

High-Resolution Directivity Measurements of an Artificial Head and Mouth Shaped to Three Vowels

PAUL LUIZARD,^{1,3,*}  JULIE DAUDRÉ,¹ MARTIN SCHNEIDER,²
(paul.luizard@tu-berlin.de) (julie.daudre@proton.me) (martin.schneider@neumann.com)

DAVID ACKERMANN,^{1,4}  AND STEFAN WEINZIERL¹ 
(david.ackermann@hs-duesseldorf.de) (stefan.weinzierl@tu-berlin.de)

¹*Audio Communication Group, Technische Universität Berlin, Berlin, Germany*

²*Georg Neumann GmbH, Berlin, Germany*

³*Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, 38000, France*

⁴*Institute of Sound and Vibration Engineering (ISAVE), Hochschule Düsseldorf, Düsseldorf 40476, Germany*

The directivity of the human voice has been studied since the early twentieth century using different measurement systems with progressively higher spatial resolution. Artificial heads and mouths have been used because of their ability to repeat a given sound production, hence allowing sequential measurements of directivity with a reduced number of microphones and still achieving high spatial resolution. Unlike most artificial heads, whose external geometry is abstracted, this study uses a custom 3D-printed head with detailed geometry and three different mouth openings, all based on 3D scans from magnetic resonance imaging data. The impulse response measurements were performed using a 3D robotic arm, resulting in directivity data with 5° resolution in both azimuth and elevation. The measured directivity patterns are consistent with previous research on energy distribution in space over angles (azimuth and elevation) and frequency, with a higher spatial resolution and for different mouth shapes. The resulting data set is made available in several standardized file formats to facilitate accurate voice directivity simulations in virtual acoustic environments.

Keywords: human voice, directivity measurement, artificial head, mouth simulator

1 INTRODUCTION

Artificial heads are not only used as sound receivers to emulate the directivity of the human head and pinna up to a reference point in the ear canal, they can also be augmented with a mouth opening and a loudspeaker to approximate the source directivity of the human voice. Such artificial heads have been developed in the industry with simplified head geometries (e.g., Brüel & Kjær [B&K] 4128 and 4230, Head Acoustics HMS II) or with more specific facial details (KEMAR developed by Knowles Electronics, manufactured by GRAS) and in different research laboratories. Some of these heads comply with the geometric and electroacoustic specifications related to head-and-torso simulators for measurements in telephonometry [1]. Another one used a more detailed (i.e., less abstracted) head geometry

and combines a midrange loudspeaker with two woofers in a bass-reflex design with a crossover frequency of 150 Hz to allow the head to produce sound down to 50 Hz [2]. Developed at the Institute of Technical Acoustics (ITA) in Aachen, Germany, this head is referred to as the ITA head.

Several studies have compared the directional characteristics of these artificial head-and-mouth simulators with those of human voices. The directivity of a B&K 4230 head and a custom head from the University of Parma were measured in the horizontal plane and compared with the same measurements on a human being [3]. The results presented between 250 Hz and 2 kHz show that the human voice has a slightly higher directionality than both artificial heads. Another study has compared the directivity of three artificial heads (HMS-II, the ITA head, and a B&K head) not only to the human voice but also to that of a small-diaphragm full-range loudspeaker as recommended in International Organization for Standardization 60268-16 for measuring the Speech Transmission Index [4]. Here, it was shown that these loudspeakers have a higher directionality above 800 Hz than both humans and artificial heads [5].

*To whom correspondence should be addressed: paul.luizard@tu-berlin.de

Research into the spatial characteristics of the human voice has also addressed the question of the influence of the different mouth configurations that occur with different spoken phonemes on the directivity of the voice. A study of four different mouth geometries of an artificial head [6] based on a commercial head geometry (B&K 4128C HATS) has shown that, as expected, the larger the mouth aperture is the more directional the human voice will be above 500 Hz, whereas a smaller aperture yields much less directionality at higher frequencies. Similar results were obtained, although measured only through transfer functions between a point on the axis in front of the mouth and another one on the cheek, for three different mouth sizes [7]. With another head-and-torso simulator, including different vocal tracts and mouth shapes, designed as two spheres for the head and the torso, it was shown that the lips increase the directionality of the voice above 5 kHz and that the mouth shape related to different phonemes changes the directionality, mostly in the octave bands centered around 2 and 4 kHz [8].

The effect of head-above-torso orientation has also been studied, not only for the receiver characteristics of human ears [9] but also for the source directivity of the human voice, using an artificial KEMAR head [10]. The results showed that the global difference in sound pressure level (SPL) between the frontal and other head orientations increases largely linearly with frequency, from 1 dB to 5 dB from 10° to 90° head orientation.

Most of these studies used heads with unrealistic faces, skipping face details that could, however, influence the directivity. Additionally, the results are given in terms of directivity patterns in specific planes, generally in octave bands. Hence, although these results give important insight into issues related to voice directivity, they do not provide continuous data across frequency and are not based on a realistic, detailed head geometry. This is what the present study intends to provide. As an example of the benefits yielded by a detailed artificial head, a previous comparison of a human talker and two artificial heads (the commercial Head Acoustics HMS II and the Norman head of the present article with the three mouth shapes) showed that the directivity diagrams in the horizontal plane and the directivity index across frequency of the detailed head was globally closer to the human data [11]. Hence, details of the head, as opposed to simplified head geometry, seem to play a non-negligible role in directivity accuracy.

This present paper describes the construction of an artificial head based on 3D scans of a human head and torso. This geometry was previously used to build the Fast and Automatic Binaural Impulse Response Acquisition (FABIAN) artificial head [12], which is well documented and has been involved in a number of scientific studies in the field of human spatial hearing [13–15]. Three different realistic mouth shapes were incorporated into this head, yielding three data sets corresponding to the most different vowels /a/, /i/, and /u/ at the edges of the vowel diagram [16]. The measurements were performed with a high spatial resolution of 5° in azimuth and elevation. The obtained directivity data are presented as directivity maps against frequency, and the

measured impulse responses are provided in different file formats for effective use by the community.

It is noted that this study considers three vowels specifically because only voiced phonemes can be produced with this artificial head, not plosive consonants for example. And it was decided to keep phonemes for which the mouth is open; therefore, no nasal consonants were considered. As for fricative consonants, their directivity differs essentially from vowels because of their spectral content (much higher than those of vowels), but the geometry of the face does not change much. Hence, the three most different vowels were considered, which covers a quite large aspect of speech.

