Evaluation of Binaural Noise Reduction Methods in Terms of Intelligibility and Perceived Localization

Andreas I. Koutrouvelis*, Richard C. Hendriks*, Richard Heusdens*, Steven van de Par†, Jesper Jensen‡§ and Meng Guo§
e-mails: {a.koutrouvelis, r.c.hendriks, r.heusdens}@tudelft.nl, and {jesj, megu}@oticon.com
*Circuits and Systems (CAS) Group, Delft University of Technology, the Netherlands
†Cluster of Excellence Hearing4all, University of Oldenburg, Germany
‡Dept. Electronic Systems, Aalborg University, Denmark, §Oticon A/S, Denmark

***Abstract*—In this paper, we perceptually evaluate two recently proposed binaural multi-microphone speech enhancement methods in terms of intelligibility improvement and binaural-cue preservation. We compare these two methods with the well-known binaural minimum variance distortionless response (BMVDR) method. More specifically, we measure the 50% speech reception threshold, and the localization error of all dominant point sources in three different acoustic scenes. The listening tests are divided into a parameter selection phase and a testing phase. The parameter selection phase is used to select the algorithms' parameters based on one acoustic scene. In the testing phase, the two methods are evaluated in two other acoustic scenes in order to examine their robustness. Both methods achieve significantly better intelligiblity compared to the unprocessed scene, and slightly worse intelligibility than the BMVDR method. However, unlike the BMVDR method which severely distorts the binaural cues of all interferers, the new methods achieve localization errors which are not significantly different compared to those of the unprocessed scene.**

***Index Terms*—Binaural beamforming, binaural cues, intelligibility, localization.**

I. Introduction

Binaural hearing-aid (HA) systems [1] consist of two wirelessly connected and collaborative HA devices with at least one microphone per device. In contrast, bilateral HA systems [2], [3] consist of independently working HAs. The binaural HAs can typically use a larger microphone array than bilateral HAs and, therefore, have more degrees of freedom for the beamformer. These degrees of freedom might be used to obtain a better noise reduction, or, to preserve the binaural cues of sound sources in the acoustic scene [3].

An important component in a binaural HA system is the binaural multi-microphone speech enhancement algorithm, which aims to enhance the intelligibility of the target speech signal, while at the same time to preserve the binaural cues of the acoustic scene after processing [3]. Typically, binaural multi-microphone speech enhancement methods show a trade-off between noise reduction and binaural-cue preservation. Existing binaural multi-microphone speech enhancement methods can be roughly categorized into two main groups: the spatial filtering methods (e.g., [4]–[8]) and the spatio-temporal filtering methods (e.g., [9]–[16]). The latter group typically provides a larger amount of noise reduction than spatial filtering methods, at the expense of target distortions at the output of the filter.

Only a few studies exist (e.g., [2], [15], [17], [18]) that evaluate the perceptual performance (such as intelligibility and localization) of binaural speech enhancement methods. In contrast, most studies evaluate performance using instrumental measures, e.g., predicting intelligibility (e.g., by means of STOI [19] or DBSTOI [20]) or localization accuracy (e.g., by means of interaural level and time differences errors [13], or other measures such as the ones presented in [21], [22]). Although these measures correlate well with localization and intelligibility, not all aspects of localization and intelligibility are well understood or incorporated in these measures. To understand the real trade-off between intelligibility improvement and localization accuracy, listening tests are still required.

In this paper, we evaluate two methods recently proposed in [8] and [16] by means of an intelligibility test and a localization test, and compare them with the binaural minimum variance distortionless (BMVDR) method [3]. In addition, we compare with an oracle based method [18], to get an idea of the intelligibility and perceived localization if perfect knowledge would be available. The BMVDR method provides the maximum noise reduction among all linear spatial filters, while severely distorting the binaural cues of all interferers [3]. We report the intelligibility and localization scores of self-reported normal-hearing people in several acoustic scenes.

The spatio-temporal filtering method proposed in [16] preserves the binaural cues by a binary classification of all frequency bins into target or noise-dominant bins. The classification is based on the output SNR that results by applying the BMVDR to all frequency bins. The target-dominant time-frequency bins are processed with the BMVDR, while the noise-dominant time-frequency bins are replaced with a scaled version of the corresponding unprocessed time-frequency bins.

