Analyzing the Directivity Patterns of Human Speakers

Christoph Pörschmann[1], Johannes M. Arend[1,2]
[1] TH Köln, Institute of Communications Engineering, Cologne, Germany
[2] TU Berlin, Audio Communication Group, Berlin, Germany
Email: christoph.poerschmann@th-koeln.de

Introduction

The first studies on the directional properties of human voice radiation were carried out more than 90 years ago by Trendelenburg [1], measuring directivity patterns for several vocals and fricatives in the horizontal plane. Later, Dunn and Farnsworth [2] determined directivity patterns for a spoken sentence at different distances in third-octave bands from 63 Hz up to 12 kHz, followed by Flanagan [3] who was the first to use a mannequin to determine sound radiation in the horizontal and vertical plane. Since then, voice directivity has been the subject of many studies using either human speakers or dummy heads. A specific characteristic of the human voice that cannot be analyzed with dummy heads is its dynamic directivity. To adequately determine these time-variant changes when speaking or singing, the sound radiation must be captured simultaneously for an appropriately high number of directions. In this context, Katz and D'Alessandro [4] analyzed voice directivity in the horizontal plane in angular steps of 15° for sustained vowels articulated by a professional opera singer. The study showed no systematic differences between the different vowels. Kocon and Monson [5] examined articulation-dependent effects of the voice directivity for different vocals in fluent speech. Here, the authors observed an effect of the vowel on the directivity pattern and determined the strongest directivity for an [a]. Monson et al. [6] analyzed time-variant effects in the horizontal plane and showed that the directivity varies strongly for different articulations, e.g. between voiceless fricatives. Furthermore, this study showed only minor dependencies on the articulation level.

So far only a few studies on spherical sound radiation have been published, e.g. [7, 8, 9]. To determine spherical directivities, it is advantageous to apply surrounding microphone arrays [9, 10, 11], which however are restricted to a limited number of sampling points and thus have a low spatial resolution. Consequently, methods for spatial upsampling of (sparsely) measured directivities are required. In this context, we presented the SUpDEq (Spatial Upsampling by Directional Equalization) method [12], which originally was designed for spatial upsampling of head-related transfer functions. In [13, 14] we applied the method to directivities of a dummy head with mouth simulator and showed that reasonable dense directivity sets can be obtained from sparse measurements. In this paper we now examine spatially upsampled human voice directivities and analyze individual differences between subjects as well as articulation-dependent features.

Measurements

All measurements were performed in the anechoic chamber of TH Köln, having a size of 4.5 m × 11.7 m × 2.3 m (W×D×H) and a lower cut-off frequency of about 200 Hz. We applied our surrounding microphone array, which has a basic shape of a pentakis dodecahedron with thirty-two Rode NT5 cardioid microphones located at the vertices of this shape on a constant radius of 1 m. This sampling scheme allows resolving the directivity up to a spatial order of $N = 4$ [10]. An additional Rode NT5 microphone was positioned at the front ($\phi = 0°$, $\theta = 0°$) as a reference. Four RME Octamic II devices served as preamplifiers and AD / DA converters for the 32 microphones of the array. All signals were managed with two RME Fireface UFX audio interfaces. One of these audio interfaces was also used as preamplifier and AD / DA converter for the reference microphone. Please refer to [11] for a more detailed description of the surrounding microphone array. As test stimuli we used five vocals ([a], [e], [i], [o], [u]) and three fricatives ([f], [s], [ʃ]) articulated by 13 subjects aged between 25 and 64 years (one female and 12 male). While the vocals were measured using the glissando method as proposed in [7, 9], the subjects articulated each of the fricatives for a duration of at least 3 s. Each of the articulations was measured twice per person.

Postprocessing

For each measured phoneme, we calculated the impulse response between each microphone signal of the surrounding microphone array and the frontal reference microphone. The resulting impulse responses were truncated and windowed to a final length of 128 samples at a sampling rate of 48 kHz. To mitigate the effect of reflections of the anechoic chamber below approximately 200 Hz, which is around the human fundamental frequency, we applied a low-frequency extension substituting the original low-frequency component by an adequately matched analytical description, as already proposed in [13]. Finally, we applied a distance error compensation described and evaluated in [15] to compensate for variations in distance between speaker and microphones caused by slight positioning inaccuracies of the array microphones and the placement of the human speaker in the center of the array.

