

Localization based on enhanced low frequency interaural level difference

Çalış, Metin; Heusdens, Richard; Hendriks, Richard C.; van de Par, Steven

DOI

[10.1109/TASLP.2021.3111583](https://doi.org/10.1109/TASLP.2021.3111583)

Publication date

2021

Document Version

Final published version

Published in

IEEE - ACM Transactions on Audio, Speech, and Language Processing

Citation (APA)

Çalış, M., Heusdens, R., Hendriks, R. C., & van de Par, S. (2021). Localization based on enhanced low frequency interaural level difference. *IEEE - ACM Transactions on Audio, Speech, and Language Processing*, 29, 3025 - 3039. Article 9534666. <https://doi.org/10.1109/TASLP.2021.3111583>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Localization Based on Enhanced Low Frequency Interaural Level Difference

Metin Calis , *Graduate Student Member, IEEE*, Steven van de Par, Richard Heusdens , *Senior Member, IEEE*, and Richard Christian Hendriks 

Abstract—The processing of low-frequency interaural time differences is found to be problematic among hearing-impaired people. The current generation of beamformers does not consider this deficiency. In an attempt to tackle this issue, we propose to replace the inaudible interaural time differences in the low-frequency region with the interaural level differences. In addition, a beamformer is introduced and analyzed, which enhances the low-frequency interaural level differences of the sound sources using a near-field transformation. The proposed beamforming problem is relaxed to a convex problem using semi-definite relaxation. The instrumental analysis suggests that the low-frequency interaural level differences are enhanced without hindering the provided intelligibility. A psychoacoustic localization test is done using a listening experiment, which suggests that the replacement of time differences into level differences improves the localization performance of normal-hearing listeners for an anechoic scene but not for a reverberant scene.

Index Terms—Binaural cue preservation, beamforming, hearing aids, JBLCMV, TFS, BMVDR.

I. INTRODUCTION

GOOD hearing is a vital and important part of daily life. Being hearing impaired comes with many challenging situations ranging from private to social interactions. In some cases, hearing-impaired people can find themselves in dangerous situations due to the lack of hearing. For example, when crossing in traffic. Moreover, hearing-impaired people might feel isolated in practical situations due to the inability to differentiate and understand different sound sources in complex listening environments [1]. These people can benefit by using hearing aids. However, despite being very powerful, the current generation of hearing aids is not able to completely compensate for hearing loss.

Manuscript received February 23, 2021; revised May 21, 2021 and July 17, 2021; accepted August 23, 2021. Date of publication September 9, 2021; date of current version September 24, 2021. This project was supported in part by Deutsche Forschungsgesellschaft (DFG), under Grant EXS 2177: Hearing4all, and in part by the National Science Agenda (NWA) idea generator project: Restored sound localization for hearing impaired people, NWA.1228.191.034. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Isabel Barbancho. (*Corresponding author: Metin Calis.*)

Metin Calis, Richard Heusdens, and Richard Christian Hendriks are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft 2628 CD, The Netherlands (e-mail: m.calis@tudelft.nl; r.heusdens@tudelft.nl; r.c.hendriks@tudelft.nl).

Steven van de Par is with the Department of Medical Physics and Acoustics, Carl von Ossietzky University Acoustics Group, Oldenburg 26111, Germany (e-mail: steven.van.de.par@uni-oldenburg.de).

Digital Object Identifier 10.1109/TASLP.2021.3111583

The current generation of hearing aids comes with a wireless link that enables the left and right microphone arrays to exchange information to achieve noise reduction and conservation of spatial auditory cues [2]. The microphone arrays are then combined and used in beamforming algorithms to enhance the target source while suppressing the interferers. Many beamforming algorithms have been proposed, such as the binaural minimum variance distortionless response beamformer [3] and the joint binaural linearly constrained minimum variance beamformer [4]. The former is known to be good at achieving maximum noise reduction, while the latter is known for its ability to preserve the spatial cues and reduce the noise at the same time. Besides quality and intelligibility, in many practical situations, it is also important for the suppressed interferers to sound natural and to appear from their original locations. For example, to correctly localize (interfering) point sources in the surrounding as in a traffic scenario. This leads to the typical trade-off between noise reduction and preservation of the spatial cues as examined by researchers, which led to many beamforming algorithms [5]–[7].

The sound source signals reaching the two ears contain the information that is required for the auditory system to extract and analyze the horizontal location of the sources. The auditory events are formed through two main dissimilarities that exist between the signals reaching the ears; namely, interaural time differences and interaural level differences [8]. The interaural time differences are the primary source of localization for frequencies below 1.5 kHz and mainly occur due to the differences in the time that takes for a source to reach both ears. The interaural level difference on the other hand is caused by the shadowing of the head and exploited mainly for frequencies above 3 kHz [9]. The frequencies between these ranges are the frequencies where none of the spatial cues are dominant and considered to be the range that is the worst for localization [9]. The aforementioned beamformers that preserve the location of the interferers, mainly preserve the time and level differences of the interferer after processing. In some cases, the spatial cues that are preserved are in fact not audible by hearing-impaired people [10]–[12]. This paper is part of a project that aims at exploring the suprathreshold effects of why the hearing impaired can not achieve similar performance to normal hearing people even after the audibility is established. In an attempt to answer this question, it has been hypothesized that hearing-impaired people can achieve a better performance when the auditory scene is presented in a different acoustic form than the original.

The low-frequency processing of ITDs, which is known in the literature as the temporal fine structure processing (TFS) is found to be problematic among hearing-impaired people [13]. Several reasons have been proposed in the literature to understand why low-frequency TFS processing can be impaired. The phase-locking of the auditory-nerve fibers might have become imprecise [14] or the phase shift on the basilar membrane might result in inaccurate decoding of the TFS information. The TFS information is thought to be dependent on the cross-correlation across the different places on the basilar membrane [15]. If there is a phase shift, the cross-correlation would give inaccurate TFS information reducing its reliability. Another reason proposed by the authors in [16] suggests that the broadening of the auditory filters might be the cause of the reduced TFS processing. In the light of these suggestions, the beamformers that preserve the binaural cues of the whole spectrum might be wasting the degrees of freedom that they have, to preserve the inaudible cues. Instead of designing beamformers that disregard the hearing impairment of the user, a better enhancement algorithm can be created by taking into account how well the listener utilizes the preserved binaural cues.

This paper starts with the signal model and the background information in Section II and Section III, respectively. After that, the related work is explained in Section IV. The problem formulation is explained in Section V. The following chapters cover the solution to the problem that is explained earlier. These chapters include the applied convex relaxation and parameter selection. In Section VII, the proposed method is analyzed using theoretical measures. This section includes objective intelligibility tests and localization analysis. An experiment using normal listeners has been conducted and the results are shown in Section VIII. The paper ends with a discussion and a conclusion chapter where the findings are analyzed and comments have been made about the future work that can be done.

II. SIGNAL MODEL

In our signal model, we assume the presence of two hearing aids with $M/2$ microphones at each ear, where M is an even number. The binaural enhancement algorithms apply spatial filters to the left and the right hearing aid microphones, where each filter uses the recordings of all M microphones together. It is assumed that there is one target signal and there are r interferers where the maximum number of interferers is represented as r_{\max} . The enhancement is applied in the frequency domain where k represents the frequency bin and l represents the frame index. The signal received by the j th microphone for $j = 1, \dots, M$ is given by

$$y_j(k, l) = \underbrace{a_j(k, l)s(k, l)}_{x_j(k, l)} + \sum_{i=1}^r \underbrace{b_{ij}(k, l)u_i(k, l)}_{n_{ij}(k, l)} + v_j(k, l) \quad (1)$$

where

- $s(k, l)$ denotes the target signal at the location of the source,
- $u_i(k, l)$ denotes the i th interferer signal at the location of the source,
- $a_j(k, l)$ is the acoustic transfer function (ATF) of the target signal with respect to the j th microphone,

- $b_{ij}(k, l)$ is the ATF of the i th interfering signal from the location of the source to the j th microphone,
- $v_j(k, l)$ is the additive noise at the j th microphone.

In practice, the acoustic transfer functions (ATF) are estimated [17]. However, in this work we assume the ATF to be known to avoid errors due to the ATF mismatch. In the instrumental performance evaluation in Section VII, the ATFs are generated according to a predefined setup and in the perceptual evaluation in Section VIII generic head-related transfer functions (HRTF) are used.

To simplify the notation, the indices k and l are omitted. The signal model in (1) can then be written in vector notation as,

$$\mathbf{y} = \underbrace{\mathbf{a}\mathbf{s}}_{\mathbf{x}} + \sum_{i=1}^r \underbrace{\mathbf{b}_i u_i}_{\mathbf{n}_i} + \mathbf{v} \quad (2)$$

where $\mathbf{a} \in \mathbb{C}^{M \times 1}$, $\mathbf{b}_{ij} \in \mathbb{C}^{M \times 1}$, $\mathbf{y} \in \mathbb{C}^{M \times 1}$, $\mathbf{x} \in \mathbb{C}^{M \times 1}$, $\mathbf{n}_i \in \mathbb{C}^{M \times 1}$ and $\mathbf{v} \in \mathbb{C}^{M \times 1}$ are the column vectors that hold the M realizations of a_j , b_{ij} , y_j , x_j , n_{ij} and v_j for $j = 1 \dots M$ and $i = 1 \dots r$ respectively.

The noise and the sound sources are assumed to be mutually uncorrelated. Using this assumption, the cross power spectral density (CPSD) of the received signal at the j th microphone can be written as,

$$\mathbf{P}_y = E[\mathbf{y}\mathbf{y}^H] = \mathbf{P}_x + \underbrace{\sum_{i=1}^r \mathbf{P}_{n_i}}_{\mathbf{P}_N} + \mathbf{P}_v \quad (3)$$

where

- $\mathbf{P}_x = E[\mathbf{x}\mathbf{x}^H] \in \mathbb{C}^{M \times M}$ is the CPSD matrix of \mathbf{x} .
- $\mathbf{P}_{n_i} = E[\mathbf{n}_i\mathbf{n}_i^H] \in \mathbb{C}^{M \times M}$ is the CPSD matrix of \mathbf{n}_i .
- $\mathbf{P}_v = E[\mathbf{v}\mathbf{v}^H] \in \mathbb{C}^{M \times M}$ is the CPSD matrix of \mathbf{v} ,
- $\mathbf{P}_N = \sum_{i=1}^r \mathbf{P}_{n_i} + \mathbf{P}_v \in \mathbb{C}^{M \times M}$ is the CPSD matrix of the total noise.

All the applied beamformers in this paper use the ideal voice activity detection and overcome any estimation errors which might happen in practice.

Without loss of generality, the last microphone and the first microphone are selected as the left reference microphone and the right reference microphone for the target. These are denoted as a_L and a_R , respectively. For the interferers, the left and right microphones are denoted relative to the i th interferer as $b_{i,L}$ and $b_{i,R}$ for $i = 1, \dots, r$.

The incidence angle is assumed to be 0 degrees at the mid-sagittal plane and increases clockwise to 180° degrees and decreases to -180° anticlockwise starting from the median plane.

III. BACKGROUND INFORMATION

In this section, background information is given that will be used in the following chapters. The information consists of the binaural cues and the binaural beamforming algorithms.

