Table 6 shows that certain band level differences co-vary with function words while others do not. This indicates that these band level differences are robust predictors of information content across individual talkers.

CONCLUSIONS AND NEXT STEPS

Talkers seem to use simple acoustic cues to encode specific parts of their speech as particularly information rich. It may not be surprising in itself that the talker helps the listener by marking important words acoustically. What we do find surprising is, however, the lack of technological utilization. We have not been able to identify any reports of speech transducing technology (be it telecommunication, hearing aids, or automatic speech recognition) exploiting the direct relation between simple physical properties and highly abstract linguistic content.

The authors are preparing a follow-up to the reported experiment using its results in an algorithm for prediction of information richness with extremely short time delay. The algorithm will be used for modulation of speech materials masking out low-content and high-content parts of the signal respectively. The manipulated signals will then be scored for intelligibility in a perception experiment. Hopefully, the results will pave the way for a new technology with a *flair* for speech.

REFERENCES

Black, A. W. and K. A. Lenzo (**2007**). “Building Synthetic Voices, Carnegie Mellon University” www.festvox.org/, www.cs.cmu.edu/~lenzo/.

Boersma, P. (**2001**). ”Praat, a system for doing phonetics by computer” Glot International, **5** (9/10), 341–345.

Dik, S. C. (**1997**). *The Theory of Functional Grammar, Part 1: The Structure of the Clause,* 2nd ed., Mouton de Gruyter, Berlin.

Grønnum, N. (**2009**). “A Danish phonetically annotated spontaneous speech corpus (DanPASS)” Speech Communication **51**, 594–603.

Henrichsen, P. J. (**2011**). **“**Fishing in a speech stream, angling for a lexicon” Proceedings of 18th Nordic conference of computational linguistics NODALIDA, Pedersen, B. S., Nespore, G. and Skadina, I. (Eds.), pp. 90–97.

Jurafsky, D. and Martin, J. H. (**2009**). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edition. Prentice-Hall.

Klammer, T., Schulz, M. R. and Volpe, A. D. (**2009**). *Analyzing English Grammar* 6th edition, Longman.

Moore, B. J. C. M. (**2003**). *An introduction to the psychology of Hearing*, 5th edition, Academic Press, pp. 72-75.

Uneson, M. and P. J. Henrichsen (**2011**). ”Expanding a Corpus of Closed-World Descriptions by Semantic Unit Selection” accepted for the Proceedings of Computational Linguistic Applications, Warsaw, October 2011

The influence of noise type on the preferred setting of a noise reduction algorithm

ROLPH HOUBEN[1], TJEERD M.H. DIJKSTRA[2], AND WOUTER A. DRESCHLER[1]

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

[2] *Institute for Computing and Information Sciences, Radboud University Nijmegen*

Despite the frequent application of noise reduction in hearing aids, there is little research on user preference for different settings of noise reduction. We therefore measured individual preference for noise reduction strength for speech that was embedded in background noise. In a laboratory experiment, three types of noise (speech shaped stationary noise, party babble, and traffic noise) were processed with two single-channel noise reduction algorithms. Ten normal-hearing and seven hearing-impaired subjects participated. The preference for strength of noise reduction differed between the noise types and this was consistent for the two different noise reduction algorithms. The inter-individual spread between hearing-impaired listeners was as large as between normal-hearing listeners and as a consequence we found no systematic differences between the groups. These results support earlier findings that an individual tuning of noise reduction parameters is important. Furthermore, the results suggest that it could be beneficial to adaptively change the setting of noise reduction in a hearing aid, depending on the type of background noise.

