Frequency Dependencies of the Main Radiation Direction of the Human Voice

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

Introduction

For communication in acoustically difficult situations, the directivity of the human voice plays an important role, for example, when we face a person in a meeting or casual conversation. While the directivity in the horizontal plane has been investigated in numerous investigations, only a few studies analyzed spherical voice radiation. The first one has been published in 1939 by Dunn and Farnsworth [1] and determined spherical directivity patterns for a spoken sentence and at different distances in octave or half-octave bands from 63 Hz up to 12 kHz for a spoken sentence. Their data allows resolving specific aspects of voice directivity in vertical directions, such as the frequency-dependent main radiation direction (MRD). Chu and Warnock [2] provided mean and standard deviations in third-octave bands of measurements for 40 subjects determined at 92 positions on a spherical grid showing that the MRD of the human voice varies over frequency. While the radiation is directed slightly downwards for most of the frequency bands, at around 1 kHz the MRD is elevated upwards by about 30°. Recently, Leishman et al. [3] published results of measurements with six subjects carried out with a resolution of 5° in the horizontal and vertical planes supporting these findings. Because in all these studies the directivity patterns were averaged over a whole sentence, the data do not provide information about phoneme or articulation dependencies.

In Pörschmann and Arend [4, 5], we determined full-spherical directivity patterns of 23 phonemes of various groups for 13 persons using a surrounding spherical microphone array with 32 microphones and studied the directivity patterns as well as the directivity index. The results revealed statistically significant differences between the groups of phonemes and between some of the phonemes within each group. The present work extends this analysis by studying the MRD for the different phonemes showing that the MRD is slightly downwards in a wide frequency range, which is in line with earlier research, e.g. Marshall and Meyer [6]. In agreement with Dunn and Farnsworth [1] and Chu and Warnock [2], our results also show that at around 1 kHz, the main radiation rapidly changes and is slightly directed upwards to about 15°.

Accordingly, when considering voice radiation in room acoustics or virtual acoustic environments, not only the horizontal directional characteristics must be considered, but as well the vertical ones accounting for these specifics of the MRD.

Materials

This work employs the datasets from our previous studies [4, 5], which are available open-source as supplementary material to these publications[1]. The corresponding measurements were carried out in the anechoic chamber of TH Köln applying a surrounding spherical microphone array [7] with a basic shape of a pentakis dodecahedron and 32 Rode NT5 cardioid microphones located at the vertices on a constant radius of 1 m, allowing to resolve the directivity up to a spatial order of $N = 4$ [8]. An additional Rode NT5 microphone was positioned at the front ($\phi = 0°$, $\theta = 0°$), used as a reference and for spectral equalization in postprocessing.

None of the 13 subjects (2 female and 11 male) who participated in the measurements sang professionally, but some of them did so as amateurs. They self-reportedly had no speech defect. The procedure included two measurements for each of the following articulations:

- **Vowels**: [a], [e], [i], [o], [u]
- **Plosives**: [p], [t], [k], [b], [d], [g]
- **Unvoiced fricatives**: [f], [s], [ʃ], [x], [h]
- **Voiced fricatives**: [z], [v]
- **Nasals**: [m], [n], [ŋ]
- **Voiced alveolars**: [l], [r]

To determine the directivity patterns, we chose an appropriate voice excitation signal for each phoneme. For all voiced phonemes, we applied the glissando method [9, 10], which means the subjects sang the phoneme with an increasing pitch over at least an octave. For the plosives and the unvoiced fricatives, the participants repeated the respective phoneme three times for each measurement. As an exception, for the unvoiced [x], the subjects continuously articulated the phoneme for at least three seconds. Postprocessing, as described in Pörschmann and Arend [11, 4, 5], contained the following steps. To eliminate the influence of room modes that become prominent at low frequencies, a low-frequency extension replaced the measured signal by an analytical signal. Furthermore, the inaccuracies in positioning the subjects in the center of the microphone array were compensated by an appropriate recentering. The most important postprocessing step was a spatial upsampling of the measured sparse datasets to a dense grid with 2702 sampling points on a Lebedev sampling scheme applying the SUpDEq [11] method.