A. Binaural Cues

The auditory system uses the time and level differences of the signals reaching the ears to determine the horizontal location of the sound [8]. Following the convention in [5] and [18], the

time and level differences can be defined using the interaural transfer function (ITF). The input and output ITF for the target signal is defined as

$$ITF_x^{in} = \frac{a_L}{a_R}, \quad ITF_x^{out} = \frac{\mathbf{w}_L^H \mathbf{a}}{\mathbf{w}_R^H \mathbf{a}}. \quad (4)$$

The ITF for the interferer is defined as

$$ITF_{u_i}^{in} = \frac{b_{i,L}}{b_{i,R}}, \quad ITF_{u_i}^{out} = \frac{\mathbf{w}_L^H \mathbf{b}}{\mathbf{w}_R^H \mathbf{b}}. \quad (5)$$

The interaural level difference is defined to be the squared magnitude of the ITF. The ILD before processing can be represented as

$$ILD_x^{in} = |ITF_x^{in}|^2, \quad ILD_{u_i}^{in} = |ITF_{u_i}^{in}|^2. \quad (6)$$

The output ILD is defined similarly, that is,

$$ILD_x^{out} = |ITF_x^{out}|^2, \quad ILD_{u_i}^{out} = |ITF_{u_i}^{out}|^2. \quad (7)$$

The interaural time differences are defined to be the phase of the ITF normalized by the angular frequency, w [19]. For the input ITD, this can be shown as

$$ITD_x^{in} = \frac{\angle ITF_x^{in}}{2\pi f}, \quad ITD_{u_i}^{in} = \frac{\angle ITF_{u_i}^{in}}{2\pi f}. \quad (8)$$

Similarly, the output ITD is defined as

$$ITD_x^{out} = \frac{\angle ITF_x^{out}}{2\pi f}, \quad ITD_{u_i}^{out} = \frac{\angle ITF_{u_i}^{out}}{2\pi f}. \quad (9)$$

In some cases, the interaural time differences are represented using the phase information [20]. The interaural phase differences for the sources before processing can then be defined as

$$IPD_x^{in} = \angle ITF_x^{in}, \quad IPD_{u_i}^{in} = \angle ITF_{u_i}^{in}. \quad (10)$$

After the processing, the IPDs can be written as

$$IPD_x^{out} = \angle ITF_x^{out}, \quad IPD_{u_i}^{out} = \angle ITF_{u_i}^{out}. \quad (11)$$

If the beamformer preserves the ITF of the source after processing, then both the ILD and ITD will also be preserved. On the other hand, a beamformer might be written such that at specific frequencies only the ILD or ITD cues are preserved, discarding the other binaural cue.

B. Binaural Beamforming Algorithms

Some beamformer algorithms such as the binaural minimum variance distortionless response (BMVDR) aim at reducing the noise power as much as possible while other beamformers such as the joint binaural linearly constrained minimum variance algorithm (JBLCMV) aim at achieving noise reduction while preserving the spatial cues of the interferers. The BMVDR improves the listening comfort by providing the maximum noise suppression without preserving any of the spatial cues of the interferers. All the binaural cues of the interferers collapse to the target's direction forcing the interferers to sound from the same direction as the target. The optimization problem of the BMVDR is represented by the following expression, that is,

$$\begin{aligned} \min_{\mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}} \quad & \mathbf{w}_L^H \mathbf{P}_N \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P}_N \mathbf{w}_R \\ \text{s.t.} \quad & \mathbf{w}_L^H \mathbf{a} = a_L, \quad \mathbf{w}_R^H \mathbf{a} = a_R. \end{aligned} \quad (12)$$

The optimization problem given in (12) has the following closed form solutions [4], [21], which can be expressed as

$$\hat{\mathbf{w}}_L = \frac{\mathbf{P}_N^{-1} \mathbf{a} a_L^*}{\mathbf{a}^H \mathbf{P}_N^{-1} \mathbf{a}}, \quad \hat{\mathbf{w}}_R = \frac{\mathbf{P}_N^{-1} \mathbf{a} a_R^*}{\mathbf{a}^H \mathbf{P}_N^{-1} \mathbf{a}}. \quad (13)$$

If the left and right spatial filters \mathbf{w}_L and \mathbf{w}_R are merged into one vector $\mathbf{w}_{\text{BMVDR}} = [\mathbf{w}_L^H \quad \mathbf{w}_R^H]^H$, the optimization problem (12) can be written jointly as

$$\min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \quad \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \quad (14)$$

$$\text{s.t.} \quad \mathbf{w}^H \mathbf{\Lambda}_A = \mathbf{f}_A^H, \quad (15)$$

where

$$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_N \end{bmatrix}, \quad \mathbf{\Lambda}_A = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix}, \quad \mathbf{f}_A = \begin{bmatrix} a_L \\ a_R \end{bmatrix}. \quad (16)$$

The closed-form solution to the jointly written BMVDR problem [4] is given by

$$\mathbf{w}_{\text{BMVDR}} = \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda}_A (\mathbf{\Lambda}_A^H \tilde{\mathbf{P}} \mathbf{\Lambda}_A)^{-1} \mathbf{f}_A. \quad (17)$$

When (17) is used, all the interfering sound sources will sound from the same location as the target source. This prevents a possible increase in the intelligibility due to the spatial release from masking (SRM) [22]. The SRM suggests that the sounds that are spatially separated have higher intelligibility than sources that are colocated. The JBLCMV beamformers solve this problem by introducing additional constraints to preserve the spatial information of the interferers. This comes up with a trade-off with regards to possible noise reduction which can be provided since the feasible set of the spatial filters to perform the noise reduction will shrink with the additional constraints.

The JBLCMV framework is given by the following formulation,

$$\min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \quad \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \quad (18)$$

$$\text{s.t.} \quad \mathbf{w}^H \mathbf{\Lambda} = \mathbf{f}^H, \quad (19)$$

where

$$\mathbf{\Lambda} = [\mathbf{\Lambda}_a | \mathbf{\Lambda}_b] \quad (20)$$

$$= \begin{bmatrix} \mathbf{a} & \mathbf{0} & \mathbf{b}_1 b_{1R} & \dots & \mathbf{b}_m b_{mR} \\ \mathbf{0} & \mathbf{a} & -\mathbf{b}_1 b_{1L} & \dots & -\mathbf{b}_m b_{mL} \end{bmatrix} \in \mathbb{C}^{2M \times (2+r)}, \quad (21)$$

$$\mathbf{f}^H = [\mathbf{f}_a^H | \mathbf{f}_b^H] \quad (22)$$

$$= [a_L^* \quad a_R^* | 0 \quad 0 \quad \dots \quad 0] \in \mathbb{C}^{1 \times (2+r)}. \quad (23)$$

In addition to the target distortionless constraints $\mathbf{w}^H \mathbf{\Lambda}_a = \mathbf{f}_a$, there are additional constraints $\mathbf{w}^H \mathbf{\Lambda}_b = \mathbf{f}_b$, which preserve the cues of the interferers. These additional constraints preserve the ITF of the interferers forcing both ITD and ILD to be preserved in the whole spectrum.

Assuming that there are r interferers and M microphones, the degrees of freedom that are left for the JBLCMV to do noise reduction is $2M - 2 - r$. Here the distortionless target constraint reduces the total degrees of freedom by two whereas, interferer cue preservation reduces the total degrees of freedom

by r as there is one equality constraint per interferer. On the other hand, the BMVDR has $2M - 2$ degrees of freedom left to do noise reduction as there are only two target distortionless constraints. This enables the BMVDR to have a larger domain to minimize noise power which provides a better noise reduction capability. For relaxed methods such as presented in [3], the degrees of freedom are not straightforward due to the inequality constraints. However, by checking the feasible set of the optimization problem, a comparison can still be done. The authors of [3] relax the interferer cue preservation constraint providing a user-controlled trade-off between noise reduction and cue preservation. The relaxation provides a feasible set between the JBLCMV and the BMVDR. Hence, the output noise power becomes bounded between the JBLCMV and the BMVDR.

IV. RELATED WORK

In [23], a method to enhance the low-frequency ILDs was introduced. The authors first solved the phase ambiguity problem, which might occur for frequencies below 1500 Hz. The resulting unambiguous ITD values were plugged into the ILD-to-ITD function measured for high-frequency tones which were obtained from [24]. The resulting ILDs were smoothed and sent to the bilateral hearing aids to replace the low-frequency ITD cues. The authors did not find any improvement in intelligibility. However, they found an improvement of localization for the speech stimulus but not for the broadband noise, lowpass filtered noise, or lowpass filtered AM noise.

In [25] and [26], the authors analyzed the localization performance of bimodal listeners when the ILD in the available dynamic range was enhanced. In the former, the signals were noise-vocoded at one ear and lowpass filtered at the contralateral ear to simulate the bimodal hearing. The authors first measured the ILD of the full-band received signals on the hearing aid and cochlear implant devices and used the root-mean-square ratio of the signals to enhance the low-frequency content. It was found that the localization ability of the normal-hearing listeners under a simulated hearing setup was improved for a broadband noise by 14° but not for the telephone alerting signal. In the latter study, the authors created an artificial ILD versus angle function which overcame the non-monotonicity of the ILD signals around 60° . This was done by using white noise as the source and the resulting natural ILD versus angle functions of the six bimodal listeners' hearing devices that were placed on a mannequin's head. This function was obtained by using the full-band spectrum of the received signal which was used irrespective of the stimulus spectrum. The resulting ILD-to-angle function was transformed into a monotonic relation by extrapolating the curve at the point when the non-monotonicity started. The authors reported a 4° to 10° improvement for the horizontal localization performance for the bimodal listeners. Following this line of work, the authors in [27] designed a beamformer that attenuated the sources coming from the contralateral direction as opposed to traditional beamformers which attenuated the sources coming from the rear. The authors reported an improvement of horizontal localization and speech intelligibility for bimodal listeners.

Spatial release from masking (SRM) using the enhanced low-frequency ILD cues was investigated by [28]. Spatial release

from masking is the improvement in speech reception thresholds when the target and the distractor change from the same to different locations. Normal hearing people benefit around 20 dB SRM, whereas the hearing impaired benefit less [28]. Apart from the horizontal separation which benefits the intelligibility of the signals, in [29], [30] the authors realized that distance cues also improve the SRM. In the former, one sound source was fixed to one meter and the other source was moved closer to the listener. It was found that a better target-to-masker ratio could be achieved for the better ear. In the latter, three experiments were conducted which assessed the effect of distance cues on spatial segregation. It was found out that the intelligibility of the target source could be improved due to the spatial release from masking. Following this line of work, the authors in [28] investigated the use of low-frequency ILDs separately on the spatial release from masking. Maximum low-frequency ILDs were applied and it was found out that HI can benefit from an additional increase of SRM.

This work investigates the effect of near-field low-frequency distance cues on horizontal localization. It is different from [23] because of the formulation of how the enhanced ILD cues are generated. In addition to this, a low-frequency enhancing beamformer is proposed and its performance is analyzed. The non-monotonicity of the ILD cues is avoided by limiting the low-frequency range that is going to be used for ILD enhancement, unlike [26] where an artificial angle versus ILD function is generated empirically.

V. PROBLEM FORMULATION

The duplex theory proposed by [31] suggests that the ITD cues are effective for frequencies below 1500 Hz and ILD cues are effective for frequencies above 1500 Hz. Although these ranges are perceptually the most important frequencies for the processing of pure tone ILD and ITD cues, the sensitivity of the human ear to these binaural cues extends beyond this scope. In the high frequencies, human listeners have been observed to be sensitive to the slowly-varying envelope fluctuations of broadband signals [32]. In the low frequencies, the sensitivity to the fast-varying temporal fine structure of the signal is observed [33]. The sensitivity to the ILD is observed in the whole region.

