Aalborg Universitet

Personalized signal-independent beamforming for binaural hearing aids

 H. Moore, Alastair; Mark de Haan, Jan; Syskind Pedersen, Michael; A. Naylor, Patrick; Brookes, Mike; Jensen, Jesper

Link to publication from Aalborg University

Personalized signal-independent beamforming for binaural hearing aids

Alastair H. Moore, Jan Mark de Haan, Michael Syskind Pedersen, Patrick A. Naylor, Mike Brookes, and Jesper Jensen

ARTICLES YOU MAY BE INTERESTED IN

JASA
THE JOURNAL OF THE
ACOUSTICAL SOCIETY OF AMERICA

WHY PUBLISH WITH US?

Personalized signal-independent beamforming for binaural hearing aids

Alastair H. Moore,[1,a)] Jan Mark de Haan,[2] Michael Syskind Pedersen,[2] Patrick A. Naylor,[1] Mike Brookes,[1] and Jesper Jensen[3,b)]

[1]*Department of Electrical & Electronic Engineering, Imperial College, London, SW7 2AZ, United Kingdom*
[2]*Oticon A/S, 2765 Smørum, Denmark*
[3]*Department of Electronic Systems, Aalborg University, 9220 Aalborg Øst, Denmark*

The effect of personalized microphone array calibration on the performance of hearing aid beamformers under noisy reverberant conditions is studied. The study makes use of a new, publicly available, database containing acoustic transfer function measurements from 29 loudspeakers arranged on a sphere to a pair of behind-the-ear hearing aids in a listening room when worn by 27 males, 14 females, and 4 mannequins. Bilateral and binaural beamformers are designed using each participant's hearing aid head-related impulse responses (HAHRIRs). The performance of these personalized beamformers is compared to that of mismatched beamformers, where the HAHRIR used for the design does not belong to the individual for whom performance is measured. The case where the mismatched HAHRIR is that of a mannequin is of particular interest since it represents current practice in commercially available hearing aids. The benefit of personalized beamforming is assessed using an intrusive binaural speech intelligibility metric and in a matrix speech intelligibility test. For binaural beamforming, both measures demonstrate a statistically signficant ($p < 0.05$) benefit of personalization. The benefit varies substantially between individuals with some predicted to benefit by as much as 1.5 dB. © *2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/ licenses/by/4.0/).* https://doi.org/10.1121/1.5102173

I. INTRODUCTION

Multichannel speech enhancement is important in many applications, including telecommunications, robot audition, and hearing aids. Signal-dependent beamformers adapt their filter weights according to the observed signals and so have the potential to be always-optimal according to some specified design criterion. However, errors in estimating the signal and noise statistics, for example, due to inaccurate voice activity detection,[1] may lead to degraded performance. A common approach to implementing a signal-dependent beamformer is to use a generalized sidelobe canceller (GSC) comprising a signal independent beamformer, a blocking matrix, and an adaptive noise canceller.[2] Signal-dependent beamformers are sensitive to signal cancellation due to steering errors and multipath propagation.[3,4]

Signal-independent beamformers, also known as fixed beamformers, use *a priori* knowledge or assumptions about the source direction and noise characteristics to determine the filter weights. They are computationally efficient and robust at low signal-to-noise ratio (SNRs) but suboptimal if the sound scene or array characteristics differ from those used in the filter design.[5] For example, using a free-field propagation model to describe the acoustic transfer function (ATF) between a source and a head-mounted array degrades

performance.[6] The effect of steering errors can be inferred from the shape of the beampattern's main lobe, which in general gets narrower with the number of microphones and their spacing.[5] Differences in array element sensitivity, due for example, to manufacturing tolerances and component ageing, have been mitigated through robust offline beamformer design[7] and adaptive approaches.[8,9] A number of studies have demonstrated a benefit in performance when microphone arrays are individually calibrated rather than by using an idealized model of the geometry.[10–12] This calibration accounts for any sensitivity or frequency response variations between microphones and also acoustic differences in the scattering effect of the array enclosure.

In the context of binaural hearing aids, it is common to design beamformers assuming an average head, as typified by an acoustic mannequin. To investigate the potential benefit of individual calibration, or personalization, in the context of binaural hearing aids requires an understanding of the variability of the ATF when a particular pair of hearing aids is worn by different people. We refer to this ATF as the hearing aid head-related transfer function (HAHRTF), or equivalently its time domain representation, the hearing aid head-related impulse response (HAHRIR). Most available databases[13–17] of HAHRIR measurements have used either a single mannequin or a single human[17] and so are unsuitable for investigating such differences; a notable exception is a recent database[18] that includes measurements of 16 subjects and 3 dummy heads. Comparing measurements between

a)Electronic mail: alastair.h.moore@imperial.ac.uk
b)Also at: Oticon A/S, 2765 Smørum, Denmark

these databases is also not helpful since each uses a different hearing aid device. Measurements of conventional, two-channel head-related impulse response (HRIRs) are typically made in the ear-canal or at the blocked entrance to the ear canal. Databases of such measurements, for example, CIPIC,[19] LISTEN,[20] and ARI,[21] suggest that there are large differences between individuals. These differences arise from differences in the head geometry and direction-dependent resonances of the pinnae and primarily affect localization accuracy in the horizontal and vertical dimensions, respectively. Quantifying the extent to which the use of personalized or generic HAHRTFs in hearing aid beamforming affects intelligibility for a human listener is important in determining whether commercially available hearing aids could benefit from personalized processing.

In this article we describe a newly collected and publicly available database of 46 HAHRIRs measurements which will, for the first time, allow the benefits of individual calibration of hearing aid arrays to be investigated. There are many ways in which inter-individual variability in the measured ATFs could be analyzed. Also, there are many algorithms and signal processing tasks which may benefit from individual array calibration. In this article, we focus on signal-independent beamforming because it is widely used in practice.[22]

The remainder of this article is organized as follows. Section II briefly describes the acoustic measurements and post-processing. Section III presents the beamforming problem. Section IV presents the method by which the newly acquired database is used to simulate noisy, reverberant speech signals and to design matched and mismatched beamformers for its enhancement. The performance of the beamformers is evaluated and analyzed using signal-based metrics in Sec. V. In Sec. VI, the validity of this analysis is confirmed using a headphone-based matrix speech intelligibility test in which 11 participants heard a virtual representation of unprocessed and enhanced sound scenes where all stimuli were created using the participant's own HAHRIRs. Finally, the article is concluded in Sec. VII.

II. DATABASE OF ACOUSTIC MEASUREMENTS

Acoustic measurements were performed in an acoustically treated listening room of dimensions $7.9 \times 6.0 \times 3.5$ m with a reverberation time of 250 ms. Loudspeakers (Genelec 8050A) were arranged on the surface of a sphere of radius 1.9 m in three equally spaced horizontal rings. Following the AES standard[23] spherical co-ordinates system, loudspeakers were located at elevation, $\theta = 0°$, azimuth, $\phi \in \{0°, 22.5°,…, 337.5°\}$ and at $\theta \in \pm45°$, $\phi \in \{30°, 90°,…, 330°\}$. An additional loudspeaker at $\theta = 90°$ was at a radius of 1.6 m giving a total of $1 + 6 + 16 + 6 = 29$ loudspeakers. For brevity of notation, directions are expressed as unit vectors, $\mathbf{u}$, in Cartesian coordinates according to

$$\mathbf{u} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}. \tag{1}$$

Microphones were embedded in hearing aid shells (Oticon Epoq) with two behind the ear (BTE) microphones spaced 10 mm apart with an additional microphone in the ear canal secured by a generic vented silicone dome, as shown for the left ear in Fig. 1. As far as possible, a typical positioning of the hearing aid devices was obtained by allowing participants to insert the in-ear microphones in the ear canal themselves. However, to avoid snagging of the cables, the BTE devices were placed over the ear by the experimenter, and, since the same transducers were used for all participants, the wire lengths adjoining in-ear microphones to the shells were not customized for each individual. Participants were seated on a chair positioned at the centre of the loudspeaker array and the height of the chair adjusted such that the BTE microphones were in the horizontal plane of the loudspeaker array, thereby ensuring the middle ring of loudspeakers corresponds to an elevation of 0°.