[1] The datasets are available at `https://doi.org/10.5281/zenodo.7452117`.

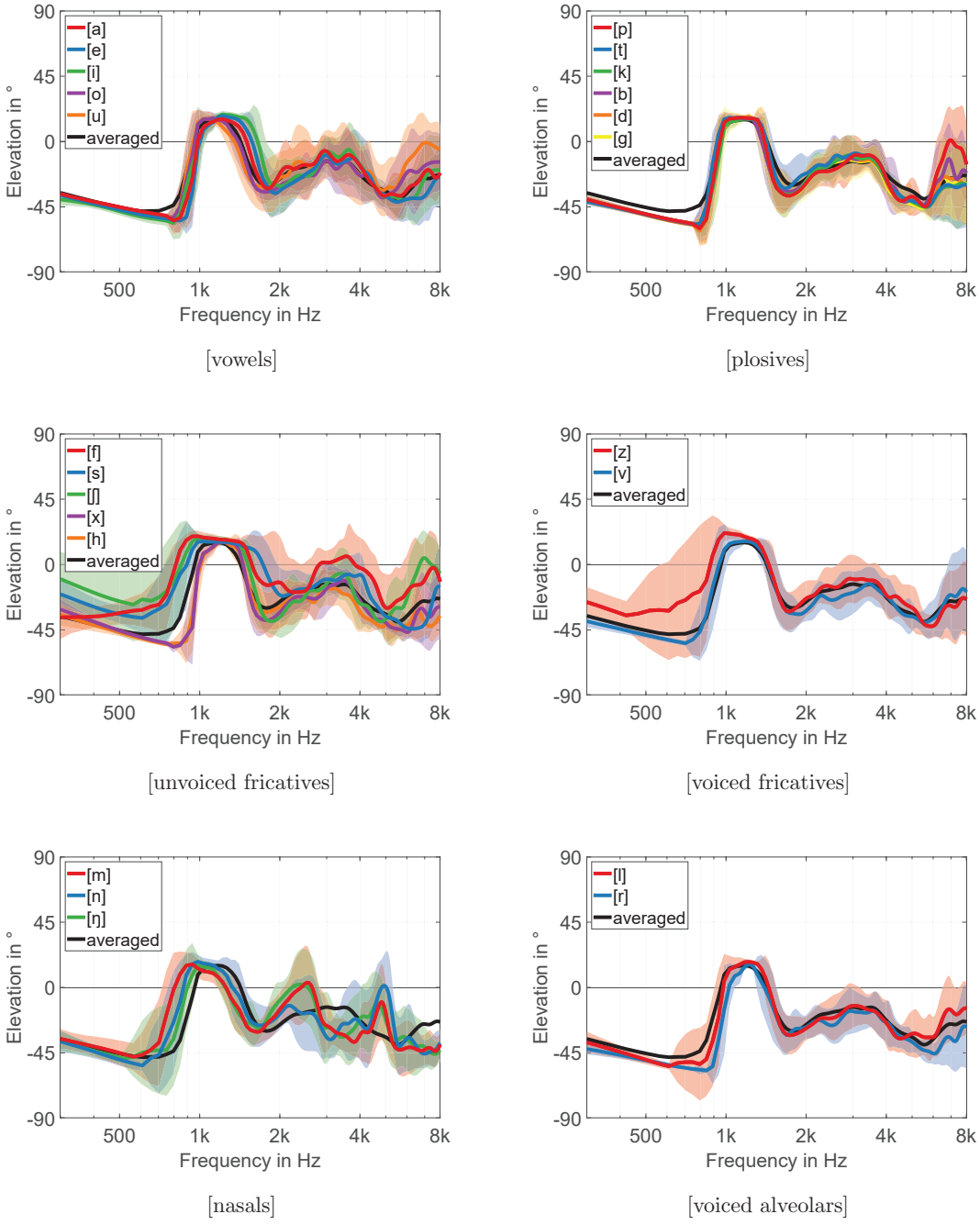

Figure 1: Mean values and standard deviations of the MRD (twelfth-octave smoothed) for the groups of phonemes. For comparison in each plot the MRD averaged over all phonemes is shown in black.