2 HEAD CONSTRUCTION

The detailed aspect of the head is achieved by using both molding of a real person's face and magnetic resonance imaging scans of mouths and vocal tracts as a basis for its design. On the one hand, the FABIAN head geometry was chosen because it provides a morphologically authentic human head [12]. On the other hand, the mouth geometry for each vowel was chosen from the data set of mouths and vocal tracts [17] from Aalto University. A detailed description of the head design and construction, including diverse versions that were eventually not chosen, can be found in a master's thesis [18], and a shorter description was presented in a conference [19].

2.1 Digital Design

The 3D meshes were created and developed using the Blender software. For each vowel, the meshes of FABIAN and the corresponding mouth were blended together, resulting in a specific external head shape. To account for a series of constraints, such as the loudspeaker located inside the final head, the space for its cables (power supply and audio signal), and the screws that hold the whole structure together, the head had to be divided into several parts. The final version could then be assembled with only one part to be changed between measurements: the sliding mouth and chin, fixed by two screws. The shoulders were separated from the torso so that the whole system can be stored in a compact case.

Exploded views of the head and its various parts can be seen in ([19], Fig. 1). Note that tapered pins and holes have been included in the 3D model of each part so that they can be connected in the desired position and screwed to the 5-mm-thick aluminum plate under the torso.

2.2 Practical Construction

Selective laser sintering was used as the 3D printing method [20]. A laser selectively sinters the particles of a polymer powder, fusing them together and building a part layer by layer. The different parts were lightly sanded as needed, depending on how well they fit together. It is noted that the parts are designed as hollow forms to reduce the weight of the whole structure. Only the shoulders are made of full material. The assembled head with the three different mouth parts can be seen in Fig. 1. The compact active

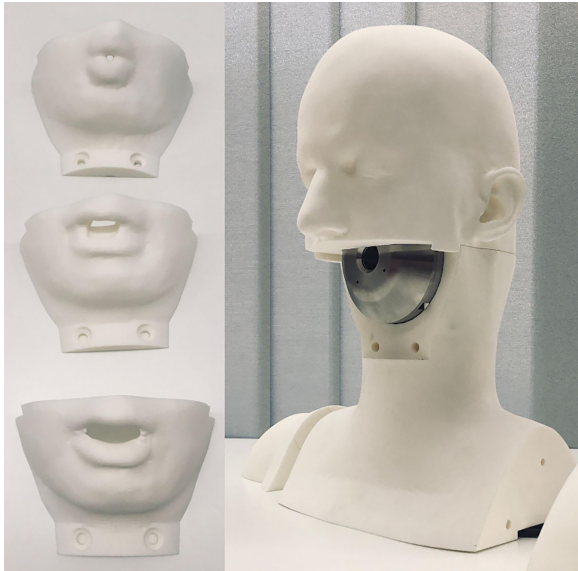


Fig. 1. Photographs of the head with its loudspeaker and the three mouth geometries: /u/, /i/, and /a/ (top to bottom). These also appear in [19].

loudspeaker (B&K 4227A¹) was placed in the special cavity that was designed as a tight fit around the loudspeaker, except in front of it, toward the mouth opening.

The frequency responses of the loudspeaker alone and of the artificial head with the different mouths are shown in Fig. 2, where the signals are normalized by the highest value of magnitude spectra at 1 kHz. First (a) shows the frequency response as measured all together, and then (b) shows the signals obtained with the mouths, equalized by the response of the loudspeaker alone (i.e., without any element of the head around it), allowing for a direct estimation of the influence of each mouth. The loudspeaker has a transmission range from 100 Hz to 15 kHz, where the sensitivity does not drop by more than about 20 dB from the maximum [see Fig. 2(a)]. The frequency response of the loudspeaker alone shows the highest amplitude and a somewhat nonflat response [graph in Fig. 2(a)]. The latter was further used as a reference to equalize the measured signals when the upper part of the head and each mouth shape was mounted around the loudspeaker. The obtained frequency responses [see the Fig. 2(b)] show the effect of each mouth geometry on the emitted sound, including peaks and notches.

The responses with the head structure mounted around the loudspeaker show more peaks and notches at frequencies higher than 1 kHz, resulting in variations of about 25 dB for /a/, 20 dB for /i/, and 30 dB for /u/ in the range 500 Hz to 10 kHz. The specific effect of each mouth on the sound radiated in front of the head and behind it (dashed lines) can be seen in the bottom graph of Fig. 2(b). These are due to the resonances that occur in the three different small cavities created by the mouth shapes between the loudspeaker and the printed inner surface around the mouth.

In terms of directivity, however, it is not necessary to compensate for this change in amplitude over frequency because the interest here is in the directivity induced by each mouth shape, which is normalized in amplitude with respect to the frontal direction in each frequency band considered.

It should be noted that the sound level obtained on average between 100 Hz and 10 kHz with the /a/ and /i/ mouths are comparable, 5 dB below the loudspeaker alone, whereas the /u/ mouth yielded sound levels about 5 dB lower than the other mouths. This is probably due to the size of the mouth opening, as can be seen in Fig. 1, which is smaller for /u/, causing a higher radiation impedance and thus less sound energy propagating outside the head. Otherwise, the recorded sound levels above 10 kHz are comparable for the three mouths.

3 MEASUREMENT

3.1 Setup

The measurements took place at the Institute for Acoustics and Audio Techniques in Aachen, Germany. In a class 1 semi-anechoic chamber with a limit frequency of 100 Hz, the artificial head was attached to a 3D robotic arm² with motors enabling controlled rotations along two orthogonal axes (i.e., the *phi* [azimuth] and *theta* [elevation] angles with a precision of 1°. The measurements were carried out with an equiangular resolution of 5° in azimuth and elevation, resulting in 2,522 measurement points. It should be noted that this measurement system is based on the so-called *frontal convention*³ (i.e., the obtained sphere of measurement points has a pole in front of the source and the other behind it, not above and below as classically represented in the *elevation convention*). The head was positioned so that the mouth aperture was at the center of the sphere induced by the movements of the robotic arm because it was considered to be the acoustic center of the sound source. A class 1 microphone (GRAS one-quarter-inch 46BF microphone with preamplifier) connected to an amplifier (B&K 2610) was located on the floor 8 m away from the head.