Spatial Upsampling

In a next step, we performed spatial upsampling of the post-processed measurements using the SUpDEq method adapted for voice directivities [12, 13]. The basic idea is

as follows: The sparsely measured directivity is equalized by a direction-dependent spectral division with a set of rigid sphere transfer functions (hereinafter called equalization dataset), yielding an equalized sparse directivity dataset. Generally speaking, the equalization dataset represents a simplified directivity only featuring the basic form of a spherical head, but without any information on the specific shape of the mouth opening or the form of e.g. the cheekbones. As mouth opening position for the equalization dataset, we chose Ω_e as $\phi = 0°$ and $\theta = -25°$ [16]. The equalization removes energy from high spatial orders and therefore order-truncation and spatial aliasing errors are significantly reduced when transforming the equalized sparse dataset to the SH domain by a spatial Fourier transform (SFT). After the SFT, spatial upsampling (spherical harmonics interpolation) is performed by an inverse SFT on a dense grid. Finally, a de-equalized dataset is obtained by a directional multiplication with a set of rigid sphere transfer functions according to the dense grid. The described processing was done with Matlab using the SUpDEq toolbox [12].

We already evaluated the SUpDEq method for a dummy head with mouth simulator and showed that the approach leads to directivity patterns much closer to a reference than common order-limited SH interpolation without any preprocessing [13, 14]. Our studies revealed that array measurements with a spatial order of $N = 4$ are sufficient to generate a decent full-spherical dense directivity set, with an error averaged over the entire sphere below 4 dB for frequencies up to 8 kHz. Thus, it can be assumed that a surrounding microphone array with a number of 32 microphones can be used to reliably determine human voice directivities as well.

Results

In a first step, we examined the directivities in the horizontal and vertical plane averaged over all subjects for the articulations of vocals (Fig. 1) and fricatives (Fig. 2). We refrained from plotting the directivities for frequencies below 1 kHz as in this frequency range both individual and articulation-dependent differences can be almost neglected. As can be seen in the plots, there are only slight differences between the different articulations in the octave bands of 1 kHz and 2 kHz. However, the variations between the different vocals generally increase with frequency. For example, in the frequency bands of 4 kHz and 8 kHz, the directivity in the horizontal plane is stronger for an [a] than for the other vocals. This is in line with the findings of [5]. For fricatives we also observed significant differences in the horizontal plane. For example, in the frequency bands of 2 kHz and 4 kHz, the directivity for an [s] is more directional than for an [ʃ] or an [f]. This is consistent with the results of [6], who found that in the horizontal plane the directivity of an [s] is narrower than of an [f]. In general, these differences diminish in the vertical plane as both for vocals and fricatives only minimal differences occur, even at higher frequencies.

In a next step we determined the directivity index (DI), which can be calculated as

$$DI = 10\,log_{10}\frac{4\pi|p_{\Omega_e}|^2}{\int_0^{2\pi}\int_0^{\pi}|p(\phi,\theta)|^2 sin\theta d\theta d\phi}, \qquad (1)$$

with p_{Ω_e} the sound pressure measured at a defined distance in the main direction.

Phoneme	1 kHz	2 kHz	4 kHz	8 kHz	Type
[a]	3.35	7.35	9.11	10.96	vocals
[e]	2.50	7.71	8.74	9.92	
[i]	1.45	6.73	8.44	9.19	
[o]	3.57	6.98	7.39	8.91	
[u]	3.93	6.43	6.94	8.03	
[f]	2.52	5.75	6.60	7.70	fricatives
[s]	3.33	6.86	8.95	8.85	
[ʃ]	3.01	6.08	7.06	8.98	

Table 1: Directivity index in dB averaged over subjects for the different articulations in the $1\,kHz$, $2\,kHz$, $4\,kHz$, and $8\,kHz$ octave bands.

For vocals, the DI decreases with a decreasing size of the mouth opening. The largest mouth opening can be observed for an [a] and results thus in the highest DI. In contrast, the smallest DI was determined for an [u]. For the fricatives, maximal and minimal DIs were measured for an [s] and an [f] respectively. In general, the differences in the DIs between the articulations are quite small and do not exceed 3 dB in any of the octave bands.

Finally, we examined interindividual differences between the subjects. Fig. 3 shows the mean values and the standard deviations for two exemplary phonemes, an [a] and an [f]. As can be observed from the plot, the standard deviations are rather small. In the frontal hemisphere they do not exceed ±2 dB for frequencies up to 4 kHz and ±3 dB at 8 kHz. In general, the standard deviation tends to be higher for fricatives than for vowels, especially in the horizontal domain.

Discussion

In general, the differences tend to be maximal for rearward directions, both due to variations of the different articulations or due to interindividual differences. Here the propagation around the head results in a complex pattern with constructive and destructive interferences changing rapidly for small directional changes, especially towards higher frequencies. Even though this behavior is generally the same for all articulations and all participants, the detailed structure varies.