The spatial filtering method proposed in [8] uses additional inequality constraints in the BMVDR optimization problem to preserve the binaural cues of all interferes. The inequality constraints are functions of anechoic pre-determined head-related transfer functions (HRTFs), which are considered as known and are acoustic-scene-independent, but user-dependent [8].

Section II reviews the binaural speech enhancement methods that we evaluate. Section III shows the evaluation proce-

This work was supported by the Oticon Foundation and NWO, the Dutch Organisation for Scientific Research.

dure and its results. Section IV gives concluding remarks.

II. Overview of the Evaluated Methods

In this section, we briefly review the binaural speech enhancement methods that we evaluate in this paper. For more details, the reader is referred to the associated papers.

A. BMVDR

The BMVDR spatial filter [3] provides the maximum noise reduction compared to all the other spatial filters. It preserves the binaural cues of the target, but distorts the binaural cues of all other sound sources, and makes them identical to the target's binaural cues. The BMVDR consists of two MVDR spatial filters [23] sharing the same microphone array, but using two different reference microphones, one on each HA. The two optimization problems, associated with the two MVDR spatial filters, minimize the total output noise power under the constraints that the target signal is preserved without any distortion at the two reference microphones. As such the binaural cues of the target signal are preserved, but the binaural cues of the interferers are not, since there are no constraints for them in the optimization problems.

B. Relaxed Binaural LCMV with Pre-determined HRTFs

The relaxed binaural linearly constrained minimum variance (LCMV) with pre-determined HRTFs is a spatial filtering method introduced in [7], [8]. This method uses additional inequality constraints in the BMVDR optimization problem to preserve the binaural cues of pre-selected azimuths and/or elevations around the head [8]. The inequality constraints can be relaxed as desired using a relaxation parameter, $0 \leq c \leq 1$. The maximum amount of relaxation (i.e., $c = 1$) results in the BMVDR filter as a special case. The trade-off between noise suppression and binaural-cue preservation of this method depends not only on c, but also on the number of pre-determined HRTFs. In this paper, we only vary the c-value and not the number and locations of the pre-determined HRTFs. More specifically, we always use anechoic pre-determined HRTFs (from the database in [24]) associated with 24 uniformly spaced locations in the horizontal plane on a circle around the head with a distance of 3 m from the center of the head.

C. BMVDR with Thresholding

The BMVDR with thresholding method is a spatio-temporal filtering method introduced in [16]. First, the BMVDR filter is applied to all time-frequency bins and, next, the output narrowband SNR, of all these enhanced time-frequency bins, is estimated. A time-frequency bin is considered target-dominant, if the output SNR of a time-frequency bin is above a certain threshold τ. Otherwise, the time-frequency bin is considered as noise-dominant. The noise-dominant time-frequency bins are suppressed identically, so that the interaural time and level differences are not changed in order to preserve the binaural cues of the noise. In particular, if the residual noise is inaudible after applying the BMVDR method, its binaural cues need not be preserved and, therefore, maximum possible noise suppression is achieved. If the noise in some time-frequency bins dominates the target after processing, the BMVDR output is not beneficial and, thus, the BMVDR output is replaced by a scaled-down version of the unprocessed scene to suppress the interferers and preserve the binaural cues of the acoustic scene. Since this scaling reduces both the target and noise components, the target signal will be distorted.

D. Ideal Binaural Target Enhancement

This is an oracle-based method that consists of the unprocessed acoustic scene with an SNR equal to the SNR output of the BMVDR method [18]. This method achieves the same amount of noise suppression as the BMVDR while perfectly preserving the binaural cues of the complete acoustic scene.

III. Experiments

To evaluate the methods presented in Section II, we conducted an intelligibility test, which measures the 50% speech reception threshold (SRT), and a localization test, which measures the binaural localization error of the dominant point sources in the acoustic scene. Both tests are divided into two different phases; a parameter selection phase and a testing phase. The acoustic scenes in the testing phase are different from the one in the parameter selection phase. This is done to examine the robustness in different acoustic scenes with respect to the chosen parameter settings. The main purpose of the parameter selection phase is to obtain the c and τ parameters for the relaxed binaural LCMV and the BMVDR with thresholding methods, respectively, to be used in the testing phase. The testing phase examines the performance of all methods in the remaining two acoustic scenes.