Microphone signals were amplified using custom pre-amplifiers. Measurement signal output and acquisition was performed with 24-bit precision at 44.1 kHz sample rate using a Ferrofish DANTE A32 interface. A direct loopback connection between an output and input was used to measure the internal delay of the measurement system which was removed in post-processing.

Impulse responses, $g_{m,i}(t)$, between loudspeaker i and microphone m were measured using the exponential sine sweep method[24] using a 370 ms sweep between 50 Hz and 16 kHz. The sweep duration was chosen to give an acceptable compromise between the time taken to make the measurements, and hence the risk of head movements, and the SNR of the impulse responses. For the subset of microphones and directions used in this study (see Sec. II A), the average SNR of the direct path component was 55 dB. All 29
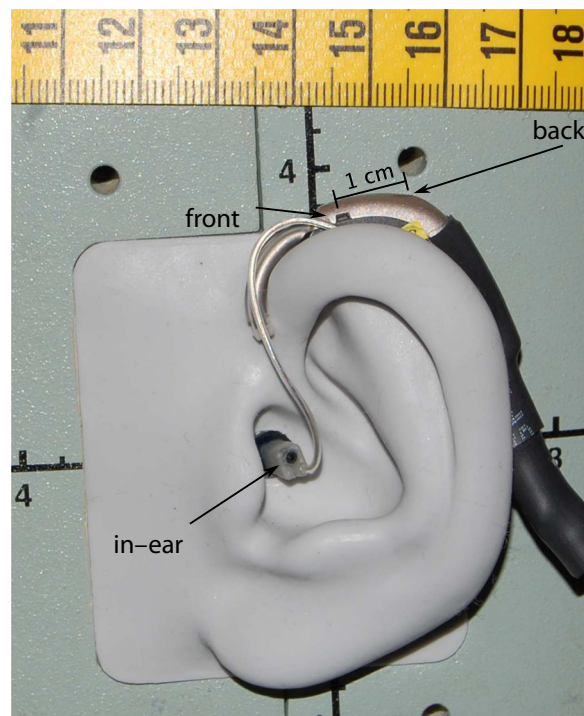

FIG. 1. (Color online) Hearing aid shell mounted on left ear of HATS with front, back, and in-ear microphone locations annotated.

directions were measured in succession with 370 ms silence between each sweep giving a total measurement time of about 20 s. The measurements were repeated with the participant rotated by $-15°$ and $-30°$ so that direct path impulse responses are available with 7.5° resolution on the horizontal plane and 30° resolution at elevations $±45°$. For each rotation, the measurements were repeated three times to allow for consistency checks. Only one measurement of each direction/rotation combination is retained in the final database. In all, the acoustic measurements took about 10 min.

A total of 46 sets of HAHRIRs were made. This consists of 27 males, 14 females, and 4 mannequins, of which one, a head and torso simulator (HATS), was measured twice. The other three mannequins were three-dimensionally (3D) printed head models of real people mounted on an artificial torso.

While it is well known that head-related transfer function (HRTFs) have a complicated dependency on the fine structural details of the outer ear, head, and torso of an individual,[19,25] a partial fit of generic HRTFs can be obtained using gross measurements of head size to control interaural cues.[26,27] Hypothesizing that a similar approach may also be possible for partially-personalized hearing aid beamforming without the need for individual acoustic measurements, the height, depth, and width of each participant's head was measured using calipers and the circumference was measured using a tape measure, since these measurements could rapidly be made in an audiology clinic. Table I shows the distribution of head circumferences for the human (i.e., non-mannequin) participants. For comparison, the head circumference of the HATS is 55.9 cm.

Prior to each measurement session, calibration measurements were made, without the chair present, from each loudspeaker to a G.R.A.S 46AE reference microphone and amplifier set positioned at the centre of the array. Similar calibration measurements were also made for the left and right hearing aids separately. Compared to the mean sensitivity, microphones varied by $≤0.4$ dB. With the exception of the overhead loudspeaker, loudspeaker sensitivities varied by $≤0.8$ dB and propagation delay from each loudspeaker to the center of the array varied by $≤93$ μs.

Since the focus of this study is the acoustic differences between individuals and the signals emitted from each loudspeaker are uncorrelated, the experiments reported in Secs. V and VI were performed without any further calibration.

A. Study database

A subset of the raw impulse responses was selected, comprising those between the 16 horizontal plane

TABLE I. Distribution of head circumference measurements for human participants. Circumference of HATS is 55.9 cm.

Percentile	Circumference (cm)		
	Male	Female	Human
0	56.4	54.6	54.6
25	58.7	56.1	56.7
50	59.8	56.4	58.9
75	60.4	57.3	60.1
100	62.7	59.2	62.7

loudspeakers and the 4 BTE microphones, with the chair in its initial, front-facing, position. The complete measured hearing aid room impulse response (HARIR), $g_{m,i}(t)$, between the ith loudspeaker and the mth microphone may be decomposed into the sum of two components

$$g_{m,i}(t) = h_m(\mathbf{u}_i, t) + \tilde{h}_{m,i}(t), \tag{2}$$

where $h_m(\mathbf{u}, t)$, the HAHRIR, is the impulse response at time t due to a plane wave with direction of arrival (DOA) $\mathbf{u}$ and $\tilde{h}_{m,i}$ includes all the later-arriving room reflections. For the purposes of this study, $h_m(\mathbf{u}_i, t)$ is obtained from $g_{m,i}(t)$ by cropping the first 10.7 ms. A raised cosine fade out is applied to the last 1 ms of the cropped response. Cropping a reverberant measured impulse response to obtain a quasianechoic measurement is possible because the response to the direct path wavefront decays rapidly and is already approximately 30 dB below the peak response before the first reflection from the room arrives.

B. Complete database

Whilst it was not necessary for the current study, for some applications, such as simulating the experience of listening in a virtual sound environment,[28] very precise matching of the loudspeaker array is desirable. Therefore, for the convenience of future users, a calibrated database which compensates for transducer sensitivity differences and time of arrival offsets has also been produced.

The complete database of as-meaured HARIRs, calibration impulse responses, study database and calibrated database are publicly available.[29] Futhermore, MATLAB scripts to produce the study and calibrated databases from the as-measured database are also publicly available.[30]

III. PROBLEM FORMULATION

Expressed in the frequency domain, the observed signal, $Y_m(\omega)$, at the mth microphone in an array is

$$Y_m(\omega) = X_m(\omega) + V_m(\omega), \tag{3}$$

where $X_m(\omega)$ is the signal due to the desired source and $V_m(\omega)$ is the unwanted signal due to reverberation, acoustic noise, and sensor noise. The signals for an array of M microphones are expressed in vector notation as

$$\mathbf{y}(\omega) = \mathbf{x}(\omega) + \mathbf{v}(\omega), \tag{4}$$

where $\mathbf{y}(\omega) = [Y_1(\omega) \quad \cdots \quad Y_M(\omega)]^T$ and $(\cdot)^T$ denotes the transpose; the vectors $\mathbf{x}(\omega)$ and $\mathbf{v}(\omega)$ are similarly defined.

For a target source signal, $S(\omega)$, incident from direction $\mathbf{u}_j$, the M channel observation at the array is $\mathbf{x}(\omega) = \mathbf{h}(\mathbf{u}_j, \omega)S(\omega)$, where $\mathbf{h}(\mathbf{u}, \omega) = [H_1(\mathbf{u}, \omega) \quad \cdots \quad H_M(\mathbf{u}, \omega)]^T$, and $H_m(\mathbf{u}, \omega)$ is the Fourier transform of $h_m(\mathbf{u}, t)$.

