2001) or an arbitrary grid point, see Table 1. An example of grid points is shown is Figure 3, section 3.2.

Overall, the results for a two-dimensional optimization algorithm indicated that the procedure was viable and very efficient compared to round-robin procedures. The test-retest variability was small. Up to 80% of the listeners selected in retest nearly identical settings as in the first measurement, with a grid size of 5 dB (Kuk and Pape, 1992).

Franck, Dreschler, and Lijzenga (2004) extended the two-dimensional technique to include a third dimension. They extended the grid size from 3x3 to 9x9x9 points. To improve the algorithm they introduced an adaptive step-size and a repeat button. Their results showed that when non-linear processing was used, there were multiple local optima with the result that the search performed poorly. Furthermore, Franck et al. concluded that an adaptive step-size was needed to circumvent problems with small changes in parameter values that did not produce perceptible changes. Additionally, too small a step size led to low reliability. In another experiment, Franck et al. (2007) used the optimization algorithm of Franck et al., 2004 with a commercially available hearing aid. Again, they found that the perceptual differences between different parameter values were small. The minimum step-size had to be chosen to be quite large to obtain an acceptable reliability.

In conclusion, previous research shows that:

- interactive fitting with an adaptive procedure may become an important generic tool in hearing aid fitting, and it seems viable;
- the modified simplex procedure is very efficient;
- an adaptive step-size can be used to improve test-retest reliability;
- the perceptual differences might be small where non-linear processing is used, and this can impair the reliability of the modified simplex procedure.

Based on these results from literature we designed a procedure in which the main focus was on improving the reliability and the efficiency of the information from the listener's response that forms the input to the adaptive procedure. That is, we tried to extract accurate data from the subjects in an efficient way.

Unfortunately, the just noticeable difference (jnd) for some of the parameters of hearing-aid processing as this affects speech is not known beforehand because it can depend on factors such as the hearing loss of the subject and the parameter under investigation. Moreover, the jnd can be influenced by the interaction between parameters. For instance the jnd for time constants with compression hearing-aids might be smaller for larger compression ratios.

In our procedure the original paired-comparison paradigm was replaced by a paradigm that automatically estimates the just noticeable perceptual differences for the hearing-aid algorithms being used. This measure of just noticeable perceptual differences was used to determine the step size of the optimization algorithm. This was presumed to have the advantage that the step size can be changed more reliably and that the test duration and test accuracy are improved.

In addition, we removed the requirement that all possible parameter values are placed neatly in a grid. That is, we changed the set-up from categorical to semi- continuous. This has the advantage that the point distance (grid width) does not need to be determined beforehand.

3 Materials and methods

3.1 Paradigm

In order to provide the optimization algorithm with reliable input, and to simultaneously obtain an estimate of the minimally noticeable perceptual differences, we designed a two-step paradigm.

The first step consisted of a discrimination task. In this three-alternative forced-choice paradigm three stimuli were presented to the listener. Of these three stimuli, one stimulus differed physically from the other two. The subject’s task was to detect which stimulus differed from the other two ('odd-ball paradigm'). All three presentations used the same original sentence (but they were differently processed by the hearing-aid algorithm). In line with the experiments of Franck et al. (2004), the text of the sentence was shown on the screen. This was done to prevent the last presentation to be preferred due to increased intelligibility caused by repetition of the same sentence. Figure 1 shows the computer screen that was shown to the subjects.

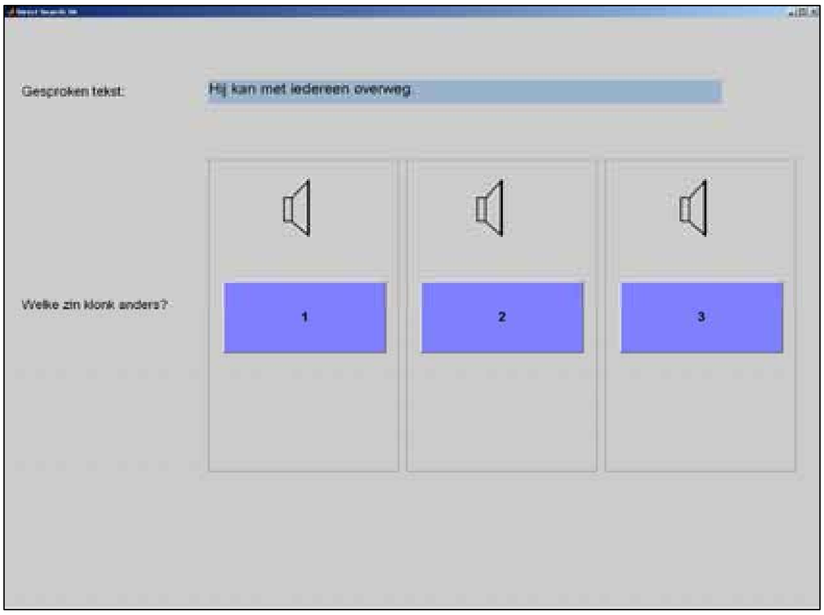

Figure 1: Screen capture of the judgment task (3AFC). The sentence was presented three times. One of the presentations was differently processed than the other two. The text at the top of the screen shows the text of the sentence.

If the subject issued an incorrect response, he/she was notified that the wrong stimulus was chosen, and the second step was skipped. If the response was correct, the user was presented with the judgment task of the second step. In a two-alternative unforced-choice the user was asked if the selected sound sample (from the first step) was preferred over the other two: "Did you prefer this sound sample? Yes, No, Equal". The sound fragments were not presented again for this second step. Figure 2 shows the subject screen.