The deficiency of low-frequency ITD processing up to 1000 Hz was observed among hearing-impaired people [34]–[36], where the authors found a correlation between increasing age and the inability to process the low-frequency temporal fine structure. During the preparation of this paper, a binaural beat listening experiment with a panel of 15 hearing-impaired people was carried out. The stimuli used in the experiment consisted of two tones with a slightly different frequency in the left and right ear, approximately around 2 Hz. The subjects were presented with three intervals, one of which contained the binaural beat stimulus with different frequencies in the left and right ear, while the other intervals contained identical sinusoids with some randomized frequency to make sure that the differences in pitch could not be used. A 2-up 1-down adaptive staircase method was used where 2 successive correct answers resulted in a step up in frequency and one false answer to a step down in frequency. The

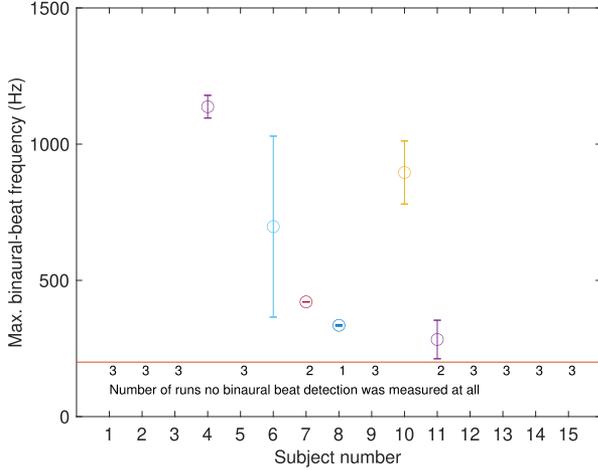


Fig. 1. The mean and standard deviation of the maximum binaural beat frequency that the panel of hearing-impaired people can hear.

initial step was 150 Hz, after two reversals 75 Hz, and after two reversals 50 Hz. At the 50 Hz step-size a total of 6 reversals were then measured to serve for calculating a mean threshold value. Subjects were instructed to select the modulated or moving interval. They could practice one or two runs in advance of formal data collection. For all but one subject at least three formal threshold measurements were done, only for the subject with the high binaural beat frequency (subj. 4), only two thresholds were collected. It was found out that 9 out of 15 subjects were unable to hear a binaural beat even at a low frequency like 200 Hz, while all but one subjects could not hear binaural beats in the frequency range that normal hearing people can still hear binaural beats, which is defined to be around 1200 Hz [37]. The results of the experiment can be seen in Fig. 1.

In the frequency range where the temporal fine structure cues are dominant, it is still possible to differentiate the ILD cues. The just noticeable difference of ILD is approximately 1 dB in the whole spectrum [8]. The ILD in the low-frequency region can still be heard and claimed to be the dominant cue for distance perception [38]. These ILD cues do not occur naturally for the far-field signals. In the far-field, the ILD cues are around 0 dB for frequencies below 500 Hz [39]. Since the diameter of the head is significantly smaller than the wavelength, the head does not shadow the contralateral ear and hence, level differences do not occur.

In the near-field region defined to be less than 1 m [38], low-frequency ILDs can reach as high as 20-30 dB. When a source comes closer to the head, the level differences in the whole spectrum increase, including the low-frequency region. As head shadowing is not observed in this region, the near-field ILDs and the source angle create an injective relation which can be utilized to boost the horizontal localization performance.

Binaural beamformer algorithms such as the binaural linearly constrained minimum variance algorithms spend the degrees of freedom that they have to preserve both the ILD and ITD cues in the whole spectrum [40]. If the ITD cues are not heard by the hearing-impaired person, the degrees of freedom which can be used for noise reduction will be wasted for preserving the

inaudible ITD cues. Since the low-frequency ITD cues are not processed by the hearing impaired, we will investigate whether they can artificially be replaced by the near-field ILD cues in the same region. This research project, therefore, aims to answer the following question: *Can the enhancement of low-frequency ILDs overcome the loss of spatial information induced by the deficient temporal fine structure processing?*

VI. PROPOSED METHOD

In this section, a beamformer is introduced to enhance the low-frequency ILDs of the interferers while keeping the target distortionless. The section starts with a brief overview of the problem, continues with the applied convex relaxation and the methodology behind choosing a particular scaling factor that determines to which extend the low-frequency ILDs are enhanced.

In the low-frequency region, the interaural level differences will be artificially introduced, while in the high frequency region, both the ILD and ITD will be preserved. This can be expressed with the following optimization problem:

$$\begin{aligned}
 f < f_c \quad & \min_{\mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}} \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \\
 \text{s.t.} \quad & \mathbf{w}_L^H \mathbf{a} = a_L \quad \mathbf{w}_R^H \mathbf{a} = a_R, \\
 & \left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} \right|^2 - c \left| \frac{b_{i,L}}{b_{i,R}} \right|^2 = 0, \\
 f > f_c \quad & \min_{\mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}} \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \\
 \text{s.t.} \quad & \mathbf{w}_L^H \mathbf{a} = a_L \quad \mathbf{w}_R^H \mathbf{a} = a_R, \\
 & \mathbf{w}_L^H \mathbf{b}_i b_{i,R} - \mathbf{w}_R^H \mathbf{b}_i b_{i,L} = 0, \quad (24)
 \end{aligned}$$

for $i = 1, \dots, r' \leq r_{\max}$. This means that for all the frequencies above the cut-off frequency f_c , the JBLCMV will be applied which is explained in (18). For the frequencies below f_c , the proposed method will be applied which enhances the low-frequency ILD while leaving the target undistorted. The enhancement of the ILDs is represented by the last constraint

$$\left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} \right|^2 - c \left| \frac{b_{i,L}}{b_{i,R}} \right|^2 = 0.$$

Without the scaling factor c , the input ILD of each interferer will be preserved at the output. By including the factor c , we aim at enhancing the ILDs of the interferers. This factor will depend on the direction of the interferer and the near-field transformation. The sound sources that are close to the head have a higher ILD and the sound sources that are away from the head have a lower ILD in the low-frequency region. In addition to this, the magnitude of the ILDs increases as the sound source reaches to 90 or -90 degrees starting from the mid-sagittal plane. A description of this relation for different frequencies can be seen in Fig. 2.

While there is a closed-form solution for the JBLCMV, there is no closed-form solution for the introduced low-frequency enhancement beamformer due to its non-convexity. In the following section, the optimization problem for $f < f_c$ will be relaxed using a semi-definite relaxation approach [41].

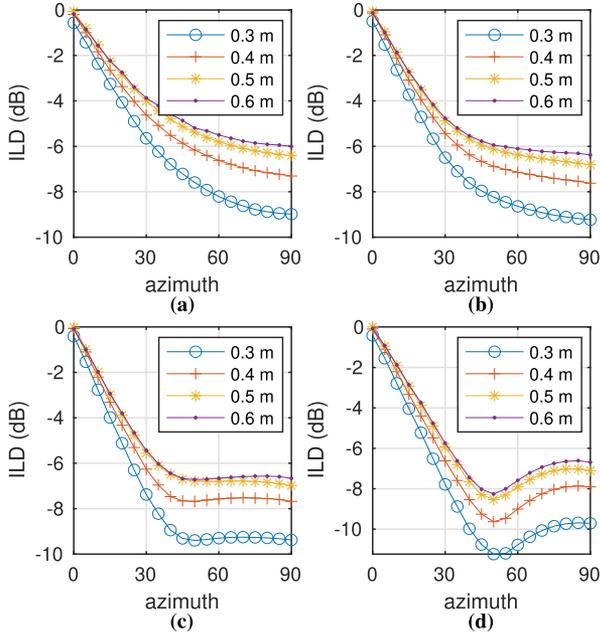


Fig. 2. ILD versus azimuth plotted using the HRTF taken from [48] for four different frequencies: (a) ILDs at 800 Hz, (b) ILDs at 1000 Hz, (c) ILDs at 1200 Hz, and (d) ILDs at 1500 Hz.

A. Convex Relaxation

The problem at hand preserves the target signal at the left and right ears. In addition to that, the ILD of the interferers has been enhanced by the scaling factor c .

The ILD enhancement constraint in (24) can be extended and written as

$$\begin{aligned}
 0 &= \left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} \right|^2 - c \left| \frac{b_{i,L}}{b_{i,R}} \right|^2 \\
 &= \mathbf{w}_L^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_L |b_{i,R}|^2 - c \mathbf{w}_R^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_R |b_{i,L}|^2 \\
 &= \underbrace{\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix}}_{\mathbf{w}^H \in \mathbb{C}^{2M \times 1}} \underbrace{\begin{bmatrix} \mathbf{b}_i \mathbf{b}_i^H |b_{i,R}|^2 & \mathbf{0} \\ \mathbf{0} & -c \mathbf{b}_i \mathbf{b}_i^H |b_{i,L}|^2 \end{bmatrix}}_{\mathbf{M}_i \in \mathbb{C}^{2M \times 2M}} \underbrace{\begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix}}_{\mathbf{w}}.
 \end{aligned} \tag{25}$$

Using the expansion in (25), the main problem given in (24) can be written compactly as

$$\begin{aligned}
 \min_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} & \underbrace{\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix}}_{\mathbf{w}^H \in \mathbb{C}^{1 \times 2M}} \underbrace{\begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{bmatrix}}_{\tilde{\mathbf{P}} \in \mathbb{C}^{2M \times 2M}} \underbrace{\begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix}}_{\mathbf{w} \in \mathbb{C}^{2M \times 1}} \\
 \text{s.t.} & \underbrace{\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix}}_{\mathbf{w}^H} \underbrace{\begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix}}_{\Lambda_a \in \mathbb{C}^{2M \times 2}} = \underbrace{\begin{bmatrix} a_L & a_R \end{bmatrix}}_{\mathbf{f}_a^H \in \mathbb{C}^{1 \times 2}}, \\
 & \underbrace{\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix}}_{\mathbf{w}^H} \mathbf{M}_i \underbrace{\begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix}}_{\mathbf{w}} = 0.
 \end{aligned} \tag{26}$$

If vector notation is used, the optimization problem given in (26) can be written as

$$\begin{aligned}
 \min_{\mathbf{w}} & \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\
 \text{s.t.} & \mathbf{w}^H \Lambda_a = \mathbf{f}_a^H, \\
 & \mathbf{w}^H \mathbf{M}_i \mathbf{w} = 0.
 \end{aligned} \tag{27}$$

The optimization problem in (27) is not convex due to the quadratic equality constraint $\mathbf{w}^H \mathbf{M}_i \mathbf{w}$. Since \mathbf{M}_i is not positive semi-definite, the expression $\mathbf{w}^H \mathbf{M}_i \mathbf{w}$ is not convex. On the other hand, the objective function is convex as $\tilde{\mathbf{P}}$ is positive semi-definite. The linear equality $\mathbf{w}^H \Lambda_a = \mathbf{f}_a^H$ is convex. If the equality constraint at the last line of (27) is written as two inequality constraints, the problem at hand becomes a Quadratically Constrained Quadratic Program (QCQP), which is NP-hard [42]. We can use semi-definite relaxation [43] and reformulation-linearization techniques [44] to transform the optimization problem from (27) to a relaxed convex problem which can be solved in polynomial time using interior-point solvers such as CVX in MATLAB.

The semi-definite relaxation can be applied to transform (27) into a semi-definite program (SDP). Two important matrix properties will be used to execute the transformation.