Defining $m = \alpha$ to be the reference channel, and the relative transfer function (RTF), $\mathbf{d}_\alpha(\mathbf{u}, \omega)$, with respect to the reference channel as

$$\mathbf{d}_\alpha(\mathbf{u}, \omega) \triangleq \mathbf{h}(\mathbf{u}, \omega)/H_\alpha(\mathbf{u}, \omega), \tag{5}$$

J. Acoust. Soc. Am. **145** (5), May 2019

Moore *et al.* 2973

gives $\mathbf{x}(\omega) = \mathbf{d}_\alpha(\mathbf{u}_j,\omega)X_\alpha(\omega)$. Substituting into Eq. (4) gives

$$\mathbf{y}(\omega) = \mathbf{d}_\alpha(\mathbf{u}_j,\omega)X_\alpha(\omega) + \mathbf{v}(\omega), \qquad (6)$$

in which the clean signal at the microphones is described in terms of the RTF, $\mathbf{d}_\alpha(\mathbf{u}_j,\omega)$, and the clean signal, $X_\alpha(\omega)$, observed at the reference microphone.

The aim of beamforming is to obtain an estimate, $Z_\alpha(\omega)$, of $X_\alpha(\omega)$, that is the observation of the target source at the reference microphone which is free from reverberation and noise, by applying a spatial filter, $\mathbf{w}_\alpha(\omega)$, according to

$$Z_\alpha(\omega) = \mathbf{w}_\alpha^H(\omega)\mathbf{y}(\omega), \qquad (7)$$

where $(\cdot)^H$ denotes the conjugate transpose. The minimum variance distortionless response (MVDR) beamformer solution[31,32] to this estimation problem is

$$\mathbf{w}_\alpha(\omega) = \frac{\mathbf{R}_v^{-1}(\omega)\mathbf{d}_\alpha(\mathbf{u}_j,\omega)}{\mathbf{d}_\alpha^H(\mathbf{u}_j,\omega)^H\mathbf{R}_v^{-1}(\omega)\mathbf{d}_\alpha(\mathbf{u}_j,\omega)}, \qquad (8)$$

where $\mathbf{R}_v(\omega) = E\{\mathbf{v}(\omega)\mathbf{v}^H(\omega)\}$ is the noise covariance matrix and $E\{\cdot\}$ denotes the expected value. Regularization of Eq. (8) can improve beamformer robustness[33,34] but is not considered in this study. Assuming $\mathbf{d}_\alpha(\mathbf{u}_j,\omega)$ and $\mathbf{R}_v(\omega)$ are known precisely, substituting Eq. (8) into Eq. (7) gives

$$Z_\alpha(\omega) = X_m(\omega) + \mathbf{w}_\alpha^H(\omega)\mathbf{v}(\omega), \qquad (9)$$

which indicates the desired signal is passed undistorted and the beamformer filters (i.e., reduces) the unwanted noise. The extent of noise reduction that can be achieved depends on the number of microphones, the interchannel coherence of the noise and reverberation, and the DOA of the target.

In *bilateral* beamforming, the left and right hearing aids are considered as two independent arrays, each obtaining an estimate of the desired signal at its own reference channel using only the local $M = 2$ microphones. In *binaural* beamforming, the spatial filter associated with each hearing aid again obtains an estimate of the desired signal at its own reference channel, but the two hearing aids are treated as a single array with $M = 4$ microphones. Therefore, in both bilateral beamforming (denoted "2:2") and binaural beamforming (denoted "4:4"), the MVDR solution ensures that, provided $\mathbf{d}_\alpha(\mathbf{u}_j,\omega)$ is known, the signals will retain the correct binaural cues for the target source. In contrast, the binaural cues associated with the noise will not be preserved.[35] In the case of the bilateral beamformers, the residual noise at each ear depends only on the microphone signals at that ear and so the coherence between the noise at each ear is no higher than the original microphone signals. For the binaural beamformer, the enhanced signals at each ear are two different weighted combinations of the same $M = 4$ microphone signals and so the noise coherence is increased to unity.[36]

In practice, estimates of $\mathbf{d}_\alpha(\mathbf{u}_j,\omega)$ and $\mathbf{R}_v(\omega)$ are obtained using calibration measurements and assumptions about the spatial distribution of the noise, or online using the received signals.[37–39]

The remainder of this paper is focused on investigating the impact of using mismatched ATFs for MVDR beamforming, where the common assumption of isotropic noise is used to calculate $\mathbf{R}_v(\omega)$ and the DOA of the desired source is assumed to be known *a priori*.

IV. METHODS

A. Simulated acoustic scene

Microphone signals are generated which simulate those encountered by a particular individual in the listening room described in Sec. II. Specifically, the time domain microphone signals, $y_m^{(l)}(t)$, for microphone index m and participant index l, are calculated as

$$y_m^{(l)}(t) = \beta g_{m,j}^{(l)}(t) * s(t) + \frac{1}{\sqrt{I}}\sum_{i=1}^{I} g_{m,i}^{(l)}(t) * n_b(t - \Delta_i), \qquad (10)$$

where $g_{m,i}^{(l)}(t)$ is the full HARIR for participant index l from direction index $i = \{1,\dots,I\}$ measured at microphone index m, j denotes the direction index of the target source, $s(t)$, $n_b(t)$ is a babble signal with the same long-term average speech spectrum (LTASS) and power as $s(t)$, Δ_i is a time offset associated with direction i, β is a scalar gain parameter, and $*$ denotes convolution. Sensor noise is not included in the simulated signals.

The babble signal in Eq. (10) is composed of concatenated utterances from each of four male and four female talkers from the IEEE sentences corpus all overlayed to form 8-talker babble and Δ_i is an arbitrary offset, randomly selected for each direction in each simulation, to select a different section of noise. All $I = 16$ measured source directions, spaced $22.5°$ apart on the horizontal plane, are used to create the noise field. The resulting background noise is therefore approximately isotropic around a circle in the horizontal plane but, since $g_{m,i}^{(l)}(t)$ contains the natural reverberation of the room, there is reflected sound energy arriving from all directions including above and below the horizontal plane. Note that the target source, $s(t)$, is also filtered by the full reverberant response for the room and so the simulated sound scene is representative of a real listening environment.

The current study considers scenarios in which the target source is either to the front of the listener ($\phi = 0°$), denoted frontal target, or else towards the listener's left side ($\phi = 67.5°$), denoted lateral target. The reported SNR, $20\log_{10}(\beta)$ dB, represents the ratio of desired source energy to noise energy input into the room. As such any direction-dependent change in level observed at any of the microphones due to the direction-dependent filtering of the HARIR is retained in the simulated stimuli. Further details of the target source material specific to the numerical and human evaluations are described in Secs. V and VI, respectively.

B. Beamformers

All beamformers are designed with *a priori* knowledge of the DOA of the target source and under the assumption that the noise field is cylindrically isotropic, i.e., where uncorrelated noise sources are uniformly distributed around

a circle in the horizontal plane. The *a priori* knowledge of the target DOA represents a realistic use case where the listener can choose to turn to directly face the target (frontal target) or independently steer the beamformer's look direction to be non-frontal (lateral target). The assumption of noise field isotropy is common[5] in signal-independent beamformers and cylindrical isotropy is appropriate since active sound sources tend to be on, or near, the horizontal plane of the listener and floors and ceilings tend to be more absorbent than walls.[40] As in most real rooms, the cylindrical isotropy assumption is slightly incorrect since there are reflections from the floor and ceiling and the contributions of these, along with reflections from the walls, are neither necessarily equally distributed in azimuth, nor uncorrelated with each other. This intentional mismatch between the assumed and simulated noise field ensures that the results of the current study are representative of a real-world use case.

Beamformers are implemented as linear time invariant finite impulse response (FIR) filters whose coefficients are designed in advance based on measured HAHRIRs and a simulated cylindrically isotropic noise field. The filter weights are designed in the frequency domain such that they depend only on the narrowband interchannel covariance of the simulated noise, as expressed in the noise covariance matrix (NCM), rather than its power spectral density (PSD).

For each participant, indexed l', the required HAHRIRs, $h_m^{(l')}(\mathbf{u}_i, t)$, are obtained by cropping the direct path from the HARIRs, as described in Sec. II A. The discrete time Fourier transform (DTFT) of $h_m^{(l')}(\mathbf{u}_i, t)$ is denoted $H_m^{(l')}(\mathbf{u}_i, \nu)$ from which the RTF, $\mathbf{d}_\alpha^{(l')}(\nu)$, is obtained, as in Eq. (5), where ν is the frequency index.