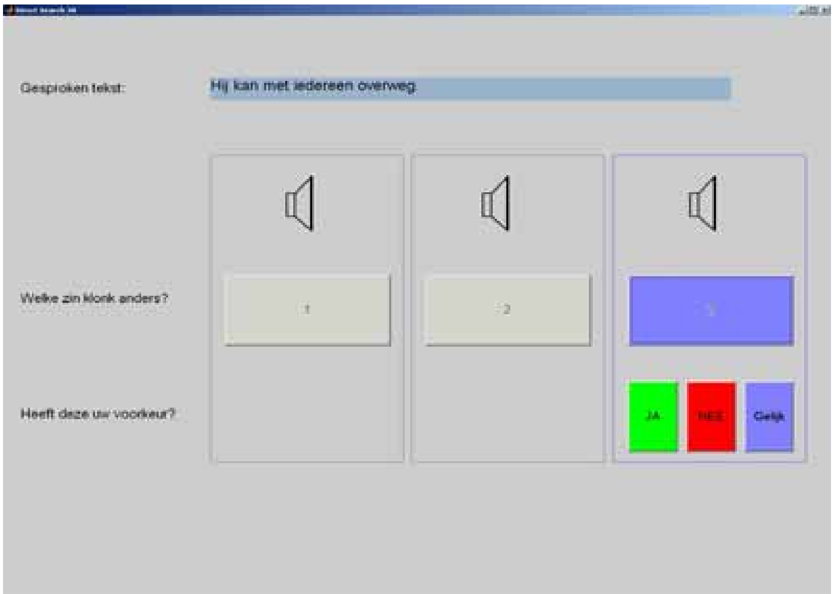

Figure 2: Screen capture for the judgment task (2AUFC). The subject had to indicate if the selected sound sample was preferred over the other two.

3.2 Optimization algorithm

Each hearing-aid processing algorithm is treated as one dimension of the optimization procedure. Each trial compares stimuli that are different in one dimension only (i.e., only one hearing-aid parameter is different between the sound samples that are compared), see Figure 3. After this comparison the other dimensions are used. For example, for an optimization system in which the gain in the low and high-frequency channels are optimized, a comparison will be made with a difference in gain for the low frequencies first (A vs. B). After this the different gain values for the high-frequencies are compared (A vs. C). A comparison with a change in both dimensions simultaneously does not occur (e.g., no comparison between A and D, that is, no comparison between a low gain in the high frequencies and a high gain in the low frequencies).

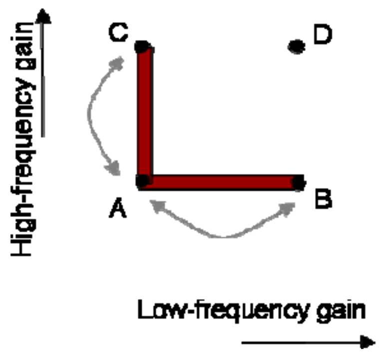

Figure 3: Schematic representation of the optimization algorithm. For each optimization dimension (i.e., for each hearing-aid algorithm) a separate comparison is made. The target setting is compared to the original (A) -> (A,B) (A,C). The comparison only uses one dimension at a time: A is never compared to D, only to B and C.

Figure 4 shows another example of the procedure.

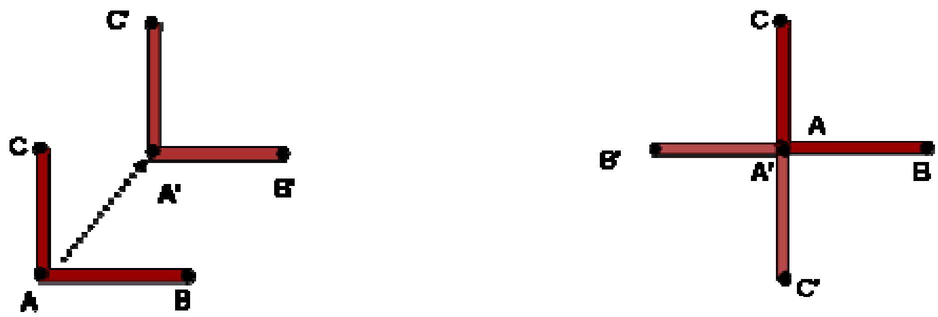

Figure 4: Two examples of a paired comparison of A with. B (dimension 1) and A with C (dimension 2). The comparisons for the next round are denoted A', B', and C'. In the left panel both B and C were preferred over A, in the next round the procedure looks in the direction of B and C. In the right panel A was preferred over both B and C, and the search direction is reversed to B’ and C’.

In order to improve the input of the above described optimization procedure, we replaced the triple paired-comparisons that were used in previous research by a two-step paradigm (section 3.1).

If the answer in the first step (discrimination) was incorrect, the step size was increased (with a factor 1.7), and the procedure went along to the next dimension. If the correct answer was given, the judgment task was presented to the subject. Based on the answer, a direction for the next comparison was determined, see Figure 4. The step size for the new comparison was then decreased factor 1/1.7), and the procedure went on to the next dimension.

A too small step size can lead to too many repetitions (inefficient). Therefore, the step size was adapted automatically. After two reversals in step size (from increase to decrease or vice versa) the step-size change factor was decreased with 0.2 (from 1.7 to 1.5 to 1.3, etc.). The minimum step-size change factor was 1.1 (or 1/1.1).

Both ours and the set-up of Franck et al. (2004, 2007) deviate from the research described in Table 1, in that a preference was accepted after just one paired-comparison. The other research repeated a paired-comparison three times; the stimulus that was preferred 2 (or 3) times out of 3 was used as input to the optimization algorithm. This repetition increases reliability but is very time consuming, especially if more than two dimensions are used. Franck et al. (2004, 2007) therefore used only one paired comparison. However, they added a repeat button that allowed the subjects to repeat all three stimuli prior to making a choice. They also used an adaptive step-size to further increase the reliability (Franck et al., 2004).

