DOI: 10.61782/fa.2023.0380

forum acusticum 2023

A 256-MICROPHONE MEASUREMENT SYSTEM TO ESTIMATE 3D DIRECTIVITY PATTERNS

P. Luizard[1,3,4*] **M. Hartenstein**[2] **H. Moingeon**[2]
C. Pinhède[1] **M. Pachebat**[1] **C. Ollivon**[2] **F. Ollivier**[2] **F. Silva**[1]
[1] Aix Marseille Univ, CNRS, Centrale Marseille, LMA, Marseille, France
[2] Sorbonne Université, CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, Paris, France
[3] Audio Communication Group, Technische Universität Berlin, Germany
[4] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

ABSTRACT

The first published attempt to measure human voice directivity in 1929 involved a single microphone and a rotating chair. Since then, a number of experiments were conducted, increasing the level of detail of the sound field representation according to the available equipment. Most experiments assumed repeatability of voice production by the human talker over an iterative process, which can hardly be guaranteed even for trained subjects. Hence using a rotating device to retrieve the 3D directivity of a human talker is questioned. All measurement positions should preferably be recorded simultaneously to ensure an optimal consistency of the obtained directivity patterns. The present experimental setup is composed of 256 MEMS microphones located on a spherical structure of 1.80 m radius. The characteristics (sensitivity, frequency range, and dynamics) of those compact devices are presented and allow for measurements of voice directivity from 50 Hz up to 20 kHz. A first step of characterization is conducted with a reproducible sound source composed of 12 Aurasound loudspeakers producing controlled directivity. Then, human talkers are recorded in the system. Far-field directivity functions are estimated using a spherical wave propagation model. The obtained results are consistent with the previous literature and provide an extended angular accuracy.

**Corresponding author*: *paul.luizard@lilo.org.*

Keywords: *Voice directivity, radiation patterns, MEMS microphones*

1. INTRODUCTION

Each sound source radiates acoustic energy in the surrounding space in a particular manner. This signature is the sound directivity of the source. The first published measurement of voice directivity was performed by means of a fixed single microphone and a human talker on a rotating chair [1]. Most research on this topic followed the paradigm of repeated measurements, i.e. used a microphone array which was rotated around the sound source while the latter would repeat its sound production, hence achieving a collection of measurement points around the source [2, 3]. Such an approach requires the ability of the sound source to repeat precisely its sound production. While this is the case for electro-acoustic equipment such as loudspeaker systems, even trained human beings cannot repeat several times exactly the same vocal production. In fact, small variations of pitch, intensity, and timber always occur. To overcome this limitation, another approach was developed, including a number of microphones distributed in the space surrounding the sound source [4]. The goal is then to measure the sound field in a single shot to avoid discrepancies related to repetition of the sound production. The spatial accuracy of such measurements depends on the number of microphones and the latest studies used 32 [5] and even 64 microphones [6, 7]. It is noted that the latter studies used a distribution of microphones along the horizontal and sagittal planes centered on the person.

The present paper describes the measurement system aimed at capturing the sound directivity of various

Table 1. Specifics from the constructor of the MEMS microphones.

Sensitivity	-26 dB FS ±1 dB
SNR	65 dBA
Current	490 µA
AOP	120 dB SPL
Sample rate	23 - 51.6 kHz

sources, including natural sources such as the human voice, by means of 256 simultaneously recorded channels from MEMS microphones distributed on a spherical structure. The results presented in this paper are restricted to the horizontal plane involving only 80 microphones, although the recorded data include the 256 channels.

2. MICROPHONE ARRAY

The microphone array is composed of 256 MEMS (Invensense ICS-43434) microphones whose specifics can be seen in Tab. 1. The standard deviation between all the frequency responses does not exceed 3 dB at 1 kHz. The microphones are arranged in 32 series of eight MEMS sliding along aluminium profiles in U shape of 5 mm section and 1 m length. The latter are fixed on a spherical structure of cylindrical steel bars of 3.6 m diameter illustrated in Fig. 1. Each series of eight MEMS is connected to a custom hub, engineered at Sorbonne Université (Paris, France), resulting in a single USB3 connection for the whole system. The system was built in the anechoic chamber at LMA laboratory (Marseille, France). A custom software was developed for the synchronous acquisition of the 256 channels at a sampling frequency of 50 kHz.