We used Beyerdynamic DT 990 PRO 250 OHM headphones for the listening tests. The average sound level of the total noise that was played via the headphones was fixed to 65 dB SPL and the target level was varied to achieve a certain SNR.

For convenience, we use the following acronyms for the compared methods in the following figures and tables: relaxed binaural LCMV (RBLCMV(c)), binaural MVDR (BMVDR), BMVDR with thresholding (BMVDR (τ)), ideal binaural target enhancement (IBTE), and unprocessed scene (UNPR).

A. Generation of Audio Signal Database

For both listening tests we created a database of unprocessed and processed 2-channel binaural signals with SNRs ranging from -28 dB to 10 dB. The unprocessed HA signals were computed using the behind-the-ear impulse response database in [24]. For the multi-microphone binaural speech enhancement methods we used the front and middle microphones from each HA to create an array of 4 microphones. After processing, we saved the 2-channel binaural output signals corresponding to the reference microphones.

We used as the target signal randomly selected Dutch-spoken sentences with a duration of about 2 s from a female talker, taken from the database in [25]. We padded these sentences at the beginning and at the end with extra zeros such that a length of 4 s was obtained and the spoken sentence

TABLE I
SUMMARY OF ACOUSTIC SCENES.

acoustic scene	point source position (degrees)			diffuse noise	recording environment	mic. noise
	female talker	male talker	music			
AC1	0	-30	90	cafeteria noise	cafeteria	yes
AC2	-30	90	-90	cafeteria noise	cafeteria	yes
AC3	0	-45	60	–	office	yes

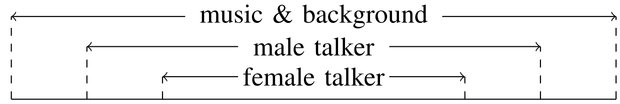

Fig. 1. Time duration of each source signal. The background signal is a cafeteria background noise and is present only in AC1 and AC2.

was exactly temporally centered within the masking noise as shown in Fig. 1. This was done in order to avoid confusion of the listener due to simultaneous initiation of all sources.

We used four different noise types to simulate the acoustic scenes: a music signal, a randomly selected English-spoken sentence from a male talker taken from the TIMIT database [26], a diffuse cafeteria noise taken from the database in [24], and microphone-self noise. We also used three different acoustic scenes, which we denote as AC1, AC2 and AC3. Table I summarizes the acoustical sources and their locations in all acoustic scenes. Note that AC1 was used for parameter selection, while AC2 and AC3 were used for the testing phase.

The female and male talkers' signals were zero-padded to have an equal length of 4 s. For the music sound source, a 4 s fragment was extracted randomly per sentence from an approximately 5 minutes long music piece. All three noise contributions were set to have equal average power at the two reference microphones, making all disturbances equally important in the acoustic scene. The input SNR, defined as the target power with respect to the total noise power, was computed by concatenating the left and right reference microphone recordings of the target and the noise signals. The sampling frequency of all signals was set to 16 kHz.

B. Subjects

In the parameter selection phase, we used 5 native speakers of Dutch for the intelligibility test, and 5 non-native speakers of Dutch for the localization test. In the testing phase, we used 14 native speakers of Dutch for the intelligibility test, and 15 non-native speakers of Dutch for the localization test. All subjects from the parameter selection phase participated in the testing phase as well. All subjects were self-reported as normal-hearing and their age range was 20-36 years.