Apart from rearward directions, the directivities are quite smooth, even towards high frequencies. Peaks and dips which were found in some other studies, e.g. [4], were hardly observed. This might be a result of the reduced truncation and aliasing errors when applying the SUpDEq method for spatial upsampling. Comparing the

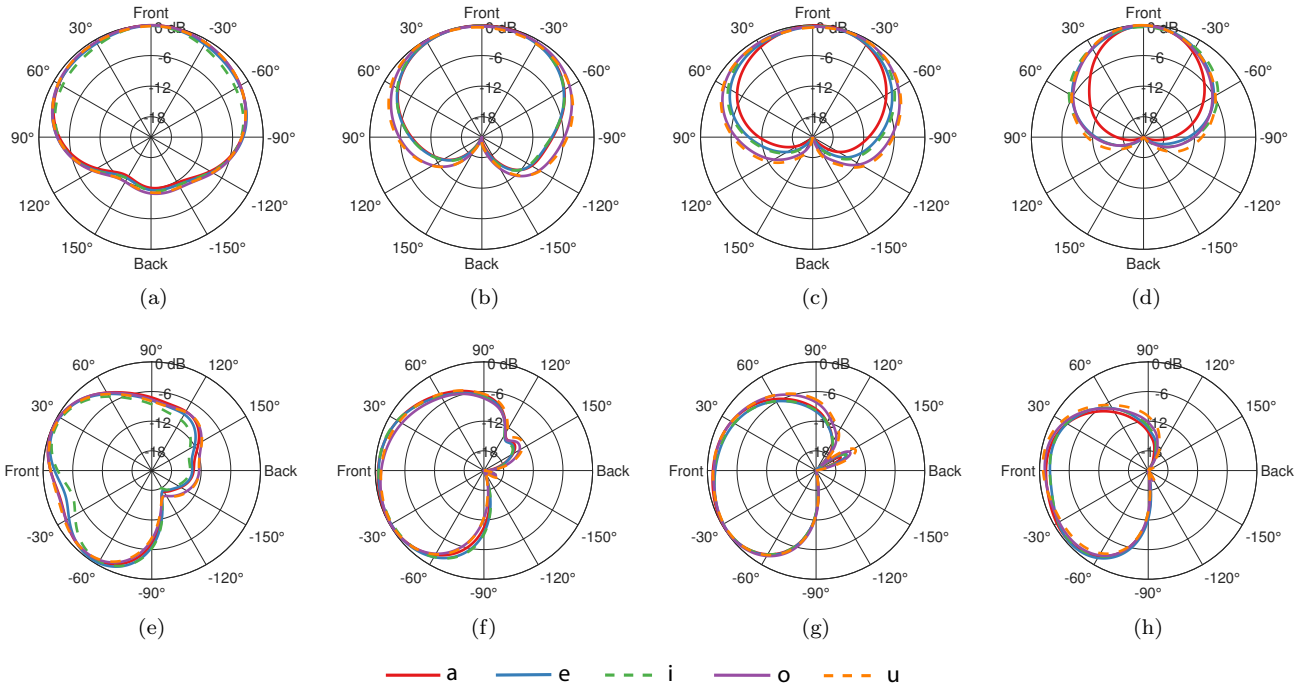

Figure 1: Directivities averaged over subjects in the horizontal (a – d) and vertical (e – h) plane determined for different vocals. (a,e): 1 kHz, (b,f): 2 kHz, (c,g): 4 kHz, (d,h): 8 kHz octave band.