By approximating the radiating mouth as a rigid circular piston and assuming that far-field conditions for directivity measurements are given for $\lambda \gg d^2/r$, where d is the diameter of the piston and r is the distance from the source [21], far-field conditions were obtained for all frequencies in the audible range.

The signal from the microphone was collected through a sound interface (RME Multiface HDSP). The whole measurement system, including the robotic arm, was controlled by the WinMF software (Four Audio, GmbH), so that the series of 2,522 points could be measured in one run.

3.2 Measured Data

At each orientation (ϕ_i, θ_j), an exponential sine sweep (20 Hz–20 kHz) of 5 s duration, was played by the loud-

¹ <https://media.hbkworld.com/m/d1646a18a2331b72/original/Mouth-Simulator-Types-4227-and-4227-A.pdf>

² <https://www.four-audio.com/en/products/elf/>

³ <https://pyfar.readthedocs.io/en/latest/classes/pyfar.coordinate.html>

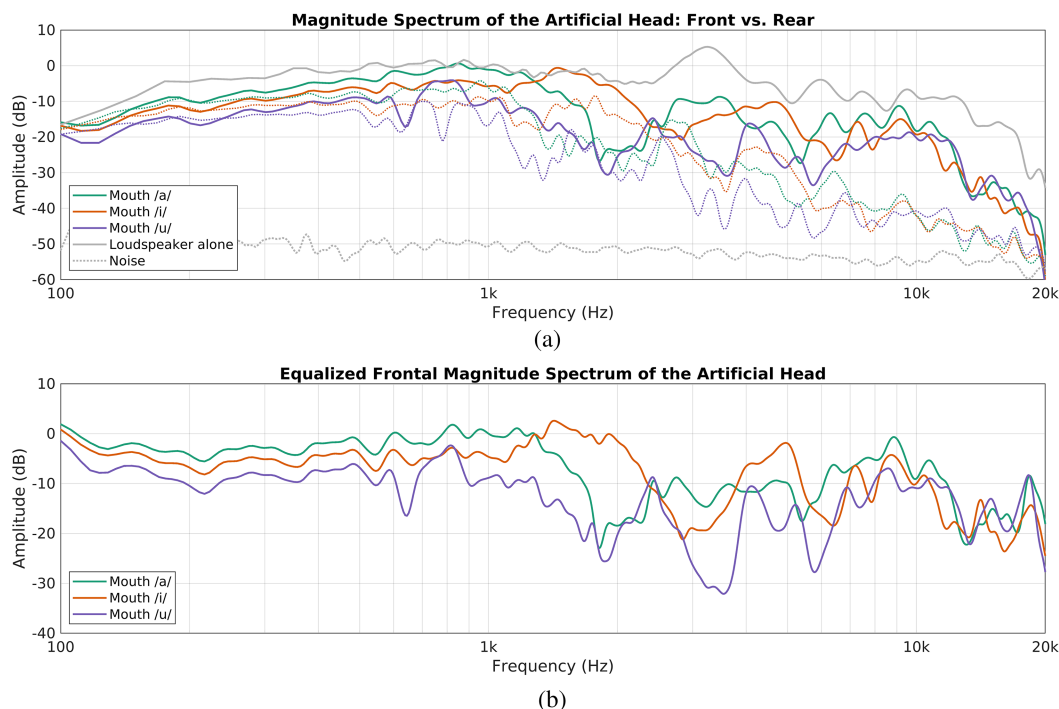


Fig. 2. The magnitude spectrum in decibels is computed from the recorded impulse responses at the front and rear of the source (solid and dashed lines, respectively). This is shown for three mouth shapes (/a/, /i/, and /u/) of the artificial head under investigation and for the loudspeaker alone and for noise (a). In (b) the graph shows the magnitude spectrum of the three mouth shapes normalized by the loudspeaker's spectrum, at each frequency bin.

speaker located inside the head and recorded by the microphone at a sampling rate of 48 kHz with a resolution of 24 bits. The amplifier and recording setup remained identical for the 3 mouths. After a series of 2,522 points, which took about 4 h, the mouth was changed and another measurement run was performed. The obtained impulse responses were stored as synology package files individually for each orientation.

4 DATA PROCESSING

4.1 Impulse Responses

The measured impulse responses were all band-pass filtered between 60 Hz and 20 kHz and truncated to remove residual noise while preserving the relevant energy information, with a final duration of about 90 ms. A representation of the data is given in Fig. 3 as a map of the root mean square (RMS) level at each orientation in both azimuth and elevation, covering the entire recorded frequency bandwidth. The global trend is the same for all vowels (i.e., a concentration of sound energy in front of the head) between -40° and 40° in azimuth, and slightly downward, between -70° and 20° in elevation. However, there are differences, especially for the /u/ vowel. The global level is slightly lower, and the frontal energy is divided into two parts, one around 0° in azimuth and elevation and the other lower, around -50° in elevation. In contrast, the frontal energy for the other vowels is more uniform. In terms of vertical direction, /i/ seems to be more directive than /a/, whose energy is more diffuse downward. This may be due to the shapes of the mouths, which can be seen in Fig. 1. Although /i/

is similarly open in both horizontal and vertical directions, sending energy mostly in the lower frontal direction due to the position of the mouth opening on the head, /a/ has a slightly larger aperture horizontally than vertically. This could increase diffraction in the vertical direction, favoring energy dispersion in that direction.

4.2 Sound Spectra

Sound spectra were computed from the impulse responses over their full length at each head orientation (i.e., every 5°). Fig. 4 illustrates the evolution of the SPL measured in the horizontal and vertical planes versus frequency. In both planes, the sound levels are quite uniform below about 1 kHz. Above 1 kHz, there are clear differences between mouth shapes. Large interference patterns can be observed in the high-frequency range, symmetrical with respect to 0° in the horizontal plane and nonsymmetrical in the vertical plane.