- 1) We have the following relation for any quadratic expression

$$\mathbf{q}^H \mathbf{Z} \mathbf{q} = \text{Tr}(\mathbf{q}^H \mathbf{Z} \mathbf{q}) = \text{Tr}(\mathbf{q} \mathbf{q}^H \mathbf{Z}). \tag{28}$$

- 2) Using Schur's complement [45],

$$\begin{aligned}
 \mathbf{Z} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{bmatrix} \geq 0 &\Leftrightarrow \\
 \mathbf{A} \geq 0, & (\mathbf{I} - \mathbf{A} \mathbf{A}^\dagger) \mathbf{B} = \mathbf{0}, \quad \mathbf{S}_1 \geq 0, \\
 \mathbf{C} \geq 0, & (\mathbf{I} - \mathbf{C} \mathbf{C}^\dagger) \mathbf{B}^H = \mathbf{0}, \quad \mathbf{S}_2 \geq 0,
 \end{aligned} \tag{29}$$

with $\mathbf{S}_1 = \mathbf{C} - \mathbf{B}^H \mathbf{A}^\dagger \mathbf{B}$ the generalized Schur complement of \mathbf{A} in \mathbf{Z} and $\mathbf{S}_2 = \mathbf{A} - \mathbf{B} \mathbf{C}^\dagger \mathbf{B}^H$ the generalized Schur complement of \mathbf{C} in \mathbf{Z} . \mathbf{A}^\dagger is the pseudo-inverse of \mathbf{A} .

Let $\mathbf{W} = \mathbf{w} \mathbf{w}^H$. Using (28), the optimization problem in (27) becomes

$$\begin{aligned}
 \text{Problem 1:} & \min_{\mathbf{w}, \mathbf{W}} \text{Tr}(\mathbf{W} \tilde{\mathbf{P}}) \\
 \text{s.t.} & \mathbf{w}^H \Lambda_a = \mathbf{f}_a^H, \\
 & \text{Tr}(\mathbf{W} \mathbf{M}_i) = 0, \quad i = 1, \dots, r, \\
 & \mathbf{W} = \mathbf{w} \mathbf{w}^H.
 \end{aligned} \tag{30}$$

The optimization problem given in (30) is not convex due to the last equality constraint $\mathbf{W} = \mathbf{w} \mathbf{w}^H$. Removing this rank 1 constraint makes the problem convex [41],

$$\begin{aligned}
 \min_{\mathbf{w}, \mathbf{W}} & \text{Tr}(\mathbf{W} \tilde{\mathbf{P}}) \\
 \text{s.t.} & \mathbf{w}^H \Lambda_a = \mathbf{f}_a^H, \\
 & \text{Tr}(\mathbf{W} \mathbf{M}_i) = 0, \quad i = 1, \dots, r, \\
 & \mathbf{W} \geq \mathbf{w} \mathbf{w}^H.
 \end{aligned} \tag{31}$$

Finally using (29), (31) can be written in the standard SDP form as,

$$\begin{aligned} \text{Problem 2: } \quad & \min_{\mathbf{w}, \mathbf{W}} \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\ \text{s.t. } \quad & \mathbf{w}^H \boldsymbol{\Lambda}_a = \mathbf{f}_a^H, \\ & \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \quad i = 1, \dots, r, \\ & \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & \mathbf{1} \end{bmatrix} \geq 0. \end{aligned} \quad (32)$$

The optimization problem in (32) is convex and can be solved in polynomial time. Due to the relaxation introduced to the equality constraint $\mathbf{W} = \mathbf{w}\mathbf{w}^H$, the new problem in (32) does not necessarily give the same solution as (30). Let the optimal arguments be $\hat{\mathbf{w}}^*$ and $\hat{\mathbf{W}}^*$. If $\hat{\mathbf{W}}^* = \hat{\mathbf{w}}^* \hat{\mathbf{w}}^{*H}$, the solution to (32) is the optimal value for (30), which is equal to the solution of the original problem in (24).

Let the optimal value for (30) be p_1^* and the optimal value for (32) be p_2^* . Due to the relaxation that is introduced, the feasible set of (32) is larger than the feasible set of (30). For this reason, p_2^* lower bounds p_1^* as there is a larger set for the minimization problem in (32) which encapsulates the set of (30). Using reformulation-linearization techniques, we can further tighten this bound by introducing redundant constraints to the problem in (31). The target distortionless equality constraints in (31) can be reformulated using [44] as follows. The target distortionless constraint is given by

$$\mathbf{w}^H \boldsymbol{\Lambda}_a = \mathbf{f}_a^H. \quad (33)$$

If we multiply left and right by \mathbf{w} , (33) becomes

$$\begin{aligned} \mathbf{w}\mathbf{f}_a^H &= \mathbf{w}\mathbf{w}^H \boldsymbol{\Lambda}_a \\ \mathbf{w}\mathbf{f}_a^H &= \mathbf{W}\boldsymbol{\Lambda}_a. \end{aligned} \quad (34)$$

In addition to (34), another redundant constraint can be added to tighten the bound even further. The target distortionless constraint in (33) can be reformulated as,

$$\begin{aligned} 0 &= \mathbf{w}^H \boldsymbol{\Lambda}_a - \mathbf{f}_a^H \\ &= (\mathbf{w}^H \boldsymbol{\Lambda}_a - \mathbf{f}_a^H)(\mathbf{w}^H \boldsymbol{\Lambda}_a - \mathbf{f}_a^H)^H \\ &= \mathbf{w}^H (\boldsymbol{\Lambda}_a \boldsymbol{\Lambda}_a^H) \mathbf{w} - \mathbf{w}^H \boldsymbol{\Lambda}_a \mathbf{f}_a - \mathbf{f}_a^H \boldsymbol{\Lambda}_a^H \mathbf{w} + \mathbf{f}_a^H \mathbf{f}_a \\ &= \text{Tr}(\mathbf{W}\boldsymbol{\Lambda}_a \boldsymbol{\Lambda}_a^H) - \mathbf{w}^H \boldsymbol{\Lambda}_a \mathbf{f}_a - \mathbf{f}_a^H \boldsymbol{\Lambda}_a^H \mathbf{w} + \mathbf{f}_a^H \mathbf{f}_a. \end{aligned} \quad (35)$$

If the feasible set of (31) is tightened by the addition of (34) and (35), the optimization problem becomes

$$\begin{aligned} \text{Problem 3: } \quad & \min_{\mathbf{w}, \mathbf{W}} \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\ \text{s.t. } \quad & \mathbf{w}^H \boldsymbol{\Lambda}_a = \mathbf{f}_a^H, \\ & \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \\ & \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & \mathbf{1} \end{bmatrix} \geq 0, \\ & \text{Tr}(\mathbf{W}\boldsymbol{\Lambda}_a \boldsymbol{\Lambda}_a^H) - \mathbf{w}^H \boldsymbol{\Lambda}_a \mathbf{f}_a - \\ & \mathbf{f}_a^H \boldsymbol{\Lambda}_a^H \mathbf{w} + \mathbf{f}_a^H \mathbf{f}_a = 0, \\ & \mathbf{W}\boldsymbol{\Lambda}_a = \mathbf{w}\mathbf{f}_a^H, \end{aligned} \quad (36)$$

for $i = 1 \dots r \leq r_{\max}$. Let the optimal value for the optimization problem in (36) be p_3^* . Due to the additional constraints, we have

$$p_1^* \geq p_3^* \geq p_2^*. \quad (37)$$

The feasible set is tightened and the approximation is improved.

B. Cut-Off Frequency

The dynamic range of the ILD cues in the near field is significantly larger than the ILD cues in the far field for frequencies below 1000 Hz. For example, at 500 Hz, the ILD increases from 4 dB to 19 dB as a source at 90° approaches the head from 1 m to 0.12 m [38]. We can introduce these low-frequency ILD cues artificially in an attempt to overcome the lack of low-frequency ITD processing, which can be seen among hearing-impaired people [46], [47].

The injective nature of the ITDs provides a reliable cue for frequencies below 1500 Hz. This reliability has to be provided if the ITDs are intended to be replaced by the ILDs.

To understand which frequency should be used as a cut-off frequency for the beamformer algorithms, an analysis is done using the head-related transfer functions recorded using a KEMAR manikin at the Technical University of Berlin [48]. Four distances are selected: 0.3, 0.4, 0.5, and 0.6 meters. The ILDs are calculated using four frequencies: 800, 1000, 1200 and 1500 Hz. The results can be seen in Fig. 2. It is observed that for frequencies 1200 and 1500 Hz, the head shadow is visible and the ILD to frequency function is not injective. For 1000 Hz, the ILD is monotonic but a flattening in the same range can be observed. For this reason, 800 Hz is selected to be the cut-off frequency for the low-frequency ILD enhancement.

C. Scaling Factor

The scaling factor c in (24) can be decided through a look-up table which can be generated using the near-field head-related transfer functions [39], [48], [49]. Instead, the spherical model introduced by [50] and later verified by [51] can be used to generate the near-field distance cues from the far-field HRTF. This method has been proposed by [52] and has been shown psychoacoustically to give accurate estimates.

The distance variation function (DVF) proposed by [52] assumes the head can be considered as a rigid sphere with a radius a . The ears are assumed to be located at 100° away from the mid-sagittal plane. The rigid sphere model estimates the pressure at a point on the surface of a sphere as

$$p(a, w, \theta, r) = -kr \sum_{m=0}^{\infty} (2m+1) \frac{h_m(kr)}{h'_m(ka)} P_m(\cos(\theta)) e^{-ikr}, \quad (38)$$

where $h_m(kr)$ is a spherical Hankel function of the first kind of order m and $h'_m(ka)$ is the first derivative at radius a , $k = \frac{w}{c}$ is the wavenumber, c is the speed of sound, $P_m(\Lambda)$ is the Legendre polynomial of degree m , θ is the angle between a vector that starts from the center of the sphere and ends at the source location and the vector that starts at the center of the sphere and ends at a location on the surface of the sphere, r is the distance of the source to the center of the sphere and w is

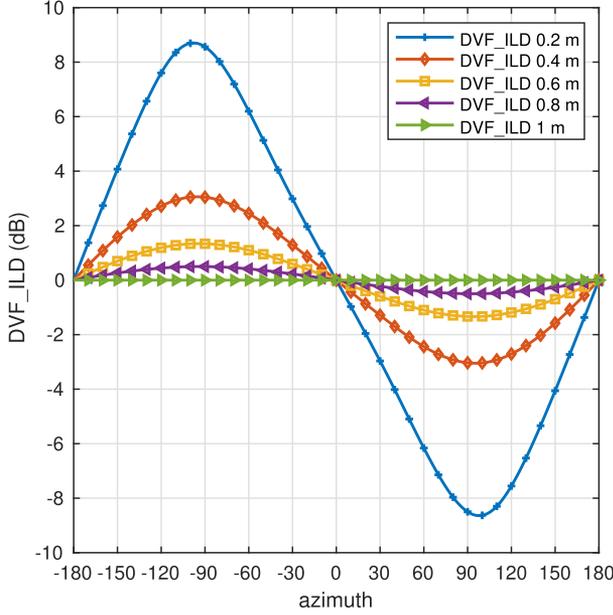


Fig. 3. The scaling factor DVF_ILD calculated for 5 different distances.