C. Intelligibility Test

The target sentences (not necessarily meaningful) were part of a matrix test consisting of 5 words each, with the correct grammatical structure name, verb, number, adjective and noun. The sentences and the noise realizations were randomly selected from the database. Using a graphical user interface (GUI), the listeners had access to a 10×5 matrix with each column consisting of the 10 candidate words used to construct the sentences. The sentences were played only once to the subjects, after which they had to select from each column the word that was understood. We used the one-down-one-up adaptive staircase method [27] to find the SRT-50 scores (i.e., the SNR at which the subject scored 50% correct) for each method and subject. The adaptive track started with an initial SNR of 10 dB and an initial step-size of 8 dB. For each new reversal the step-size was halved until it became 1 dB. After this, the procedure continued until 8 more reversals were completed. Finally, the median of the last 8 reversals was computed as the SRT-50 score of every subject. Every subject had a 2-3 minutes training session before the official test, to get familiar with the GUI. Per subject, the SRT-50 was computed once for each algorithm.

D. Localization Test

In order to perform the localization test, we implemented the GUI as depicted in Fig. 2. There is a question on the top of the screen which asks the subject to identify the perceived direction of a specific source. The subjects were asked to listen to the algorithms by pressing the buttons on the right-hand side as many times they wanted and then identified the angle by pressing one of the circles on the image. There are 6 buttons in total on the right-hand side (for the testing phase) as shown in Fig. 2, because there are five competing methods and one reference signal, which is the point source in question in isolation. For the testing phase, the user pressed the 'next experiment' button $3 \times 2 \times 2 = 12$ times (i.e., there are 12 pages in total) to find the azimuths of all the point sources (3 in total) in the acoustic scenes AC2 and AC3 for two repetitions.

The algorithms were presented in a random order and in the testing phase the acoustic scenes were also presented in random order between different pages and within the same page. Moreover, the input signals to all presented algorithms had an overall SNR of -5 dB, in order to clearly hear all dominant point sources after processing. Finally, the localization errors were computed with respect to a reference signal azimuth (and not the true azimuth of the source). This is because, the HA signals were constructed using a single set of HRTFs from [24], which are different than the HRTFs of the subjects. Thus, the subjects will, typically, perceive binaural cues differently from each other. Since the localization test is to verify preservation of binaural cues, it is better to check how close the binaural cues after processing are to those before processing for each subject. Finally, since there are two repetitions, we played each reference source signal twice and calculated the average response on this as the reference location. Finally, we averaged all localization errors across all sources in the acoustic scene per algorithm.

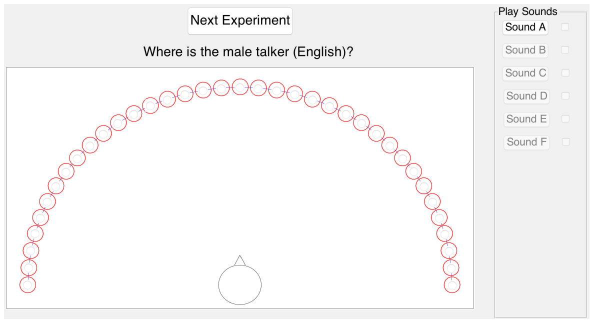

Fig. 2. Graphical user interface of localization test.

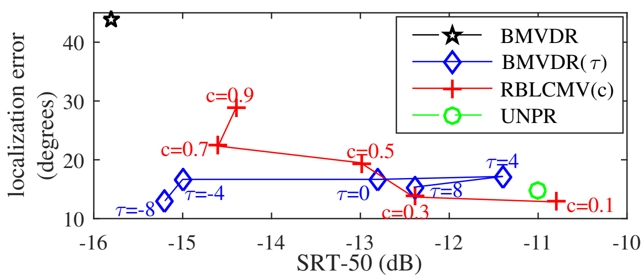

Fig. 3. Parameter selection phase: Trade-off between localization error (degrees) and SRT-50 (dB).

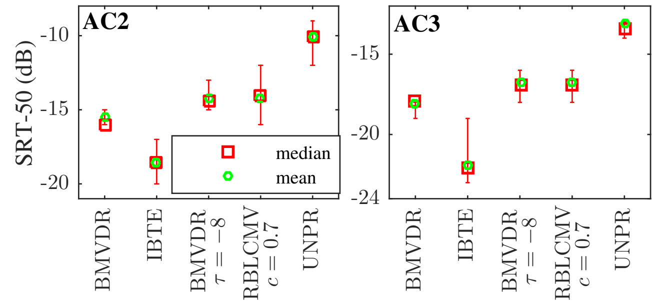

Fig. 4. Testing phase: SRT-50 (dB) statistics.