INTRODUCTION

For many hearing-aid users their aid does not work well in noisy environments. In an attempt to improve listening to speech in a noisy background, most modern hearing aids have a noise reduction algorithm (NR). There are many different noise reduction implementations and strategies (Hoetink *et al*., 2009) and these processing differences can lead to differences in the sound perception by the user (Brons *et al*., 2011). Besides differences between NRs, there can also be differences in subjective preference between users. The question is whether the factory default noise reduction settings, geared to the average end user, can be improved by individualization. Unfortunately, previous research on optimizing noise reduction parameters for individual listeners is scarce since most published research is focused on the comparison between distinct algorithms (Bentler *et al*., 2008; Loizou and Kim, 2011) or only compares *on* versus *off* (Bentler *et al*., 2008), and does not focus on optimization of a single algorithm. In a previous investigation (Houben *et al*., 2011) we varied the strength of a spectral subtraction noise reduction algorithm and

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

measured which strength was preferred by the participants. The ten normal-hearing participants differed significantly in the strength that they preferred, suggesting that users might benefit from individualization of noise reduction.

To further determine if individualization of NR is viable we need to know if the found differences between listeners also hold up in different background noises and with different noise reduction algorithms. Moreover, for application in hearing aids, we need to establish if there is a difference in preference between normally hearing and hearing-impaired participants. Here, we expand upon the previous research by investigating if the individual preference for the strength of a spectral subtraction noise reduction algorithm differs between: A. normal-hearing and hearing-impaired listeners, B. different background noises, C. different noise reduction algorithms.

METHODS

Experimental Design

The prime variable under investigation was the maximum reduction of sound energy (G) by the noise reduction algorithm. With this variable we could manipulate noise reduction strength. We chose G because it strongly influences the trade-off between the amount of residual noise and unwanted distortions in the speech signal. A higher value of G corresponds to less residual noise, but also goes along with a higher degree of distortion. The settings for the two NR algorithms were chosen in such a way that the values were evenly spread across the range of settings relevant for clinical use. This led to the following conditions: a reference condition with no processing (i.e. G=0 dB), and for NR1 G= 4, 9, 11, and 19 dB, and for NR2 G= 4, 8, 12, and 16 dB). Within each noise and each algorithm, every G value was compared to all other values (stimuli were not compared to themselves), and each run was done twice in a single session. It is important to note that the retests were not exactly identical as the speech material was different for each comparison. Preference was measured with paired comparisons (a two alternative forced choice paradigm). Each stimulus pair consisted of one sentence that was processed with two different values of G. For stimulus pairs with stimuli of similar quality, subjects have been shown to have a bias for the second stimulus (Wickelmaier and Choisel, 2006; Arehart *et al.*, 2007). To minimize this bias the text of the sentence was shown on the screen so that the subjects knew the content of the sentence beforehand. Subjects could listen to both alternatives as often as they liked. The subjects' task was to choose between the two alternatives based on the question: "Imagine that you will have to listen to these signals all day. Which sound would you prefer for prolonged listening?".

Subjects

Ten normal hearing (NH, average age 26±5 Yrs) and seven hearing-impaired (HI, average age 43±18) subjects participated in this study. All subjects had Dutch as their native language, and they were naïve participants. All normal-hearing subjects had hearing thresholds of 15 dB or better for each audiometric frequency from 125 Hz to 8 kHz. All hearing impaired subjects were experienced hearing-aid wearers

(>1 yr) and they had symmetrical hearing loss. Their average hearing loss is shown in Figure 1.

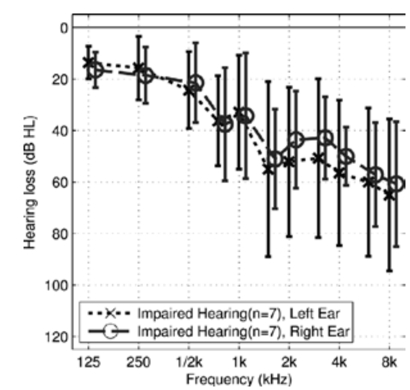

Fig. 1: Average hearing loss of the seven participants with impaired hearing. Mean values are shown and error bars denote the inter-subject standard deviation.