the angular frequency [50]. The DVF is calculated as

$$DVF = \frac{p_n(\alpha, w, \theta, d_n)}{p_f(\alpha, w, \theta, d_f)}, \quad (39)$$

where p_n stands for the near field pressure on the surface of the sphere and p_f stands for the pressure on the surface of the sphere for a far-field source. Although the authors used an individualized head radius, an average radius 8.75 cm is used in this study. If the DVF is calculated at the desired frequencies, the near-field HRTF can be calculated by

$$HRTF(d_n) = DVF \times HRTF(d_f), \quad (40)$$

where d_n is the near-field distance and d_f is the far-field distance. In addition to this, the far-field ILD cues can be transformed into near-field ILD cues by using (40). Given that the left ear head-related transfer function is named as $HRTF_L(d)$ and the right ear head-related transfer function as $HRTF_R(d)$, the ILDs can be calculated as

$$\begin{aligned} HRTF_L(d_n) &= DVF_L \times HRTF_L(d_f), \\ HRTF_R(d_n) &= DVF_R \times HRTF_R(d_f), \\ \underbrace{\frac{HRTF_L(d_n)^2}{HRTF_R(d_n)^2}}_{ILD_n} &= \underbrace{\frac{DVF_L^2}{DVF_R^2}}_{DVF_ILD} \times \underbrace{\frac{HRTF_L(d_f)^2}{HRTF_R(d_f)^2}}_{ILD_f}, \\ ILD_n &= DVF_ILD \times ILD_f, \end{aligned} \quad (41)$$

where ILD_n stands for the near-field ILDs, ILD_f stands for the far-field ILDs, and DVF_ILD stands for the scaling factor, which relates the near-field ILD cue to the far-field ILD cue. The scaling factor DVF_ILD is calculated until 800 Hz for the distances 0.2 m, 0.4 m, 0.6 m, 0.8 m, 1 m and shown in Fig. 3. Since the far-field distance is assumed to be 1 m, DVF_ILD is 0 dB for 1 m. It reaches in magnitude to 8 dB for a source at

0.2 m and at 90 degrees to the left or right of the mid-sagittal plane. The final problem becomes

$$\begin{aligned} f < 800: \quad & \min_{\mathbf{w}, \mathbf{W}} \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\ & \text{s.t. } \mathbf{w}^H \mathbf{\Lambda}_a = \mathbf{f}_a^H, \\ & \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \\ & \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & 1 \end{bmatrix} \geq 0, \\ & \text{Tr}(\mathbf{W}\mathbf{\Lambda}_a\mathbf{\Lambda}_a^H) - \mathbf{w}^H \mathbf{\Lambda}_a \mathbf{f}_a - \\ & \mathbf{f}_a^H \mathbf{\Lambda}_a^H \mathbf{w} + \mathbf{f}_a^H \mathbf{f}_a = 0, \\ & \mathbf{W}\mathbf{\Lambda}_a = \mathbf{w}\mathbf{f}_a^H, \\ f \geq 800: \quad & \min_{\mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}} \mathbf{w}_L^H \mathbf{P}\mathbf{w}_L + \mathbf{w}_R^H \mathbf{P}\mathbf{w}_R \\ & \text{s.t. } \mathbf{w}_L^H \mathbf{a} = a_L \quad \mathbf{w}_R^H \mathbf{a} = a_R, \\ & \mathbf{w}_L^H \mathbf{b}_i b_{iR} - \mathbf{w}_R^H \mathbf{b}_i b_{iL} = 0, \end{aligned} \quad (42)$$

where $c = DVF_ILD$, which is generated using (41), f stand for the frequency and $i = 1 \dots r \leq r_{\max}$. In the following sections, the optimization problem given in (42) is represented with the abbreviation ILD_d where d stands for the near field distance. For example, if enhancement with respect to 0.2 m is used, it is represented as $ILD_{0.2}$.

VII. PERFORMANCE MEASURES

In this section, the intelligibility and the localization performance of the proposed algorithm are investigated using objective measures. For the intelligibility, the speech intelligibility in bits metric is used [53]. For the spatial cue preservation performance, the absolute value of the difference between the input cue and the output cue is used.

We create a synthetic scenario with 8 interferers, where the first five are selected to be random sentences from the TIMIT database speech corpus [54] and the others are different noise types such as non-stationary noise, speech shaped noise, and babble noise. The interferer $u_1, u_2, u_3, u_4, u_5, u_6, u_7$ and u_8 are located at $-20^\circ, 20^\circ, -40^\circ, 40^\circ, -60^\circ, 60^\circ, -90^\circ, 90^\circ$ degrees respectively. There is one target signal s , which is located at 0° .

The anechoic head-related function from the Oldenburg database [55] is used to generate the acoustic transfer functions. The sampling frequency is selected to be 16 kHz to cover the most energetic components for speech signals. Each signal is windowed with a square-root-Hann window with a frame length of 12.5 ms and 50% overlap. The signals are concatenated such that there is a 20 s of voice active signal duration. For the application of the beamformer algorithms, a 256-point short-term fast Fourier transform (STFT) is applied to each frame. After processing the signals in the frequency domain, each frame is converted back to the original time domain by multiplying with a square-root Hann window and taking the inverse Fourier

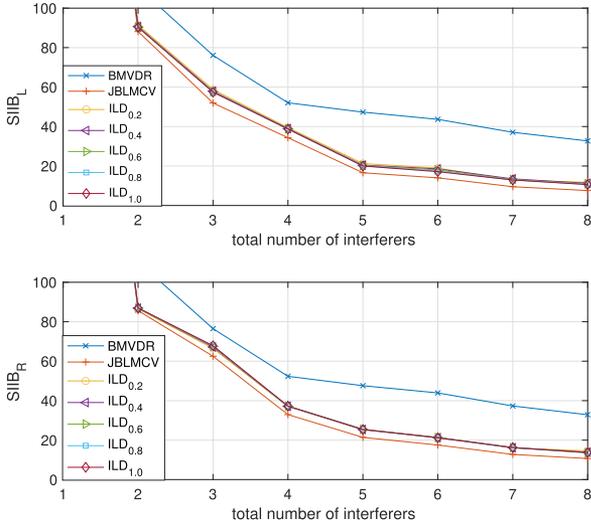


Fig. 4. Speech intelligibility in bits index for the BMVDR, JBLCMV, and five different enhancements of the proposed method.

transform. The frames are overlapped accordingly and the time domain signal is obtained.

Each interferer is scaled to be 0 dB with respect to the target signal so that cues related to audibility can be avoided. In addition, white gaussian noise at an SNR of 50 dB with respect to the target source is added to the received signals after they have been processed by the respective acoustic transfer functions to imitate the microphone's self-noise.

A. Speech Intelligibility in Bits

The SIIB index uses the mutual information between the clean signal and the degraded signal to assess intelligibility. The algorithm incorporates the time-frequency dependencies in the speech signal and was found to be effective for speech degraded by noise and processed by enhancement algorithms [53]. A higher bit index represents higher intelligibility whereas a lower bit index represents lower intelligibility. In this section, the BMVDR and the JBLCMV are compared with the different enhancements of the proposed method using the SIIB measure. The results can be seen in Fig. 4 for the signals at the left and the right ear that are represented as $SIIB_L$ and $SIIB_R$ respectively. The BMVDR has the highest SIIB metric with 13 bits difference compared to the proposed methods and the JBLCMV has the lowest SIIB metric with 5 bits difference compared to the proposed methods at four interferers.

The proposed methods share a slightly better performance compared to JBLCMV where the individual differences between different enhancement amounts can be neglected. It can be deduced that the applied low-frequency ILD enhancement does not affect intelligibility. Since the low-frequency ITD cues are not preserved, the optimization problem has more degrees of freedom to do noise reduction. Hence, 5 bits of improvement has been observed compared to JBLCMV. It should be noted that the applied convex relaxation overestimates the optimal value,

which is shown in (37). Although in our analysis we observed that most of the time $W = ww^H$ holds, the noise reduction capability is still an overestimation of the original problem given in (24). In conclusion, the intelligibility of the signals is not disrupted with the low-frequency ILD enhancement.

B. Localization Performance

ITD and ILD cues are the binaural cues that are important for horizontal localization. The ILD is defined as (6) and ITD is defined as (8). If the ILD at the input is the same as the ILD at the output, it can be said that the perfect preservation of the ILD has been achieved. However, if there is a mismatch between the input and output ILD, the ILD is not perfectly preserved. Depending on the frequency and the magnitude of the error, the horizontal location of the source might be different. The ILD error is defined to be

$$ILD_x^{err} = ||ITF_x^{out}|^2 - |ITF_x^{in}|^2|, \quad (43)$$

$$ILD_{u_i}^{err} = ||ITF_{u_i}^{out}|^2 - c|ITF_{u_i}^{in}|^2|. \quad (44)$$

where c is the scaling factor that is chosen using (41) for each interferer $i = 1 \dots r \leq r_{max}$.

The interaural phase difference error is defined in (10). The IPD^{err} is represented for the target and the interferer as

$$IPD_x^{err} = \frac{\angle ITF_x^{out} - \angle ITF_x^{in}}{\pi}, \quad (45)$$

$$IPD_{u_i}^{err} = \frac{\angle ITF_{u_i}^{out} - \angle ITF_{u_i}^{in}}{\pi}, \quad (46)$$

where $IPD^{err} \in [0, 1]$.

The localization performance of the two variants of the original problem in (24), namely (32) and (36), will be examined for a different number of interferers. This investigation will be focused on $f < 800$ Hz to analyze the performance of the proposed ILD enhancement beamformer performance separate from the JBLCMV. The reader is invited to [20] for an analysis of the LCMV framework. For each frequency bin less than 800 Hz, the error is calculated using (43) and (45). The results are averaged and transformed to the dB scale. The Fig. 5 includes the ITF error of the target source, ITD, and IPD errors of the interferers for the JBLCMV, the BMVDR, and the two variants of the proposed method with five different enhancements, which is $ILD_{0.2}$, $ILD_{0.4}$, $ILD_{0.6}$, $ILD_{0.8}$ and $ILD_{1.0}$. The IPD and ILD error of the JBLCMV was approximately -130 dB, which is represented by the lowermost line due to visualization purposes. It can be seen that for all cases, the target signal is kept undistorted. The mean IPD error of the proposed beamformers is similar to BMVDR as none of them have any constraint about the preservation of the IPD, unlike the JBLCMV which has an error of -130 dB. The ILD error is higher when the interferer number is greater than 1 when Problem 2 is used compared to the algorithm in Problem 3 due to the additional constraints.