Based on these dense spherical datasets, the MRD was analyzed in the vertical plane. For this, the dense datasets were interpolated to 180 directions in steps of 1° ($\theta = -90°$ to $90°$, where positive angles point upwards; $\phi = 0°$), which was done by transforming the upsampled dataset to the spherical harmonics (SH) domain at an SH order of $N = 35$ and then resampling the datasets using the inverse SH transform. Then, a frequency-dependent value of the MRD was calculated for each of the measured datasets. Furthermore, the MRD gain was determined, which describes the level difference of the radiation towards the MRD and the frontal direction.

Results

To analyze the frequency and phoneme dependencies, the mean values and the standard deviations of the MRD are shown in Fig. 1. The plots show that for most of the frequencies, the MRD is slightly downwards in a range between $\theta = -15°$ to $-45°$, confirming the results of previous studies [1, 6]. Furthermore, the plots reveal that for all phonemes, the MRD sharply rises to $\theta \approx +15°$ at $f \approx 800\,\mathrm{Hz}$ and falls back to $\theta \approx -20°$ at $f \approx 1.6\,\mathrm{kHz}$. A more detailed comparison of the curves for the different phonemes shows that the deviations within the groups of vowels, nasals, and especially plosives are small.

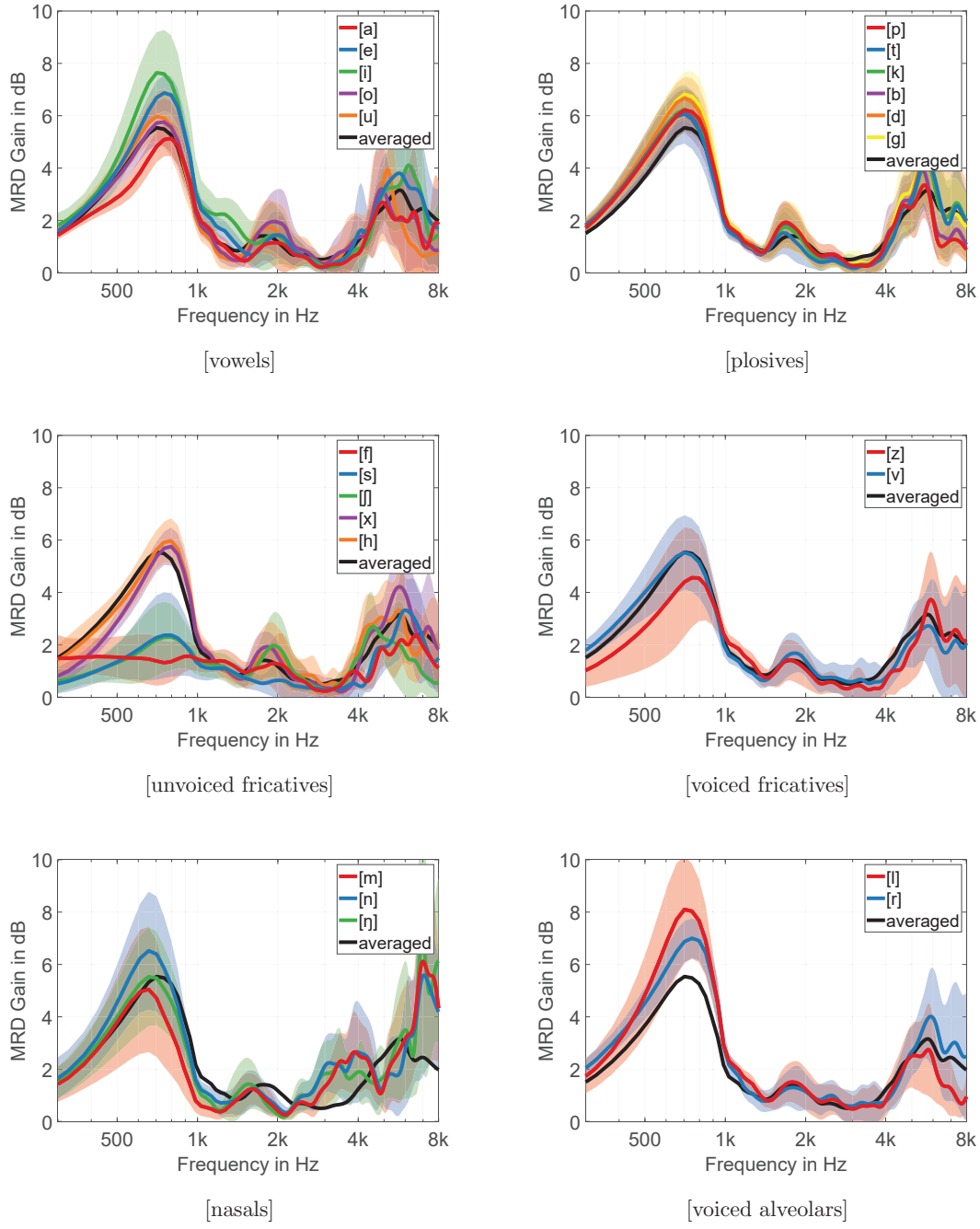

Figure 2: Mean values and standard deviations of the MRD gain related to the frontal direction for the groups of phonemes (twelfth-octave smoothed). For comparison in each plot the MRD gain averaged over all phonemes is shown in black.

Above 4 kHz, the differences between the phonemes increase, which could be due to the higher spatial complexity of the directional characteristics with dips at different frequencies and directions. For the nasals, the MRD curves are slightly shifted toward lower frequencies, so that the sharp increase from a downward to an upward MRD begins at about 700 Hz. The MRD of the voiced [z] is similar to that of the voiceless fricatives [s], and the [ʃ]. However, the standard deviations of the [z] below 1 kHz are much larger than for all other phonemes. Fig. 2 shows the mean values and the standard deviations of the MRD gain. For most of the phonemes, there is a prominent peak at $f \approx 700$ Hz. For the [s] and the [ʃ], there is only a moderate increase and for the [f], an increase can hardly be observed. Between 1 kHz and 4 kHz, the MRD gain stays below 2 dB for all phonemes. Above 4 kHz, only for the nasals the MRD gain shows a significant increase reaching 6 dB. For all other phonemes, the MRD gain remains below 4 dB which is in line with our observations in Pörschmann and Arend [5], where we found for the nasals in this frequency range the lowest directivity index of all phonemes.

Discussion

This study examined the frequency dependencies of the MRD of human voice radiation and investigated to what extent the radiation towards the MRD is stronger than in the frontal direction. Our results can be compared with other studies, but all of them were obtained by averaging over a complete sentence and thus cannot resolve phoneme dependencies. Dunn and Farnsworth [1] performed measurements in octave bands for $f < 500\,\mathrm{Hz}$ and half-octave bands for $f \geq 500\,\mathrm{Hz}$ at a distance of 1 m and a vertical resolution of 45°. They found that in the half-octave bands between 500 Hz and 1000 Hz, the main direction is downward ($\theta = -45°$) and between 1000 Hz and 1400 Hz upward ($\theta = +45°$), which is consistent with our results. It is worth noting that the authors observed a similar trend for near-field measurements. Marshall and Meyer [6] showed the vertical voice directivity in octave bands with center frequencies between 125 Hz and 8 kHz and found the MRD at about $\theta = -20°$. Because this study provides data only in octave bands, the abrupt changes in MRD around 1 kHz are not resolved. However, in the 1 kHz octave band, the strength of the radiation is similar for $\theta = -20°$ and $\theta = +45°$. Chu and Warnock [2], who performed measurements in one-third octave bands, found downward MRDs ($-42° \leq \theta \leq -13°$) for the third-octave bands with center frequencies between 315 Hz and 800 Hz and an MRD upwards ($\theta = 31°$) in the frequency bands with center frequencies of 1 kHz and 1.25 kHz, which is also in line with our results. Similarly, the results of Moreno and Pfretzschner [12] show MRDs upwards and downwards in these frequency bands, but with slightly different absolute values, probably due to the varying sampling grid.

The MRD gain as depicted in Fig. 2 shows a prominent peak for most of the phonemes at $f \approx 700\,\mathrm{Hz}$. Similarly, the spherical directivity index has a characteristic dip in this frequency range [4, 5], and both are direct counterparts. Since the directivity index is calculated with respect to the frontal direction, it decreases when the radiation in other directions increases. It is worth mentioning that the dip in directivity index is less prominent for the unvoiced fricatives [f], [s], and [ʃ], and so is the MRD gain for these phonemes (cf. [4] and Fig. 2).

Although the described behavior of the MRD can be determined from literature data, it has not yet been explored in detail and no causes for the increase in MRD gain at $f \approx 700\,\mathrm{Hz}$ have been identified. Since our results show that the behavior is similar for a large number of phonemes, it is unlikely to be caused by properties of specific phonemes, e.g., different mouth opening sizes and radiation through the mouth and nostrils in nasals. In this context, Birkholz et al. [13] recently simulated the frontal radiation of a modeled human mouth and torso, which showed a sharp dip at $f \approx 800$ Hz and a rise to a (local) maximum at $f \approx 1.3\,\mathrm{kHz}$. Their comparative simulations of a spherical cap in a spherical baffle did not show these peaks and dips. Accordingly, shoulder and torso influences could be a cause for these effects, which needs to be further analyzed in future research.

Conclusion

In this study, the main radiation direction (MRD) of the human voice for a variety of phonemes has been analyzed. The results show a characteristic pattern of MRD across frequency that differs only slightly between phonemes. The results emphasize the importance of determining voice directivity not only in the horizontal plane but also for different elevation angles or for a spherical sampling grid. Furthermore, the results show that the assumption of a slightly downward MRD of the human voice is reasonable in a wide frequency range, but is not valid between 800 Hz $\leq f \leq 1.6\,\mathrm{kHz}$.

References

[1] 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, doi: 10.1121/1.1915975.

[2] Chu, W. T. and Warnock, A. C. C., "Detailed Directivity of Sound Fields Around Human Talkers," Technical Report NRC-IRC-15212, National Research Council of Canada, Ottawa, 2002, doi:10.4224/20378930.

[3] Leishman, T. W., Bellows, S. D., Pincock, C. M., and Whiting, J. K., "High-resolution spherical directivity of live speech from a multiple-capture transfer function method," *The Journal of the Acoustical Society of America*, 149(3), pp. 1507–1523, 2021, doi:10.1121/10.0003363.

[4] Pörschmann, C. and Arend, J. M., "Investigating phoneme-dependencies of spherical voice directivity patterns," *The Journal of the Acoustical Society of America*, 149(6), pp. 4553 – 4564, 2021, doi:10.1121/10.0005401.

[5] Pörschmann, C. and Arend, J. M., "Investigating phoneme-dependencies of spherical voice directivity patterns II: Various groups of phonemes," *The Journal of the Acoustical Society of America*, 153(1), 2023, doi:10.1121/10.0016821.

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

[7] 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.

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

[9] Kob, M. and Jers, H., "Directivity measurement of a singer," *The Journal of the Acoustical Society of America*, 105, p. 1003, 1999, doi:10.1121/1.425813.

[10] Brandner, M., Blandin, R., Frank, M., and Sontacchi, A., "A pilot study on the influence of mouth configuration and torso on singing voice directivity," *The Journal of the Acoustical Society of America*, 148(3), pp. 1169 – 1180, 2020, doi:10.1121/10.0001736.

[11] Pörschmann, C. and Arend, J. M., "A Method for Spatial Upsampling of Voice Directivity by Directional Equalization," *Journal of the Audio Engineering Society*, 68(9), pp. 649–663, 2020, ISSN 15494950, doi:10.17743/jaes.2020.0033.

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

[13] Birkholz, P., Ossmann, S., Blandin, R., Willbrandt, A., Krug, P. K., and Fleischer, M., "Modeling speech sound radiation with different degrees of realism for articulatory synthesis," *IEEE Access*, pp. 95008–95019, 2022, doi:10.1109/ACCESS.2022.3204816.