Individual factors in speech recognition with binaural multi-microphone noise reduction: Measurement and prediction

TOBIAS NEHER[1,3,*], JACOB ADERHOLD[1,3], DANIEL MARQUARDT[2,3], AND THOMAS BRAND[1,3]

[1] *Medizinische Physik, Oldenburg University, Oldenburg, Germany*

[2] *Signal Processing Group, Oldenburg University, Oldenburg, Germany*

[3] *Cluster of Excellence Hearing4all, Oldenburg, Germany*

Multi-microphone noise reduction algorithms give typically rise to large signal-to-noise ratio improvements, but they can also severely distort binaural information and thus compromise spatial hearing abilities. To address this problem Klasen *et al*. (2007) proposed an extension of the binaural multi-channel Wiener filter (MWF), which suppresses only part of the noise and, in this way, preserves some binaural information (MWF-N). The current study had three aims: (1) to assess aided speech recognition with MWF(-N) for a group of elderly hearing-impaired listeners, (2) to explore the impact of individual factors on their performance, and (3) to test if outcome can be predicted using a binaural speech intelligibility model. Sixteen hearing aid users took part in the study. Speech recognition was assessed using headphone simulations of a spatially complex speech-in-noise scenario. Individual factors were assessed using audiometric, psychoacoustic (binaural), and cognitive measures. Analyses showed clear benefits from both MWF and MWF-N and also suggested sensory and binaural influences on speech recognition. Model predictions were reasonably accurate for MWF but not MWF-N, suggesting a need for some model refinement concerning supra-threshold processing.

INTRODUCTION

Recently, hearing aids have become available that can wirelessly exchange audio signals across the user's head. This has opened up possibilities for 'binaural' signal processing, such as multi-microphone noise reduction, which can lead to large signal-to-noise ratio (SNR) improvements but also to distortions of binaural information (e.g., Doclo *et al.*, 2010). Because binaural information plays an important role for speech understanding in complex listening situations (e.g., Bronkhorst, 2015) and because hearing-aid users can differ substantially in terms of their residual binaural hearing abilities (e.g., Neher *et al.*, 2011; 2012), it is of interest to relate individual factors to benefit, or lack thereof, from this type of processing.

*Corresponding author: tobias.neher@uni-oldenburg.de

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

The purpose of the current study was to address this issue for one type of multi-microphone noise reduction: binaural multi-channel Wiener filtering (MWF). MWF perfectly preserves the binaural cues of the target signal, but undesirably changes the binaural cues of the noise to those of the target (e.g., Doclo *et al*., 2006). To address this problem, Klasen *et al*. (2007) proposed an extension of MWF, which suppresses only part of the noise and, in this way, retains some binaural information (MWF-N). For a group of young normal-hearing participants, van den Bogaert *et al*. (2008) found that MWF-N improved localisation while speech recognition was unaffected.

In the current study, we aimed to extend this research by pursuing the following three aims:

1. To assess aided speech recognition with MWF(-N) for elderly hearing aid users
2. To explore the influence of individual factors on their performance
3. To investigate if outcome can be predicted using a state-of-the-art binaural speech intelligibility model

METHODS

Speech stimuli

Our speech stimuli were based on recordings from the Oldenburg sentence test (Wagener *et al*., 1999). To simulate a realistic complex listening situation, we convolved these recordings with pairs of head-related impulse responses, which were measured in a reverberant cafeteria using a head-and-torso simulator equipped with two behind-the-ear hearing aid shells (Kayser *et al*., 2009). Specifically, we used the measurements made with the front and rear microphones of each hearing aid shell and a frontal source at a distance of 1 m from, and at the same height as, the head-and-torso simulator. For the interfering signal, we used a (spatially complex) recording made in the same cafeteria with the same setup during a busy lunch hour. During the measurements, we presented this signal at a nominal sound pressure level of 65 dB and mixed it with the target sentences, the level of which we adjusted to produce a given SNR.