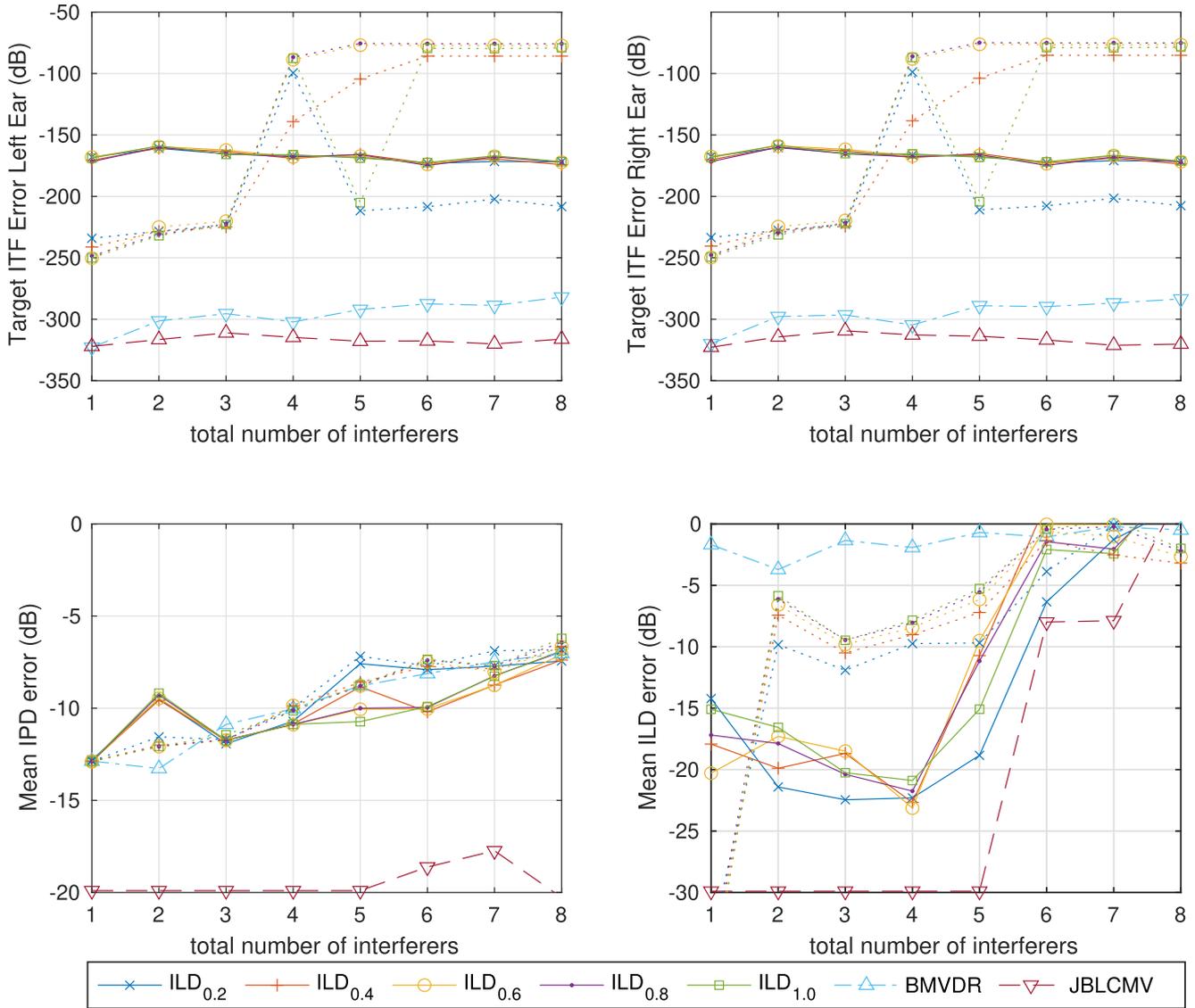


Fig. 5. The interaural transfer function of the target and the interaural level and phase differences of the interferers plotted for the JBLCMV, BMVDR and five different enhancements of two variants of the proposed method. The solid lines represent the optimization problem in (36) and the dotted lines represent the variant in (32) for 5 different enhancements $ILD_{0.2}$, $ILD_{0.4}$, $ILD_{0.6}$, $ILD_{0.8}$, and $ILD_{1.0}$. The dashed line with upside down marker represent the JBLCMV method and the dot-dashed line with upside down marker represent the BMVDR method.

VIII. LISTENING EXPERIMENT

An experiment is designed to assess the localization performance of the proposed method. The proposed method is compared to two reference methods, these are, the JBLCMV and the BMVDR. Each scenario consists of four signals. Two of these signals are a male and a female speaker selected randomly from the TIMIT database [56]. The third signal is a piano piece and the fourth signal is a cellphone vibration, which is low-pass filtered at 800 Hz. The female speaker is assigned as the target signal whereas, the three other signals are assigned as interferers. A comparison of the power spectral densities of the stimuli used in the experiment can be seen in Fig. 6. After convolving the audio signals with the anechoic head related transfer function, the power spectral densities of the resulting binaural signals can

TABLE I
SUMMARY OF THE ACOUSTIC SCENES

| Acoustic scene | Source position | | | | T60 |
|--------------------|-----------------|-------------|------------|---------------------|--------|
| | Female talker | Male talker | Piano tune | Cellphone vibration | |
| Anechoic (Scene 1) | 0 | -90 | 65 | -30 | 50 ms |
| Office (Scene 2) | 0 | 75 | 15 | -45 | 300 ms |

be seen in Fig. 7. The summary of the acoustic scenes can be seen in Table I.

Each signal is processed with the BMVDR, the JBLCMV, and the three variations of the proposed method. The first variation

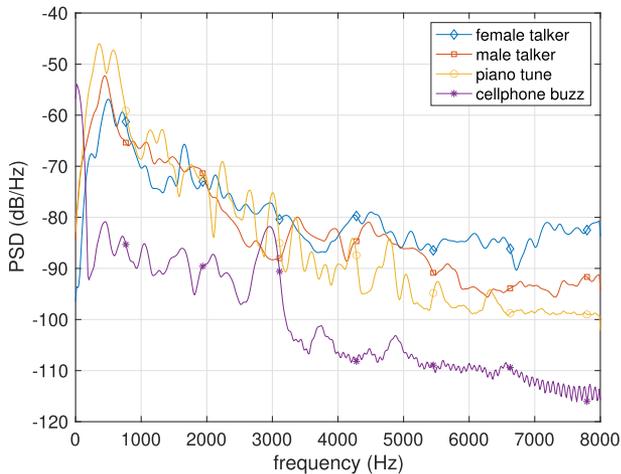


Fig. 6. The power spectral density of the stimuli used in the experiment before convolution with the head related transfer functions.

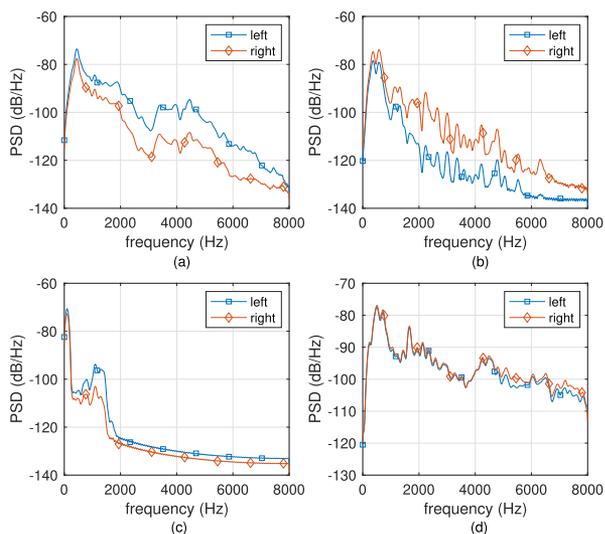


Fig. 7. The power spectral density of the stimuli used in the experiment for the anechoic scene. The signals are (a) male speaker, (b) piano tune, (c) cellphone vibration, and (d) female speaker.

preserves the natural low-frequency ILD cues up to 800 Hz while leaving the ITD cues unpreserved in this range. This method is abbreviated as ILD_1 . The second variation artificially introduces ILDs in the low-frequency region up to 800 Hz with respect to 0.6 m distance. This method is abbreviated as $ILD_{0.6}$. The third variation artificially introduces ILDs in the low-frequency region up to 800 Hz with respect to 0.2 m distance. This method is abbreviated as $ILD_{0.2}$. The enhancement scale is calculated according to (41) and applied per frequency bin. In all three variations, the JBLCMV beamformer is used for frequencies higher than 800 Hz.

The head-related transfer functions are selected using the middle and rear microphone recordings of the behind the ear (BTE) hearing aid [55]. There are two microphones on each of the left and right hearing aid totaling to four. Two different environments are used to understand the effect of reverberation. The first scene is the anechoic environment where the reverberation

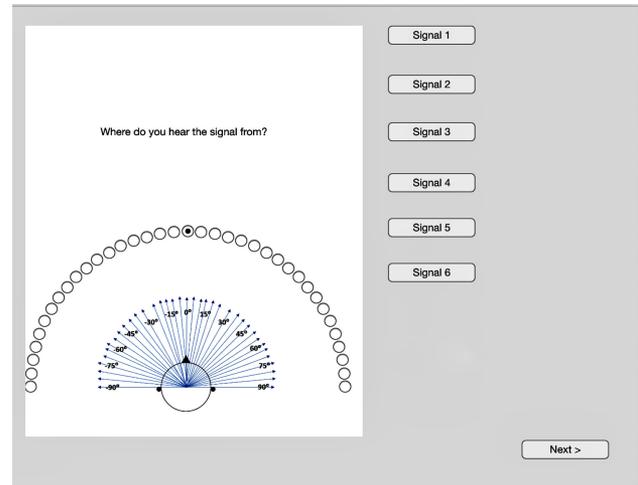


Fig. 8. Graphical user interface for the listening experiment.

is minimal. The second scene is the office environment where the reverberation is greater than the anechoic environment. The anechoic head-related impulse response has a T60 of 50 ms and the office head-related impulse response has a T60 of 300 ms where the early reflections are considered.

An azimuth is assigned for each target and the interferer signal. The target signal is always kept at 0° while the interferers are assigned to other angles in the frontal plane. The graphical user interface depicted in Fig. 8 is created. The user can click on a button as many times as they desire until they are confident with its location. There are 6 buttons in one page, which correspond to JBLCMV, BMVDR, $ILD_{0.2}$, $ILD_{0.6}$, $ILD_{1.0}$ and the unprocessed version of the same signal. We consider a single source scenario where each button plays only one signal. The signals are 4 seconds long. The microphone self noise is simulated as an additive white Gaussian noise. Each recording has a 50 dB SNR due to the self noise. The target and the interferers are processed such that they share the same power. The anechoic and the office scene recordings are played in a random order to prevent any bias. Each signal is presented twice. Since generic head related transfer functions are used, a mismatch is expected between the actual source locations and the interpreted locations. We included the unprocessed signals to account for this mismatch. The locations assigned to the unprocessed signals are used as a reference azimuth for the calculation of the localization error.

Participants are asked to attend the experiment online. This causes a few intricacies that need to be handled. Firstly, the participants are asked to do the test in a silent room to avoid the interference of other sounds. Secondly, the participants are informed to finish the whole experiment without changing the volume after getting used to it initially. This prevents any cues related to increased audibility. Last but not least, a guideline is prepared to make sure that the headphones are placed correctly and the system is working.

A. Results

In total, 21 people aged between 20 – 27 participated in the experiment. None of the participants had any reported hearing problems. Since it was not possible to control the environment

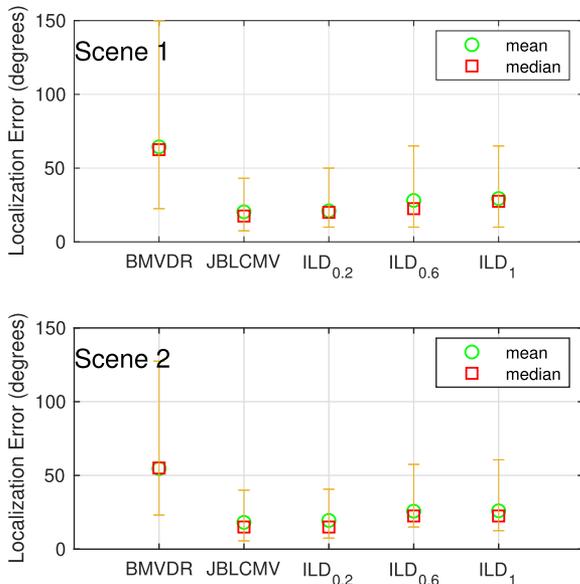


Fig. 9. Mean, median, 0.25, and 0.75 quantiles of the localization error of all the participants at two different scenes, averaged over the interferers.

or the tools that the participants were using, a variance test is done to find if there are outliers. The variance is calculated using the results of the unprocessed case and its repetitions. If the participants assigned significantly different angles to the same unprocessed signals (twice repeated), they are considered outliers. Only one participant is found to be three standard deviations higher than the others. This participant's result has been omitted from the analysis reducing the number of participants to 20.