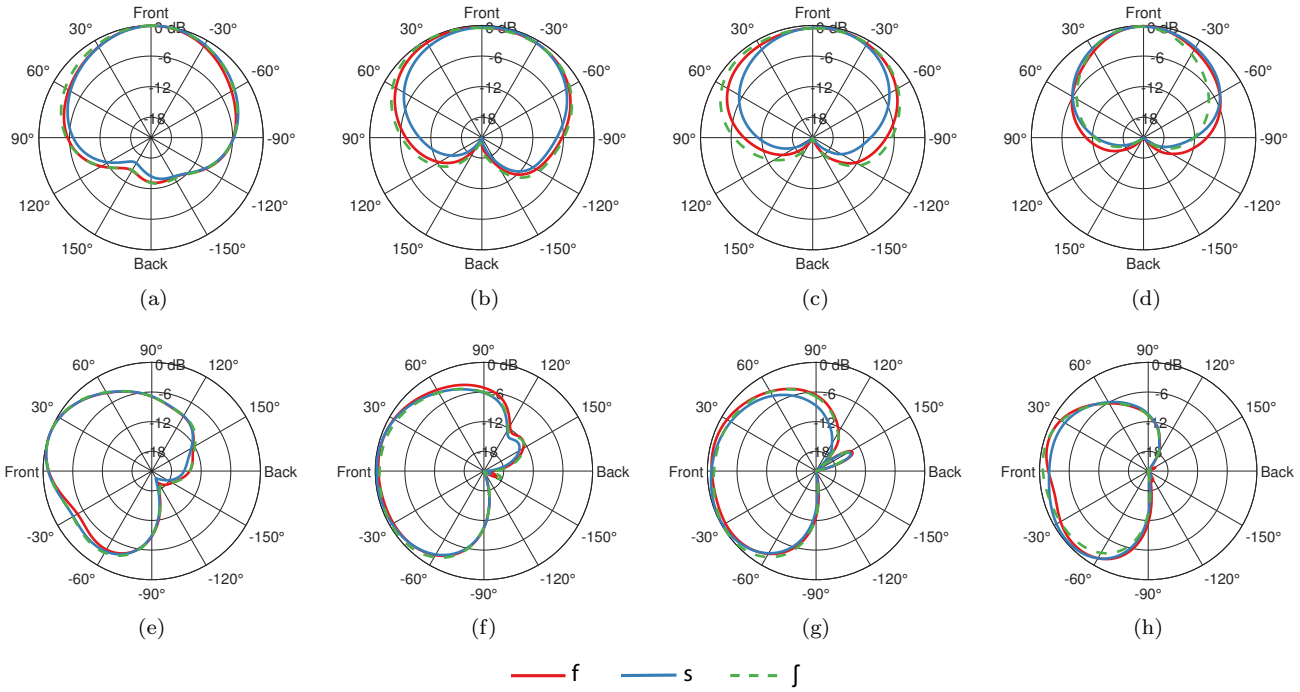

Figure 2: Directivities averaged over subjects in the horizontal (a – d) and vertical (e – h) plane determined for different fricatives. (a,e): 1 kHz, b,f): 2 kHz, (c,g): 4 kHz, (d,h): 8 kHz octave band.

directivities of fricatives and vocals did not reveal systematic differences. They generally have a similar shape and the DIs are in the same range.

The studies of [8, 17] are mostly in line with our results. In the frontal hemisphere, differences are within the standard deviation of our measurements. For rearward directions larger differences occur. This can be explained by the limited spatial resolution of the measurement data in these studies, which did not allow to resolve the exact contour of the directivity. Unfortunately no datasets of spherically measured voice directivities are available, so that no direct comparison is possible. However, if made

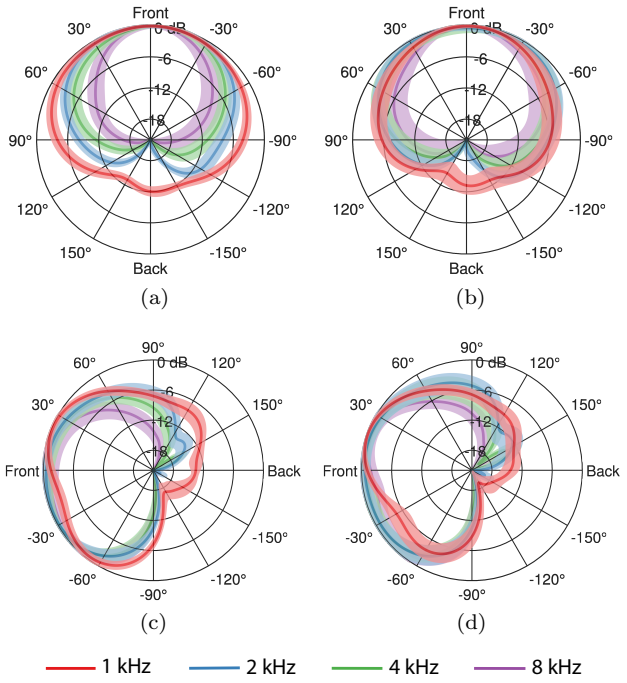

Figure 3: Mean directivities and standard deviations in the horizontal (a,b) and vertical (c,d) plane determined for an [a] (a,c) and an [f] (b,d) in the octave band of 1 kHz, 2 kHz, 4 kHz, and 8 kHz.

publicly available, other datasets, e.g. those of Brandner et al. [9] might be used for comparison to our studies.

