

# Interactive fitting of hearing aids

ROLPH HOUBEN AND WOUTER DRESCHLER

*Clinical and Experimental Audiology, Academic Medical Center, Amsterdam, The Netherlands*

This paper describes a pilot experiment with an experimental procedure for individualized fine-tuning of hearing-aid fittings. The preference of a hearing-aid user is determined by a comparison between different sound samples. This preference is fed to a mathematical optimization system (a modified simplex procedure) that efficiently chooses the next comparison. The system should eventually converge to the optimal hearing-aid setting. In order to improve the input of the optimization algorithm, a two-step paradigm was developed. The listeners were presented with three sound samples. First, they had to determine which sample differed from the other two (three-alternative forced choice with an ‘odd-ball paradigm’). This step was included to increase the reliability of the second step in which the listeners had to judge which of the different samples sounded better (two-alternative unforced choice). The new paradigm was tested with speech in noise (SNR=+5 dB) for three parameters of a simulated four-channel compression hearing aid. The results indicate that the new paradigm was much less time-consuming than those previously described in literature. However, for both normal hearing and hearing-impaired subjects, the minimally perceived differences between hearing-aid settings were large compared to the values relevant for clinical practice. This indicates that either the subjects did not hear the differences between the samples or that the task was too difficult. Based on user comments, it is likely that the unfavourable results were largely caused by the odd-ball paradigm. A slight adaptation of the procedure might lead to better results in the future.

## INTRODUCTION

Levitt and co-workers (Levitt *et al.*, 1986) were the first to use a mathematical optimization algorithm for the fitting of hearing aids. They adapted the simplex algorithm for use with paired comparisons of subjective quality judgments. This work was followed by numerous other studies, notably from Kuk and colleagues (for instance Kuk and Pape, 1992), from Dirks *et al.*, (1993), and from co-workers of Levitt (Preminger *et al.*, 2000). Most previous research into the modified simplex procedure was two-dimensional: the gain in a low and high-frequency channel was varied independently. To improve the reliability, all paired comparisons were done three times. Overall, the results for a two-dimensional optimization algorithm indicated that the procedure was very efficient and that the test-retest variability was small (with a grid size of about 5 dB). About 80% of the listeners selected in retest nearly identical settings as in the first measurement (Kuk and Pape, 1992). Franck *et al.* (2004) used three dimensions: noise reduction, spectral enhancement, and spectral lift. In another experiment Franck

*et al.* (2007) used noise reduction, temporal signal enhancement and amplitude compression (in which the optimization parameter was the number of frequency channels). To keep experimental time limited, they used only one paired comparison (in stead of three). This increased the test-retest variability, but that was subsequently improved by the use of an adaptive step-size.

From the above mentioned previous research we can conclude that interactive fitting with an adaptive procedure seems viable and that it may become an important tool in hearing-aid fitting. But test duration and test accuracy need to be improved.

## **MATERIALS AND METHODS**

### **Paradigm**

Based on the results of previous research we designed a new procedure. The main focus was on speeding up the user input procedure while improving the reliability. That is, we tried to extract accurate data from the subjects in an efficient way. In our procedure the original three paired-comparisons were replaced by a paradigm that automatically estimates the just-noticeable perceptual differences for the hearing-aid algorithms used. This was used to determine the step size of the optimization algorithm. A possible advantage is a more reliable change in step size which could increase the test-retest reliability. In addition, we removed the requirement that all possible parameter values are placed in a fixed grid. That is, we changed the set-up from categorical to semi-continuous. This has the advantage that the point distance (grid size) does not need to be determined beforehand.

Our paradigm consists of two steps. The first step is a discrimination task. In this three-alternative forced-choice test, one stimulus differs physically from the other two. The subject's task is to detect which stimulus differs from the other two ('odd-ball paradigm'). All three presentations are based on the same unprocessed sentence. In line with the experiments of Franck *et al.* (2004), the text of the sentence is shown on the screen. This is done to prevent the last presentation to be preferred due to increased intelligibility caused by repetition of the same sentence. If the subject issues an incorrect response, he/she is notified that a wrong stimulus was chosen and the second step will be skipped. If the response was correct, the user will be presented with the judgment task of the second step (without presenting the two stimuli again).

The user is then asked if he/she prefers the selected sound sample over the other two: "Do you prefer this sound sample? Yes, No, Equal" (two-alternative unforced-choice).