Our two-step paradigm also presented only one paired comparison. However it differs from the set-up of Franck et al. (2004, 2007) in that we replaced the 'repeat' button by a discrimination task (first step), see section 3.1.

In addition, we removed the requirement for a fixed grid, so that a continuum of settings was possible.

3.3 Experimental design

The hearing-aid fitting was optimized based on user preference. The user was asked which sound sample was preferred, no further direction was given. Afterwards the users were asked what their decision criterion was, e.g., intelligibility, clarity, etc.

The optimization algorithm was used for three different sets of hearing-aid algorithms, see Table 2.

Table 2. Experimental design.

Set	Dimension 1	Dimension 2	Dimension 3
A	low-freq gain	high-freq gain	-
B	broad-band gain	slope of the gain	-
C	low-freq gain	high-freq gain	broad-band compression/expansion ratio

The first two sets (A and B) are two-dimensional and they optimized the gain for linear amplification only (compression ratio was 1). For set A this was done separately for the low frequencies ('bass') and for the high frequencies ('treble'). For set B, the broad-band gain ('volume') was optimized as well as the difference between the high and the low frequencies ('sound colour'). This set should result in exactly the same amplification (end point) as for set A, however, the perceptual effect of the 'route' to this end point is different. The comparison of A and B offers a good test of the dependence of the outcome on rotation of dimensions.

Set C was three dimensional and optimized the same gain parameters as set A ($Gain_{low}$/$Gain_{high}$) complemented with optimization of the compression ratio (CR). The same CR was used for all four frequency-channels, and could vary between 0.2 dB/dB (expansion) and 10 dB/dB (compression). The initial fitting of the emulated hearing-aid was done according to the NAL-RP prescription rule for a speech input-level of 65 dB SPL. This input level was chosen because for this level NAL-RP is nearly equal to the commonly used NAL-NL1 for non-linear hearing aids (Dillon, 2001). The initial compression ratio was either 1.5 dB/dB or 2.5 dB/dB.

The modified simplex procedure requires two settings that can be compared. Therefore the start point for sets A and B was quasi-randomly chosen around the initial fitting. Either [-3 dB,+3 dB] or [+3 dB,-3 dB] around NAL-RP for [dimension 1 , dimension 2]. The settings [+3,+3] and [-3,-3] did not occur. The starting point for compression ratio was chosen around 2 dB/dB (randomly chosen between either 1.5 dB/dB or 2.5 dB/dB).

All sentences were selected and presented at random. The presentation order of the dimensions was randomized after each step. The experiment ended after three reversals were obtained for each of the dimensions. After the optimization procedure had ended, the initial fitting was compared to the reached end point. In order to estimate the test-retest variance, all measurements were made in duplicate at random during the same session.

All experiments were conducted in a soundproof room. The subjects removed (both) their hearing aids prior to the experiment. Since the headphones (see section 3.6) provided high passive attenuation of external noise (approximately 29 dB at 1 kHz according to specifications from manufacturer), masking at the contra lateral ear was not used. Subjects received written and oral instructions. Every session started with a practice session to familiarize subjects with the experimental procedure. After the experiment the participants were asked to comment on the procedure. The entire experiment lasted about 2 hours, including breaks, and was conducted in a single visit.

3.4 Materials

Speech materials consisted of Dutch sentence material for measurement of the Speech Reception Threshold in noise, developed by Versfeld et al. (2000). From this material only the female speech was used. The total number of sentences was 507.

A babble background noise was constructed by concatenating all sentences and subsequently placing ten sequences on top of each other. The starting point was chosen randomly to prevent the prosody of the individual sentences to line up and to be audible in the noise. The constructed noise matched the long-term average speech spectrum (LTASS) of the speaker. The dynamic range of the original speech material was 26±6 dB (the standard deviation was calculated from the dynamic range of all the separate sentences); the dynamic range of the background noise was approximately 9 dB (both wide band measurements between the 1st and 99th percentile, applying an integration time constant of 125 ms).

Speech was embedded in the background noise. The noise extended from 0.5 seconds before speech (onset), to 0.5 second after speech (offset). The onset was used to allow any initial gain adjustments of the compressor to even out before speech was presented and to attract subjects' attention to the stimulus. The offset prevented an abrupt ending after the speech had finished, thus giving the compressor time to fully change the gain in accordance with its time constants. The time between two consecutive sentences was approximately 1.2 second.

3.5 Subjects

Six normal-hearing (NH) adult subjects participated in this study, as well as five adults with sensorineural hearing-loss (HI). All included subjects completed the entire experiment; none of them quitted early. The inclusion criterion for the normal hearing subjects was a hearing loss that was less than 20 dB HL at all audiometric frequencies from 125 to 8000 Hz. The inclusion criterion for the hearing impaired subjects was a sensorineural hearing loss larger than 40 dB HL for any frequency in the range of 1 to 4 kHz, but no thresholds beyond 80 dB HL in this range. All subjects were native Dutch speakers.

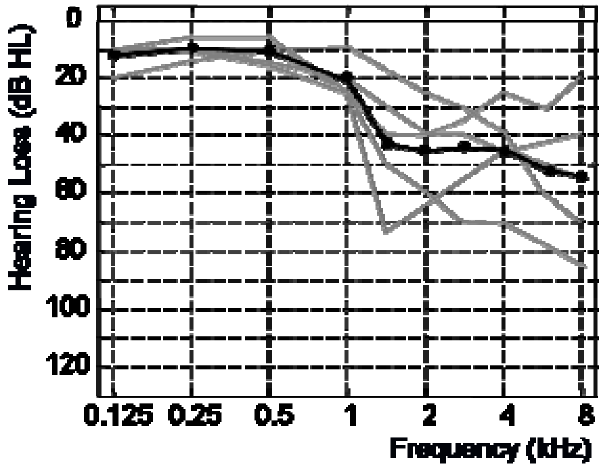

Figure 5: Pure-tone thresholds of the hearing-impaired subjects (plotted in gray). The black line shows the average threshold.