MWF(-N) processing

The MWF(-N) processing we tested mimicked that of van den Bogaert *et al*. (2008). There were two main algorithmic parameters: μ and η. μ determines the strength of spectral post-filtering and thus trades off noise reduction against speech distortion. It was set to 1 here to result in standard MWF. η is a scaling factor between 0 and 1 that determines how much of the unprocessed input signal is mixed back into the noise-reduced output signal. For $\eta = 0$, nothing of the input is mixed back into the output, resulting in standard MWF with full noise suppression but no binaural cue preservation. For $\eta = 1$, the input is mixed completely into the output, resulting in full binaural cue preservation but no noise suppression. In the current study, we tested the three η-settings also tested by van den Bogaert *et al*. (2008): 0, 0.2, and 1. In the following, we will refer to these as the MWF, MWF-N, and reference

conditions. Furthermore, as in the study of van den Bogaert *et al*. we used a perfect voice activity detector (i.e., we assumed access to the clean speech signal).

To quantify the physical effects of MWF(-N) we estimated the resultant speech-weighted SNR improvement (ΔAI-SNR) as a function of the input SNR. As expected, ΔAI-SNR increased with higher input SNRs (see Fig. 1). Furthermore, ΔAI-SNR was up to 0.8 dB larger for MWF than for MWF-N. Figure 1 also shows the SNRs during the speech recognition measurements (see below). Across participants, ΔAI-SNR amounted to 2.2 dB (σ = 0.5 dB) for MWF-N and to 2.7 dB (σ = 0.7 dB) for MWF.

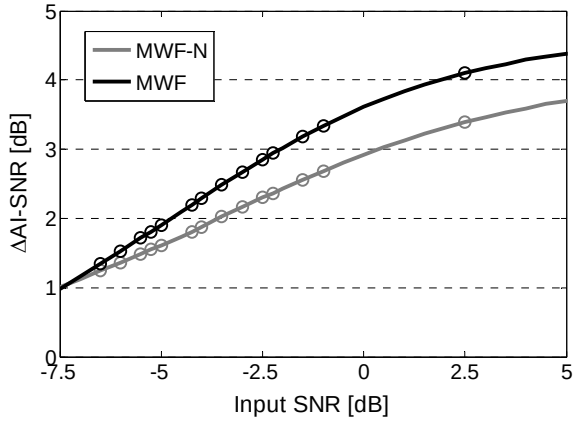

Fig. 1: ΔAI-SNR for MWF (black) and MWF-N (grey) as a function of input SNR. Circles denote individual test SNRs.

In addition, we estimated the interaural coherence (IAC) of our speech stimuli for the three processing conditions with the help of the auditory model of Dietz *et al*. (2011). The IAC can be interpreted as a measure of binaural complexity. As expected, binaural complexity decreased with MWF-N and especially MWF, i.e., the stimuli became increasingly interaurally correlated (see Fig. 2).

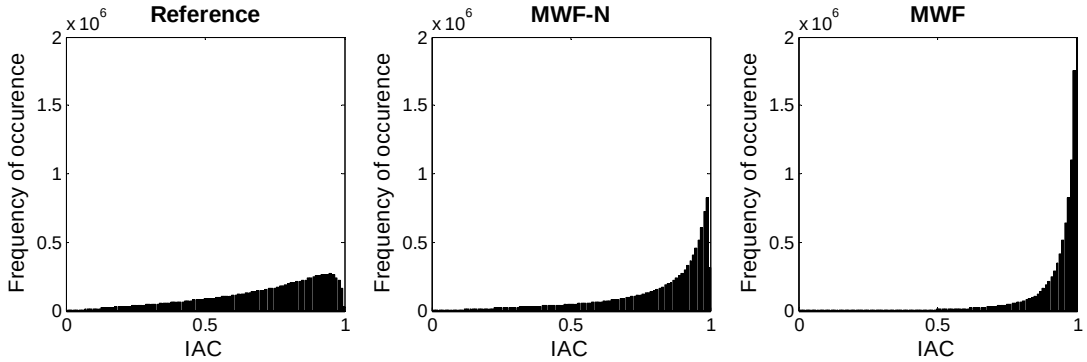

Fig. 2: Histograms of the estimated IAC for an example speech stimulus with an input SNR of −4 dB for the reference, MWF-N, and MWF conditions.