The NCM for a cylindrically isotropic noise field is obtained by simulating the microphone signals which would be observed in such a field according to

$$v_m^{(l')}(t) = \sum_{i=1}^{I} h_{m,i}^{(l')}(t) * n_{w,i}(t), \tag{11}$$

where, similar to Eq. (10), noise sources are arranged at equally-spaced angles around the horizontal plane. In Eq. (11), only the direct path sound propagation is included and the individual noise source signals, $n_{w,i}(t)$, are independent, identically distributed, realizations of white Gaussian noise of 1 s duration. The frequency domain NCM, $\mathbf{R}_v^{(l)}(\nu)$, is obtained as

$$\mathbf{R}_v^{(l)}(\nu) = \frac{1}{\mathcal{L}} \sum_{\ell=1}^{\mathcal{L}} \mathbf{v}_m^{(l')}(\nu, \ell) \mathbf{v}_m^{(l')H}(\nu, \ell), \tag{12}$$

where $V_m^{(l')}(\nu, \ell)$ is the short time Fourier transform (STFT) of $v_m^{(l')}(t)$, ℓ is the time index and $\mathcal{L}$ is the number of frames. Frames are 10.7 ms long, to match the length of $h_m^{(l')}(\mathbf{u}_i, t)$, and non-overlapping, such that each frame is an independent sample of the noise process.

The frequency domain filter weights, $\mathbf{w}_\alpha^{(l')}(\nu)$, are obtained as in Eq. (8). The inverse DTFT transforms the conjugated beamformer weights back into the time domain, a circular shift is applied to ensure causality and a Hamming

window applied to ensure there are no discontinuities in the final FIR filters. This time domain post-processing of the beamformer weights avoids the possible introduction of artefacts due to sharp spectral features. Note that direct application of the frequency domain filter weights is avoided since the resulting filters are inexact,[41,42] time-variant and, in general, lead to STFT coefficients for which there is no realizable real-valued time domain signal.[43,44]

A beamformer is personalized when $l' = l$, that is the same individual's measurements are used in Eq. (10) to simulate the microphone signals and to design the beamformer weights which process them. It should be emphasised that the personalized signal-independent beamformers investigated in this study are perfectly fit to the head and torso acoustics of a particular individual but not to the acoustics of the encountered sound scene.

A beamformer is generic when $l' \neq l$. Of particular interest is the generic beamformer obtained from a HATS designed in accordance with ANSI standard S3.36,[45] since such mannequins are widely used.

In addition, this study investigates the effect of personalization, the effect of binaural ($M = 4$) vs bilateral ($M = 2$) beamforming, and the effect of target direction.

V. EVALUATION USING SIGNAL-BASED METRICS

In this section, the inter-individual differences between HAHRIRs are investigated in terms of signal-based metrics of the resulting MVDR beamformer performance. An illustrative example of directivity patterns of the different beamformers is presented to give some initial insight. Subsequently, a systematic study of the effect of beamformer personalization is conducted using the modified binaural short-time objective intelligibility (MBSTOI) measure[46,47] to predict the expected intelligibility improvements offered by alternative beamformers.

A. Directivity patterns

The normalized A-weighted[48] directivity pattern, $B_\alpha(\mathbf{u})$, is the power output from a beamformer in response to an A-weighted source plane wave with DOA $\mathbf{u}$ relative to the power at the reference microphone when the same A-weighted source plane wave is incident from the front, $\mathbf{u}_0$, i.e.,

$$B_\alpha(\mathbf{u}) = \frac{\int_\omega E\left\{ |\mathbf{w}_\alpha^{(l')}(\omega)^H \mathbf{h}^{(l)}(\mathbf{u}, \omega) S(\omega)|^2 \right\} d\omega}{\int_\omega E\left\{ |\mathbf{e}_\alpha^T \mathbf{h}^{(l)}(\mathbf{u}_0, \omega) S(\omega)|^2 \right\} d\omega}, \tag{13}$$

where $S(\omega)$ is an A-weighted source signal and $\mathbf{e}_\alpha$ is an $M \times 1$ microphone channel selection vector with a 1 in the αth element and zeros elsewhere. When $l' = l$ the beamformer is said to be personalized, denoted "Per," whereas when $l' \neq l$ it is said to be generic, denoted "Gen." The baseline in both cases, denoted "Ref," is the power of the input signal at the reference channel (front-left microphone), which is given by $\int_\omega E\{|\mathbf{e}_\alpha^T \mathbf{h}^{(l)}(\mathbf{u}, \omega) S(\omega)|^2\} d\omega$. In the specific case $l' = 28$, the generic beamformer is referred to as "HATS."

J. Acoust. Soc. Am. **145** (5), May 2019

Moore *et al.* 2975

Each plot in Fig. 2 shows A-weighted directivity patterns at the left ear for five different beamformer configurations when steered towards a frontal target (top row) or a lateral target (bottom row). Each column shows the directivity pattern for a different individual. The "Ref" directivity patterns indicate the natural directivity due to the acoustics of the head and, for any given individual, are independent of the target direction. Comparison of the "Ref" directivity patterns in each column reveals substantial variation between individuals, which is consistent with the literature.[18] In the direction of the target, the response in the "Per" conditions is always identical to that of the "Ref" condition because of the distortionless constraint. According to the normalization in Eq. (13), this corresponds to 0 dB for the frontal target (top row) and an individual-dependent level for the lateral target (bottom row). For all individuals, the directivity pattern for the bilateral beamformers between 300° and 60° is very similar to the "Ref" condition with increased sensitivity on the ipsilateral side. Conversely, the binaural beamformers are more symmetric between the left and right sides. In general, for a frontal target position, the personalized binaural beamformer directivity patterns are much sharper than the personalized bilateral directivity patterns. This is not the case for the lateral target.

Individual s42 (left column) is the same mannequin as in s28 but measured on a different day. Any difference in the directivity pattern between the "Per" and "HATS" conditions for s42 is therefore representative of the variation which might be expected due to, for example, reseating the hearing aids. It also gives an indication of the intrinsic variability due to the measurement setup. For the frontal target, the directivity patterns in "Per" and "HATS" conditions have a very similar shape. The largest differences, 1 and 3 dB for bilateral and binaural, respectively, occur between 120° and 150°. Whilst the bilateral response at 0° is 0 dB, the binaural response is −1 dB. It is possible that this difference is due to a small difference in head alignment during the two sets of measurements. Since bilateral beamformers have a broader main lobe they are also more robust to steering errors.[5] For the lateral target, performance for s42 is essentially identical in the "Per" and "HATS" conditions, suggesting that the beamformers are robust to small differences in the array manifold. The directivity patterns are also the same for bilateral and binaural beamforming, suggesting that in this configuration the contralateral microphones do not contribute substantially.

Participant s34 has a head circumference most similar to s28 (56.0 cm cf. 55.9 cm). Nevertheless, comparison of the "Ref" directivity between s34 and s42 shows differences of up to 6 dB at 180°. The result is that the "HATS" directivity patterns are distorted compared to the "Per" directivity patterns. Performance for the binaural beamformer steered to the front is particularly bad with 1.5 dB of attenuation in the target direction and 3 dB less suppression than the "Per" beamformer over most other angles.