The pure-tone audiograms of the hearing-impaired participants are shown in Figure 5. The average age of the NH participants was 32 (±6) year and of the HI subjects it was 58 (±11) year. The ear used was the one that satisfied the inclusion criteria. If both ears were suitable, the ear with the smallest hearing loss was used in order to avoid listening by the contralateral ear. Subjects participated on a voluntary basis and received a reimbursement of travelling expenses and a small financial compensation of 25 Euro.

3.6 Hearing-aid emulation

Signal processing was carried out off-line using a personal computer with MATLAB Release 14 (The Mathworks, 2005). All processing was done at a sample rate of 44100 kHz and a bit depth of 24 (the original speech materials were available at 16 bits per sample). The noise masker and the speech signal were summed before further processing.

Frequency-channel filtering

The emulated hearing-aid had four frequency-channels. This was constructed by letting the input signal pass through four elliptical band-filters (see Table 3). The filters were applied twice (non-causal filtering): after the first filtering the filters were applied again to the time-reversed signals to remove any phase distortion introduced by the first filtering stage (Smith, 1997).

Table 3: Characteristics of frequency-channel filtering. Effectively the order was twice as large, because the filters were applied twice.

Channel	Low cut-off frequency	High cut-off frequency	Filter order
1	90	250	3
2	250	707	4
3	707	2000	4
4	2000	8000	4

Amplitude compression

Compression was used in the experiment with set C only (see section 3.3), and preceded the linear amplification stage (input dependent compression or AGC-i). The signal was compressed independently in all four channels. The amplitude envelope in each channel was calculated for an entire stimulus (full sentence) by means of a Hilbert transform (Oppenheim and Schafer, 1989) and subsequent low-pass filtering (Butterworth 2e order, non-causal filtering, cut-off frequency of 80 Hz). These envelopes were then logarithmically compressed (Braida et al., 1979). A gain signal was constructed for the signal in each channel by calculating the ratio of the instantaneously compressed envelope and the original envelope. In order to introduce attack and release times, each gain signal passed through two first-order low-pass filters. One filter acted only on rising signals (attack-time filter), and resulted in a logarithmic decay. The other filter acted only on falling signals (release-time filter), and resulted in a gain that rose logarithmically. After subjecting the gain signals to the attack and release filters, the signals in the channels were multiplied sample by sample by their respective gain signal. In each channel the compressed signals were then amplified to the long-term rms values of the original input signals in that channel. After this the signal in the channels was amplified in accordance with the gain requirements of the adaptive optimization procedure.

The compression threshold was chosen at 47.7, 46.9, 37.6, and 40.1 dB SPL for the four channels respectively. These levels correspond to a broad-band threshold level of 52 dB for our speech materials, as prescribed by NAL-NL1. The attack and release times were 4 and 40 ms, respectively. Compression was applied prior to the gain from the other optimization dimensions, so this gain did not influence the compression threshold.

Linear amplification

The optimization algorithm used two parameters to change the gain. For sets A and C these two parameters (gain in the low and the high frequencies, respectively) corresponded directly to the frequency channels 1 and 4 (see Table 3). The gain in channels 2 and 3 was linearly interpolated on a dB scale.

For set B, The first dimension (broad-band gain) corresponded directly to the gain in all four channels. For the second dimension (difference between the gain in the high and the low frequencies) the highest frequency channel (4) received a gain of plus half the value of the second dimension, whereas the lowest channel (1) received minus half. For example, the optimization point of [6,6] corresponded to a gain of +3 dB (6 dB -3 dB) in channel 1 and of +9 dB (6 dB + 3 dB) in channel 4. The gain in channels 2 and 3 was linearly interpolated on a dB scale.

Finally, the outputs of all four channels were summed.

Post processing

The fitting rules NAL-RP and NAL-NL1 prescribe an insertion gain, i.e., the difference between the sound pressure level in the ear canal under aided and unaided conditions (Dillon, 2001). In order to present the stimuli at the prescribed levels a filter is needed that mimics a free-field signal when a headphone is used. We therefore calculated the Free Field to Headphone transfer function for our Sennheiser HDA 200 headphones. This was done by measuring the frequency response on a B&K Head and Torso simulator (HAT) for both the free field and the headphone condition. The response at the eardrum of the HAT was measured with an Aurical measurement system (Madsen Electronics). And from these responses the compensation filter was calculated that was applied to the sound signal.

Lastly, a software safety limiter made sure that none of the processed signals had a root mean square value higher than 25 dB above NAL-RP (full program stop above this level).

Digital-analogue conversion

The digital stimuli were converted to the analogue domain using a 'RME Fireface 800' 24-bit sound card. The output of the sound card was connected to a Behringer Audio Interactive Dynamics Processor model Autocom pro-xl MDX 1600. This device was used as an additional safety measure to prevent too high sound levels in the unlikely event of a system failure. Only the peak limiter was in use, all other processing features were turned off. The unit was calibrated to have an insertion gain of 0 dB, and did not influence the sound signal during normal operation. The very fast limiter (attack time was near zero, compression ratio was infinity) was calibrated to limit sound at 115 dB(A). During the experiments, none of the signals reached this level.

The signal was amplified by a Tucker Davis MA2 microphone amplifier followed by a Tucker Davis headphone buffer HB6. All signals were presented monaurally through Sennheiser HDA200 circumaural headphones.