Participants and individual factors

Sixteen experienced hearing-aid users with symmetrical, gently sloping sensorineural hearing impairments participated in the experiment. Their mean age was 74 yr (range: 56-86 yr). Their mean pure-tone average hearing loss from 500 Hz to 4 kHz (PTA) was 46 dB HL (range: 38-53 dB HL), while from 125 Hz to 750 Hz (PTA_{LF}) it was 30 dB HL (range: 17-41 dB HL).

To characterise our participants' binaural hearing abilities we performed binaural masking level difference (BMLD) measurements (test and retest) at 500 Hz with a broadband noise masker. In addition, we performed interaural phase difference frequency range (IPD_{FR}) measurements (test and retest). These measurements mark the highest frequency for which a participant is still able to detect interaural phase changes of 180° in a sinusoidal stimulus (e.g., Neher *et al*., 2011). Furthermore, we administered the reading span test (RST; Carroll *et al*., 2015) to our participants to also determine their working memory capacity.

Speech recognition measurements

Because we were interested in *aided* speech recognition performance we spectrally shaped the speech stimuli in accordance with the NAL-RP prescription rule (Byrne *et al*., 1991). We started our measurements with three training runs and then determined the individual speech reception threshold (SRT_{ind}) for the reference condition. In all subsequent measurements, we then kept the SNR fixed at the SRT_{ind}. In this manner, we obtained speech recognition rates (in percent correct) for our three processing conditions.

Binaural speech intelligibility model

For the prediction of the participants' speech scores we used the binaural speech intelligibility model (BSIM) of Beutelmann *et al*. (2010). BSIM combines a multichannel equalization cancellation stage according to Durlach (1963) with the Speech Intelligibility Index (SII; ANSI, 1997). In the current study, we individualised BSIM based on the hearing thresholds of each participant and carried out the predictions based on the amplified speech stimuli. Furthermore, because we measured speech recognition rates at a fixed SNR for each processing condition (rather than one SRT per processing condition) we restricted the predictions to the computation of SII (rather than SRT) values and related these to the speech scores of our participants.

RESULTS

Individual factors

Analysis of the test-retest data showed that the BMLD and IPD_{FR} measurements were reliable (both $r > 0.7$, $p < 0.01$). Figure 3 provides an overview of the BMLD, IPD_{FR}, and RST data. Averaged across participants, the BMLD was 11.2 dB (range: 4-20 dB), while the IPD_{FR} was 770 Hz (range: 342-1196 Hz). In terms of RST performance, the participants were on average able to recall 37.4% of all target words (range: 28-52%). Altogether, the BMLD and IPD_{FR} data were in good

agreement with the literature, while the RST data exhibited less spread toward the 'poor' end (cf. Neher *et al*., 2011; Santurette and Dau, 2012).

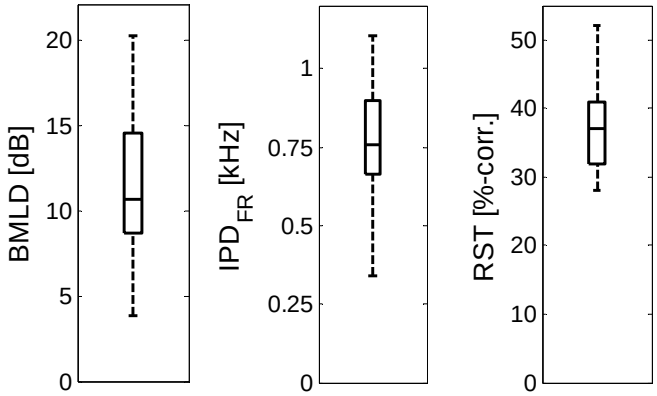

Fig. 3: Boxplots of the BMLD, IPD_{FR}, and RST data.

Speech recognition

Analysis of the SRT_{ind} data revealed a mean threshold of −3.7 dB SNR and a range of almost 10 dB (see Fig. 1). Figure 4 shows the speech scores for the three processing conditions. In the reference condition, participants could recognise 52.7% of the target speech. In the MWF-N and MWF conditions, they were able to recognise 80.5% and 78.4%, respectively. An analysis of variance with post hoc comparisons confirmed highly significant differences between the reference condition and MWF(-N), while MWF-N and MWF did not differ from each other.