Regarding the different vowels, it seems that the general patterns of /a/ and /i/ are quite similar compared with /u/ in both horizontal and vertical planes. The latter shows more asymmetry between the left and right sides at 1 kHz. Polar directivity diagrams of these data appear in ([19], Fig. 4) where vowels /a/ and /i/ show lateral lobes in the 2 and 4 kHz octave bands, whereas /u/ has a more omnidirectional behavior in the horizontal plane. At higher frequency, in the octave band 8 kHz, the frontal region presents stronger directivity for /a/ than for /i/ and /u/, probably because of the larger aperture of the /a/ mouth. In the vertical plane, similar trends can be observed, where vowel /u/ has generally

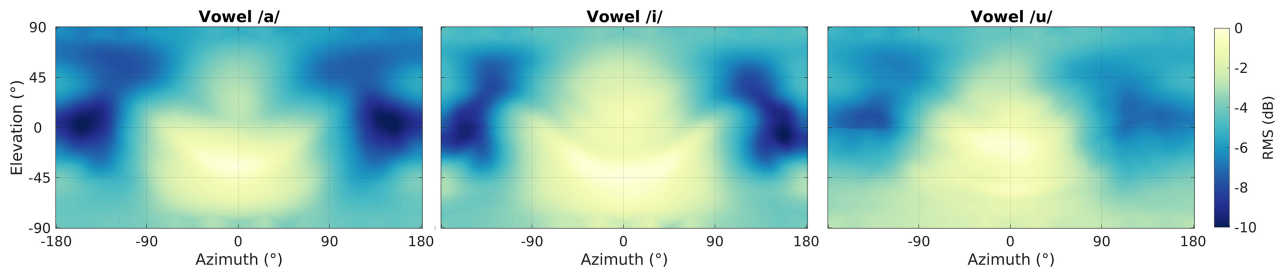


Fig. 3. RMS level map in decibels for the three vowels at all elevations and azimuth angles. The amplitude is normalized to 0 dB based on the maximum RMS level of each vowel.

weaker directivity from 2 kHz and above than the two other vowels.

It also appears that certain specific frequency ranges carry more energy than others (e.g., around 3 kHz for the vowel /a/). This could be due to resonances that occur in the oral cavity associated with each vowel, similar to resonances in a vocal tract. However, because the shape of these cavities is not physiologically valid, these resonances occur at random frequencies and are not related to the formants of each vowel [16]. Because the volume where the loudspeaker is located is very small, adjusted to the loudspeaker, it is impossible to add a damping material that would lower these resonances. However, they do not alter the directivity information because the data are normalized to 0°, and the relevant interference patterns across angles are still visible.

This is confirmed by the signal-to-noise ratio (SNR) values of the sound radiated in front of the head that are above 20 dB up to 12 kHz (see the difference between the signals and the noise level in Fig. 2), which means that the data present a dynamic range large enough to get information of directivity up to this frequency. Even behind the head, the SNR values above 20 dB up to 15 kHz for /a/ and /i/ and above 10 dB up to 3 kHz for /u/ allow one to analyze the directivity of this sound source.

The influence of technical components (i.e., the loudspeaker) can be compensated for by weighting the normalized directivity according to the timbre of real human voices. For this purpose, each of the three directivity patterns measured with the different mouth shapes was normalized to the frontal radiation pattern and then equalized with the spectral envelope of the corresponding human vowels recorded in the frontal direction. As a result, the radiation characteristics in the frontal direction for the respective vowels examined by the artificial head have an identical frequency response to the sound radiation of a human speaker. The weighted directivities are shown in Fig. 5, where the formants typical of each vowel appear, contrary to the unequalized data shown in Fig. 4.

4.3 Error Analysis

To evaluate the difference in sound radiation between an artificial head and a human head, it is useful to consider the radiated sound power. To this end, the directivities were averaged in octave frequency bands according to International Electrotechnical Commission 61260-1 [22] and equalized

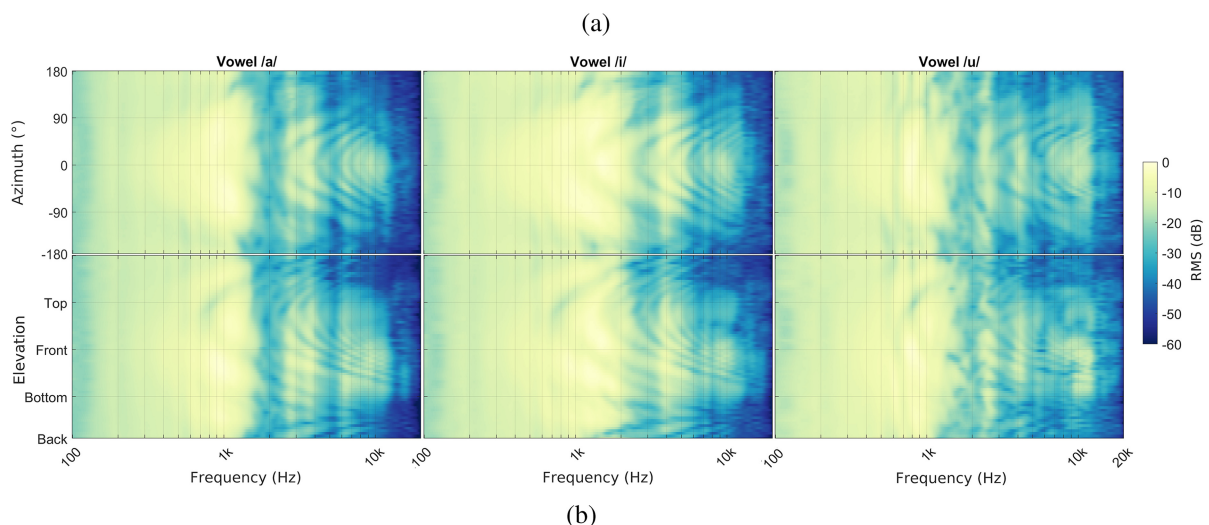


Fig. 4. Map of the SPL in decibels in the horizontal (a) and vertical median plane (b) at all angles and frequencies for the three vowels. The amplitude is normalized to 0 dB based on the maximum SPL level of each vowel and plane.

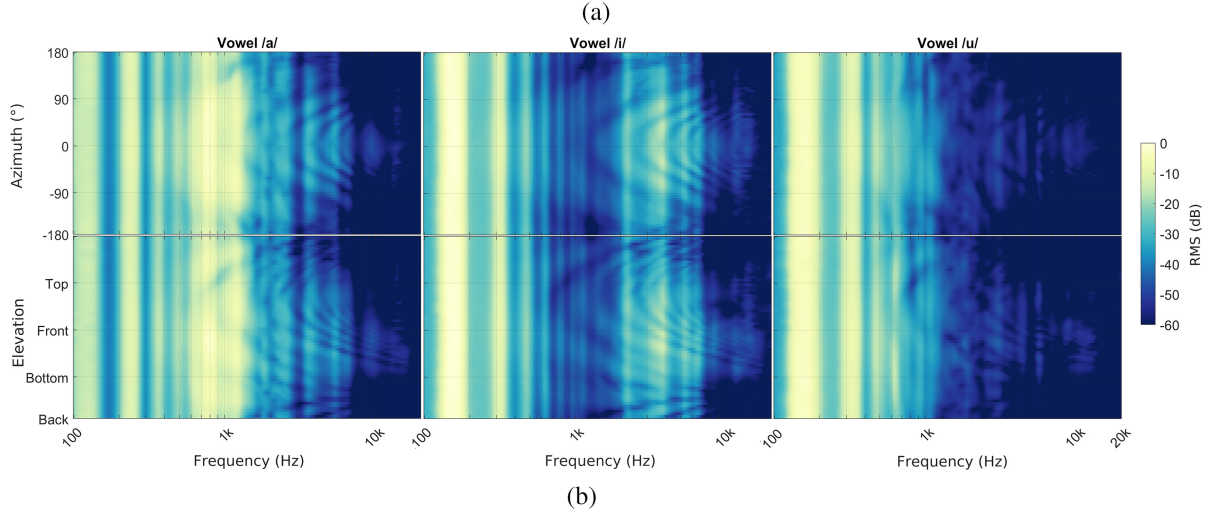


Fig. 5. Map of the SPL in decibels equalized by the vowels of a human speaker in the horizontal (a) and vertical median plane (b) at all angles and frequencies. The amplitude is normalized to 0 dB based on the maximum SPL level of each vowel and plane.

so that they emit the same amount of energy in all directions within each band (i.e., to

$$\bar{p}_{\text{diff},q,m} = \frac{\bar{p}_{q,m}}{\sqrt{\sum_{q=1}^Q \bar{p}_{q,m}^2 \cdot w'_q}}, \quad (1)$$

where $\bar{p}_{q,m}$ corresponds to the octave band-averaged sound pressure for the Q sampling points of the measurement grid with $q \in \{1, 2, \dots, Q\}$, whereas m are the center frequencies of the octave bands with $m \in \{1, 2, \dots, M\}$, $M = 10$, and w'_q are the normalized area weights of the measurement grid with $\sum w'_q = 1$. This representation of the radiation patterns is called *diffuse* equalization in the following.

The relative error in radiated sound power Ψ_m can be calculated as the summed area weighted relative differences of the squared sound pressures of artificial head $\bar{p}_{\text{diff},A,q,m}$ and the reference, the human head $\bar{p}_{\text{diff},H,q,m}$, related to the summed area weighted squared sound pressure of the reference:

$$\Psi_m = \frac{\sum_{q=1}^Q \left| \bar{p}_{\text{diff},A,q,m}^2 - \bar{p}_{\text{diff},H,q,m}^2 \right| \cdot w'_q}{\sum_{q=1}^Q \bar{p}_{\text{diff},H,q,m}^2 \cdot w'_q}. \quad (2)$$

Note that an error of $\Psi_m = 0.5$ states that the artificial head emits 50% of the sound power in incorrect directions, whereas a value of $\Psi_m = 0$ shows that the radiation pattern of the artificial head fully matches the reference.

The global error measure for the artificial head is shown in Fig. 6.

5 DISCUSSION

Comparing this measured data to previously published research can be made regarding two axes: other artificial heads and measurements performed with human voice. The specificity of the present head is that its mouth shapes correspond to three different vowels, which is not the case of most heads used in previous research. The horizontal direc-

tivity of a commercial artificial head (Head Acoustics HMS II), the present Norman head, and a human who repeated the vowels several times (referred to as Averaged Speaker) were measured in the horizontal plane [11]. As illustrated in Fig. 6, the first row (a) shows the directivity patterns for vowel /i/ in three octave bands. It appears that the two artificial heads present similar results, whereas the average human speaker is slightly different, and the global patterns are still retrieved by the artificial heads.

The second row (b) shows the values of the error quantifier described in SEC. 4.3, which computes the distance to the human speaker. It appears that, depending on the octave bands, each artificial head obtained a lower error, suggesting that none of them were more suited than the other to this particular human speaker. Further comparisons over a larger set of talkers, including various genders and ages should be run to assess whether the mouth shape specific to different vowels provides directivity patterns more related to certain people than the abstracted artificial head.

One of the few previous studies that investigated voice directivity by including different arbitrary mouth sizes of a simplified artificial head performed measurements of transfer functions in the near field between a point on the horizontal axis facing the mouth and another one along the cheek of the manikin [7]. It appeared that the larger the mouth opening, the higher the sound level projected in front of the head, which is confirmed by the measurements of the present paper. Unfortunately no angle-dependent measurement was then performed.

The directivity maps presented in Fig. 4 show specific interference patterns that were already observed in artificial head measurements [23–25, 8]. However, measurements of voice directivity of human talkers showed that the directivity patterns differ with the pronounced phoneme [26, 27], and the different mouth shapes of the present artificial head yield directivity patterns closer to the actual behavior of human voice directivity [11]. In particular, in the frequency region where most acoustic energy is produced in vowels (300 Hz–2 kHz), the main characteristics of the three

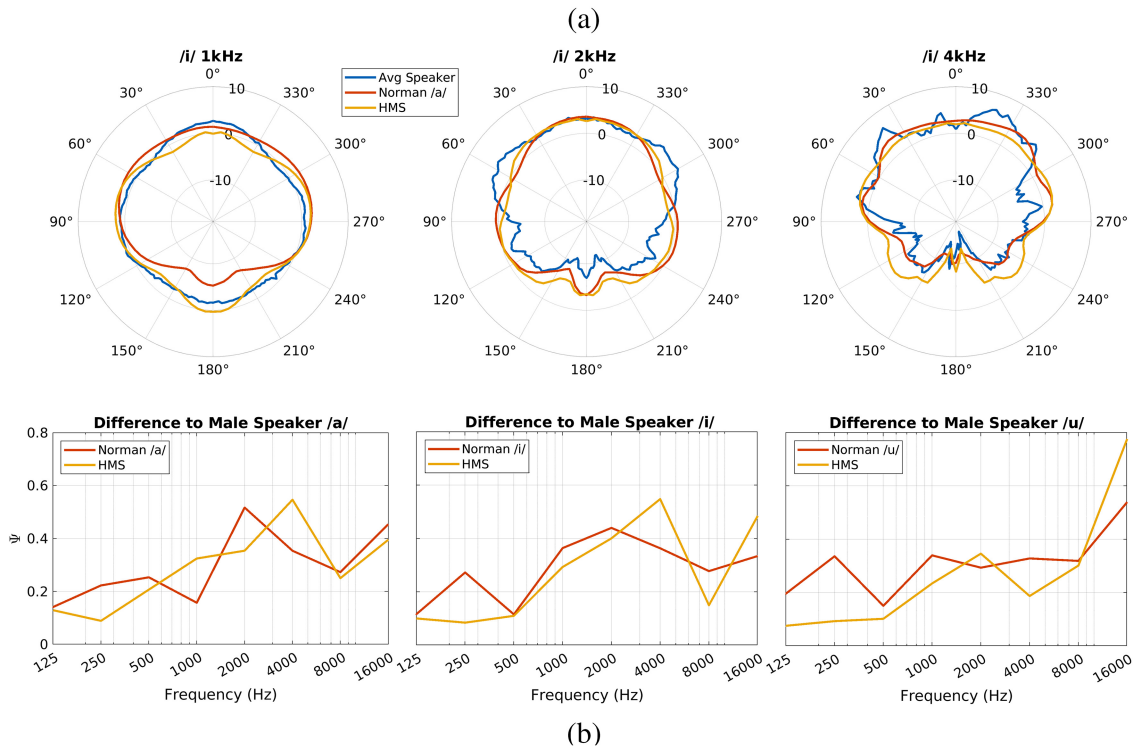


Fig. 6. Diffusely equalized polar patterns according to Eq. (1) of horizontal radiation of average /i/ for the human talker, /i/ for the Norman head, and the general HMS head, in three octave bands (a). The global error measure Ψ over all radiation directions in the horizontal plane between the Norman head and a human talker for the vowels /a/, /i/, and /u/ and between the HMS head and the human talker (b).

vowels are comparable to the most comprehensive (various phonemes and talkers measured in 3D) measurements of human talkers to date (see Fig. 4 of [27]; i.e., the energy of /u/ is very much localized around 1 kHz, the energy level of /i/ is high, up to 2 kHz, and the energy of /a/ is more diffuse across angles). In line with most previous research [28], the energy is globally projected downward in the vertical plane (Fig. 3).

A major limitation of this artificial head is the absence of a physiologically valid vocal tract between the loudspeaker and the mouth opening. In fact, the small cavity that is specific to each mouth shape has an influence on the tonal color (i.e., sound spectrum) of the emitted sound, as illustrated in Fig. 2. As mentioned in SEC. 4.2, this coloration is rather arbitrary and does not correspond to the formants of each vowel, although it is noted that the directivity of this head is not affected by the different spectra, provided that the SNR is sufficiently high.

A solution to overcome this issue is to apply an equalization across frequency so that each of the three vowels would present a typical spectrum with regular formants in front of the head. Then in other directions, the influence of the head geometry modifies this typical spectrum according to the directivity of this sound source. Applying such a filter yields the directivity maps in Fig. 5, where the specific formants of each vowel clearly appear. This shows that the present artificial head, in addition to the analysis of the human voice directivity, can also be used to synthesize specific vowels, including their precise spatial characteristics, for instance, in the context of virtual spatial audio environments.

6 FILE FORMATS OF THE AVAILABLE DATA

The available data cover the two points of the presented results, namely impulse responses and frequency responses in each orientation around the head and for each vowel. Several file formats are proposed, all standardized. Two of them are open source and another one is proprietary. This data is available through the link provided in the Appendix A.1.

6.1 Spatially Oriented Format for Acoustics

The data were stored in spatially oriented format for acoustics (SOFA) [29] using the Sofar package of the Pyfar suite [30]. For each vowel, the 2,522 temporal impulse responses and the corresponding complex frequency responses were stored in two separate SOFA files. The SOFA conventions⁴ used were *GeneralFIR* and *FreeFieldDirectivityTF*, respectively. In all cases, the azimuth and elevation angles and the sampling rate, the frequency vector, and other information were included.

SOFA files are composed of many fields of metadata, such as the measurement environment, application programming interface used to implement the files, azimuth, elevation, distance to the source in the field (*ReceiverPosition*) in spherical coordinates, or the title and a short description of the file contents. For the time data, the field *Data* contains the impulse responses with dimensions $M \times R \times N$, where $M = 1$, $R = 2,522$ (receivers), and $N = 3,744$

⁴ https://www.sofaconventions.org/mediawiki/index.php/SOFA_conventions

(time bins). For the frequency data, the *Data* field contains the real and imaginary parts of the complex frequency responses separately. Each of these subfields has dimensions $M \times R \times N$, where $M = 1$, $R = 2,522$ (receivers), and $N = 1,873$ (frequency bins). The corresponding frequency vector is the field *N* at the same level as *Data* in the SOFA file.

6.2 Open Directional Audio File Format

The open directional audio file format (openDAFF) [31] is open-source and dedicated to directional audio content such as the directivity of sound sources. For each vowel, two files are proposed: one containing the 2,522 time impulse responses and the other the corresponding frequency responses averaged over one-third octave bands. The openDAFF is used in the acoustic simulation environment RAVEN [32] to carry the spatial characteristics of simulated sound sources. The files were generated from the measured impulse responses using the ITA toolbox [33] and AKTools [34] in MATLAB.

6.3 Generic Loudspeaker Library

The final format used to store the data is the Generic Loudspeaker Library (GLL), a proprietary format adapted to the EASE⁵ acoustic simulation environment. The 2,522 impulse responses are stored for each of the three vowels, resulting in three different files. Other file formats such as the Common Loudspeaker Format/Common Instrument Format format (used by the acoustic simulation environments ODEON⁶ and CATT-Acoustics⁷) can be derived using AFMG's SpeakerLab Pro software.⁸

7 CONCLUSION

The data presented are a collection of impulse responses produced by an artificial head with a detailed shape and three different mouths corresponding to the three vowels /a/, /i/, and /u/. The measurements were performed in anechoic conditions, with a 3D arm rotating the artificial head around its mouth in 5° steps in both azimuth and elevation, resulting in the highest-resolution (5° resolution in azimuth and elevation) and most complete (three vowels) voice directivity data set published to date. It is available in the form of time impulse responses and their complex frequency responses in several standardized file formats, namely SOFA, DAFF, and GLL. They can be used in virtual acoustic environments for speech simulation and for speech intelligibility and quality studies in various fields, such as architectural design or speech transmission evaluation (e.g., in telecommunications, gaming, or museography). If only one vowel of the present data set has to be chosen to represent an average voice production, it is recommended to use

the vowel /i/ because its aperture is a kind of compromise between the large /a/ and the very small /u/.

8 ACKNOWLEDGMENT

This work was funded by the German Research Foundation (DFG LU 2499/2-1). The authors would like to thank Samuel Jerichow and Ben Jastram (3D-Lab, Technische Universität [TU] Berlin) for their advice on the final design stages and 3D printing of the head; Dirk Mietle, Markus Wittchen, and Steffen Ziebarth (Georg Neumann) for their advice on head construction and the final assembly; Anselm Goertz (Institute for Acoustics and Audio Techniques [IFAA] and Four Audio) for access to the facility and help with the directivity measurements; Fabian Brinkmann (Audio Communication Group, TU Berlin) for his help with data postprocessing; and Stefan Feistel (AFMG) for enabling us to generate the GLL files.

9 REFERENCES

- [1] ITU, "Head and Torso Simulator for Telephonometry," Recommendation ITU-T P58 (2023 Mar.).
- [2] M. Kob and H. Jers, "Directivity Measurement of a Singer," *J. Acoust. Soc. Am.*, vol. 105, no. 2-Supplement, p. 1003 (1999 Feb.). <https://doi.org/10.1121/1.425813>.
- [3] F. Bozzoli and A. Farina, "Directivity Balloons of Real and Artificial Mouth Simulators for Measurement of the Speech Transmission Index," presented at the *115th Conv. of the Audio Engineering Society* (2003 Oct.), paper 5953.
- [4] International Electrotechnical Commission, "Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility By Speech Transmission Index," IEC 60268-16 (2020 Sep.).
- [5] J. Struve, *Directivity of Real and Artificial Speakers for Room Acoustic Measurements*, Master's thesis, Technische Universität Berlin, Berlin, Germany (2018 Oct.).
- [6] K. Stewart and D. Cabrera, "Effect of Artificial Mouth Size on Speech Transmission Index," in *Proc. of the 14th Intl. Congress on Sound and Vibration* (Cairns, Australia) (2007 Jul.).
- [7] T. Halkosaari, M. Vaalgamaa, and M. Karjalainen, "Directivity of Artificial and Human Speech," *J. Audio Eng. Soc.*, vol. 53, no. 7/8, pp. 620–631 (2005 Jul./Aug.).
- [8] R. Blandin, J. Geng, and P. Birkholz, "Investigation of the Influence of the Torso, Lips and Vocal Tract Configuration on Speech Directivity Using Measurements From a Custom Head and Torso Simulator," *Acta Acust.*, vol. 7, p. 39 (2023 Jul.). <https://doi.org/10.1051/aacus/2023035>.
- [9] F. Brinkmann, R. Roden, A. Lindau, and S. Weinzierl, "Audibility and Interpolation of Head-Above-Torso Orientation in Binaural Technology," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 5, pp. 931–942 (2015 Aug.). <https://doi.org/10.1109/JSTSP.2015.2414905>.
- [10] S. Bellows and T. W. Leishman, "Effect of Head Orientation on Speech Directivity," in *Proc. of INTER-SPEECH*, pp. 246–250 (Incheon, South Korea) (2022 Sep.).

⁵ <https://www.afmg.eu/en/ease>

⁶ <https://odeon.dk>

⁷ <https://www.catt.se>

⁸ <https://www.afmg.eu/en/ease-speakerlab-pro>

- [11] P. Luizard, P. Abehsera-Morell, J. Daudré, S. Bellows, and B. F. G. Katz, “Human Talkers VS Artificial Heads: A Comparison of Their Directivity in Dense Plane Microphone Arrays,” in *Proc. of the 50th German Acoustics Conf. (DAGA)* (2024 Mar.). <https://hal.science/hal-05039997>.
- [12] A. Lindau, S. Weinzierl, and H. Maempel, “FABIAN—An Instrument for Software-Based Measurement of Binaural Room Impulse Responses in Multiple Degrees of Freedom,” in *Proc. of 24th Tonmeister-tagung—VDT Intl. Conv.* (Dusseldorf, Germany) (2006 Nov.).
- [13] F. Brinkmann, A. Lindau, S. Weinzierl, et al., “A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations,” *J. Audio Eng. Soc.*, vol. 65, no. 10, pp. 841–848 (2017 Oct.).
- [14] J. W. Lee, S. Lee, and K. Lee, “Global HRTF Interpolation via Learned Affine Transformation of Hyper-Conditioned Features,” in *Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5 (Rhodes Island, Greece) (2023 Jun.).
- [15] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, “A Cross-Evaluated Database of Measured and Simulated HRTFs Including 3D Head Meshes, Anthropometric Features, and Headphone Impulse Responses,” *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 705–718 (2019 Sep.).
- [16] International Phonetic Association, *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (Cambridge University Press, Cambridge, UK, 1999).
- [17] D. Aalto, O. Aaltonen, R.-P. Happonen, et al., “Large Scale Data Acquisition of Simultaneous MRI and Speech,” *Appl. Acoust.*, vol. 83, no. 1, pp. 64–75 (2014 Sep.). <https://doi.org/10.1016/j.apacoust.2014.03.003>.
- [18] J. Daudré, *Norman - Talking Head. Development of a Speaking Artificial Head to Study the Directivity of the Human Voice*, Master’s thesis, Technische Universität Berlin, Berlin, Germany (2024 Jun.).
- [19] J. Daudré, P. Luizard, M. Schneider, and S. Weinzierl, “Norman - Talking Head, Development and Evaluation of a Speaking Artificial Head to Study the Directivity of the Human Voice,” in *Proc. of German Acoustical Society Conf. (DAGA)* (Hannover, Germany) (2024 Mar.).
- [20] R. Kudelski, J. Cieslik, M. Kulpa, P. Dudek, K. Zagorski, and R. Rumin, “Comparison of Cost, Material and Time Usage in FDM and SLS 3D Printing Methods,” in *Proc. of XIIIth Intl. Conf. on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, pp. 12–14 (Lviv, Ukraine) (2017 Apr.).
- [21] International Electrotechnical Commission, “Sound System Equipment - Part 4: Microphones,” IEC 60268-4 (2018 Sep.).
- [22] International Electrotechnical Commission, “Electroacoustics - Octave-Band and Fractional-Octave-Band Filters - Part 1: Specifications,” IEC 61260-1 (2014 Feb.).
- [23] G. Fischer, C. Schneiderwind, and A. Neidhardt, “Comparing the Directivity of a Mouth Simulator and a Simple Physical Model,” in *Proc. of the 45th Annual Meeting on Acoustics (DAGA)* (Rostock, Germany) (2019 Mar.).
- [24] M. Brandner, R. Blandin, M. Frank, and A. Sontacchi, “A Pilot Study on the Influence of Mouth Configuration and Torso on Singing Voice Directivity,” *J. Acoust. Soc. Am.*, vol. 148, no. 3, pp. 1169–1180 (2020 Sep.). <https://doi.org/10.1121/10.0001736>.
- [25] C. Pörschmann, T. Lübeck, and J. M. Arend, “Impact of Face Masks on Voice Radiation,” *J. Acoust. Soc. Am.*, vol. 148, no. 6, pp. 3663–3670 (2020 Dec.). <https://doi.org/10.1121/10.0002853>.
- [26] P. Kocon and B. B. Monson, “Horizontal Directivity Patterns Differ Between Vowels Extracted From Running Speech,” *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. EL7–EL12 (2018 Jul.). <https://doi.org/10.1121/1.5044508>.
- [27] C. Pörschmann and J. M. Arend, “Investigating Phoneme-Dependencies of Spherical Voice Directivity Patterns,” *J. Acoust. Soc. Am.*, vol. 149, no. 6, pp. 4553–4564 (2021 Jun.). <https://doi.org/10.1121/10.0016821>.
- [28] C. Pörschmann and J. M. Arend, “On the Impact of Downward-Directed Human Voice Radiation on Ground Reflections,” *Acta Acust.*, vol. 8, p. 12 (2024 Mar.). <https://doi.org/10.1051/aacus/2024002>.
- [29] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, “Spatially Oriented Format for Acoustics 2.1: Introduction and Recent Advances,” *J. Audio Eng. Soc.*, vol. 70, no. 7/8, pp. 565–584 (2022 Jul./Aug.).
- [30] Pyfar, “Pyfar: Python Packages for Acoustic Research,” <https://pyfar.org> (accessed 1 Dec. 2024).
- [31] F. Wefers, “A Free, Open-Source Software Package for Directional Audio Data,” in *Proc. of the 36th German Annual Conf. on Acoustics (DAGA)* (Berlin, Germany) (2010 Mar.).
- [32] D. Schröder and M. Vorländer, “RAVEN: A Real-Time Framework for the Auralization of Interactive Virtual Environments,” in *Proc. of Forum Acusticum*, pp. 1541–1546 (Aalborg, Denmark) (2011 Jun./Jul.).
- [33] M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, and M. Vorländer, “The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing,” in *Proc. of the 43rd Annual German Congress on Acoustics (DAGA)*, vol. 6, pp. 222–225 (Kiel, Germany) (2017 Mar.).
- [34] F. Brinkmann and S. Weinzierl, “AKTools—An Open Software Toolbox for Signal Acquisition, Processing, and Inspection in Acoustics,” presented at the *142nd Conv. of the Audio Engineering Society* (2017 May), paper 309.

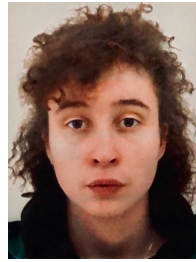
A.1 APPENDIX

The files described in this article are accessible under the link <https://doi.org/10.14279/depositonce-22511> and can be used under Creative Commons CC BY-NC-SA 4.0 license.

THE AUTHORS



Paul Luizard



Julie Daudré



Martin Schneider



David Ackermann



Stefan Weinzierl

Paul Luizard received an M.Sc. degree in acoustic engineering in 2010 and a Ph.D. degree in room acoustics in 2013 from Sorbonne Université in Paris, France. He then investigated human voice production through the vibration of vocal folds in Grenoble, France, and the vocal adaptation of singers to the acoustics of performance rooms in Berlin, Germany. He focuses now on voice directivity as a research associate in research projects in France and Germany. He is a member of the French (Société Française d'Acoustique [SFA]) and German (Deutsche Gesellschaft für Akustik [DEGA]) acoustic societies.

Julie Daudré studied physics at the French German University, visiting the universities of Nancy, Luxemburg, and Saarbrücken. For her bachelor thesis, she got into nondestructive control of materials using ultrasounds. Seeking broader perspectives, Julie continued her education in Germany, where she pursued with the Master in Audio Communication and Technology at the Technische Universität Berlin. Her master's thesis consisted in developing the Norman head presented in this article.

Martin Schneider joined the microphone development department of Georg Neumann GmbH, Berlin, in 1992 after obtaining his diploma in electrical engineering from Technische Universität Berlin. Since 2008, he has lectured

on electroacoustics at the Hochschule für Musik, Detmold, at Tonmeister Institute, and the Microphone Lab at Technische Universität Berlin at the Institute for Audio Communication.

David Ackermann received his M.Sc. in Audio Communication and Technology in 2015 and his Ph.D. (Doctor rerum naturalium [Dr. rer. nat.]) in 2024 from Technische Universität Berlin. He is a Professor of Audio Engineering at the University of Applied Sciences Düsseldorf, where his research focuses on virtual and musical acoustics and the further development of binaural reproduction methods for live streaming of 3D audio.

Stefan Weinzierl is head of the Audio Communication Group at the Technische Universität Berlin. His research is focused on audio technology, virtual acoustic reality, room acoustics, and musical acoustics. With a diploma in physics and sound engineering, he received his Ph.D. in musical acoustics. He is coordinating a master program in audio communication and technology at Technische Universität Berlin and has coordinated international research consortia in the field of virtual acoustics (Simulation and Evaluation of Acoustical Environments [SEACEN]) and music information retrieval Artist to Business to Business to Consumer Audio Branding System [ABC DJ].