### **Optimization algorithm**

The optimization algorithm is based on the modified simplex algorithm such as used by Franck *et al.* (2004). The procedure can optimize several hearing-aid algorithms in a single measurement session. For this, each hearing-aid processing algorithm is treated as one dimension of the optimization procedure. Since a too small difference in parameter values can lead to too many repetitions (inefficient), the step size is adapted

automatically. If the answer from the first task (discrimination) was incorrect, the step size is increased (with a factor 1.7), and the procedure continues with the next dimension. If the correct answer was given, the judgment task is presented to the subject. The step size for the new comparison will then be reduced by a factor 1/1.7, and the procedure goes on to the next dimension. After two reversals in step size (from increase to decrease or vice versa) the step-size change factor will be decreased with 0.2 (from 1.7 to 1.5 to 1.3, with a minimum of 1.1).

### Experimental design

The hearing-aid fitting was optimized based on user preference: the user was asked which sound sample he/she preferred; no further direction was given. We used three different sets of hearing-aid parameters, see Table 1.

Set	Dimension 1	Dimension 2	Dimension 3
A	low-freq gain ( $gain_{low}$ )	high-freq gain ( $gain_{high}$ )	-
B	broad-band gain ( $gain_{overall}$ )	slope of the gain ( $gain_{slope}$ )	-
C	low-freq gain ( $gain_{low}$ )	high-freq gain ( $gain_{high}$ )	compression ratio (CR)

**Table 1:** Experimental design: three sets of hearing-aid parameters.

The first two sets (A and B) are two-dimensional and they optimized the gain for linear amplification only. For set A this was done separately for the low frequencies ( $gain_{low}$ , 'bass') and for the high frequencies ( $gain_{high}$ , 'treble'). For set B, the broad-band gain ( $gain_{overall}$ , 'volume') was optimized together with the difference in gain between the high and the low frequencies ( $gain_{slope}$ , 'sound colour'). Set B can result in exactly the same amplification as set A, however, the perceptual effect of the 'route' to the end point is different. Set C was three dimensional and optimized the same gain parameters as set A, supplemented 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 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. 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 done twice. All measurements were conducted in a single session of about 2 hours.

### Materials and subjects

Speech materials consisted of the female speech of Dutch sentence material for measurement of the Speech Reception Threshold in noise, developed by Versfeld *et al.* (2000). A speech babble background noise was constructed from the materials by concatenating all (female) sentences and subsequently randomly placing ten sequences on top of each other. All measurements were conducted at a signal to noise ratio of +5 dB. Six normal hearing (NH) adult subjects participated in this study, as well as five

adults with sensorineural hearing-loss (HI). For the latter group, the pure-tone average hearing-loss at 1, 2, and 4 kHz was  $38 \pm 11$  dB HL. All subjects were native Dutch speakers. The average age of the NH participants was 32 ( $\pm 6$ ) year and of the HI subjects 58 ( $\pm 11$ ) year.

### Hearing-aid emulation

The signals were processed off-line using a personal computer with MATLAB (The Mathworks, 2005). All subjects (both NH and HI) listened through the emulated hearing-aid. The four frequency channels were constructed by elliptical band-filters (frequency range 90 Hz to 8 kHz, cross-over frequencies at 250, 707, and 2000 Hz). Compression (used for set C only) preceded the linear amplification stage (input-dependent compression or AGC-i). The signal was compressed independently in all four channels. For a description of the compression algorithm, see Houben (2006). Compression threshold was chosen at 48, 47, 38, and 40 dB SPL for the four channels respectively. These levels correspond to the threshold levels prescribed by NAL-NL1. The attack and release times were 4 and 40 ms, respectively.

The hearing-aid gain was optimized with two parameters. For sets A and C these two parameters ( $\text{gain}_{\text{low}}$  and  $\text{gain}_{\text{high}}$ ) corresponded directly to the frequency channels 1 (low) and 4 (high). The gain in channels 2 and 3 was linearly interpolated on a dB scale. For set B, the first dimension ( $\text{gain}_{\text{overall}}$ ) corresponded directly to the gain in all four channels. For the second dimension ( $\text{gain}_{\text{slope}}$ ) a slope in the gain was realized by an addition of half of the gain difference to the highest frequency channel (4) and subtraction of half the gain difference from the gain in the lowest channel (1), with linear interpolation (on a dB scale) for channels 2 and 3. Finally, the outputs of all four channels were summed. The signals were 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. In order to present the stimuli with the prescribed insertion gain, a filter is needed that mimics a free-field signal when a headphone is used. We therefore measured the frequency response on a B&K Head and Torso simulator (HAT) for both the free field and the headphone condition, and we calculated the free field to headphone transfer function. The digital stimuli were converted to the analogue domain using a 'RME Fireface 800' sound card. The output of the card was connected to a Behringer Audio Interactive Dynamics Processor, model 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.