3. CONTROLLED SOUND SOURCES

As a first step, a sound source with controlled directivity [8] was used to test the ability of the system to retrieve expected sound fields. It is a 12-channel compact spherical loudspeaker array with 12 individual drivers (Aurasound NSW2-326-8A) uniformly distributed on a sphere of 15 cm diameter. The drivers can be controlled independently, i.e. they can simultaneously play different channels. The source was positioned at the center of the sphere, hence within the horizontal *equatorial* plane. The

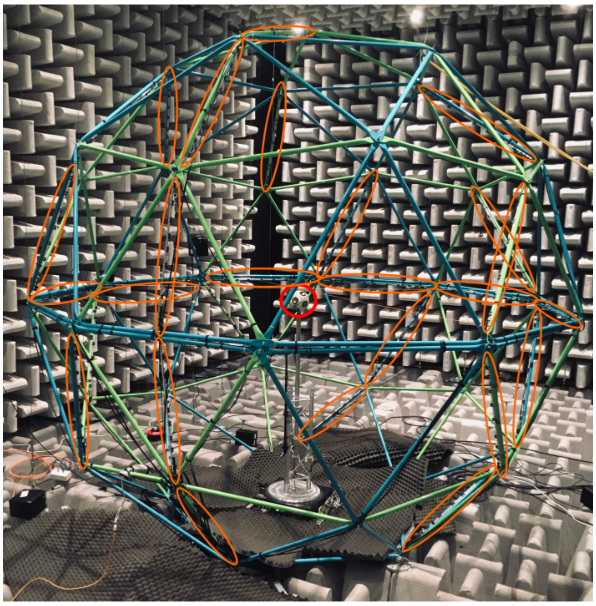

Figure 1. Picture of the spherical structure. The source is highlighted in red and the series of eight MEMS in orange.

excitation signal was an exponential synchronized sine sweep [9] from 20 Hz to 20 kHz of 10 s duration.

Two configurations of the sound source were used:

- *Omnidirectional sound source*, with the 12 drivers playing the same signal together.
- *Spherical cap*, with only one driver playing while the others, facing a MEMS that represented 0° in the horizontal plane.

During each measurement, the excitation signal was sent to the source and also directly to the recording interface so that it would follow the same path as the recordings themselves through the acquisition chain. A deconvolution of the latter signal with the sine sweeps recorded by the MEMS enabled to compute the 256 impulse responses of the MEMS microphones. The sound pressure level was further estimated from these impulse responses. These values were processed in the circular harmonics domain [10] at order 6 based on magnitude spectra. They are presented in Fig. 2 with self-maximum normalized amplitude, filtered in octave frequency bands for each source configuration.

The main characteristics of each configuration are retrieved. First, the monopole aspect of the *omnidirectional*

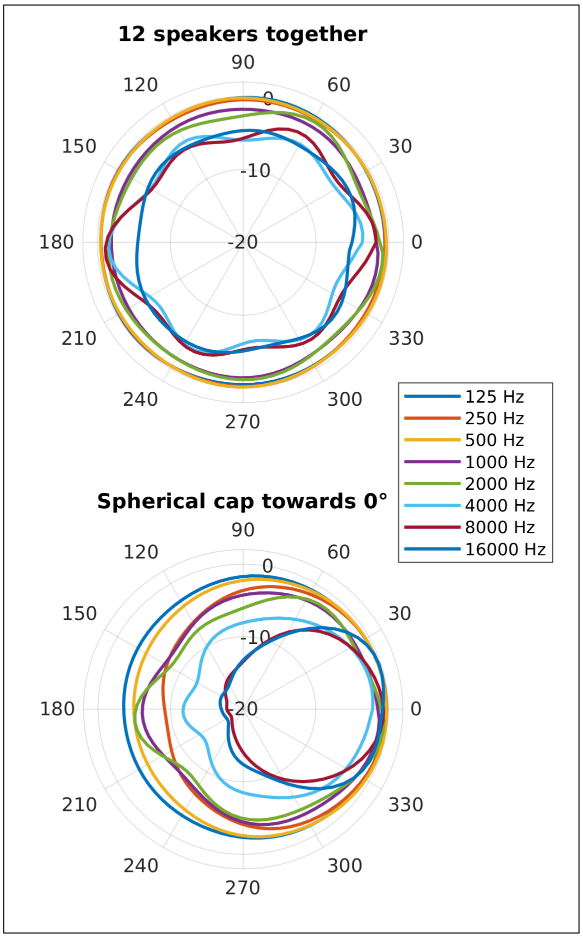

Figure 2. Directivity patterns normalized by the maximum value in the horizontal plane of the controlled source. Top: 12 drivers playing together. Bottom: Only one driver playing to the front, i.e. spherical cap condition.