The averaged results of the experiment can be seen in Fig. 9. Two-way ANOVA with repeated measures is used with audio signals as within-group factors and different beamforming algorithms as between-group factors to compare the localization performances of the beamformers with the spatial cue preservation constraint, that is, the JBLCMV and the proposed method with different enhancements. Each participant chooses the angle of a sound source twice independently. The answers are averaged to reduce the variance per participant. Three hypotheses have been postulated. The first hypothesis denoted as H_0^a examines if there is a significant difference between the methods: the JBLCMV, $ILD_{0.2}$, $ILD_{0.6}$, and $ILD_{1.0}$. This hypothesis assesses how close the JBLCMV and the proposed methods are with respect to localization performance. Since the JBLCMV preserves both the ILD and ITD cues in the whole spectrum, it is expected to perform the best. If there is a significant difference between the JBLCMV and the proposed methods, it can be deduced that the localization is hindered due to the lack of ITD preservation in the low-frequency region. The second hypothesis, denoted as H_0^b , examines if any of the sound sources have a significantly different localization outcome. The third hypothesis, denoted as H_0^c , examines the interaction between the beamformers and the sound sources. The former hypothesis assesses if the localization performance is different for a low-pass filtered sound, speech sound or a broadband piano piece. The latter hypothesis, on the

TABLE II
SUMMARY OF THE TWO-WAY ANOVA RESULTS FOR THREE HYPOTHESES AT DIFFERENT ACOUSTIC SCENES

| Hypothesis | Acoustic Scene | F | Prob>F |
|------------|----------------|-------|---------|
| H_0^a | Scene 1 | 4.6 | 0.0038 |
| | Scene 2 | 3.01 | 0.0300 |
| H_0^b | Scene 1 | 3.46 | 0.0330 |
| | Scene 2 | 12.26 | <0.0001 |
| H_0^c | Scene 1 | 0.46 | 0.8382 |
| | Scene 2 | 0.94 | 0.4620 |

TABLE III
POST-HOC ANALYSIS OF THE SOURCES USING TUKEY'S TEST

| Beamforming Algorithms | | Acoustic Scene | p-value |
|------------------------|-------------|----------------|---------|
| $ILD_{0.2}$ | $ILD_{0.6}$ | Scene 1 | 0.0822 |
| | | Scene 2 | 0.2038 |
| $ILD_{0.2}$ | $ILD_{1.0}$ | Scene 1 | 0.0242 |
| | | Scene 2 | 0.1640 |
| $ILD_{0.2}$ | JBLCMV | Scene 1 | >0.9999 |
| | | Scene 2 | 0.9964 |
| $ILD_{0.6}$ | $ILD_{1.0}$ | Scene 1 | 0.9688 |
| | | Scene 2 | 0.9995 |
| $ILD_{0.6}$ | JBLCMV | Scene 1 | 0.0793 |
| | | Scene 2 | 0.1302 |
| $ILD_{1.0}$ | JBLCMV | Scene 1 | 0.0232 |
| | | Scene 2 | 0.1020 |

other hand, explores if any of the beamformer performances are correlated with the audio signals. The summary of the results is given in Table II.

For the anechoic scene, the first hypothesis H_0^a is rejected due to $[F(3, 228) = 4.6; p \approx 0.0038]$, which suggests that at least one of the algorithms is significantly different from the others. A follow-up test is done at the end of the section to understand which of the methods are significantly different.

For the office environment, the results are similar to the first scene. H_0^a is rejected due to $[F(3, 228) = 3.01; p = 0.03]$. This suggests that the beamformer algorithms the JBLCMV, $ILD_{0.2}$, $ILD_{0.6}$, and $ILD_{1.0}$ perform differently also for scenarios with higher reverberation. The second hypothesis is rejected as $[F(2, 228) = 12.26; p < 0.0001]$. This suggests that at least one of the sources is significantly different from the other sources. The last hypothesis H_0^c is accepted as $[F(6, 228) = 0.94; p = 0.46]$. There is no significant interaction between any of the algorithms and the type of the source.

In both scenes, it has been found that there is a significant difference between the performance of the JBLCMV and the proposed method with different enhancements. To understand which ones are significantly different from the JBLCMV, post-hoc analysis is done using Tukey's test [57]. The summary of the results can be seen in Table III.

The Tukey's significance test for scene 1 shows that there is a significant difference between the JBLCMV and $ILD_{1.0}$ with $p = 0.0232$. Moreover $ILD_{0.2}$ and $ILD_{1.0}$ are also found to be significantly different with $p = 0.0242$. There is

no significant difference captured among the other methods. As both the JBLCMV and $ILD_{0.2}$ are significantly different from $ILD_{1.0}$, it can be deduced that the loss of information induced by the low-frequency ITD cues has been overcome by the enhancement of the ILDs in the low-frequency range.

Tukey's significance test for scene 2 however, does not show any significant differences among the proposed methods although a similar trend can be observed. The p-value between the JBLCMV and $ILD_{1.0}$ is 0.1020, between $ILD_{0.2}$ and $ILD_{0.6}$ is 0.1640 and between $ILD_{0.2}$ and the JBLCMV is $p = 0.9964$. The significant improvement of horizontal localization performance that is observed for the anechoic scene is not observed for the reverberant scene.

Since no interaction has been found between the proposed beamformers and the audio signals, further analysis has not been conducted for these cases.

IX. DISCUSSION

This project aims to understand if the low-frequency ITD information can be exchanged with the ILD information in the same range. The intelligibility and the localization performance of the proposed method have been theoretically examined in Section VII. A psychoacoustic analysis is done using a listening experiment to understand if indeed there is any improvement in localization performance as reported in the literature.

The proposed method targets hearing-impaired people with a low temporal fine structure processing ability. There are several methods in the literature such as [58], which focuses on measuring the TFS sensitivity. By understanding the auditory capabilities of the hearing impaired, those who will benefit from the proposed method can be selected. In addition to this, cochlear hearing loss is known to cause a deficiency in temporal fine structure processing [59]. The hearing-impaired people with cochlear hearing loss has been known to be utilizing the envelope time differences and the level differences in the high frequencies. This makes the proposed method, which preserves the phase and the levels in the high frequency while replacing the low-frequency ITD information with ILD a suitable method for those who suffer from cochlear hearing loss.

In our experiment, we used normal hearing subjects to understand if the low-frequency ILD cues can be used to overcome the lack of ITD cues in the same region. In an attempt to imitate the deficiency of TFS processing, we have introduced low-frequency ITD errors during beamforming. From the binaural study given in Fig. 1, we have found out that some of the hearing-impaired people are able to hear some of the low-frequency ITD cues. We have created a scenario where the frequencies below 800 Hz do not contain reliable ITD information, whereas the frequencies above 800 Hz do. With the introduction of near-field ILD cues below 800 Hz, we have observed improved localization performance. The hearing-impaired people share a similar condition where six of the attendees were able to hear some of the ITD cues in the low-frequency region whereas, nine of them could not hear any at all. We believe the listening test on normal hearing people generalizes to the hearing-impaired people, where the attendees were asked to localize a sound

source with deficient ITD cues. However, the same performance that is observed for normal hearing people might not be observed for hearing-impaired people. This should be further tested.

The results in Section VII-A suggest that the low-frequency ITD information is not directly related to intelligibility or its effect below 800 Hz can be compensated by the information that exists above 800 Hz. The intelligibility metric SIIB shows that the enhancement of the low-frequency ILDs does not disrupt the intelligibility of the signals as the proposed method has better intelligibility compared to JBLCMV. In the literature, a high spatial release from masking is reported when the low-frequency ILDs are introduced. This effect has not been observed in the intelligibility metrics and should be examined in future work.

Section VII-B covers the localization performance of the proposed method compared to the BMVDR and the JBLCMV using the binaural cues. It can be seen that Problem 3 given in (36) has overall better performance. The desired low-frequency ILDs can be reached with less error compared to Problem 2 given in (32). In addition to the interferer localization properties, the target is left distortionless for any number of interferers as expected. There is no objective localization measure that assesses the localization ability of the hearing impaired that includes auditory deficiencies. For this reason, the binaural cues preservation performance is examined in an attempt to show the performance of the optimization problem at solving the original problem.

The results in Section VIII suggest that it is possible to replace the ITD information using low-frequency ILDs for the anechoic scene. In the averaged results, which can be seen in Fig. 9, the localization performance of $ILD_{0.2}$ is close to the JBLCMV. On the other hand, $ILD_{0.6}$ and $ILD_{1.0}$ have a higher variance and a higher mean error compared to $ILD_{0.2}$ and the JBLCMV. Although a similar trend has been observed for the reverberant scene, a significant difference has not been observed according to the statistical tests. The reason could be the worsened localization performance of normal hearing listeners at reverberant environments or the incoherence between the proposed ILD enhancement function DVF_ILD and the naturally occurring ILD enhancement in reverberant environments. We have used a spherical head model to calculate the DVF_ILD , which does not capture the reflection patterns of reverberation. A more sophisticated enhancement function might give significant results.

As the magnitude of the ILD cues in the low-frequency region increases, the localization performance improves. This behaviour can be explained by the decrease in the audible angle with increasing ILD magnitude. It might be possible to differentiate the difference between the 90° and 60° when enhancement with respect to 0.2 m is applied whereas, it might not be possible when enhancement with respect to 0.6 m is applied. In addition, the enhancement amount is expected to be dependent on the sensitivity of the hearing-impaired listener to ILD differences. On the other hand, a higher enhancement might impair the distance perception of the hearing-impaired user since the low-frequency ILD cues are mainly used as distance cues. This relation will be examined in the future work to understand the trade-off.

X. CONCLUSION

In this research project, a step towards manipulating the acoustic scene according to the hearing loss has been taken. The low-frequency ITD cues, which some of hearing-impaired people have shown problems of processing, are transformed into ILD cues using a near-field transformation. The listening test suggests that the localization ability of the normal listeners can be improved when enhancement with respect to 0.2 m has been applied for the anechoic scene. This amount is expected to be user-dependent and can be tailored using psychoacoustic analysis. An intelligibility and localization experiment on hearing-impaired people with deficient TFS processing is left for future work.