Participant s32 is the least similar to the HATS in terms of head circumference with the biggest head (62.7 cm) of all
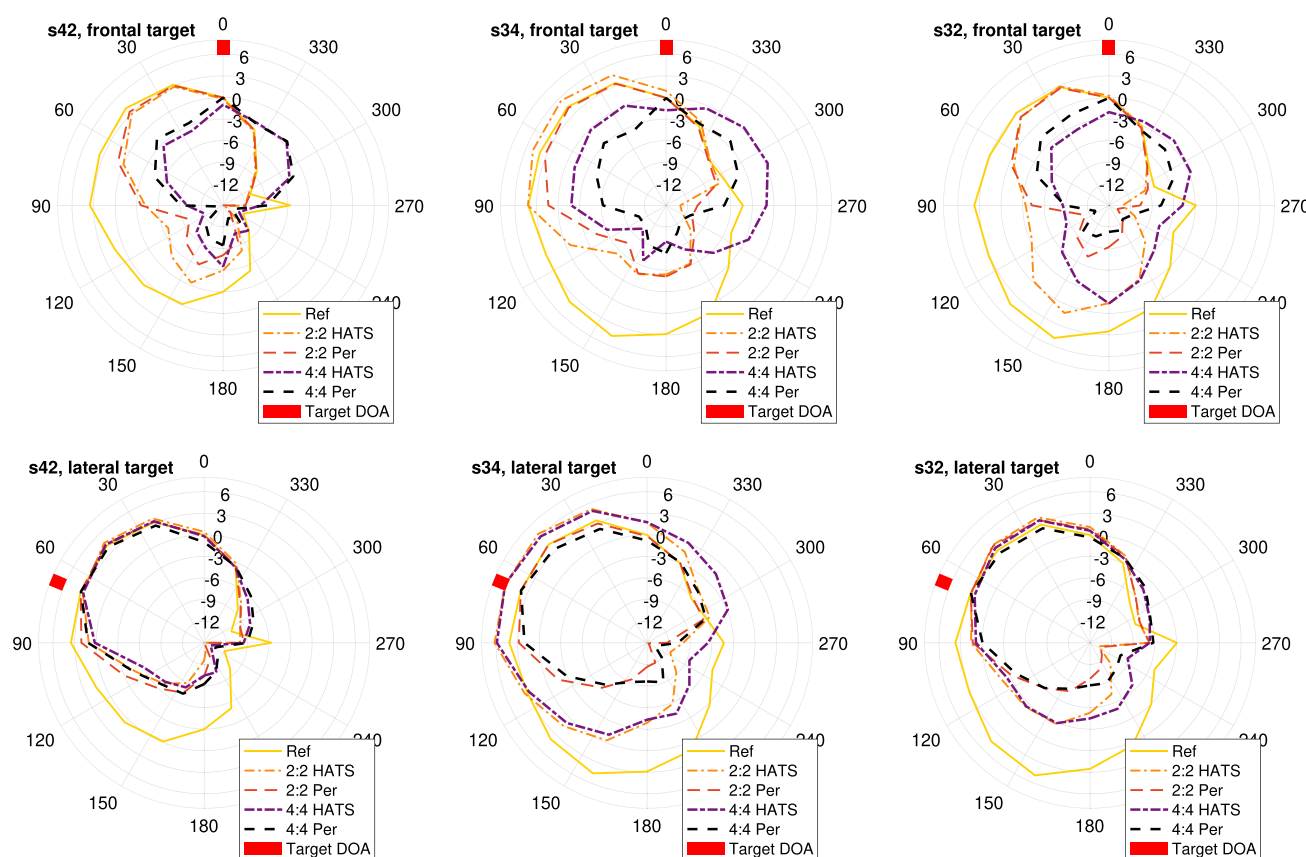

FIG. 2. (Color online) A-weighted directivity patterns of five beamformer configurations steered to (upper row) the front and (lower row) the side for (left column) s42, which is HATS remeasured, (middle column) s34, who has the head circumference most similar to the HATS, and (right column) s32, who has the largest head circumference measured.

tested individuals. Again, the "4:4 HATS" directivity pattern is <0 dB at $0°$ for the frontal target, but in this case more suppression is achieved that with the "4:4 Per" beamformer between 22.5 and $67.5°$. However, this is offset by substantially less attenuation between $112.5°$ and $337.5°$.

With the lateral target, for s32 and s34 there is negligible difference between bilateral and binaural beamforming, but a clear effect of personalization. For some directions towards the rear, the benefit of personalization is 4–5 dB.

In general, the analysis of directivity patterns suggests that, for the frontal target, binaural beamforming always gives a benefit over bilateral beamforming, regardless of personalization, but in both cases, personalization leads to more compact directivity patterns and avoids signal attenuation in the target direction. For the lateral target binaural beamforming offers little, if any, advantage over bilateral, but personalization offers a substantial improvement, particularly in suppression of sound arriving from the rear.

B. Effect on predicted intelligibility

The directivity patterns presented in Sec. V A confirm that the spatial response of the beamformers is consistent with expectations. However, they do not account for the reverberation properties of the room and they do not consider the effect of interaural coherence on binaural hearing. Therefore, directivity patterns do not allow one to easily assess the effective speech intelligibility improvement experienced by a listener in practice. To address this, we (i) simulate noisy reverberant listening conditions, and (ii) assess the effect of hearing aid beamformers using an intrusive measure of predicted binaural speech intellgibility.

1. Method

Reverberant microphone signals for each participant, l, in the database are simulated as described in Sec. IV A. Bilateral and binaural beamformers are designed for each participant, l', in the database, as described in Sec. IV B. Consistent with the steering vectors used, the desired signal at each ear is $\beta h_\alpha^{(l)}(\mathbf{u}_j, t) g_{m,j}^{(l)}(t) * s(t)$ where, as in Eq. (5), α is the index of the reference channel at that ear. That is, the direct path component of the target source as observed at the reference microphones. For each combination of signals and beamformers, the predicted intelligibility of the enhanced binaural signals are computed using the MBSTOI metric[46,47] for input SNRs -17, -14, -11, and -8 dB. It was shown[46,47] that the MBSTOI is able to predict well the intelligibility of speech signals in combined additive noise and reverberation, similar to the ones in this study.

For each combination tested, a logistic model was fitted to the monotonically increasing MBSTOI vs SNR relationship. The SNR at which the model achieved a MBSTOI value of 0.25, compared to the SNR at which the unprocessed noisy signal achieved the same value, was taken to be a measure of the equivalent improvement in input SNR in dB. The MBSTOI value of 0.25 was chosen as the midpoint between the best-case value at -17 dB (0.21) and the worst-case value at -8 dB (0.29). The choice of reference is not critical since the slopes of the logistic models are very similar. This measure of equivalent SNR improvement allows an intuitive way to compare the predicted binaural intelligibility benefit of personalized ($l' = l$) beamformers and all possible non-personalized ($l' \neq l$) beamformers.
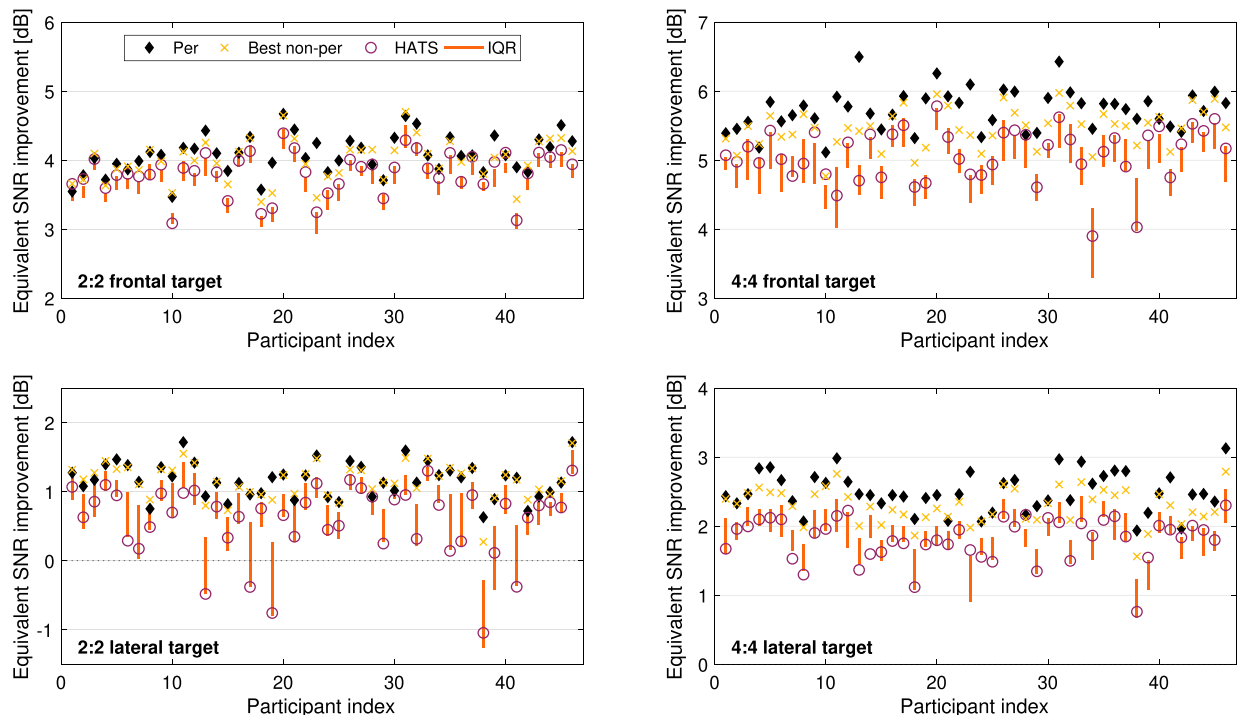

FIG. 3. (Color online) Improvement in MBSTOI, expressed as equivalent improvement in SNR, due to personalized beamforming, denoted "Per," compared to "Ref" condition for (left col) bilateral and (right col) binaural with (top row) frontal target and (bottom row) lateral target. For comparison, the improvements obtained with the HATS beamformer, the best non-personalized beamformer, denoted "Best non-per" and the interquartile range, denoted "IQR," over all possible non-personalized beamformers are also shown. Note that the ordinate axes have the same scale but are offset.

J. Acoust. Soc. Am. **145** (5), May 2019

Moore *et al.*   2977

2. Results and discussion

The equivalent SNR improvement for bilateral and binaural beamformers is shown in the left and right columns of Fig. 3, respectively, in response to a frontal (top row) and lateral (bottom row) target. The shift in ordinate axes between plots reflects the overall trends that binaural beamformers offer greater benefit than bilateral beamformers and that greater benefit is obtained for a frontal target than a lateral target. The particular focus of this study is the relative benefit of personalized beamforming. In general, the benefit obtained using the personalized beamformers is greater than for non-personalized beaformers, as summarized by the interquartile range (IQR). This is particularly apparent for the binaural beamformers. However, there are some combinations of individual, beamformer, and target for which the best non-personalized beamformer, denoted "Best non-per" performs better than the personalized one. This is more frequent for bilateral beamformers, with 14 occurrences for the frontal target and 22 occurrences for the lateral target, than for binaural beamformers, with four occurrences each for the frontal and lateral targets. This non-optimal performance of the personalized beamformer in some cases is possibly due to the small mismatch between the simulated conditions and the model assumptions. An alternative explanation, given the small differences involved, is that it is a limitation of the MBSTOI metric. In any case, it should be noted that the best-performing non-personalized beamformer is different in each case.

Of particular interest is performance using the HATS-derived beamformer, since a HATS is intended to represent an average person. In all but two cases (s1 and s40 with bilateral beamformer and frontal target) the improvement obtained using the personalized beamformer is greater than the benefit of the HATS beamformer. It is therefore expected that personalized beamforming will improve speech intelligibility compared to generic beamforming using filters derived from the HATS.

In Fig. 4, the predicted relative benefit of personalized beamformers compared to the HATS beamformers are shown. The boxplots show the overall distribution and the crosses to the right of each box are the predicted benefits only for those individuals who also participated in the matrix speech intelligibility test reported in Sec. VI. The median predicted benefit of personalization for bilateral beamformers is from 0.3 to 0.4 dB and the median equivalent predicted SNR benefit of personalization for binaural beamformers is from 0.6 to 0.7 dB. In all cases, these medians are significantly ($p < 0.05$) different to 0 dB. In addition to this overall result, it is interesting that some individuals are predicted to benefit substantially more, up to 2.0 and 1.8 dB, for personalized bilateral and binaural beamforming, respectively.

The results presented in this section suggest that HATS-derived beamformers may be close to optimal for some individuals but far from it for others. One might expect the extent of the benefit to be related to, for example, the similarity of an individual's head size to that of the HATS. However, dividing the population into two equally-sized groups based on head size did not lead to a significantly better improvement for the
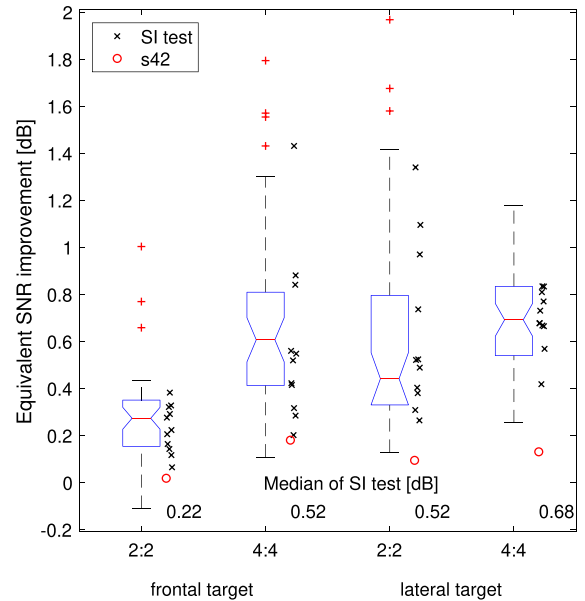

FIG. 4. (Color online) Improvement in MBSTOI, expressed as equivalent improvement in SNR, due to personalized beamforming compared to improvement due to s28, HATS, beamformers. Boxes show quartiles of the distribution for all individuals, excluding the HATS (s28 and s42), whiskers extend to include all points within 1.5 IQR of the quartiles and + symbols indicate more extreme values. Notches indicate 95% confidence interval of median. Crosses to the right of each box show individual values corresponding to the people who participated in the experiment reported in Sec. VI. Circles show the value for s42, which is the same HATS used to design the beamformer but measured on a different day.

smaller group, which is more similar to the HATS head. Therefore, the psychophysical validation of the effect of beamformer personalization on speech intelligibility, reported in Sec. VI, treats all participants as belonging to a single group.

VI. PSYCHOPHYSICAL VALIDATION

A. Experiment design

Eleven native Danish speakers (six male, five female) whose HAHRTFs were measured as described in Sec. II participated. All had normal hearing according to pure tone audiograms measured prior to the experiment. All stimuli were generated specifically for the individual, as detailed in Sec. IV A, such that each participant heard noisy reverberant hearing aid signals as would be experienced by that individual in the measurement room. Five processing conditions were tested, as listed in Table II.

Stimuli were presented over Sennheiser HD650 headphones without equalization. Following the procedure of Experiment 3 in Ref. 49, in each trial the participant listened to a single presentation of a sentence from the Dantale II corpus[50] and used a graphical user interface to select the heard words. For each of the five words in a sentence, the subjects were offered a choice between ten possible words and the option to pass (if the word had not been heard at all). Stimuli were generated according to Eq. (10), with the speech level scaling parameter, β, set to give SNRs prior to enhancement of −17, −14, −11, and −8 dB. These SNRs were determined in pilot experiments to elicit word accuracy rates

TABLE II. Processing conditions used in matrix speech intelligibility test.

Label	Description
Ref	Unprocessed noisy reverberant signals
2:2 HATS	Bilateral - weights derived for mannequin
2:2 Per	Bilateral - weights derived for individual
4:4 HATS	Binaural - weights derived for mannequin
4:4 Per	Binaural - weights derived for individual

which approximately span the informative range of word accuracy scores over all test conditions. The playback level was set such that a noise-only stimuli, i.e., $\beta = 0$, was presented at approximately 65 dB sound pressure level (SPL).

A block of trials consisted of a single presentation of all experiment conditions in random order. Each experiment condition was repeated ten times, giving a total of 11 participants $\times$ 5 processing conditions $\times$ 2 target DOAs $\times$ 5 words per sentence $\times$ 4 SNRs $\times$ 10 repetitions = 22 000 words. Responses were collected over two self-paced sessions lasting less than 1 h each. Participants were encouraged to take at least one break per session. The first session began with a short training phase where participants familiarized themselves with the response interface and types of stimuli and were invited to adjust the playback level if desired. No feedback was given on the accuracy of responses either during training or the main experiment.

B. Results

Figure 5 shows the distribution of average word accuracy across participants for each of the five test conditions separately for frontal and lateral targets. It is immediately apparent that all the beamformers offer substantial improvement over the reference condition. This is particularly true with the frontal target where performance in the reference condition is worse than with the lateral target. This is consistent with both the literature[51,52] and the numerical results presented in Sec. V.
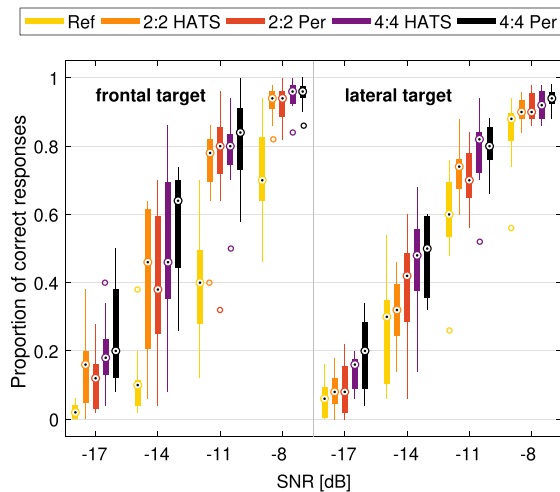

FIG. 5. (Color online) Distribution of proportion of correct words for matrix speech intelligibility test. Each box represents 11 means, each averaged over 50 words. Filled circles and open circles represent median and outlier values, respectively.

Since the focus of this study is the effect of personalized beamformer design versus non-personalized beamformer design, the reference condition is excluded from further analysis. Statistical significance of word accuracy between conditions is addressed using a mixed-effects logistic regression as implemented in the "lme4" package[53] for R.[54] Fixed factors of target direction (frontal, lateral), configuration (bilateral, binaural), and personalization (personal, HATS) were coded with treatment contrasts while SNR was continuous. The random effect of participant identity was modelled as an independent offset. Starting from the null model, the likelihood ratio test was used to sequentially add significant ($p < 0.05$) main effects and interactions. At each stage simplification of the model by pruning terms was also tested.

The final model includes significant effects of SNR and target direction and significant interactions between SNR and configuration and between configuration and personalization. Table III details the final model coefficients and standard errors in logit units. The positive coefficients indicate that increasing SNR increases intelligibility as does binaural beamforming with personalized filters. The negative coefficient for the lateral target condition indicates that, after beamforming, lateral targets are less intelligible than frontal targets. The negative coefficient for the interaction of binaural beamforming and SNR indicates that the relative benefit reduces as SNR increases. This is consistent with the fact that intelligibility reaches a ceiling at higher SNRs even without enhancement and so the relative benefit reduces. Note that there is no interaction between target direction and either the configuration or personalization of the beamformer.

To give an intuitive sense of the mixed-effects logistic regression model behaviour, Fig. 6 shows the model response as a function of input SNR in terms of proportion correct for a frontal target (top) and lateral target (bottom). Improvements due to the significant fixed effects can be interpreted in terms of a left-shift of the 50% speech reception threshold (SRT). Here binaural beamforming reduces (improves) the SRT by 0.90 dB compared to bilateral beamforming and personalization adds a further benefit of 0.40 dB over the HATS derived beamformer. It should be stressed that the 0.40 dB benefit of personalization is the average predicted benefit over the population of test participants and is statistically significant at the 5% level; some individuals stand to benefit more than this relatively modest amount. It can be seen that the effect of the interaction between SNR and beamformer configuration is to reduce the relative benefit of binaural beamforming as SNR increases, since the probability of a correct response approaches 1 in all conditions.

TABLE III. Mixed-effects logistic regression model coefficients for matrix speech intelligibility test (all effects significant at 5% level).

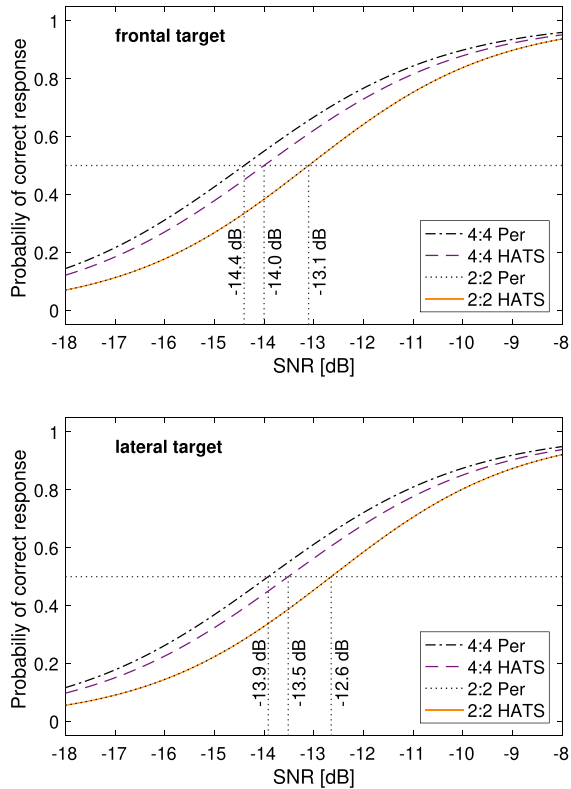| Factor	Value	Standard error
snr	0.530	0.008
doa (lateral)	−0.241	0.040
snr:config (4:4)	−0.034	0.004
config (4:4):person (Per)	0.199	0.055

FIG. 6. (Color online) Fitted model predictions for (top) frontal target and (bottom) lateral target. Note that for the lateral target, all curves are shifted to the right by 0.5 dB compared with the frontal target. Also, since the model predicts no significant difference between "2:2 Per" and "2:2 HATS," their curves are identical.

Comparing the psychophysical experiment results reported in this section to the numerical results obtained in Sec. V B, the predicted intelligibility improvements were reasonably close. Considering only those participants who took part in the speech intelligibility test, as represented by the crosses in Fig. 4, the median benefit of personalization for binaural beamformers was predicted to be 0.52 dB and 0.68 dB for frontal and lateral targets, respectively, and 0.4 dB was achieved in practice. In the case of bilateral beamforming, no benefit of personalization was observed in the psychophysical experiment, whereas the numerical results predicted a median benefit of 0.22 and 0.52 dB for frontal and lateral targets, respectively.

VII. CONCLUSIONS

Using a newly collected and publicly available database of HAHRIRs, the effect of beamformer personalization on model-based beamforming has been studied.

An analysis of directivity patterns suggests that, for the frontal target, binaural beamforming always gives a benefit over bilateral beamforming, regardless of personalization, and that for both bilateral and binaural beamforming, personalization leads to more compact directivity patterns and avoids signal attenuation in the target direction. For the lateral target, binaural beamforming offers little, if any, advantage over bilateral beamforming, but personalization offers a substantial improvement, particularly in suppressing sound arriving from the rear.

Predicted speech intelligibility using the MBSTOI measure suggests that the benefit of personalized beamforming compared to using HATS-derived beamformers is equivalent to a 0.3–0.4 dB increase in SNR. For binaural beamformers, the equivalent benefit of personalized beamforming is from 0.6 to 0.7 dB. For some individuals, the benefit is predicted to be as much as 2.0 dB.

In a matrix speech intelligibility test, binaural beamforming gave an average benefit of 0.9 dB over bilateral beamforming. In the binaural case, personalization of the beamformers provided an additional 0.4 dB benefit over the HATS beamformer.

[1]B. Cornelis, M. Moonen, and J. Wouters, "Speech intelligibility improvements with hearing aids using bilateral and binaural adaptive multichannel Wiener filtering based noise reduction," J. Acoust. Soc. Am. 131(6), 4743–4755 (2012).