source, i.e. a similar amount of sound energy being received at each point surrounding the source, even up to the 2 kHz octave band. Above that limit, the directivity pattern is not flat anymore and the lobes produced by each loudspeaker emerge, but a sort of circular shape remains. Second, the classical directivity patterns of the *spherical cap* source, being omnidirectional at low frequency and becoming more directional with increasing frequency. At about 1 kHz and above, the source cannot be considered omnidirectional anymore. More complex directivity patterns can be observed at the rear of the source, i.e. at the opposite side of the driver, at higher frequency.

4. MEASURING THE HUMAN VOICE

The next step was to perform measurements of directivity of the human voice. A participant was placed in the spherical structure. The mouth was located as close as possible to the center of the sphere by means of nylon strings attached to opposite points of the structure and crossing just above the center. The participant had to remain with the crossing just before the mouth. A series of phonemes was pronounced independently from each other. This study focuses on three vowels at the extreme points of the vowel diagram [11], namely /a/, /i/, and /u/, as well as three voiceless fricative consonants /f/, /s/, and /ʃ/.

Figure 3 shows the directivity patterns in the horizontal plane of the vowels and consonants in the octave bands centered on 125 Hz, 1 kHz, and 8 kHz. It is noted that only the 80 microphones located in the horizontal plane were involved here, i.e. these measurements were performed with an average angle of 360°/80 = 4.5° between consecutive MEMS. The signals from the MEMS were processed in the circular harmonics domain at order 6.

Results show different behavior depending on the frequency band, the vowel, and the consonant. At 125 Hz, all vowels are close to omnidirectional sources. At 1 kHz, differences arise with /a/ and /u/ being more directed towards the front than /i/. In the octave band centered on 8 kHz, only the vowels /a/ and /i/ had enough energy (although at low level, see Fig. 4) to present a classical directivity pattern of 10 dB higher energy to the front as compared to the back. /u/ had so low energy in this relatively high frequency range that no particular directivity pattern could be observed, hence a circular shape.

The consonants show different patterns: at low frequency (125 Hz), no particular pattern appears; at mid frequency (1 kHz), the patterns are similar although more interference can be seen at rear for /f/; and at high frequency (8 kHz), most of the energy is sent towards the front and almost no energy towards the rear, following a sort of cardioid shape. It is noted that these directivity patterns present strong asymmetry, especially for /f/ which is due to the particular sound production, not averaged over several recordings nor talkers.

Additional information can be gathered with Fig. 4 which presents a map of sound energy level in the plane of azimuth angles against frequency. The vowels appear in the two top rows, the only difference being the frequency range, up to 3 kHz in row 1 and up to 10 kHz in row 2.