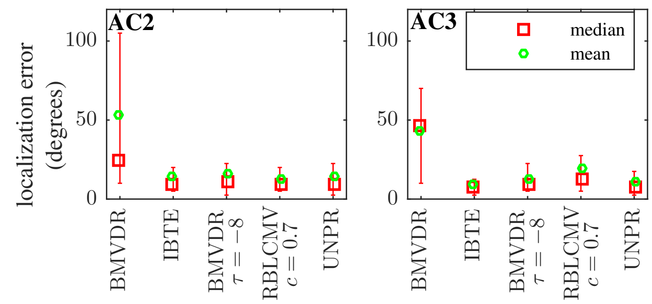

Fig. 5. Testing phase: localization error (degrees) statistics.

E. Parameter Selection Phase Results

In the parameter selection phase, we compared all methods from Section II except for the ideal binaural target enhancement method. The comparisons were made only for scene AC1. For the relaxed binaural LCMV method, we tested all values of the parameter c from the set $c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, and for the BMVDR with thresholding, we tested all values of the parameter τ from the set $\tau \in \{-8, -4, 0, 4, 8\}$ dB. Fig. 3 shows the trade-off curves of the two methods with the SRT-50 scores on the x-axis and localization error on the y-axis, parameterized by the tested τ and c parameter. For both the SRT-50 and the localization error, the final score was calculated as the mean across different subjects. The mean localization-error scores were also computed across different sources and repetitions.

As expected (see Section II-B), as c increases, the relaxed binaural LCMV method, in most cases, has an increased localization error and an increased intelligibility. The BMVDR thresholding method has a steady localization error for all tested τ values, while it provides a large intelligibility improvement for small τ values. In Fig. 3, two reasonably good parameter choices for the two methods are the ones with the largest intelligibility improvement and as small localization error as possible, i.e., $c=0.7$ and $\tau=-8$ dB. We used only these two parameter choices for the testing phase (Section III-F).

F. Testing Phase Results

In the testing phase, we compared all methods from Section II in scenes AC2 and AC3. Fig. 4 shows the median and mean SRT-50, and the 0.25 and 0.75 quantiles. Fig. 5 shows the median and mean localization error, and the 0.25 and 0.75 quantiles. We also performed two t-tests to determine, if the compared methods are significantly different in terms of intelligibility improvement and localization error. The p-values of the intelligibility t-test are given in Table II and III for acoustic scenes AC2 and AC3, respectively. It is clear from the p-values that the relaxed binaural LCMV ($c=0.7$) and the BMVDR thresholding ($\tau=-8$) are not significantly different from each other. The intelligibility of both proposed methods is significantly better compared to the unprocessed scene and significantly worse compared to the BMVDR. The p-values of the localization t-test are given in Table IV and V for acoustic scenes AC2 and AC3, respectively. In both scenes, the proposed methods have a significantly better localization than BMVDR. Moreover, in scene AC2, the two proposed methods are not significantly different from the ideal target enhancement and for scene AC3 the BMVDR ($\tau = -8$) is not significantly different from the ideal binaural target enhancement or the unprocessed noisy scene. This means that the proposed methods indeed preserve the correct locations of the sources in most cases, while significantly improve the intelligibility with respect to the unprocessed scene.

IV. Conclusion

In this paper, we perceptually evaluated two recently proposed binaural speech enhancement methods in terms of intelligibility improvement and localization error. Both methods

TABLE II
T-TEST P-VALUES FOR INTELLIGIBILITY TEST IN AC2.

Method	BMVDR	IBTE	BMVDR ($\tau = -8$)	RBLCMV ($c = 0.7$)	UNPR
BMVDR ($\tau = -8$)	0.0149	0	1	0.9177	0
RBLCMV ($c = 0.7$)	0.0401	0	0.9177	1	0

TABLE III
T-TEST P-VALUES FOR INTELLIGIBILITY TEST IN AC3.