## RESULTS

### Test-retest

The test-retest standard deviation (i.e., the measurement error) of the duplicate measurements is shown in Table 2. The results for set B ( $\text{gain}_{\text{overall}}/\text{gain}_{\text{slope}}$ ) were converted to  $\text{gain}_{\text{low}}/\text{gain}_{\text{high}}$ . An analysis of variance (ANOVA) showed no significant difference between NH and HI ( $p=0.3$ ). Therefore, the data was pooled over NH and HI. The ANOVA showed that the test-retest data for  $\text{gain}_{\text{low}}$  and  $\text{gain}_{\text{high}}$  did not differ

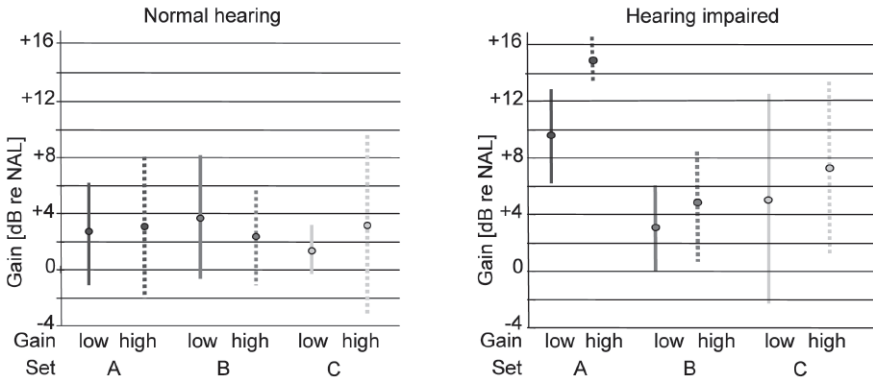
significantly between the three sets ( $p=0.2$ ).

	$\sigma$ gain <sub>low</sub> (dB)	$\sigma$ gain <sub>high</sub> (dB)	$\sigma$ CR (dB/dB)
Set A gain <sub>low</sub> /gain <sub>high</sub>	5.0	5.8	
Set B gain <sub>overall</sub> /gain <sub>slope</sub>	3.3	6.7	
Set C gain <sub>low</sub> /gain <sub>high</sub> /CR	9.8	4.0	1.6

**Table 2:** Test-retest standard deviations. The data for set B (gain<sub>overall</sub>/gain<sub>slope</sub>) is converted to gain<sub>low</sub>/gain<sub>high</sub>. CR means compression ratio.

### Average end-points

Fig. 1 shows the averaged end-points and inter-individual standard deviations for the NH and the HI subjects. The figure shows that only the data for set A of the hearing impaired group differed from the other data. The end points for set A for HI were significantly higher than the end points for set A for NH ( $p<0.01$ ). Additionally, the end points for set A for HI were significantly higher than for set B for HI ( $p<0.01$ ). The results for set B and C did not differ between NH and HI ( $p=0.1$ ,  $p=0.2$  for B and C, respectively). The end points for compression ratio (not shown in the figure) were  $2.0\pm 1.7$  dB/dB for NH and  $1.7\pm 0.6$  dB/dB for HI. These did not differ significantly ( $p=0.6$ ).

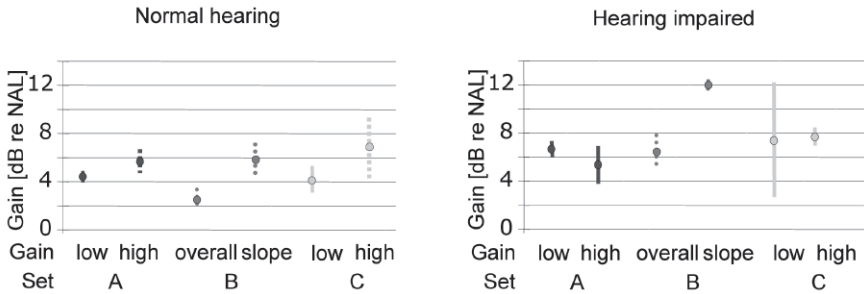


**Fig. 1:** Average end-points for gain. The left and right panel show data for normal hearing and hearing-impaired subjects, respectively. End point data for set B (gain<sub>overall</sub>/gain<sub>slope</sub>) was converted to gain<sub>low</sub>/gain<sub>high</sub>. Error bars represent one standard deviation.

### Step size

Fig. 2 shows the average step-size and the inter-individual standard deviation. For the gain, the average step-size was larger for the HI subjects than for the NH subjects. This was significant for all three sets ( $p<0.05$  for all sets). The step size for compression ratio was  $5.4\pm 3.1$  dB/dB for NH and  $4.4\pm 3.1$  dB/dB for HI and did not differ significantly between these groups ( $p=0.9$ ). The step size of set A and set B cannot be compared directly to each other since the overall gain for set B has more influence on

the stimuli because all frequencies get amplified, whereas for set A only the low or the high frequencies. For NH subjects the step size for gainoverall (set B) was significantly smaller than the step size for gainlow and gainhigh from set A (1.9 dB and 3.1 dB, respectively and  $p < 0.01$  for both). For the HI subjects the differences in step size between the sets were not significant.