4 Results

4.1 Test-retest

All measurements were conducted in duplicate. The test-retest standard deviation is shown in Table 4. The data for set B ($Gain_{overall}$/$Gain_{slope}$) is expressed as the gain in the low and high frequencies. An analysis of variance (ANOVA) of the test-retest data showed no significant difference in the test-retest standard deviation between the normal hearing and the hearing impaired subgroups ($p=0.3$). The test-retest data in Table 4 is therefore pooled over normal hearing and hearing impaired participants.

Table 4: Test-retest standard deviations. The data for set B ($Gain_{overall}$/$Gain_{slope}$) is converted to $Gain_{low}$/$Gain_{high}$. CR stands for compression ratio.

	σ $gain_{low}$ (in dB)	σ $gain_{high}$ (in dB)	σ CR (in dB/dB)
Set A $Gain_{low}$/$Gain_{high}$	5.0	5.8	-
Set B $Gain_{overall}$/$Gain_{slope}$	3.3	6.7	-
Set C $Gain_{low}$/$Gain_{high}$/CR	9.8	4.0	1.6

An ANOVA on the gain measurements of all sets (σ $gain_{low}$ and σ $gain_{high}$) showed that the test-retest standard deviation did not differ significantly between the sets ($p=0.2$).

4.2 End points

4.2.1 Average end points

Figures 6 and 7 show the averaged end-points for the normal hearing and the hearing-impaired subjects, respectively. The given gain values are relative to the gain as prescribed by NAL-RP.

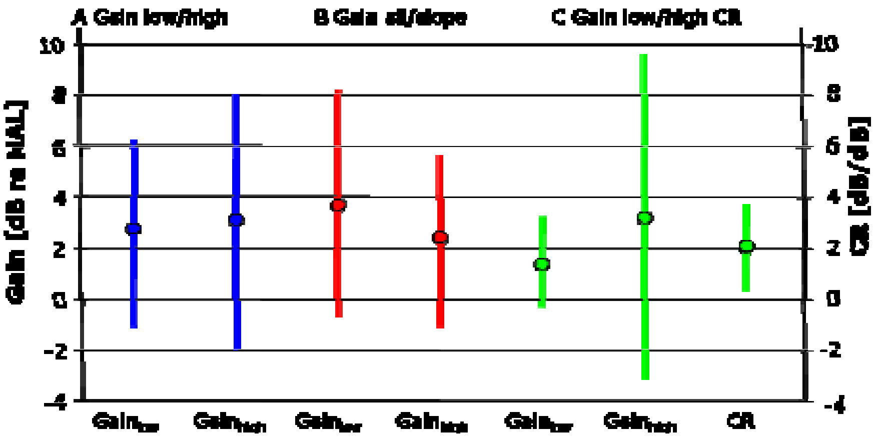

Figure 6 Average end-points for the normal-hearing subjects. The data for set B ($Gain_{overall}$/$Gain_{slope}$) is converted to $Gain_{low}$/$Gain_{high}$. Error bars represent one standard deviation.

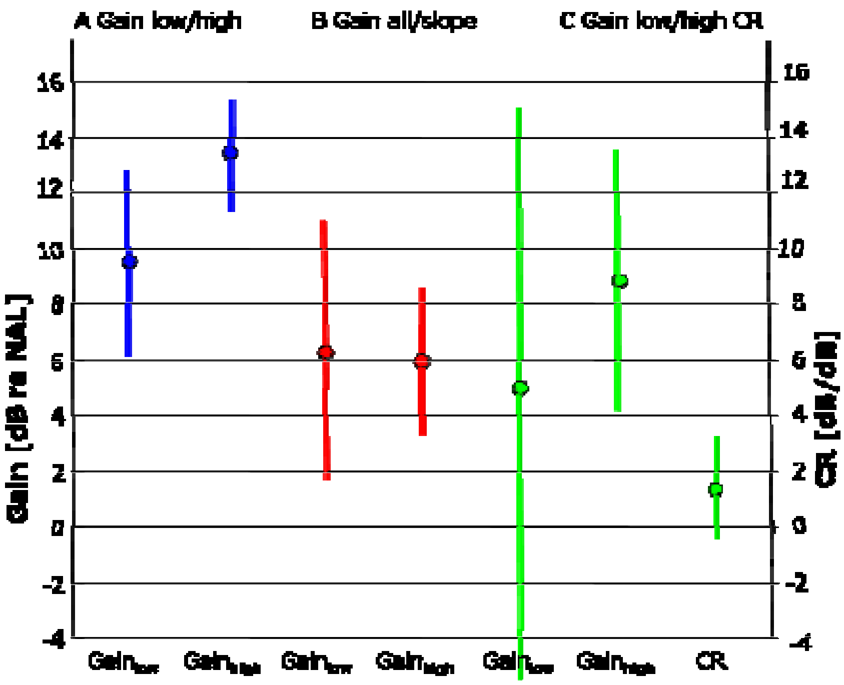

Figure 7 Average end-points for the hearing-impaired subjects. The data for set B ($Gain_{overall}$/$Gain_{slope}$) was converted to $Gain_{low}$/$Gain_{high}$. Error bars represent one standard deviation.

The figures show that the end points for set A were higher for the HI than for the NH (i.e., the optimization resulted in more gain than prescribed by NAL-RP). This effect was significant ($p<0.01$). The results for sets B and C did not differ between NH and HI ($p=0.1$, and $p=0.2$ for sets B and C, respectively).

For the hearing-impaired group the end points for set A were significantly higher than for set B (5.3 dB, $p<0.01$). For the normal-hearing group this effect was absent (-0.1 dB, $p=0.97$).