Stimuli

Speech (sentences from the female corpus of Versfeld *et al*., 2000) was embedded in three different background noises: unmodulated speech shaped noise (Dreschler *et al*., 2001, track 1), party babble (Bjerg and Larsen, 2006, track 39), and traffic noise (Bjerg and Larsen, 2006, track 7, selection of a passing car). Prior to processing, the speech and noise were mixed at +5 dB(A). We chose a random starting point of the noise for each stimulus (unfrozen noise). For the normal-hearing participants the presentation level of the stimuli was 68 dB(A) for each ear. For the hearing-impaired participants, we amplified the sound signals according to the NAL-RP fitting rule (Dillon, 2001). All stimuli were presented bilaterally and the left and right sound signals received the same amplification, based on the ear with the smallest hearing loss. This approach was possible because the asymmetry in the hearing loss was small. The speech-in-noise was processed by two noise reduction algorithms, implemented in Matlab. The first algorithm (NR1) is a modulation-based spectral-subtraction noise reduction algorithm that was used before to measure individual preference for the strength of noise reduction (Houben *et al*., 2011). The algorithm has low-complexity and low-latency and it is representative for the type of noise reduction applied in current generations of hearing aids. The second algorithm (NR2) was also a spectral-subtraction algorithm. However, NR2 is more complex because it augments modulation detection with a hidden Markov model of speech (Zhao *et al*., 2008). To allow the noise reduction algorithms to settle in their transient state, the first ten seconds of all stimuli were discarded. Participants listened via headphones (Sennheiser HDA200) to the stimuli that were generated in Matlab and that were put through a headphone buffer (Tucker-Davis Technologies HB6).

Statistical Analyses

The raw dichotomous data were analysed with a model that we developed for the analysis of paired comparison data with a continuous experimental variable (Houben *et al*., 2011). Briefly, the model assumes a trade-off between speech distortion and residual background noise. It models this trade-off for individual listeners by combining a quadratic utility model with logistic regression. We will refer to this model as the QUL model. With the QUL model we can obtain the value of G that corresponds to the participant's highest preference. Confidence intervals around the individually preferred value of G were obtained by bootstrapping the models' residuals (n=1000). We applied the QUL model to the data of all 17 subjects to calculate for each subject the preferred G per noise type (3 levels) and per algorithm (2 levels), leading to 102 values of preferred G. From this data, statistical significances were inferred with repeated measures analysis of variance in Matlab. For subjects 1 through 4 no data is available for party babble.

RESULTS

The QUL model fit the data well. The p-values indicated that the model did not deviate significantly from a model that fits the data perfectly (a saturated model). The p-values were calculated with a X^2-test on the model deviance (df=17) and all p were larger than 0.2 with a median of 0.93. To further assess the goodness of fit we calculated the positive classification rate that represents the number of responses that were correctly described by the model. Classification rates can range between 0 and 100% and 100% indicates a perfect score. For our data, the positive classification rate had a median of 90% and ranged from 60% to 100%, indicating good model fits. QUL estimates for the individually preferred values of G are shown in Figure 1 (grouped data is shown in Figure 2).