[2]L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. 30(1), 27–34 (1982).

[3]C. Chen and P. P. Vaidyanathan, "Quadratically constrained beamforming robust against direction-of-arrival mismatch," IEEE Trans. Signal Process. 55(8), 4139–4150 (2007).

[4]S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," Signal Process. 93(12), 3264–3277 (2013).

[5]A. Kuklasinski and J. Jensen, "Multichannel Wiener filters in binaural and bilateral hearing aids—Speech intelligibility improvement and robustness to DoA errors," J. Audio Eng. Soc. 65(1/2), 8–16 (2017).

[6]H. Barfuss, C. Huemmer, G. Lamani, A. Schwarz, and W. Kellermann, "HRTF-based robust least-squares frequency-invariant beamforming," in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY (October 18–21, 2015).

[7]S. Doclo and M. Moonen, "Superdirective beamforming robust against microphone mismatch," IEEE Trans. Audio Speech Lang. Process. 15(2), 617–631 (2007).

[8]I. Tashev, "Gain self-calibration procedure for microphone arrays," in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan (June 27–30, 2004), pp. 983–986.

[9]M. Buck, T. Haulick, and H.-J. Pfleiderer, "Self-calibrating microphone arrays for speech signal acquisition: A systematic approach," Signal Process. 86(6), 1230–1238 (2006).

[10]M. R. P. Thomas, J. Ahrens, and I. Tashev, "Optimal 3D beamforming using measured microphone directivity patterns," in Proceedings of the International Workshop on Acoustic Signal Enhancement (IWAENC), Aachen, Germany (September 4–6, 2012), pp. 1–4.

[11]C. T. Jin, N. Epain, and A. Parthy, "Design, optimization and evaluation of a dual-radius spherical microphone array," IEEE/ACM Trans. Audio Speech Lang. Process. 22(1), 193–204 (2014).

[12]A. Politis and H. Gamper, "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY (October 15–18, 2017), pp. 224–228.

[13]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. 2009(1), 298605 (2009).

[14]M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of the IEEE International Conference on Digital Signal Processing (DSP)*, Sydney, Australia (April 19–24, 2009), pp. 1–4.

[15]C. Oreinos and J. M. Buchholz, "Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a head and torso simulator (HATS)," Acta Acust. united Acust. **99**(5), 836–844 (2013).

[16]V. Durin, S. Carlile, P. Guillon, V. Best, and S. Kalluri, "Acoustic analysis of the directional information captured by five different hearing aid styles," J. Acoust. Soc. Am. **136**(2), 818–828 (2014).

[17]S. Harder, R. R. Paulsen, M. Larsen, S. Laugesen, M. Mihocic, and P. Majdak, "Reliability in measuring head related transfer functions of hearing aids," Acta Acust. united Acust. **101**(5), 1064–1066 (2015).

[18]F. Denk, S. M. A. Ernst, S. D. Ewert, and B. Kollmeier, "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles," Trends Hear. **22**, 233121651877931 (2018).

[19]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY (October 24, 2001), pp. 99–102.

[20]O. Warusfel, "LISTEN HRTF Database," http://recherche.ircam.fr/equipes/salles/listen/ (Last viewed 6 May 2019).

[21]"ARI HRTF Database," https://www.kfs.oeaw.ac.at/index.php?view=article&id=608&lang=en (Last viewed 6 May 2019).

[22]J. M. Kates, *Digital Hearing Aids* (Plural Publishing, San Diego, CA, 2008).

[23]AES69-2005: *AES Standard for File Exchange - Spatial Acoustic Data File Format* (Audio Engineering Society, New York, 2005).

[24]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proceedings of the Audio Engineering Society Convention*, New York (October 9–12, 2000), pp. 1–23.

[25]C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the Sydney York morphological and acoustic recordings of ears database," IEEE Trans. Multimedia **16**(1), 37–46 (2014).

[26]V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," J. Audio Eng. Soc. **49**(6), 472–479 (2001).

[27]H. Gamper, D. Johnston, and I. J. Tashev, "Interaural time delay personalization using incomplete head scans," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans (March 5–9, 2017), pp. 461–465.

[28]P. Minnaar, S. Favrot, and J. M. Buchholz, "Improving hearing aids through listening tests in a virtual sound environment," Hear. J. **63**(10), 40–44 (2010).

[29]The acoustic impulse response data described in this study are openly available from the Zenodo repository at https://doi.org/10.5281/zenodo.2660781.

[30]Code repository at https://github.com/ImperialCollegeLondon/sap-oticon-hearing-aid-impulse-responses (Last viewed 7 May 2019). Code releases published with DOI: 10.5281/zenodo.2662515.

[31]J. Capon, "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1418 (1969).

[32]M. S. Brandstein and D. B. Ward, *Microphone Arrays: Signal Processing Techniques and Applications* (Springer-Verlag, Berlin, Germany, 2001).

[33]H. Cox, R. Zeskind, and T. Kooij, "Practical supergain," IEEE Trans. Audio Speech Lang. Process. **34**(3), 393–398 (1986).

[34]J. M. Kates and M. R. Weiss, "A comparison of hearing-aid array-processing techniques," J. Acoust. Soc. Am. **99**(5), 3138–3148 (1996).

[35]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 *Proceedings International on Workshop Acoustic Echo and Noise Control (IWAENC)*, Paris, France (September 12–16, 2006).

[36]D. Marquardt, "Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques," Ph.D. thesis, University of Oldenburg, Oldenburg, Germany, 2015.

[37]S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," IEEE Signal Process. Mag. **32**(3), 18–30 (2015).

[38]S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," IEEE/ACM Trans. Audio Speech Lang. Process. **25**(4), 692–730 (2017).

[39]S. Braun, A. Kuklasiński, O. Schwartz, O. Thiergart, E. A. P. Habets, S. Gannot, S. Doclo, and J. Jensen, "Evaluation and comparison of late reverberation power spectral density estimators," IEEE/ACM Trans. Audio Speech Lang. Process. **26**(6), 1056–1071 (2018).

[40]A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," IEEE/ACM Trans. Audio Speech Lang. Process. **23**(6), 1006–1018 (2015).

[41]Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," IEEE Trans. Audio Speech Lang. Process. **15**(4), 1305–1319 (2007).

[42]Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," IEEE Signal Process. Lett. **14**(5), 337–340 (2007).

[43]D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **32**(2), 236–243 (1984).

[44]J. L. Roux, N. Ono, and S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," in *Proceedings of the Workshop on Statistical and Perceptual Audition*, Brisbane, Australia (September 21, 2008).

[45]ANSI/ASA Standard S3.36 (R2006): *Specification for a Manikin for Simulated in-Situ Airborne Acoustic Measurements* (American National Standards Institute, New York, 1969).

[46]A. H. Andersen, "Speech intelligibility prediction for hearing aid systems," Ph.D. thesis, Aalborg University, Aalborg, Denmark, 2017.

[47]A. H. Andersen, J. M. D. Haan, Z. H. Tan, and J. Jensen, "Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions," Speech Commun. **102**, 1–13 (2018).

[48]ANSI/ASA S1.42:2001: *Design Response of Weighting Networks for Acoustical Measurements* (American National Standards Institute, New York, 2001).

[49]A. H. Andersen, J. M. D. Haan, Z. H. Tan, and J. Jensen, "Predicting the intelligibility of noisy and nonlinearly processed binaural speech," IEEE/ACM Trans. Audio Speech Lang. Process. **24**(11), 1908–1920 (2016).

[50]K. Wagener, J. Josvassen, and R. Ardenkjær, "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**(1), 10–17 (2003).

[51]W. E. Kock, "Binaural localization and masking," J. Acoust. Soc. Am. **22**(6), 801–804 (1950).

[52]J. A. Grange and J. F. Culling, "The benefit of head orientation to speech intelligibility in noise," J. Acoust. Soc. Am. **139**(2), 703–712 (2016).

[53]D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–51 (2015).

[54]R Core Team, "R: A language and environment for statistical computing," http://www.R-project.org/ (Last viewed 6 May 2019).