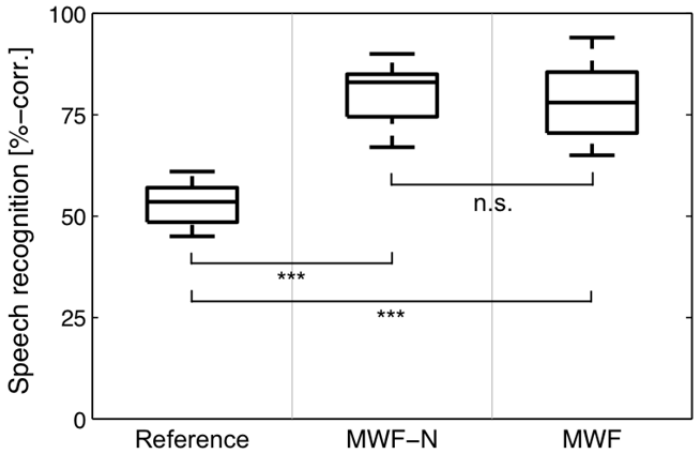

Fig. 4: Boxplots of the speech scores for the three processing conditions. *** $p < 0.001$, n.s. = non-significant.

Relations among speech outcomes and individual factors

To assess potential relations between SRT_{ind} and the individual factors we calculated a series of Pearson's *r* correlation coefficients. We observed correlations with age, PTA_{LF}, and BMLD (see Table 1). A regression model based on these three factors could account for 62.1% (adjusted R^2 = 53%) of the variance in the SRT_{ind} data ($p_{model} < 0.01$, $p_{age} > 0.05$, $p_{BMLD} < 0.05$, $p_{PTA\_LF} < 0.05$).

To assess potential relations between speech recognition (SR) with MWF(-N) and the individual factors, we calculated a series of partial correlation coefficients with ΔAI-SNR as control variable. In this manner, we controlled for the SNR-dependent effects of MWF(-N) related to speech audibility (see Fig. 1). As can be seen in Table 1, there was only a correlation between SR_{MWF-N} and RST.

	Age	**PTA_{LF}**	**BMLD**	**IPD_{FR}**	**RST**
SRT_{ind}	0.53*	0.64**	–0.61*	–0.38	–0.30
SR_{MWF-N}	–0.22	0.51	0.35	–0.23	0.62*
SR_{MWF}	0.53	0.51	–0.15	–0.18	–0.04

Table 1: Correlation coefficients for the speech scores from the reference (SRT_{ind}), MWF-N (SR_{MWF-N}), and MWF (SR_{MWF}) condition and the individual factors, with ΔAI-SNR partialled out in the case of $SR_{MWF(-N)}$. * $p < 0.05$, ** $p < 0.01$.

Outcome prediction

Figure 5 summarises the results of the outcome prediction. In each case, the abscissa shows ΔSII values [MWF(-N) – reference condition], while the ordinate shows corresponding Δ speech scores.

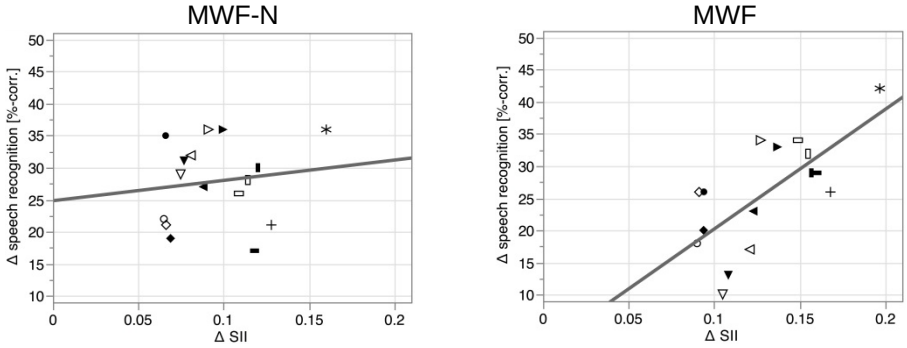

Fig. 5: Scatter plots of ΔSII values against Δ speech scores for MWF-N (left) and MWF (right). Symbols denote individual participants.

As can be seen, the accuracy was reasonably high for MWF ($r = 0.70$, $p < 0.01$) but not for MWF-N ($r = 0.14$). Partialling out ΔAI-SNR removed the correlation between the predicted and measured (relative) outcome for MWF ($r = 0.11$). Performing these predictions on short time segments of the speech stimuli and averaging across results (the "short-time BSIM"; cf. Beutelmann *et al*., 2010) did not improve the accuracy.

Altogether, these results suggest that, while BSIM is largely able to account for performance with MWF where the main effect is improved speech audibility, it fails to do so for MWF-N which due to its greater binaural complexity (see above) presumably invokes additional supra-threshold factors.

SUMMARY

With respect to the three aims outlined above, the results of the current study can be summarised as follows:

1. MWF(-N) led to significant improvements (on the order of 25%) in speech recognition performance. The benefit from MWF-N was comparable to that from MWF, despite the addition of background noise.
2. PTA_{LF} and BMLD were related to aided speech recognition in the reference condition, independent of the effects of age. For speech recognition with MWF-N, a relation with RST was found. For MWF, none of the individual factors tested here was predictive.
3. Outcome predictions were accurate for MWF, suggesting that BSIM could account for the main effect of improved speech audibility. In the case of MWF-N, outcome prediction was poor, suggesting that BSIM failed to account for certain supra-threshold effects.

Given that our study was limited to 16 participants who were tested at markedly different SNRs and that we used a perfect voice activity detector, the above findings must be regarded as preliminary. Future studies will investigate these issues in more detail, with particular emphasis on the role that individual factors play for aided outcome prediction.

ACKNOWLEDGEMENTS

This research was funded by the DFG Cluster of Excellence EXC 1077/1 "Hearing4all". We thank Christopher Hauth, Birger Kollmeier, and Simon Doclo for their support.

REFERENCES

ANSI (**1997**). "Methods for calculation of the speech intelligibility index (S3.5-1997)," New York, NY: American National Standards Institute.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). "Revision, extension, and evaluation of a binaural speech intelligibility model," J. Acoust. Soc. Am., **127**, 2479-2497.

Bronkhorst, A.W. (**2015**). "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," Atten. Percept. Psycho., **77**, 1465-1487.

Byrne, D., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe/profound hearing losses," in *The Vanderbilt Hearing Aid Report II*. Eds. G.A. Studebaker, F.H. Bess, and L.B. Beck (York Press, Parkton, NC), pp. 295-300.

Carroll, R., Meis, M., Schulte, M., Vormann M, Kießling J, and Meister H (**2015**). "Development of a German reading span test with dual task design for application in cognitive hearing research," Int. J. Audiol., **54**, 136-141.

Dietz, M., Ewert, S.D., and Hohmann, V. (**2011**). "Auditory model based direction estimation of concurrent speakers from binaural signals," Speech Comm., **53**, 592-605.

Doclo, S., Klasen, T.J., Wouters, J., Haykin, S., and Moonen, M. (**2006**). "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC), Paris, France, Sept. 12-14.

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. (**2010**). "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*. Eds. S. Haykin and K.J.R. Liu (Wiley-IEEE Press, Hoboken, NJ), pp. 269-302.

Durlach, N.I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am., **35**, 1206-1218.

Kayser, H., Ewert, S.D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (**2009**). "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," EURASIP J. Adv. Signal Process., **298605**, DOI: 10.1155/2009/298605.

Klasen, T.J., van den Bogaert, T., Moonen, M., and Wouters, J. (**2007**). "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," IEEE Trans. Signal Process., **55**, 1579-1585.

Neher, T., Laugesen, S., Jensen, N.S., and Kragelund, L. (**2011**). "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?" J. Acoust. Soc. Am., **130**, 1542-1558.

Neher, T., Lunner, T., Hopkins, K., and Moore, B.C.J. (**2012**). "Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners," J. Acoust. Soc. Am., **131**, 2561-2564.

Santurette, S. and Dau, T. (**2012**). "Relating binaural pitch perception to the individual listener's auditory profile," J. Acoust. Soc. Am., **131**, 2968-2986.

van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2008**). "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," J. Acoust. Soc. Am., **124**, 484-497.

Wagener, K., Brand, T., and Kollmeier, B. (**1999**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I-III: Design, Optimierung und Evaluation des Oldenburger Satztests", Zeitschrift für Audiologie, **38**, 4-15, 44-56, 86-95.