Conclusion

We analyzed articulation-dependent directivity patterns obtained from measurements with 13 subjects. To generate dense datasets from sparse measurements, we applied the SUpDEq method adapted for voice directivities, previously evaluated with dummy head measurements [13, 14]. The results of the study show that the proposed method can be used to analyze articulation-dependent aspects of human speaker directivities. Furthermore, in the fields of virtual and augmented reality as well as in room acoustic simulation, datasets are required to adequately integrate human voice radiation patterns in sound field synthesis. For this purpose, it must to be examined whether interindividual differences or articulation-dependent characteristics are perceptually relevant and thus need to be taken into account. Finally, when reproducing one's own voice in a virtual acoustic environment to investigate the perception of self-generated speech, its directivity also plays an essential role [11, 18, 19].

Acknowledgement

The authors thank Raphael Gillioz for supporting the measurements. The research presented here has been carried out in the project NarDasS funded by the Federal Ministry of Education and Research in Germany, support code: BMBF 03FH014IX5-NarDasS.

References

[1] Trendelenburg, F., "Beitrag zur Frage der Stimmrichtwirkung," *Zeitschrift für techn. Physik*, 11, pp. 558–563, 1929.

[2] Dunn, H. K. and Farnsworth, D. W., "Exploration of pressure field around the human head during speech," *The Journal of the Acoustical Society of America*, 10, pp. 184–199, 1939.

[3] Flanagan, J. L., "Analog Measurements of Sound Radiation from the Mouth," *The Journal of the Acoustical Society of America*, 32(12), pp. 1613–1620, 1960.

[4] Katz, B. and D'Alessandro, C., "Directivity measurements of the singing voice," in *Proceedings of the 19th International Congress on Acoustics*, 2007.

[5] Kocon, P. and Monson, B. B., "Horizontal directivity patterns differ between vowels extracted from running speech," *The Journal of the Acoustical Society of America*, 144(1), pp. EL7–EL12, 2018.

[6] Monson, B. B., Hunter, E. J., and Story, B. H., "Horizontal directivity of low- and high-frequency energy in speech and singing," *The Journal of the Acoustical Society of America*, 132(1), pp. 433–441, 2012.

[7] Kob, M. and Jers, H., "Directivity measurement of a singer," *The Journal of the Acoustical Society of America*, 105, p. 1003, 1999.

[8] Chu, W. T. and Warnock, A. C. C., "Detailed Directivity of Sound Fields Around Human Talkers," Technical report, 2002.

[9] Brandner, M., Frank, M., and Rudrich, D., "DirPat - Database and Viewer of 2D/3D Directivity Patterns of Sound Sources and Receivers," in *Proceedings of 144th AES Convention, e-Brief 425*, 1, pp. 1–5, 2018.

[10] Pollow, M., *Directivity Patterns for Room Acoustical Measurements and Simulations*, Logos Verlag Berlin, 2015.

[11] Arend, J. M., Lübeck, T., and Pörschmann, C., "A Reactive Virtual Acoustic Environment for Interactive Immersive Audio," in *Proceedings of the AES Conference on Immersive and Interactive Audio*, 2019.

[12] Pörschmann, C., Arend, J. M., and Brinkmann, F., "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(6), pp. 1060 – 1071, 2019.

[13] Pörschmann, C. and Arend, J. M., "A Method for Spatial Upsampling of Directivity Patterns of Human Speakers by Directional Equalization," in *Proceedings of the 45th DAGA*, pp. 1458 – 1461, 2019.

[14] Pörschmann, C. and Arend, J. M., "Spatial Upsampling of Voice Directivities by Directional Equalization," *submitted to Journal of the AES*, 2020.

[15] Pörschmann, C. and Arend, J. M., "How positioning inaccuracies influence the spatial upsampling of sparse head-related transfer function sets," in *Proceedings of the International Conference on Spatial Audio - ICSA 2019*, pp. 1–8, 2019.

[16] Marshall, A. H. and Meyer, J., "The directivity and auditory impressions of singers," *Acustica*, 58, pp. 130–140, 1985.

[17] Moreno, A. and Pfretzschner, J., "Human Head Directivity in Speech Emission: A new approach," *Acoustics Letters*, 1, pp. 78–84, 1978.

[18] Pörschmann, C., "One's own voice in auditory virtual environments," *Acta Acustica united with Acustica*, 87(3), pp. 378–388, 2001.

[19] Neidhardt, A., "Detection of a nearby wall in a virtual echolocation scenario based on measured and simulated OBRIRs," in *Proceedings of the AES Conference on Spatial Reproduction*, 2018.