**Fig. 2:** Step size for gain. Averaged over normal hearing (left panel) and hearing-impaired subjects (right panel). Error bars represent one standard deviation.

**Comparison to initial fitting**

In only 47% of the NH subjects the difference between the starting points and the end points could be discriminated. 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. In the HI group, 65% could discriminate the differences between the starting points and the end points. 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.

**Session efficiency**

The average duration of the sessions did not differ between the normal hearing and the hearing-impaired participants ( $p > 0.4$  for all three sets), see Table 3.

	Normal hearing	Hearing impaired		
	Reversals	Time (min)	Reversals	Time (min)
Set A gain <sub>low</sub> /gain <sub>high</sub>	4/3	5±2	5/5	6±3
Set B gain <sub>overall</sub> /gain <sub>slope</sub>	3/4	5±2	3/5	5±2
Set C gain <sub>low</sub> /gain <sub>high</sub> /CR	5/5/8	11±5	5/5/4	13±5

**Table 3:** Number of reversals and session durations.

Session duration did not differ significantly between sets A and B ( $p = 0.6$ ). Due to the extra dimension, set C took considerably longer than A and B. Even if the time per dimension is compared (2.7, 2.5, 4.0 min, for set A, B, and C, respectively), set C still took longer than A ( $p < 0.05$ ) and B ( $p < 0.01$ ). This is caused by the stop criterion that demanded that each of the dimensions had minimally three reversions. Until this cri-

terion was met, all dimensions were still tested, irrespective the number of reversions. Table 3 gives the average number of reversions for each set. The addition of an extra dimension therefore increased testing time more than proportionally.

## 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 in the previous research each comparison was made three times. In our experiment this repetition was replaced by an ‘odd-ball’ paradigm, in order to reduce the duration of the measurements. Another reason for the larger measurement error could be that we did not use a small fixed grid (e.g., 3x3 grid points with 5 dB steps). The test-retest standard deviation in this experiment was about the same size as the grid distance used by Kuk *et al.* (1992). 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.

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 in features that differ between the samples (for instance  $gain_{low}$  or CR) could have increased the adaptive step-size, and therefore the test-retest standard deviation.

Moreover, the three-alternative forced-choice paradigm is known to work well for short signals (Lijzenga, 1997), but it is perhaps less suitable for our longer signals (about 4 s). A fluctuating auditory attention during the presentation of the stimuli can cause the subjects to miss an otherwise discernible difference between the sound samples. This leads to larger step-sizes. This is in line with comments received from the participants.

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 no significant differences in average end values were found between NH and HI. This was to be expected since the initial fitting took the hearing loss of the subjects into account. The average end-points are gain values relative to the NAL prescribed gain. 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 validity of the NAL prescription 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.

The population sample included in this pilot experiment 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 before it can be applied clinically for individual hearing-aid fittings.

### **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 previously used procedures (for instance the procedure of Kuk and Pape (1992) took more than 20 minutes), and it may be expected to be acceptable for clinical use. The time needed for a 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.

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.

### **CONCLUSIONS**

Although the novel procedure is considerably faster than previous methods, the clinical applicability of the novel procedure is still limited due to large measurement errors. These errors were 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.

### **ACKNOWLEDGMENTS**

The authors wish to thank N.J. Versfeld for his valuable suggestions. 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.

## REFERENCES

- Dillon, H. (2001). "Hearing Aids," Thieme Medical Publishers, New York.
- Dirks, D. D., Ahlstrom, and J., Noffsinger, D. (1993). "Preferred frequency response for two- and three-channel amplification systems, J. 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," J. Acous. Soc. Am. **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," Intern. J. of Audiol. **46**, 388–397.
- Houben, R. (2006). "The effect of amplitude compression on the perception of speech in noise by the hearing impaired," thesis Universiteit Utrecht, The Netherlands
- Kuk, F. K., and Pape, N. M. C. (1992). "The reliability of a modified simplex procedure in hearing aid frequency-response selection," J. of Speech and Hearing Research. **35**, 418–429.
- Levitt, H., Neuman, A. C., Mills, R., Schwander, T. (1986). "A digital master hearing aid," J. of Rehabilitation Research and Development. **23**, 79–87.
- Lijzenga, J. (1997). "Discrimination of simplified vowel spectra," thesis Rijksuniversiteit Groningen, The Netherlands.
- 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," J. Acous. Soc. Am. **103**, 2273–2281.
- 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.
- 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," J. Acous. Soc. Am. **107**, 1671–1684.