Method	BMVDR	IBTE	BMVDR ($\tau = -8$)	RBLCMV ($c = 0.7$)	UNPR
BMVDR ($\tau = -8$)	0.0259	0	1	1	0
RBLCMV ($c = 0.7$)	0.0105	0	1	1	0

TABLE IV
T-TEST P-VALUES FOR LOCALIZATION TEST IN AC2.

Method	BMVDR	IBTE	BMVDR ($\tau = -8$)	RBLCMV ($c = 0.7$)	UNPR
BMVDR ($\tau = -8$)	0	0.5645	1	0.3161	0.4153
RBLCMV ($c = 0.7$)	0	0.7800	0.3161	1	0.8673

TABLE V
T-TEST P-VALUES FOR LOCALIZATION TEST IN AC3.

Method	BMVDR	IBTE	BMVDR ($\tau = -8$)	RBLCMV ($c = 0.7$)	UNPR
BMVDR ($\tau = -8$)	0	0.0515	1	0.0272	0.2366
RBLCMV ($c = 0.7$)	0	0.0001	0.0272	1	0.0014

provide a significantly better trade-off between intelligibility improvement and localization performance compared to the unprocessed scene and the reference BMVDR method. Moreover, in most cases, the two methods are not significantly different than the ideal binaural target enhancement method in terms of localization error. Moreover, the difference between the two methods is not statistically significant in most cases.

REFERENCES

[1] J. M. Kates, *Digital hearing aids*. Plural publishing, 2008.

[2] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Speech enhancement with multimicrophone wiener filter techniques in mulltimicrophone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, 2009.

[3] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[4] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 543–558, Mar. 2016.

[5] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints," *IEEE Trans. Audio, Speech, Language Process.*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.

[6] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multi-microphone noise reduction preserving binaural cues," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016.

[7] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," *IEEE Trans. Audio, Speech, Language Process.*, vol. 25, no. 1, pp. 137–152, Jan. 2017.

[8] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, J. Jensen, and M. Guo, "Binaural beamforming using pre-determined relative acoustic transfer functions," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, 2017.

[9] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[10] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multi-channel Wiener filtering based noise reduction," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, pp. 29–32.

[11] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Int. Workshop Acoustic Echo, Noise Control (IWAENC)*, Sep. 2006.

[12] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[13] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[14] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel Wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 23, no. 12, Sept. 2015.

[15] J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, "Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene," *EURASIP J. Advances Signal Process.*, 2016.

[16] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, "Binaural speech enhancement with spatial cue preservation utilising simultaneous masking," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, Aug. 2017.

[17] B. Cornelis and M. Moonen, "Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel wiener filtering based noise reduction," *J. Acoust. Soc. Amer.*, vol. 131, no. 6, pp. 4743–4755, 2012.

[18] D. Marquardt, "Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques," Ph.D. dissertation, Carl von Ossietzky Universität Oldenburg, 2015.

[19] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

[20] A. H. Andersen, J. M. de Haan, Z. H. Tan, and J. Jensen, "Predicting the intelligibility of noisy and nonlinearly processed binaural speech," *IEEE Trans. Audio, Speech, Language Process.*, vol. 24, no. 11, pp. 1908–1920, 2016.

[21] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Amer.*, vol. 116, no. 5, pp. 3075–3089, July 2004.

[22] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *ELSEVIER Speech Commun.*, vol. 53, no. 5, pp. 592–605, 2011.

[23] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[24] H. Kayser, S. Ewert, J. Annemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Advances Signal Process.*, vol. 2009, pp. 1–10, Dec. 2009.

[25] R. Houben, J. Koopman, H. Luts, K. C. Wagener, A. van Wieringen, H. Verschuure, and W. A. Dreschler, "Development of a dutch matrix sentence test to assess speech intelligibility in noise," *Int. J. Audiol.*, vol. 53, no. 10, pp. 760–763, 2014.

[26] V. Zue, S. Seneff, and J. Glass, "Speech database development at mit: Timit and beyond," *ELSEVIER Speech Commun.*, vol. 9, no. 4, pp. 351–356, 1993.

[27] H. Levitt, "Transformed updown methods in psychoacoustics," *J. Acoust. Soc. Amer.*, vol. 49, pp. 467–477, 1971.