Also, the end points for compression ratio did not differ significantly between normal hearing and hearing impaired ($p=0.6$).

In order to compare the outcomes of set A and B, figure 8 shows the end values for set A (abscissa) and set B (ordinate). Each point represents a subject. From the figure it can be seen that for the HI group the gain for set A is higher than for set B.

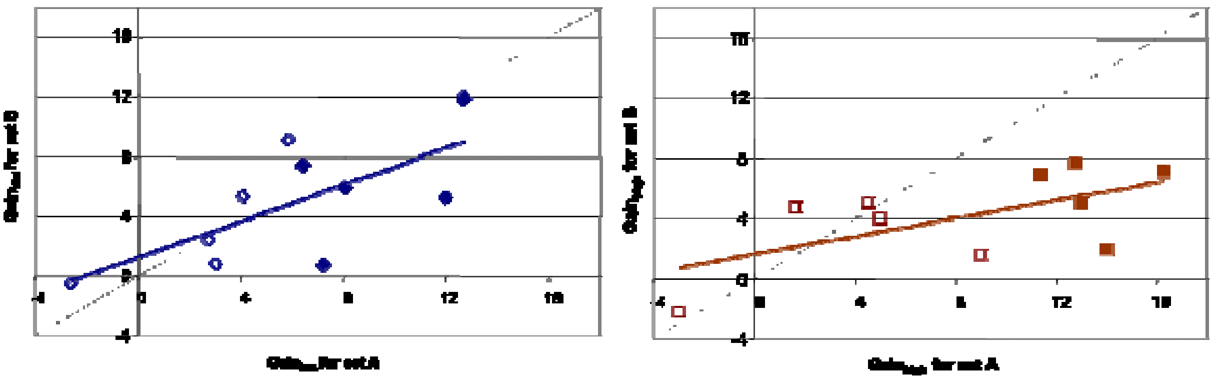

Figure 8 Comparison of outcomes of measurement set A and B. The data for set B ($Gain_{overall}$/$Gain_{slope}$) is converted to $Gain_{low}$/$Gain_{high}$. The left panel shows data for $Gain_{low}$, the right for $Gain_{high}$. Open symbols represent normal hearing participants; closed symbols show data for the hearing-impaired subjects. The dotted line shows equal gain for set A and B.

4.3 Step size

Figures 9 and 10 show the average step size for the normal hearing and the hearing-impaired subjects, respectively.

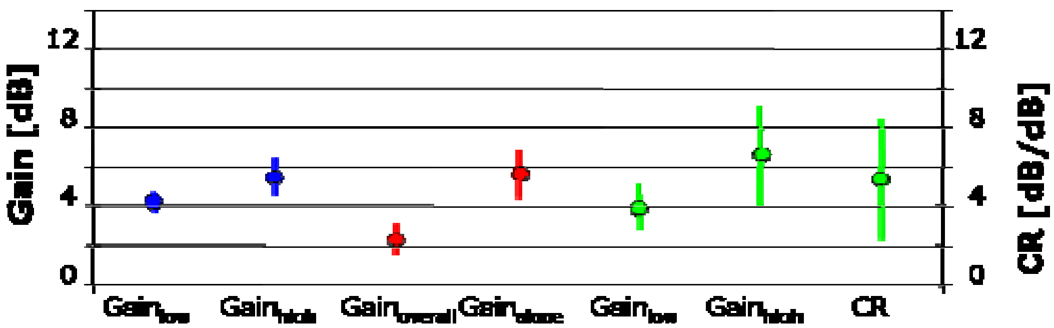

Figure 9 Average step-size for the normal-hearing subjects. Error bars represent one standard deviation.

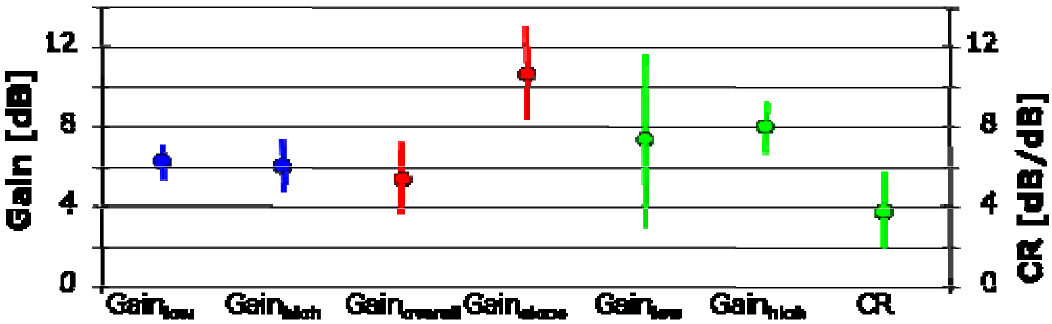

Figure 10 Average step-size for the hearing-impaired subjects. Error bars represent one standard deviation.

For the gain, the average step-size was larger for the hearing-impaired subjects than for the normal-hearing subjects. This was significant for all three sets ($p<0.05$ for all sets). The step size for compression ratio did not differ significantly between the normal hearing and the hearing-impaired groups ($p=0.9$).

The step size of set A and B cannot be compared directly. Broad-band gain (set B) has more influence on the stimulus because all frequencies get amplified, whereas for the parameters of set A only the low or the high frequencies. It is to be expected that the minimally detectable difference in $Gain_{overall}$ is smaller than $Gain_{low}$ or $Gain_{high}$. This effect is clearly visible in Figure 9 for normal-hearing subjects: the step size for $Gain_{overall}$ (set B) was 1.9 dB and 3.1 dB smaller than the step size for $Gain_{low}$ (set A) and $Gain_{high}$ (set A), respectively. These differences were significant ($p<0.01$ for both $Gain_{low}$ $Gain_{high}$). For the hearing-impaired subjects the differences in step size were much smaller (0.9 dB and 0.6 dB, for $Gain_{low}$ and $Gain_{high}$, respectively) and not significantly different ($p=0.4$ and $p=0.6$ for $Gain_{low}$ and $Gain_{high}$, respectively).

4.4 Comparison to initial fitting

After the optimization procedure had reached its end, the end point was compared to the initial NAL-RP fitting.

The difference between the starting points and the end points could only be detected in 47% of the normal-hearing subjects. In these subjects, 40% of the subjects preferred the end point and those appeared to be the subjects with higher gain at the end point than at the starting point.

The difference between the starting points and the end points could only be detected in 65% of the hearing-impaired subjects. In these subjects, 60% of the subjects preferred the end point and – again – those appeared to be the subjects with higher gain at the end point than at the starting point.

4.5 Session efficiency

Table 5 gives the average duration of the experiment.

Table 5: Number of reversals and session durations.

	Normal hearing		Hearing impaired	
	Number of reversals	Time (min)	Number of reversals	Time (min)
Set A $Gain_{low}/Gain_{high}$	4/3	5±2	5/5	6±3
Set B $Gain_{overall}/Gain_{slope}$	3/4	5±2	3/5	5±2
Set C $Gain_{low}/Gain_{high}/CR$	5/5/8	11±5	5/5/4	13±5

The average duration of the sessions did not differ between the normal hearing and the hearing-impaired participants ($p>0.4$ for all thee sets). Session duration did not differ between sets A and B ($p=0.6$). Due to the extra dimension, set C took longer than A and B. However, if the time per dimension is compared (2.7, 2.5, 4.0 min, for set A, B, and C respectively), then set C still took much longer than A ($p<0.05$) and B ($p<0.01$). This is caused by the stop criterion that prescribes that each of the dimensions has to have at leas three reversions in direction before the procedure can end. Until this criterion was met, all dimensions were still tested, irrespective the number of reversions for a specific dimension. The addition of an extra dimension therefore increases testing time more than proportionally.

4.6 User views

Participants were asked to give their opinion on the procedure. Some relevant user opinions are listed below:

- The task is clear.
- The experiment was demanding, because it requires a high level of concentration, and the audible differences were small.
- The most difficult situation was when the last two sentences were found to differ. This decision than depends on the perception of the first sentence. If this sentence was missed or not fully remembered, the subject will have to guess which of the last two sentences was the odd ball. This is unfortunate, since the subject was able to detect a difference between the last two sentences. This could have been prevented by again presenting the first sentence; however this would have increased measurement time by at least 25%.
- When the sound quality was poor, it would sometimes take long to improve.

- The decision criteria were intelligibility, clarity, loudness, natural sounding. Only 1 participant mentioned the amount of background noise as decision criterion.

- The decision criterion sometimes changed during an experiment, e.g. from comfortable loudness to sound colour.

5 Discussion

Test-retest

The measurement error (i.e., the test-retest standard deviation) for set A was 5.0 and 5.8 dB for $Gain_{low}$ and $Gain_{high}$, respectively. This test-retest is slightly larger than that from previously research that also optimized $Gain_{low}$/$Gain_{high}$ (<5 dB for 80% of subjects Kuk and Pape ,1992). One of the reasons can be the fact that their measurement of the perceptual differences was more accurate. In the previous research, each comparison was done three times, and the sound sample that was chosen 2 (or 3) times was used. In our experiment this repetition was replaced by an 'odd-ball' paradigm, in order to reduce the duration of the measurements.

Another reason for our somewhat larger measurement error could be that we did not use a small fixed grid (e.g., 3x3 grid points with 5 dB steps). Our test-retest standard deviation is about the same size as their grid distance. Additionally we used an adaptive step size that was allowed to be much larger than the grid distance of the previous experiments. Our approach might be more suitable for hearing aid parameters that span a large range of acceptable values.

Our experiment also included a compression parameter. Due to randomization of the presentation order of set A, B, and C, the subjects could not predict which type of perceptual differences were relevant for the paired comparison. This uncertainty of the features that differ between the samples could have increase the step size (jnd) and therefore the measurement error.

Moreover, the three-alternative forced-choice paradigm is known to work well for short signals (Lijzenga, 1997), but it is perhaps less suitable for longer signals. A fluctuating auditory attention during the presentation of the stimuli can lead to larger step sizes. This is in line with the comments from the participants.

The measurement error of sets B and C are not directly comparable to previous research, since the optimization experiments used different optimization parameters. However, the statistical analysis showed no significant difference between the measurement error for the gain of all three sets.

The average measurement error for compression ratio (1.6 dB/dB) is large with respect to values that are useful for clinical practice. However, in the current experiment the speech signal was presented at a comfortable listening level with a limited dynamic range of the speech levels. For this situation, where the speech was not too soft or too loud and the dynamic range of the input sounds is small, the influence of compression ratio on subjective sound quality is generally relatively minor (Neuman et al., 1998).

End points

For sets B and C there were no significant differences in average end values between NH and HI. This was to be expected since the initial fitting took the hearing loss of the subjects into account.

In contrast, for set A and HI, the average end-points differed from those for set B and C. This might be related to the perceptual effect of the optimization parameters. For set A, an increase in overall gain can only be achieved by accepting more gain for the low frequencies (‘boomier’) and for the high frequencies (‘sharper’) separately.

For all three sets, the average end-points had more gain than that prescribed by NAL. This is most likely related to our choice of input level (65 dB SPL) rather than to the amount of gain rather then the prescription of the NAL fitting rule.