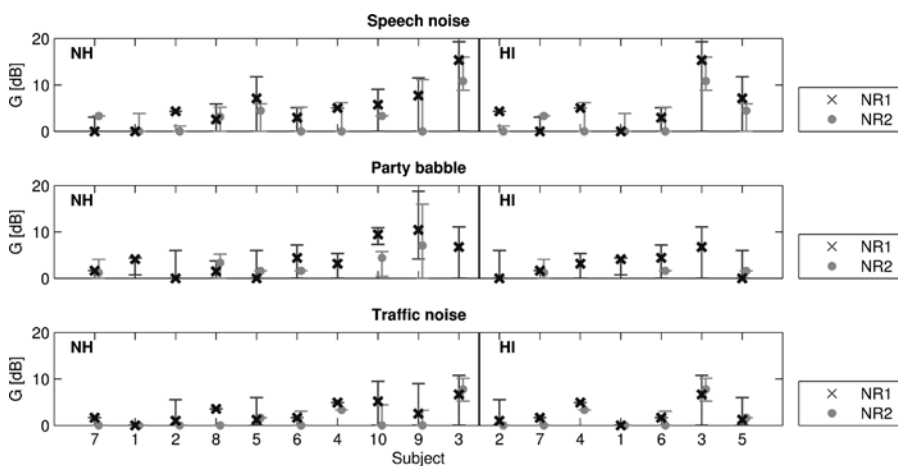

Fig. 1: Preferred value of G for each individual subject. Error bars denote within-subject 95% confidence. Each panel shows data for a noise type, and subjects are ordered per subject group, based on their average value of G.

The preferred values of G that were obtained with the QUL models were analysed with a repeated measures ANOVA[1]. "Subject" was treated as random effect and the rest of the variables were entered as fixed effects. Participants were categorised as either normal hearing or hearing impaired and this hearing loss classifier was nested in "subject". The ANOVA results are shown in Table 1. Data for the significant main effects are shown in Figure 2.

Effect	Degrees of freedom	F	p
Subject	16	3.4	0.002
Hearing loss	1	0.5	0.82
NR	1	9.4	0.001
Noise type	2	25.1	<0.001
NR * Noise type	2	3.9	0.2

Table 1: ANOVA table for repeated measures analysis of variance on the individually preferred values for G.

The main effect of subject was significant, indicating that at least one subject differed from the rest. "Hearing loss" was not significant. The reason for this is two-fold. First, the difference in preference between the two groups was small. On average, the HI group preferred a G that was only 0.3 dB larger than that of the normal-hearing group. Second, the spread of preference between individuals within both groups was large (standard deviation was 3.6 dB and 4.0 dB for the normal hearing and the hearing-impaired group, respectively).Both main effects "noise type" and "algorithm" were significant, and their interaction was not. Post-hoc analysis of "noise type" showed that the preferred value of G for speech noise and party babble did not differ significantly (Bonferroni corrected p=0.5), and that for traffic noise the results differed significantly from both speech noise (Bonferroni corrected p<0.001, preferred G for traffic noise was 2.8 dB lower than for speech noise), and party babble (Bonferroni corrected p<0.01). Post-hoc analysis on "algorithm" showed that, on average, the subjects preferred lower G for NR2 than for NR1 (the difference was 2.1 dB).

[1] The distribution of G was skewed towards 0 dB. This violates the assumptions of ANOVA. However, ANOVA is robust against this violation and inspection of residuals showed no systematic errors. Additionally, all main results were verified and confirmed with Kruskall-Wallis rank based ANOVAs..

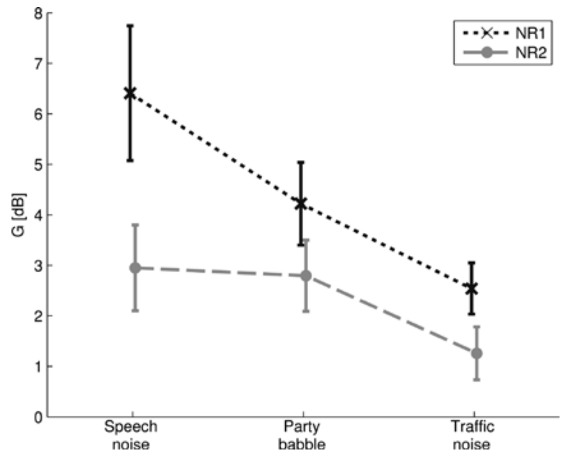

Fig. 2: Preferred value of G from the QUL model for the main effects of noise type and algorithm. Error bars denote between-subject standard error of the mean (n=17).

DISCUSSION

The results show that the preference for the strength of noise reduction does not differ between our group of normal hearing (n=10) and hearing-impaired (n=7) participants. The preferred strength of noise reduction does depend on the type of background noise. It is lower for speech in traffic noise than for both stationary speech shaped noise and party babble. Although the preference also differs between the two noise reduction algorithms, the effect of lower G for traffic noise is the same for both algorithms.

The lack of differences in preference between groups of normal hearing and hearing-impaired participants has been found before. In a large multi-center study Luts *et al.*, (2010) found no difference in preference of single channel NR over unprocessed for three groups of normal hearing (n=39), flat hearing loss (n=34), and sloping hearing loss (n=37). It seems that there is only limited effect of hearing loss on the preferred strength of noise reduction. Our interpretation is that inter-individual differences between both groups are in the same order of magnitude as within the groups. This, combined with the relatively small number of subjects typically involved in this kind of research, makes it hard to find small differences. In fact, with the knowledge gained here, we can calculate the a priori power for a hypothetical successive experiment. To reach a power of 0.9 we would need about 2000 subjects in both groups to detect significant differences between normal-hearing and hearing impaired participants. Rather than increasing statistical power by increasing the number of participants, it would probably be better to vary the types and severity of hearing loss.

The large spread in preference between subjects supports earlier findings that an individual tuning of the noise reduction parameters is important. In a previous experiment with NR1 we also found a large spread in preference between normal-hearing participants (Houben *et al*., 2011). For five out of ten subjects, the individual preference deviated significantly from that of the group average. In contrast to fitting the amplification of hearing aids, fitting of NR should be based on the personal preference rather than on hearing loss. Of course the question remains open if individualising noise reduction based on the subjective preference for the strength of noise reduction would ultimately increase user satisfaction.

The preferred value of G was lowest for traffic noise. This noise was the least stationary of the set of three noises. Spectral subtraction algorithms need to estimate the signal to noise ratio to classify sound frames into noise or speech, and this is much more difficult for background noises that change over time. Due to the inevitable errors in the noise spectrum estimation the algorithms will introduce speech distortion. For traffic noise, it is plausible that more distortion might have led to a preference for lower values of G. However, we had expected that this might be less for NR2 because this algorithm was designed to track both the noise and the speech, and thus to minimize spectrum estimation errors. Perhaps NR2 made more estimation errors, resulting in more distortion. Another explanation would be that the NR2 algorithm favoured gain reduction for speech over gain reduction for noise (Loizou and Kim, 2011), thereby influencing speech quality.

CONCLUSIONS

Individual listeners differ strongly in their preference. This confirms the previous finding that it might be beneficial to individualize noise reduction. The range of preferred values for the strength of noise reduction overlaps between groups of normal hearing and hearing-impaired participants. This suggests that individual fitting of NR should not be based on the hearing loss.

The preferred strength of noise reduction is lowest for our least stationary background noise (traffic noise). This pattern is the same for our two different spectral subtraction algorithms, but our subjects prefer less gain reduction for one of the two noise reduction algorithms. These results suggest that it could be beneficial to adaptively change the setting of noise reduction in a hearing aid, depending on the type of background noise.

AKNOWLEDGEMENTS

The authors wish to thank Koen Maaskant for his assistance in the measurements. Approval by the Medical Ethical Committee of AMC was obtained on 24 April 2008 (MEC 08/082). We gratefully acknowledge financial support from STW (Dutch Foundation for Science and Technology).

REFERENCES

Arehart, K. H., Kates, J. M., Anderson, M. C., and Harvey, L. O. (**2007**). "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners" J. Acoust. Soc. Am., **122** (2), 1150-1164.

Bentler, R., Wu, Y.-H., Kettel, J., and Hurtig, R. (**2008**). "Digital noise reduction: Outcomes from laboratory and field studies" Int. J. Audiol., **47** (8), 447-460.

Bjerg, A. P., and Larsen, J. N. (**2006**). *Recording of Natural Sounds for Hearing Aid Measurements and Fitting*, Ørsted, Denmark: Danish Technical University (DTU), Acoustic Technology.

Brons, I., Houben, R., and Dreschler, W. A. (**2011**). "Perceptual effects of noise reduction in hearing aids" in ISAAR - International Symposium on Auditory and Audiological Research.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment" International Collegium for Rehabilitative Audiology. *Audiology: official organ of the International Society of Audiology*, **40** (3), p.148.

Hoetink, A. E., Körössy, L., and Dreschler, W. A. (**2009**). "Classification of steady state gain reduction produced by amplitude modulation based noise reduction in digital hearing aids" International Journal of Audiology, **48** (7), 444–455.

Houben, R., Dijkstra, T. M. H., and Dreschler, W. A. (**2011**). "Differences in preference for noise reduction strength between individual listeners" in *130th Convention of the Audio Engineering Society*. London, pp. 1-9.

Loizou, P.C., and Kim, G. (**2011**). "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions" IEEE Transactions on Audio, Speech, and Language Processing, **19** (1), 47–56.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., *et al*. (**2010**). "Multicenter evaluation of signal enhancement algorithms for hearing aids" J. Acoust. Soc. Am., **127** (3), 1491-1505.

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.

Wickelmaier, F., and Choisel, S., (**2006**). "Modeling within-pair order effects in paired-comparison judgments" in *22nd Annual Meeting of the International Society for Psychophysics*. St. Albans, Hertfordshire, England, pp. 89–94.

Zhao, D. Y., Kleijn, W. B., Ypma, A., and de Vries, B., (**2008**). "Online Noise Estimation Using Stochastic-Gain HMM for Speech Enhancement" IEEE Transactions on Audio Speech and Language Processing, **16**(4), p.835.

Fast and intuitive methods for characterizing hearing loss

DIRK OETTING[1,2], BIRGER KOLLMEIER[2] AND STEPHAN D. EWERT[2]

[1] *Fraunhofer Institute for Digital Media Technology, Oldenburg, Germany*

[2] *Medizinische Physik, Carl-von-Ossietzky University Oldenburg, Germany*

The possibility of integrating hearing-aid technology like dynamic compression in current and future consumer audio devices raises the question how parameters of hearing supportive algorithms can be adjusted by the user to either compensate the individual hearing loss or to accommodate listening preferences. Here, three methods for measuring the auditory capacity based on loudness judgments and comparisons were evaluated. All methods used a simple interface and appear generally suited for integration in consumer audio electronics. Results of the suggested methods were compared to adaptive categorical loudness scaling [ACALOS, Brand and Hohmann, J. Acoust. Soc. Am. 112, 1597-1604 (2002)]. Gain prescriptions were derived for narrow-band loudness compensation based on the suggested methods, the clinically applicable ACALOS, and for NAL-NL2 [Keidser and Dillon, Hearing Care for Adults, 133-142 (2006)]. All loudness based procedures led to similar gains.

INTRODUCTION

Less than 20% of the mild-to-moderate hearing impaired population uses a hearing aid (Hougaard and Ruf, 2011) although the majority would benefit from hearing supportive technologies. To overcome stigma particularly for groups with mild hearing loss, hearing supportive technology could be integrated into communication and media devices (e.g., mobile phones, TVs, music players). These devices offer sufficient signal-processing power and capability to deliver high-fidelity sound quality, however, the problem of the individual fitting is still unsolved.

Standard audiometric measurements like hearing threshold appear not suited for integration in un-calibrated audio products, particularly if used in noisy environments. Moreover, like for hearing-aid fitting, knowledge about supra-threshold hearing deficits might be beneficial. Here, three fast and intuitively accessible methods, motivated by the adjustment un-calibrated video monitors based on video test images are suggested.

Instead of adjusting screen luminance levels to achieve well separable brightness impressions, sound levels were adjusted to match the well separable loudness categories “just audible”, “soft”, “comfortable”, and “loud”. The adjustment was either independent in three different frequency regions or additional loudness comparisons across frequency were included. The results were compared to laboratory measurements of the audiogram and adaptive categorical loudness scaling (ACALOS) for normal-hearing and hearing-impaired listeners.

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.