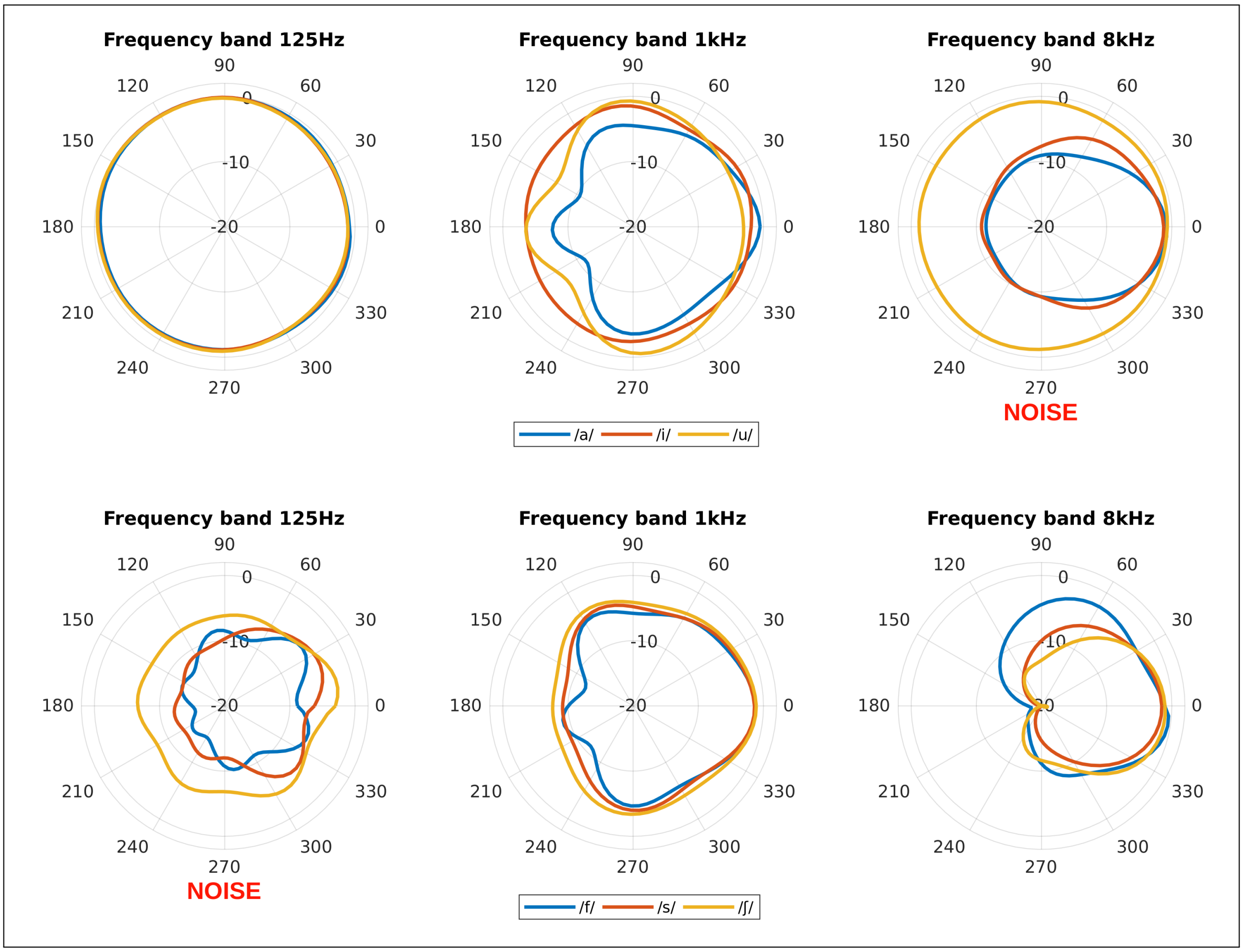

Figure 3. Directivity patterns normalized by the maximum value in the horizontal plane in frequency bands at order 6 of circular harmonics. Vowels appear in the upper row and fricative consonants in the lower row.

The first row shows the harmonics of the voice signal of the vowels while the second row presents a larger view of the spectra, highlighting the directivity of the formantic structure. It is noted that such maps for vowels are comparable to previous research [5, 12], in particular the C shape of the energy for vowel /a/ around 1 kHz. The third row shows the spectra up to 20 kHz of the consonants against the angles all around the source. One can notice that these fricative consonants are produced with energy at frequency up to 20 kHz.

5. DISCUSSION

The measurements of controlled sources have shown that the expected directivity patterns were retrieved on a broad frequency band from 100 Hz to 20 kHz, i.e. the entire range of voice production. Hence estimating directivity patterns based on measurements conducted with MEMS microphones is possible and meaningful. The obtained patterns are comparable with horizontal-plane observations observed in previous research [3–5].

The comparison of vowels and fricative consonants shows that the frequency contents is very different in each case (see Fig. 4): vowels present more energy at low frequency while these consonants have more energy at high energy, above 1 to 5 kHz depending on the consonant. The

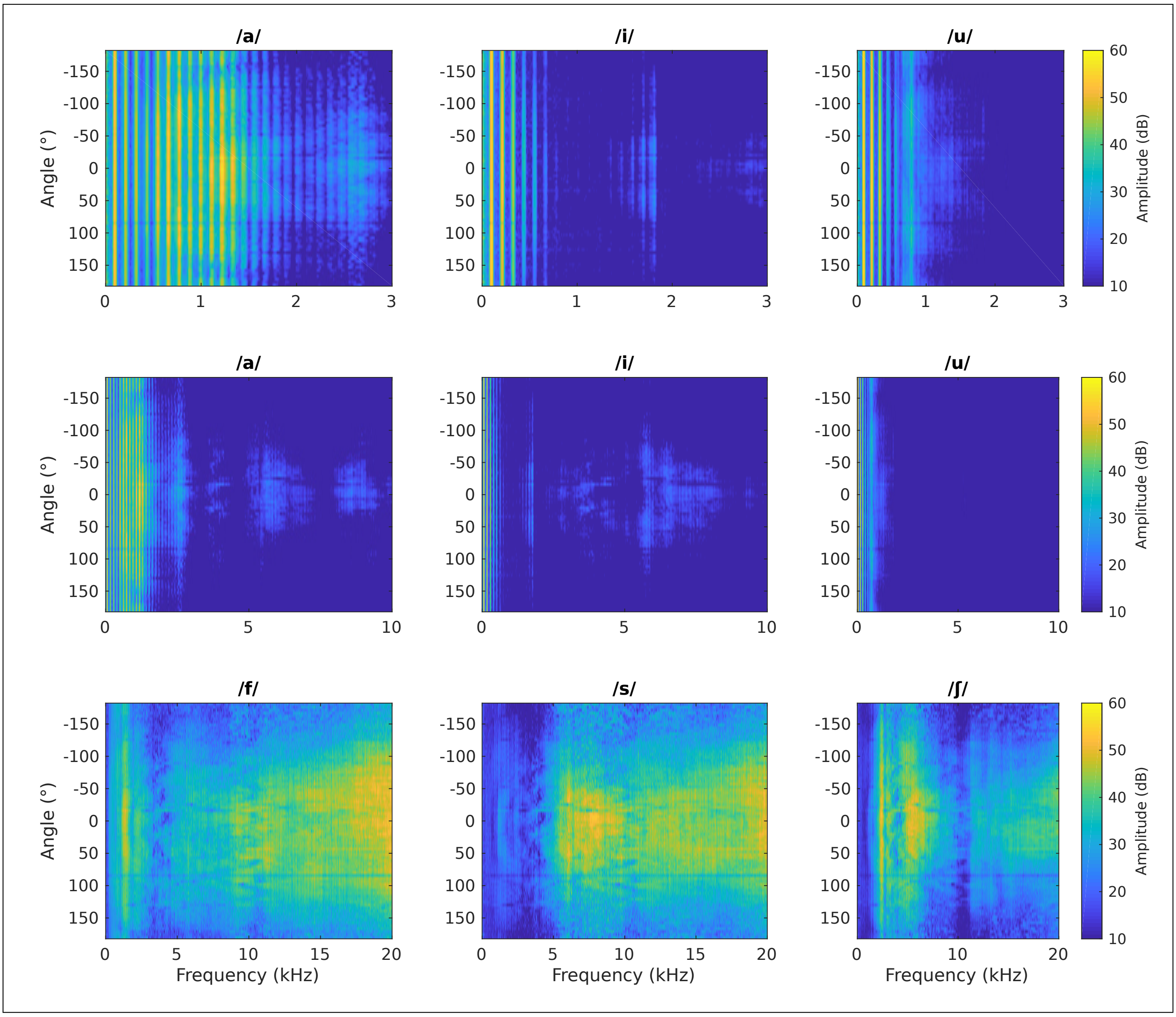

Figure 4. Sound power level (arbitrary amplitude scale) in the horizontal plane presented as angles against frequency. Vowels appear in the two upper rows (row 1: frequency range up to 3 kHz, row 2: up to 10 kHz) and fricative consonants in the lower row up to 20 kHz.

directivity patterns are then affected by this main difference: vowels have moderately strong directivity, i.e. voice is radiated around the talker, while fricative consonants have much stronger directivity, i.e. they can barely be heard behind the talker.

Between the vowels, the results present more similarity for /i/ and /u/ as compared to /a/ at mid frequency, while at high frequency, /a/ and /i/ are closer because they encompass enough energy for a directivity pattern to be observed, as opposed to /u/ whose energy has decreased by about 60 dB in this octave band. It should be noted that only three octave bands are considered in this study and a more thorough analysis might be required to generalize on and quantify these differences.

Concerning the consonants, /f/ has a slightly different behavior as compared to /s/ and /ʃ/. To a larger extent as for the vowels, the distribution of energy over the full frequency range varies between the consonants. It is known in phonetic science that fricative consonants cover a wide frequency range [13]. Yet, /f/ has less energy around 3.5 kHz, /s/ has very low energy below 5 kHz, and /ʃ/ has less energy around 10 kHz. Hence larger differences of directivity patterns would certainly be observed in narrower frequency bands, where energy is present or not, for each consonant.

6. CONCLUSION

The measurement system based on MEMS microphones was validated by means of controlled sources for which the expected directivity patterns could be retrieved. Recordings of sustained vowels and consonants were analyzed and specific directivity patterns enabled to compare these particular phonemes in terms of spatial distribution of the human voice. This measurement system presents a high number of microphones which enables to achieve a high spatial resolution. This was illustrated in Fig. 4 where detailed energy distribution around the talker is presented. Next steps include the analysis of full spoken sentences in order to study phonemes that were pronounced in ecological condition, and the analysis of all 256 available channels to reconstruct the 3D sound field around the sound source.

7. ACKNOWLEDGMENTS

This work was funded by the ANR (French National Research Agency) under the ”RAYOVOX” project ANR-21-CE42-0017-01, and by the DFG (German National Research Agency) under the ”Spatial characteristics of the human voice” project number 457863315.

8. REFERENCES

[1] F. Trendelenburg, “Beitrag zur frage der stimmrichtwirkung,” *Zeitschrift für technische Physik*, no. 11, pp. 558–563, 1929.

[2] W. T. Chu and A. Warnock, “Detailed directivity of sound fields around human talkers,” *NRC-CNRC, NRC Publications Archive Archives des publications du CNRC*, pp. 1–47, 2002.

[3] B. F. Katz, F. Prezat, and C. d’Alessandro, “Human voice phoneme directivity pattern measurements,” in *4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan*, p. 3359, 2006.

[4] B. B. Monson, E. J. Hunter, and B. H. Story, “Horizontal directivity of low-and high-frequency energy in speech and singing,” *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 433–441, 2012.

[5] C. Pörschmann and J. M. Arend, “Investigating phoneme-dependencies of spherical voice directivity patterns,” *The Journal of the Acoustical Society of America*, vol. 149, no. 6, pp. 4553–4564, 2021.

[6] M. Brandner, M. Frank, and D. Rudrich, “Dirpat—database and viewer of 2d/3d directivity patterns of sound sources and receivers,” in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.

[7] T. W. Leishman, S. D. Bellows, C. M. Pincock, and J. K. Whiting, “High-resolution spherical directivity of live speech from a multiple-capture transfer function method,” *The Journal of the Acoustical Society of America*, vol. 149, no. 3, pp. 1507–1523, 2021.

[8] A. M. Pasqual, *Sound directivity control in a 3-D space by a compact spherical loudspeaker array*. PhD thesis, Universidade Estadual de Campinas, 2010.

[9] A. Novak, P. Lotton, and L. Simon, “Synchronized swept-sine: Theory, application, and implementation,” *Journal of the Audio Engineering Society*, vol. 63, no. 10, pp. 786–798, 2015.

[10] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

[11] G. E. Peterson, “The phonetic value of vowels,” *Language*, pp. 541–553, 1951.

[12] M. Brandner, R. Blandin, M. Frank, and A. Sontacchi, “A pilot study on the influence of mouth configuration and torso on singing voice directivity,” *The Journal of the Acoustical Society of America*, vol. 148, no. 3, pp. 1169–1180, 2020.

[13] G. W. Hughes and M. Halle, “Spectral properties of fricative consonants,” *The journal of the acoustical society of America*, vol. 28, no. 2, pp. 303–310, 1956.