In about half of the measurements the subjects could distinguish the starting point from the end point. The preferred point was the point with the highest gain setting. This, again, indicates that the subjects preferred a higher speech input-level than 65 dB.

Efficiency

The measurement time was on average about 5 minutes for a single two-dimensional measurement (set A and B). This is considerably faster than the previously used procedure (for instance the procedure of Kuk and Pape, 1992, took more than 20 minutes), and is acceptable for clinical use. The time needed for the three-dimensional measurement (set C) was about 12 minutes, and this is probably about the maximally available clinical testing time for an individual fitting procedure. The addition of an extra dimension increased the testing time more than proportionally. This indicates that the procedure will most likely not be clinically applicable for more than three dimensions. However, once started, the current procedure can be done by the user without assistance from the clinician. This could allow for longer fitting sessions, and thus for more dimensions.

Improvements

The population sample was small, and the inclusion of more subjects might lead to a clearer average end-point. However, an adaptive optimization procedure is meant for individualized fitting. The test-retest results indicate that the procedure needs to be improved in order to be clinically useful for individuals.

A possible improvement would be to adapt the user interface. The sequential presentation of the three sound samples can be changed to running speech for which the user can choose when the new processing is turned on. This will make it easier to discriminate the deviant signal (‘odd ball’). Another option would be to split the dual task (discrimination and judgment) into two separate tasks: first measure the just noticeable perceptual differences and use this as a basis for the step size.

6 Dissemination and Exploitation

The current work did not yet lead to an exploitable end product. The main reason is that the procedures carry still too much uncertainty. To reduce this, the procedures become too time-consuming to be clinically applicable. The second reason is that it is hard to exploit a possible outcome of this work. If we find a successful approach, this can easily be duplicated by others without paying rights for the generally available knowledge that is used to find the optimal solution.

The work reported in this deliverable has been presented at the ISAAR meeting in 2007:

Houben, A.C.H., Dreschler, W.A., Interactive fitting of hearing aids, ISAAR conference: Auditory Signal Processing in Hearing Impaired Listeners, Helsingør, Denmark, August 29 - 31, 2007

7 Ethical issues

The study described in this deliverable has been approved by the medical ethical committee of the Academic Medical Centre in Amsterdam (reference: MEC 05/127 # 05.17.0934, dated August 3rd 2005).

8 Conclusions

Although the measurement time was clinically acceptable, the applicability of the novel procedure is limited due to the relatively large measurement error. This error was slightly larger than previously described in literature. This could be related to our trade-off of measurement time to accuracy by using a different measurement paradigm, and by the removal of a fixed measurement grid.

Suggestions for improvement have been described.

9 Literature

Braida, L. D., Durlach, N. I., Lippmann, R. P., Hicks, B. L., Rabinowitz, W. M., and Reed, C. M. (1979), "Hearing aids–a review of past research on linear amplification, amplitude compression, and frequency lowering," Journal of the American Speech and Hearing Association Monographs 19, 1–114.

Dillon, H. (2001), *Hearing Aids*, Thieme Medical Publishers, New York.

Dirks, D.D., Ahlstrom, J., Noffsinger, D. (1993), "Preferred frequency response for two- and three-channel amplification systems, Journal of Rehabilitation Research and Development", 30, 305–317.

Franck, B.A.M., Dreschler, W.A. and Lyzenga, J. (2004), "Methodological aspects of an adaptive multidirectional pattern search to optimize speech perception using three hearing-aid algorithms", The Journal of the Acoustical Society of America 116(6), 3620–3628.

Franck, B.A., Boymans, M., Dreschler, W.A. (2007), "Interactive fitting of multiple algorithms implemented in the same digital hearing aid", International Journal of Audiology, 46, 388–397.

Kuk, F.K. (1994), "A screening procedure for modified Simplex in frequency-gain response selection", Ear Hear. 15 (1), 62-70.

Kuk, F.K. and Lau. C. (1996), "Comparison of preferred frequency gain settings obtained with category rating and modified simplex procedure", Journal of the American Academy of Audiology, 7, 322–331.

Kuk, F.K. and Pape, N.M.C. (1992), "The reliability of a modified simplex procedure in hearing aid frequency-response selection", Journal of Speech and Hearing Research, 35, 418–429.

Levitt, H., Neuman, A.C., Mills, R., Schwander, T. (1986), "A digital master hearing aid, Journal of Rehabilitation Research and Development", 23, 79–87.

Lijzenga, J. (1997), *Discrimination of simplified vowel spectra*, thesis Rijksuniversiteit Groningen, The Netherlands.

Neuman, A.C., Levitt, H., Mills, R., Schwander, T. (1987), "An evaluation of three adaptive hearing aid selection strategies", The Journal of the Acoustical Society of America 82(6), 1967–1976.

Neuman, A. C., Bakke, M. H., Mackersie, C., Hellman, S., and Levitt, H. (1998), "The effect of compression ratio and release time on the categorical rating of sound quality", Journal of the Acoustical Society of America 103, 2273-2281.

Oppenheim, A. V. and Schafer, R. W. (1989), *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs.

Preminger, J.E., Neuman, A.C., Bakke, M.H., Walters, D., Levitt, H. (2000), “An examination of the practicality of the simplex procedure”, Ear and Hearing, 21, 177–193.

Smith, S. (1997), *The Scientist and Engineer’s Guide to Digital Signal Processing*, California Technical Publishing, San Diego.

Spendley, W. Hext, G.R., and Himsworth, F.R. (1962), “Sequential application of simplex designs in optimisation and evolutionary operation”. Technometrics 4, 441–461.

The Mathworks, Inc. (2005),"Matlab, R14", Natick, MA, USA

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,” Journal of the Acoustical Society of America 107, 1671–1684.