REFERENCES

- [1] A. W. Bronkhorst, "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," *Attention, Perception, Psychophys.*, vol. 77, no. 5, pp. 1465–1487, 2015.
- [2] J. Kates, *Digital Hearing Aids, Ser. G. - Reference, Information and Interdisciplinary Subjects Series*. Plural Pub., 2008. [Online]. Available: <https://books.google.nl/books?id=Pu07IQAACAAJ>
- [3] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, J. Jensen, and M. Guo, "Evaluation of binaural noise reduction methods in terms of intelligibility and perceived localization," in *Proc. 26th Eur. Signal Process. Conf.*, 2018, pp. 2429–2433.
- [4] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.
- [5] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 1, pp. 137–152, Jan. 2017.
- [6] S. Doclo, T. J. Klases, T. Van den, J. BogaertWouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel wiener filtering and interaural transfer functions," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2006, pp. 1–4.
- [7] A. W. Archer-Boyd, J. A. Holman, and W. O. Briemjoiin, "The minimum monitoring signal-to-noise ratio for off-axis signals and its implications for directional hearing aids," *Hear. Res.*, vol. 357, pp. 64–72, 2018.
- [8] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997.
- [9] W. M. Hartmann, "How we localize sound," *Phys. Today*, vol. 52, pp. 24–29, 1999.
- [10] B. Bardsley, "The use of spatial cues by hearing-impaired listeners," Ph.D. dissertation, Cardiff Univ., U.K., 2019.
- [11] C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, and B. C. J. Moore, "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Nat. Acad. Sci.*, vol. 103, no. 49, pp. 18866–18869, 2006.
- [12] N. Marrone, C. R. Mason, and G. Kidd, "Evaluating the benefit of hearing aids in solving the cocktail party problem," *Trends Amplification*, vol. 12, no. 4, pp. 300–315, Dec. 2008.
- [13] J. H. Grose and S. K. Mamo, "Processing of temporal fine structure as a function of age," *Ear Hear.*, vol. 31, no. 6, pp. 755–760, 2010.
- [14] R. L. Miller, J. R. Schilling, K. R. Franck, and E. D. Young, "Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers," *J. Acoust. Soc. Amer.*, vol. 101, no. 6, pp. 3602–3616, 1997.
- [15] M. A. Ruggero, "Cochlear delays and traveling waves: Comments on 'experimental look at cochlear mechanics': [a. dancer, *Audiology* 1992; 31: 301-312] ruggero," *Audiology*, vol. 33, no. 3, pp. 131–142, 1994.
- [16] R. Badri, J. H. Siegel, and B. A. Wright, "Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms," *J. Acoustical Soc. Amer.*, vol. 129, no. 2, pp. 852–863, 2011.
- [17] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 544–548.
- [18] D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 8648–8652.
- [19] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, "Microphone-array hearing aids with binaural output. I. Fixed-processing systems," *IEEE Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.
- [20] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multi-microphone noise reduction preserving binaural cues," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016, pp. 460–464.
- [21] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.
- [22] G. Kidd Jr, C. R. Mason, V. Best, and N. Marrone, "Stimulus factors influencing spatial release from speech-on-speech masking," *J. Acoustical Soc. Amer.*, vol. 128, no. 4, pp. 1965–1978, 2010.
- [23] B. C. Moore, A. Kolarik, M. A. Stone, and Y.-W. Lee, "Evaluation of a method for enhancing interaural level differences at low frequencies," *J. Acoustical Soc. Amer.*, vol. 140, no. 4, pp. 2817–2828, 2016.
- [24] W. Feddersen, T. Sandel, D. Teas, and L. Jeffress, "Localization of high-frequency tones," *J. Acoustical Soc. Amer.*, vol. 29, no. 9, pp. 988–991, 1957.
- [25] T. Francart, T. Van den M. Bogaert Moonen, and J. Wouters, "Amplification of interaural level differences improves sound localization in acoustic simulations of bimodal hearing," *J. Acoustical Soc. Amer.*, vol. 126, no. 6, pp. 3209–3213, 2009.
- [26] T. Francart, A. Lenssen, and J. Wouters, "Enhancement of interaural level differences improves sound localization in bimodal hearing," *J. Acoustical Soc. Amer.*, vol. 130, no. 5, pp. 2817–2826, 2011.
- [27] B. Dieudonné and T. Francart, "Head shadow enhancement with low-frequency beamforming improves sound localization and speech perception for simulated bimodal listeners," *Hear. Res.*, vol. 363, pp. 78–84, 2018.
- [28] B. Rana and J. M. Buchholz, "Better-ear glimpsing at low frequencies in normal-hearing and hearing-impaired listeners," *J. Acoustical Soc. Amer.*, vol. 140, no. 2, pp. 1192–1205, 2016.
- [29] B. G. Shinn-Cunningham, J. Schickler, N. Kopčo, and R. Litovsky, "Spatial unmasking of nearby speech sources in a simulated anechoic environment," *J. Acoustical Soc. Amer.*, vol. 110, no. 2, pp. 1118–1129, 2001.
- [30] C. Jin, V. Best, G. Lin, and S. Carlile, "Spatial unmasking of speech based on near-field distance cues," *Adv. Biomed. Eng.*, pp. 3–19, Aug. 2011.
- [31] J. W. Strutt, "XII, on our perception of sound direction," London, *Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 13, no. 74, pp. 214–232, 1907.
- [32] L. R. Bernstein and C. Trahiotis, "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise," *J. Acoustical Soc. Amer.*, vol. 95, no. 6, pp. 3561–3567, 1994.
- [33] B. C. Moore, "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," *J. Assoc. Res. Otolaryngol.*, vol. 9, no. 4, pp. 399–406, 2008.
- [34] B. C. Moore, D. A. Vickers, and A. Mehta, "The effects of age on temporal fine structure sensitivity in monaural and binaural conditions," *Int. J. Audiol.*, vol. 51, no. 10, pp. 715–721, 2012.
- [35] Y. Sergeenko, K. Lall, M. C. Liberman, and S. G. Kujawa, "Age-related cochlear synaptopathy: An early-onset contributor to auditory functional decline," *J. Neurosci.*, vol. 33, no. 34, pp. 13686–13694, 2013.
- [36] K. Hopkins and B. C. Moore, "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," *J. Acoustical Soc. Amer.*, vol. 130, no. 1, pp. 334–349, 2011.
- [37] D. R. Perrott and A. D. Musicant, "Rotating tones and binaural beats," *J. Acoustical Soc. Amer.*, vol. 61, no. 5, pp. 1288–1292, 1977.
- [38] D. S. Brungart, "Near-field auditory localization," Ph.D. dissertation, Massachusetts Inst. Technol., 1998.
- [39] G. Yu, R. Wu, Y. Liu, and B. Xie, "Near-field head-related transfer-function measurement and database of human subjects," *J. Acoustical Soc. Amer.*, vol. 143, no. 3, pp. EL194–EL198, 2018.
- [40] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.
- [41] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.
- [42] J. Park and S. Boyd, "General heuristics for nonconvex quadratically constrained quadratic programming," 2017, *arXiv:1703.07870*.

- [43] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010, doi: [10.1109/MSP.2010.936019](https://doi.org/10.1109/MSP.2010.936019).
- [44] A. Qualizza, P. Belotti, and F. Margot, "Linear programming relaxations of quadratically constrained quadratic programs," *Mixed Integer Nonlinear Programming*. New York, NY, USA: Springer, 2012, pp. 407–426.
- [45] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Baltimore, Maryland, MD, US: The Johns Hopkins Univ. Press, 1996.
- [46] A. King, K. Hopkins, and C. J. Plack, "The effects of age and hearing loss on interaural phase difference discrimination," *J. Acoustical Soc. Amer.*, vol. 135, no. 1, pp. 342–351, 2014.
- [47] B. C. Moore, *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*. World Scientific, 2014.
- [48] H. Wierstorf, M. Geier, A. Raake, and S. Spors, "A free database of head-related impulse response measurements in the horizontal plane with multiple distances," in *Proc. 130th Conv. Audio Eng. Soc.*, May 2011.
- [49] C. Pörschmann, J. M. Arend, and A. Neidhardt, "A spherical near-field hrtf set for auralization and psychoacoustic research," in *Proc. Audio Eng. Soc. Conv. 142*, May 2017. [Online]. Available: <http://www.aes.org/e-lib/browse.cfm?elib=18697>
- [50] W. M. Rabinowitz, J. Maxwell, Y. Shao, and M. Wei, "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," *Presence: Teleoperators Virtual Environ.*, vol. 2, no. 2, pp. 125–129, 1993.
- [51] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoustical Soc. Amer.*, vol. 104, no. 5, pp. 3048–3058, 1998.
- [52] A. Kan, C. Jin, and A. van Schaik, "A psychophysical evaluation of near-field head-related transfer functions synthesized using a distance variation function," *J. Acoustical Soc. Amer.*, vol. 125, no. 4, pp. 2233–2242, 2009.
- [53] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, "An instrumental intelligibility metric based on information theory," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 115–119, Jan. 2018.
- [54] J. S. Garofolo, "TIMIT acoustic-phonetic continuous speech corpus LDC93S1," 1993, doi: [10.35111/17gk-bn40](https://doi.org/10.35111/17gk-bn40).
- [55] H. Kayser, S. D. Ewert, J. Anemüller, 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. Adv. Signal Process.*, vol. 2009, no. 1, 2009, Art. no. 298605.
- [56] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NIST speech disc 1-1.1," *NASA STI/Recon Tech. Rep. N*, vol. 93, 1993, Art no. 27403.
- [57] J. W. Tukey, "Comparing individual means in the analysis of variance," *Biometrics*, vol. 5, no. 2, pp. 99–114, 1949.
- [58] D. S. Mathew, A. Sreenivasan, A. Alexander, and S. Palani, "Measuring binaural temporal-fine-structure sensitivity in hearing-impaired listeners, using the TFS-AF test," *J. Amer. Acad. Audiol.*, vol. 31, no. 02, pp. 105–110, 2020.
- [59] K. Hopkins, B. C. Moore, and M. A. Stone, "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," *J. Acoustical Soc. Amer.*, vol. 123, no. 2, pp. 1140–1153, 2008.

Metin Calis (Graduate Student Member, IEEE) received the B.Sc. degree from Bilkent University, Ankara, Turkey, in 2016 and the M.Sc. degree in a topic related to privacy preserving distributed wireless networks from the Delft University of Technology, Delft, the Netherlands, in 2019. He worked as a Research Assistant for the NWO funded project restored sound localization for hearing impaired people at the Circuits and Systems (CAS) group with Tu Delft until 2020. He is currently working toward the Ph.D. degree on tensor techniques to improve the analysis of contrast-enhanced ultrasound sequences. His main research interests include audio and acoustic signal processing and tensor decomposition methods.

Steven van de Par studied physics with the Eindhoven University of Technology, Eindhoven, The Netherlands, and received the Ph.D. degree in a topic related to binaural hearing from the Eindhoven University of Technology, in 1998. He was a Postdoctoral Researcher with the Eindhoven University of Technology, where he studied auditory-visual interaction and was a Guest Researcher with the University of Connecticut Health Center. In early 2000, he joined Philips Research, Eindhoven, to do applied research in auditory and multisensory perception, low-bit-rate audio coding and music information retrieval. Since April 2010, he holds a Professor position in acoustics with the University of Oldenburg, Germany with a research focus on the fundamentals of auditory perception and its application to virtual acoustics, vehicle acoustics, and digital signal processing. He has authored or coauthored various papers on binaural auditory perception, auditory-visual synchrony perception, audio coding and computational auditory scene analysis.

Richard Heusdens (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from the Delft University of Technology, Delft, The Netherlands, in 1992 and 1997, respectively. Since 2002, he has been an Associate Professor with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. In the spring of 1992, he joined the digital signal processing group with the Philips Research Laboratories, Eindhoven, The Netherlands. He has worked on various topics in the field of signal processing, such as image/video compression and VLSI architectures for image processing algorithms. In 1997, he joined the Circuits and Systems Group of Delft University of Technology, where he was a Postdoctoral Researcher. In 2000, he moved to the Information and Communication Theory (ICT) Group, where he became an Assistant Professor responsible for the audio/speech signal processing activities within the ICT group. He held Visiting Positions with KTH Royal Institute of Technology, Sweden, in 2002 and 2008, and was a Guest Professor with Aalborg University from 2014 to 2016. Since 2019, he is a Full Professor with the Netherlands Defence Academy. He is involved in research projects that cover subjects such as audio and acoustic signal processing, sensor signal processing, distributed optimization and security/privacy.

Richard Christian Hendriks was born in Schiedam, The Netherlands. He received the B.Sc., M.Sc. (*cum laude*), and the Ph.D. (*cum laude*) degrees in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 2001, 2003, and 2008, respectively. He is currently an Associate Professor with the Circuits and Systems (CAS) Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. His main research interest include biomedical signal processing, audio and speech processing, including speech enhancement, speech intelligibility improvement, and intelligibility modelling. In March 2010, he received the prestigious VENI grant for his proposal "Intelligibility Enhancement for Speech Communication Systems." He was the recipient of the several best paper awards, among which the IEEE Signal Processing Society Best Paper Award in 2016. He is an Associate Editor of the IEEE/ACM TRANSACTION ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and the *EURASIP Journal on Advances in Signal Processing*. He is an elected member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing.