Supplementary material for

Investigating phoneme-dependencies of spherical voice directivity patterns II: Various groups of phonemes

The Journal of the Acoustical Society of America, 153 (1), 2023

https://doi.org/10.1121/10.0016821

Christoph Pörschmann[1], **and Johannes M. Arend**[2]

[1]*Institute of Communications Engineering, TH Köln - University of Applied Sciences*

[2]*Audio Communication Group, Technical University Berlin*

christoph.poerschmann@th-koeln.de
j.arend@tu-berlin.de

Technology
Arts Sciences
TH Köln

Directivity datasets

The datasets contain measured directivity patterns from two different measurement series, the first one published by Pörschmann and Arend [1] studying vowels and fricatives, and the second one [2] studying various groups of phonemes. All directivity patterns are stored in the SOFA format [3] according to the SOFA convention "GeneralTF". The measurements took place in the anechoic chamber of TH Köln applying a surrounding spherical microphone array, which has a diameter of 2 m, and a shape of a pentakis dodecahedron with 32 cardioid microphones (Rode NT5) located at the vertices [4].

The datasets comprise two measurements (*repeat#*) for each of the 13 subjects (*subject#*). Each subject's dataset contains directivity patterns for the following groups of phonemes:

- **Vowels**: [a], [e], [i], [o], [u] (phoneme#: 1 – 5); from [1]
- **Fricatives**: [f], [s], [ʃ] (phoneme#: 6 – 8); from [1]
- **Plosives**: [p], [t], [k], [b], [d], [g] (phoneme#: 9 – 14); from [2]
- **Nasals**: [m], [n], [ŋ] (phoneme#: 15 – 17); from [2]
- **Fricatives**: [z], [v], [x], [h] (phoneme#: 18 – 21); from [2]
- **Voiced alveolars**: [l], [r] (phoneme#: 22 – 23); from [2]

The naming of the measured and postprocessed datasets with 32 sampling points is: Directivity_{*subject#*}_{*repeat#*}.sofa

Additionally, datasets which have been spatially upsampled to a dense Lebedev grid with 2702 sampling points using the SUpDEq method [5] are filed: Directivity_{*subject#*}_{*repeat#*}_upsampled.sofa

The transfer functions for each direction of the directivity pattern are stored in the data field *Data* separately for the real and imaginary parts. The data field *ReceiverPosition* contains azimuth ϕ and elevation θ for each direction and in the third field the subject's optimal head radius is given, which we determined according to Algazi et al. [6] based on measurements of each subject's head width, height, and length.

The Matlab script `Plot_Voice_Directivity_SOFA_2023.m`, which is included in the folder, reads a SOFA file and creates various directivity plots. For the sparse datasets, it performs spatial upsampling using the SUpDEq method. The script requires the SUpDEq toolbox[1].

[1] Available: https://github.com/AudioGroupCologne/SUpDEq

Comparison of measurement series

Figure 1 compares the directivity patterns measured in the first [1] and second [2] measurement series. The plots show the directivity patterns in the horizontal and vertical plane in octave bands with center frequencies between 250 Hz and 8 kHz and averaged over all subjects for the [a].

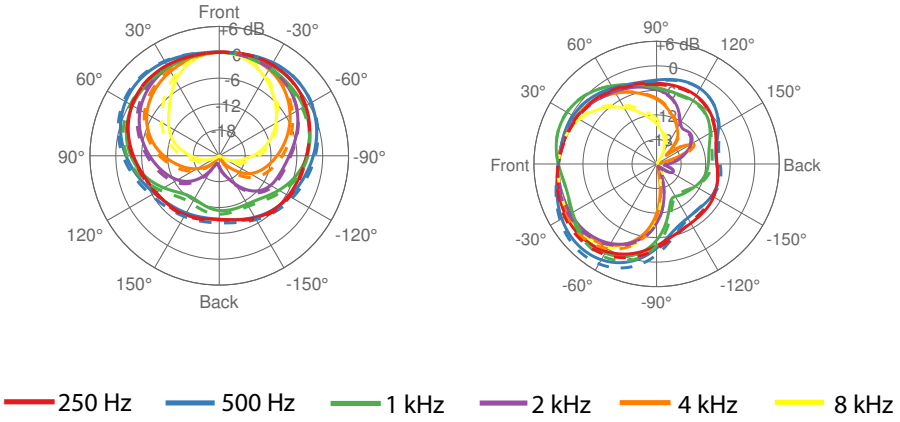

Figure 1: Directivity in the horizontal plane (left) and the vertical plane (right) of an [a] measured in the first [1] (solid) and second [2] (dashed) measurement series. Shown are the mean values in the octave bands with center frequencies between 250 Hz and 8 kHz.

Figure 2 shows the spectral deviations ΔG_{sp} between the measurements of the [a] from the two measurement series in the horizontal and vertical plane. The plot shows that for frequencies up to 4 kHz, ΔG_{sp} is nearly completely below 1 dB. Only for downward and backward sound radiation, larger values occur, e.g., at about 800 Hz. For frequencies above 4 kHz, the plots show increased values of ΔG_{sp}.

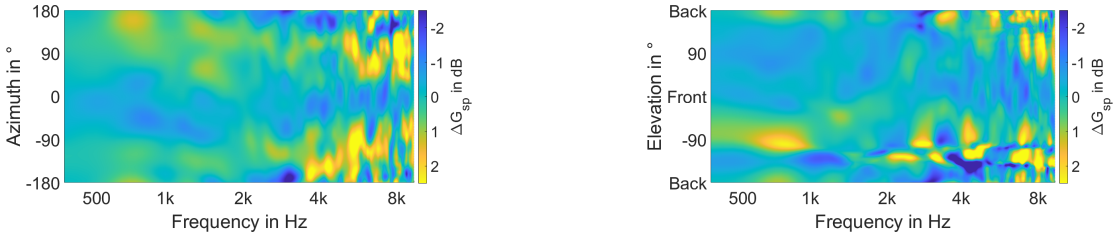

Figure 2: Spectral deviations ΔG_{sp} between the [a] articulated in the two measurement series. Horizontal plane (left) and vertical plane (right)

Figure 3 shows the mean values and the standard deviations of the DI for the articulation of the [a] in the two measurement series. The plot shows quite well that the differences are much smaller than the standard deviations in most of the frequency bands. Only for the third-octave bands between 400 Hz and 630 Hz, the differences are in the range of the standard deviations. For statistical evaluation of the differences, we analyzed the DIs of the [a] from both measurement series using a two-way repeated measures ANOVA with the within-subjects factors *frequency* and *measurement series*, corrected for slight violations of ANOVA assumptions using the Greenhouse-Geisser correction. The ANOVA yielded a significant main effect of measurement series [$F(1,12) = 25.22$,

$p_{GG} < .001$, $\eta_p^2 = .68$, $\epsilon = 1$], indicating differences between the measurements. Using paired t tests with Hochberg correction for multiple testing, we further examined in which frequency bands the DI values from the two measurement series differ significantly. The post-hoc analysis revealed significant differences in the third-octave bands centered at 400 Hz, 500 Hz, and 630 Hz (all $p < .001$), indicating differences in the DIs of the respective measurement series in the lower frequency ranges.

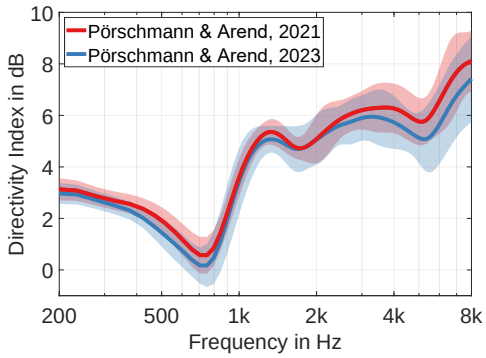

Figure 3: Mean values and standard deviations of the directivity index (DI) averaged over all subjects for the articulation of an [a] related to the frontal direction for the two different measurement series [1, 2] (third-octave smoothed).

Third-octave polar plots

Figure 4 – 15 show the directivity patterns in the horizontal and vertical plane in third-octave bands with center frequencies between 315 Hz and 8 kHz and averaged over all subjects for the six groups of phonemes.

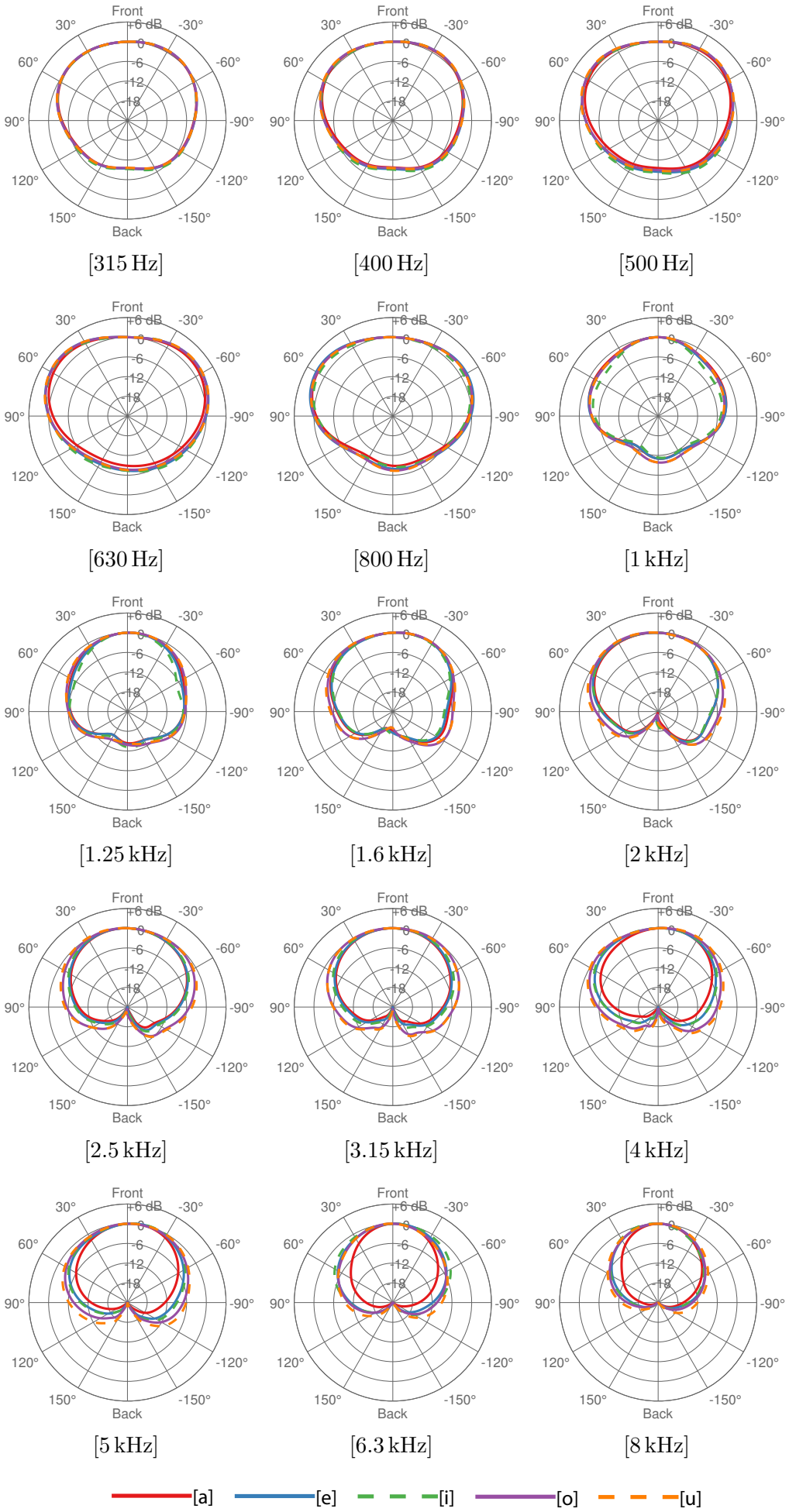

Figure 4: Polar plots of the directivity pattern in the horizontal plane, determined for the vowels. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

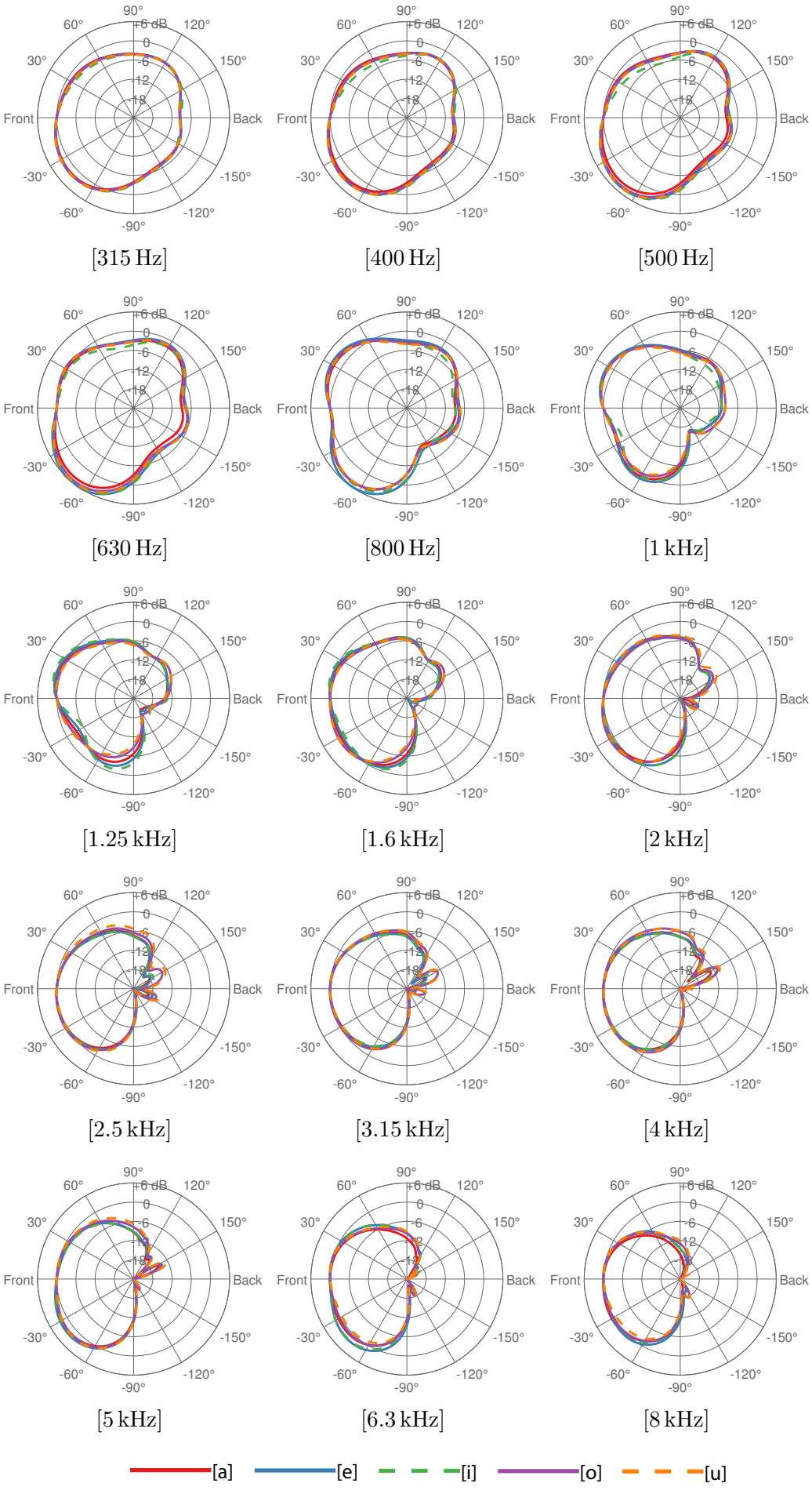

Figure 5: Polar plots of the directivity pattern in the vertical plane, determined for the vowels. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

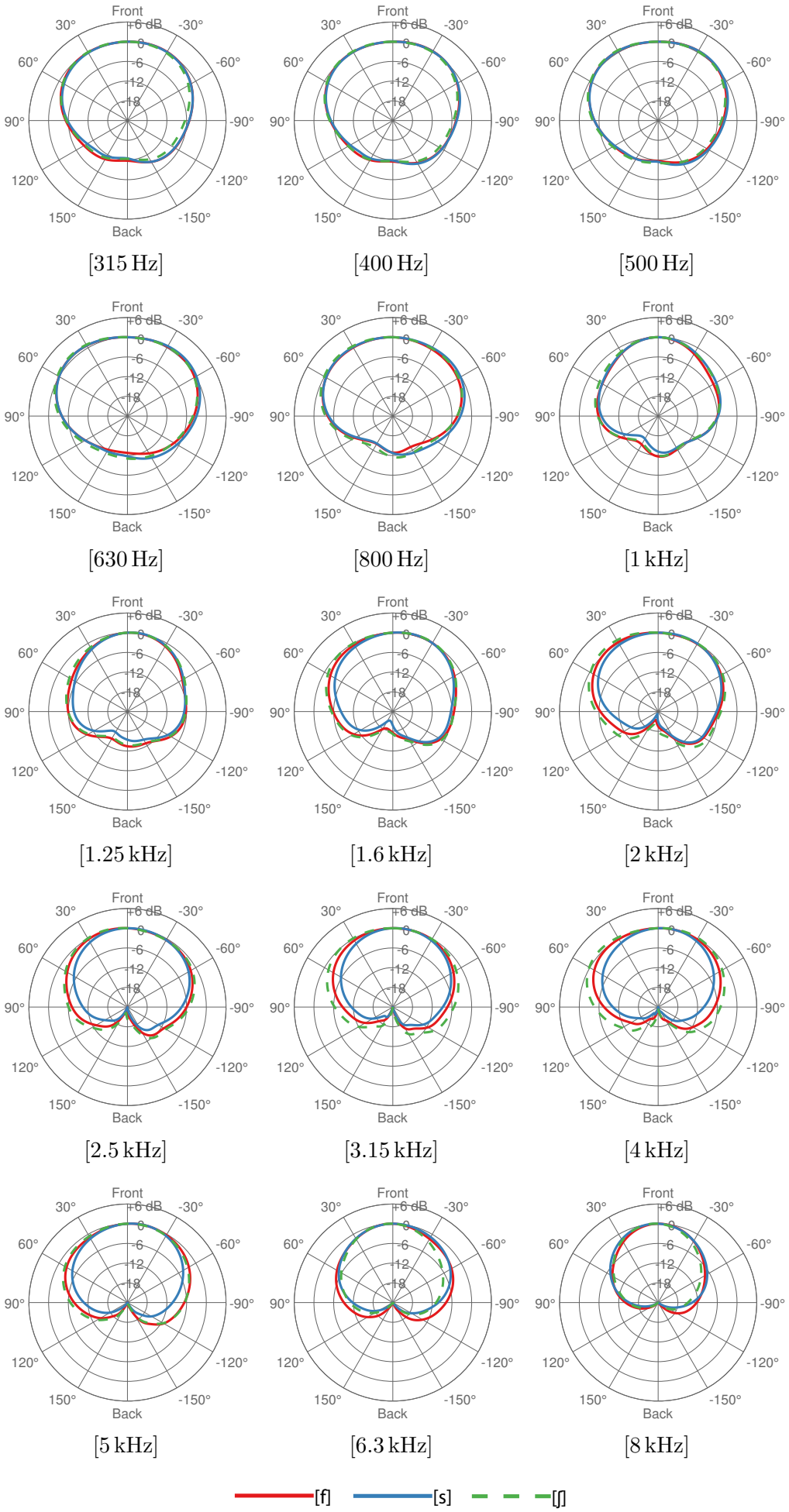

Figure 6: Polar plots of the directivity pattern in the horizontal plane, determined for the fricatives [1]. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

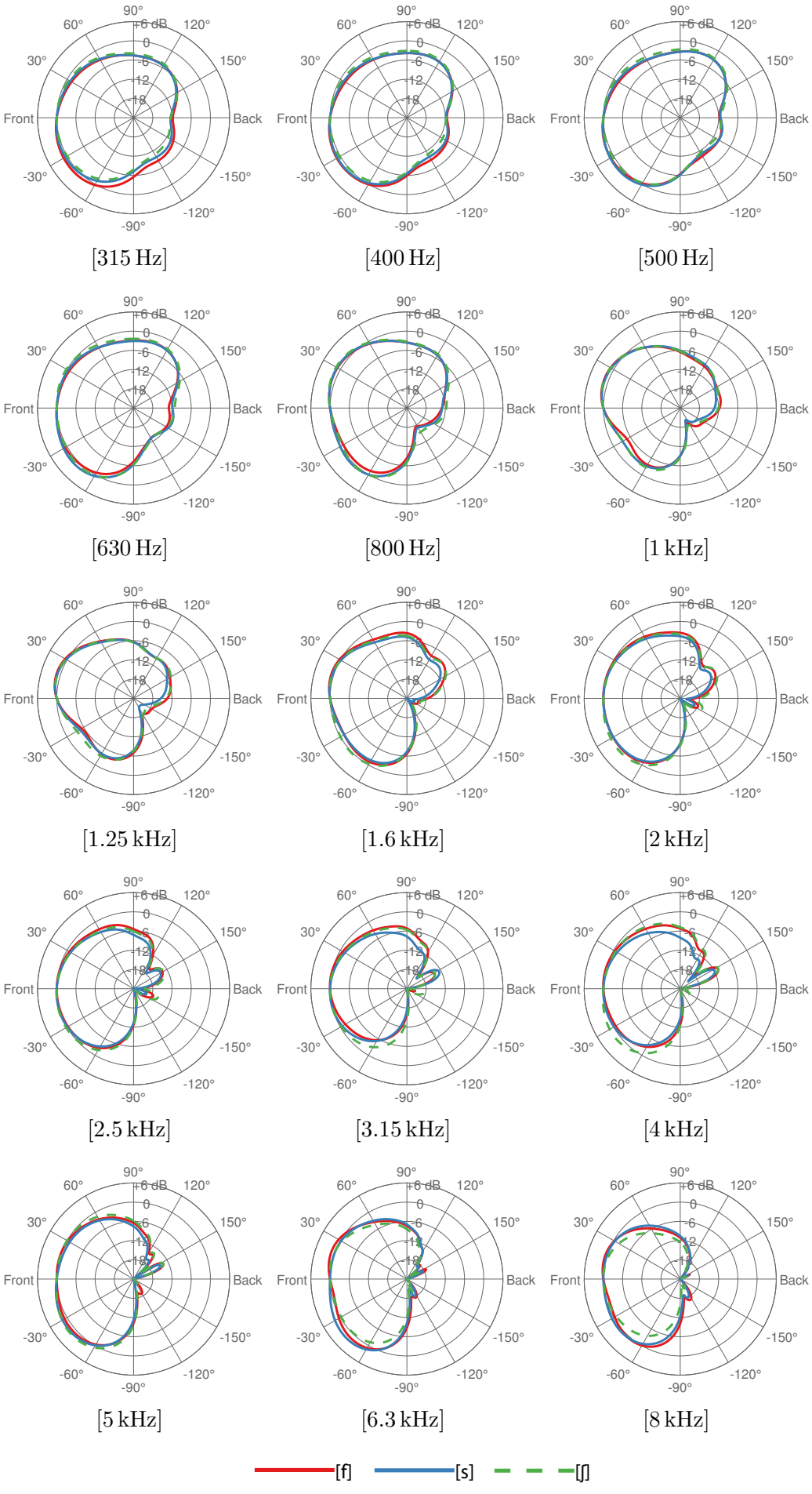

Figure 7: Polar plots of the directivity pattern in the vertical plane, determined for the fricatives [1]. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

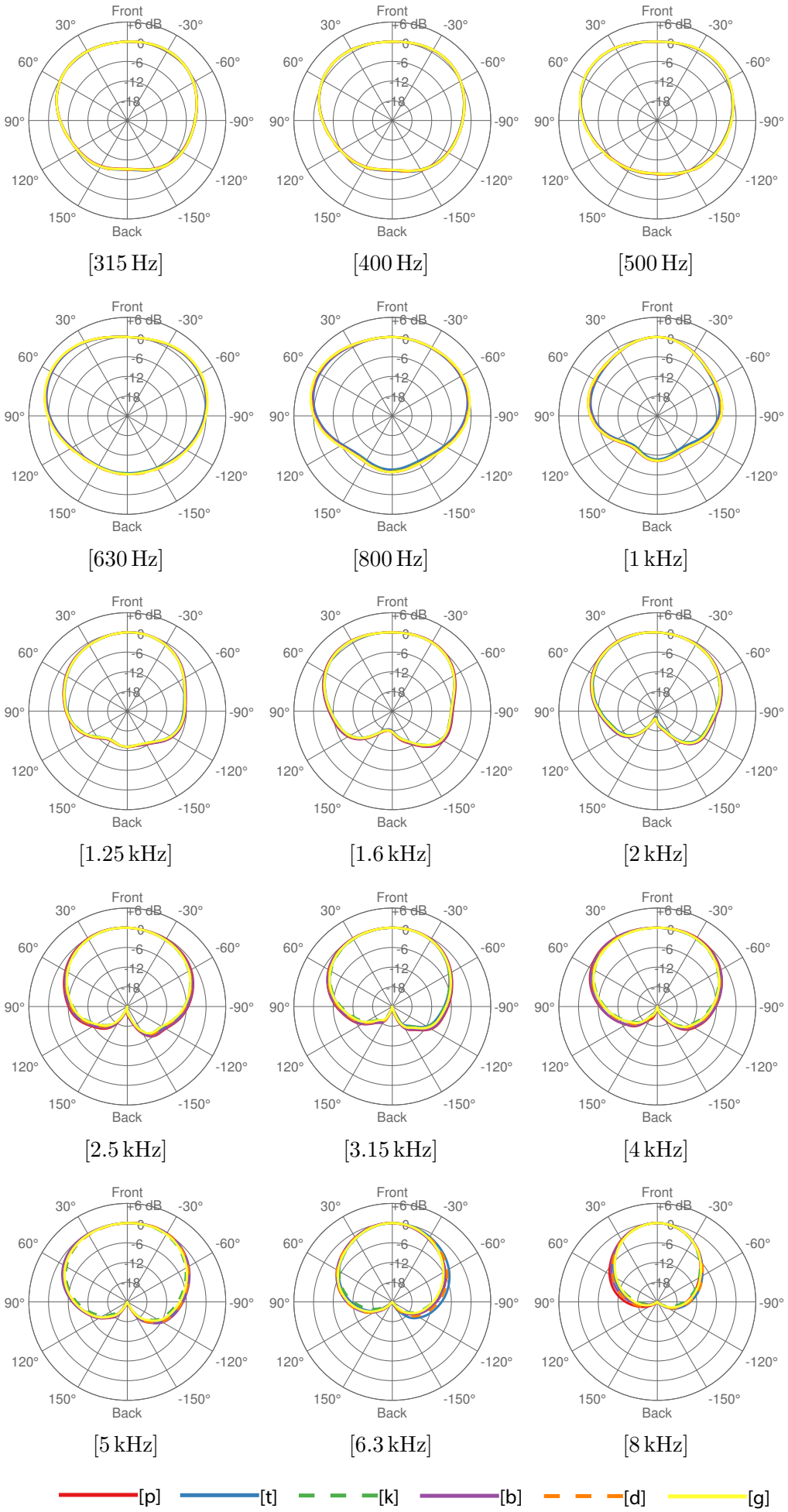

Figure 8: Polar plots of the directivity pattern in the horizontal plane, determined for the plosives. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

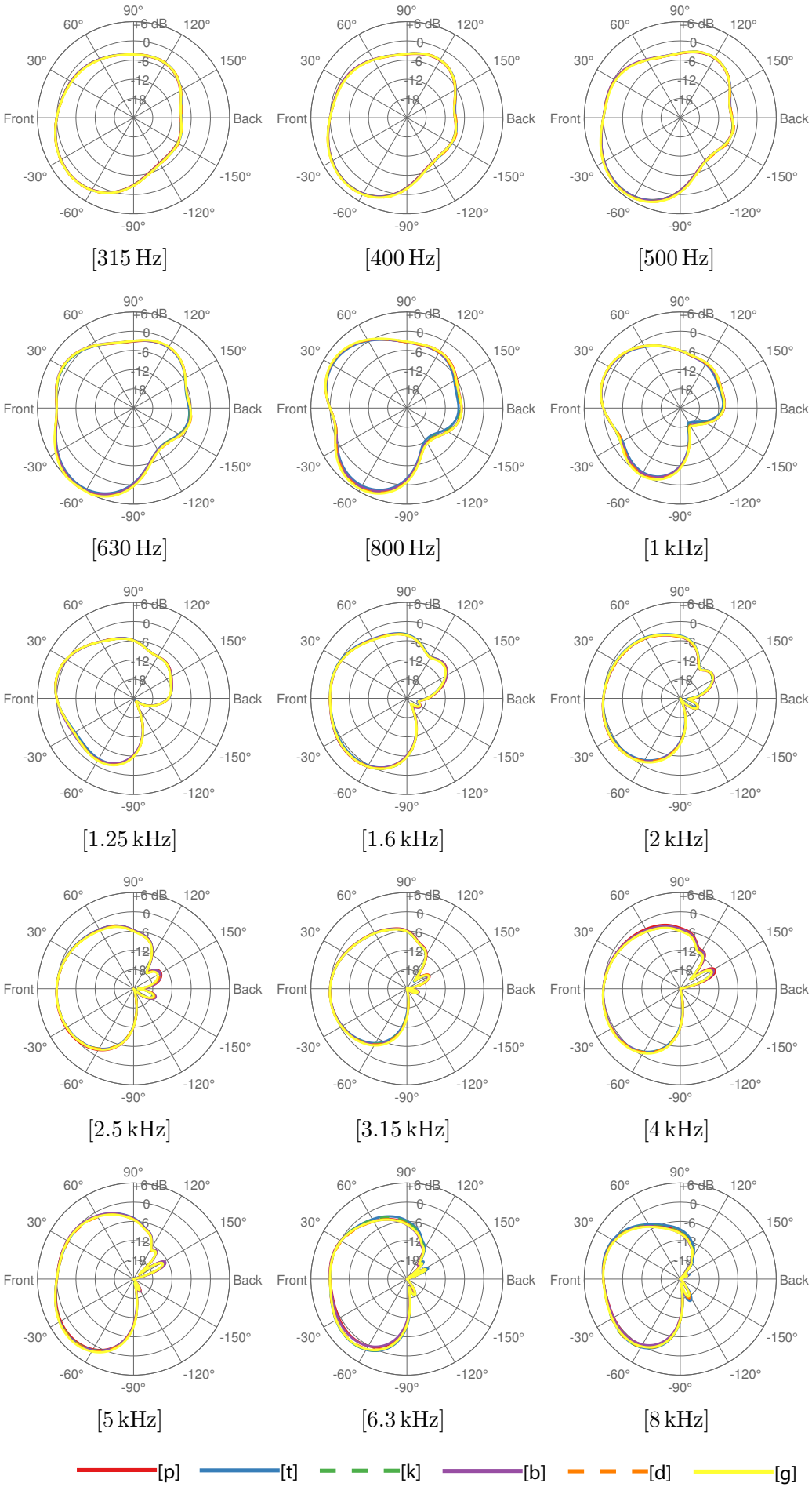

Figure 9: Polar plots of the directivity pattern in the vertical plane, determined for the plosives. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

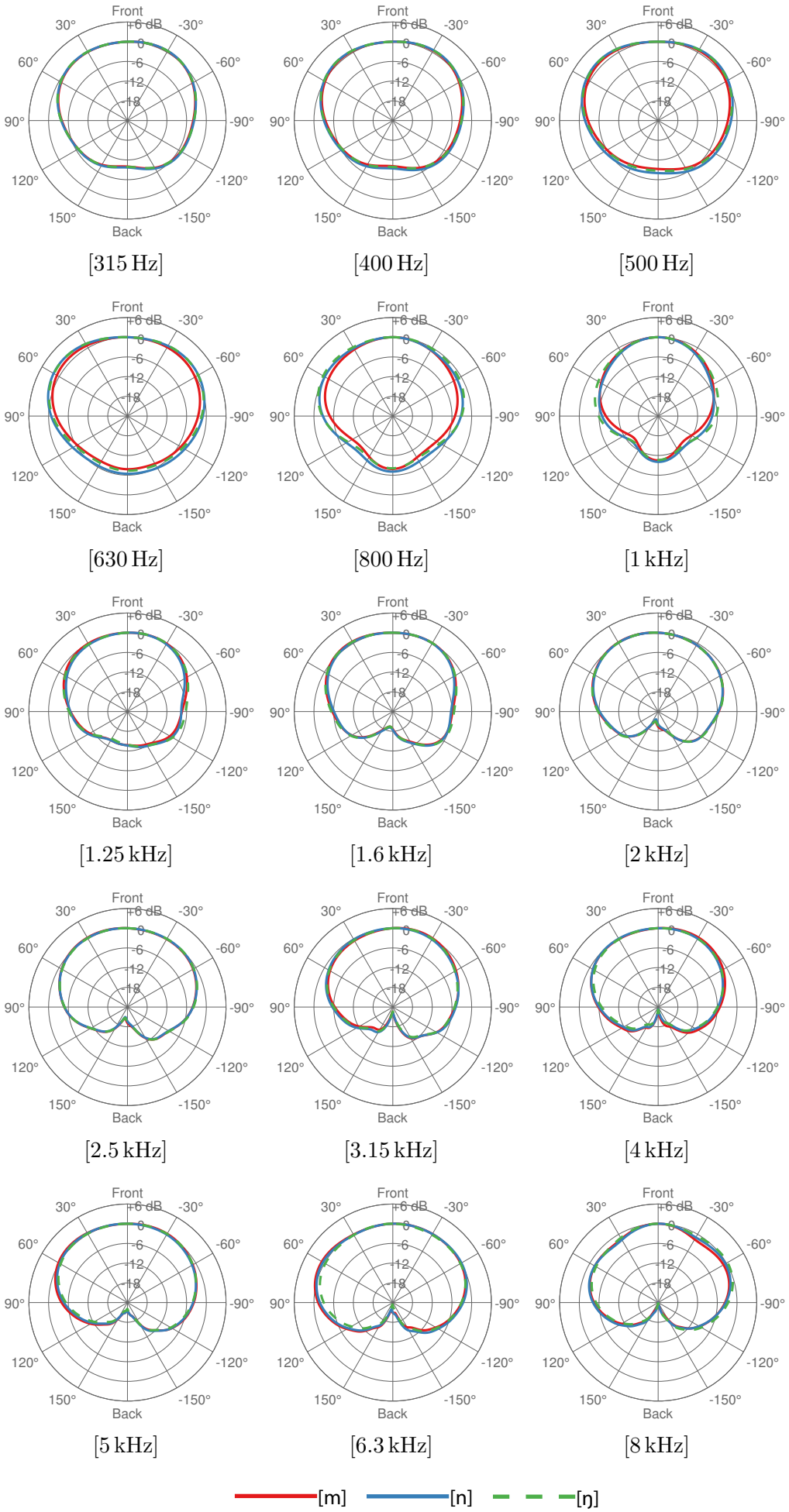

Figure 10: Polar plots of the directivity pattern in the horizontal plane, determined for the nasals. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

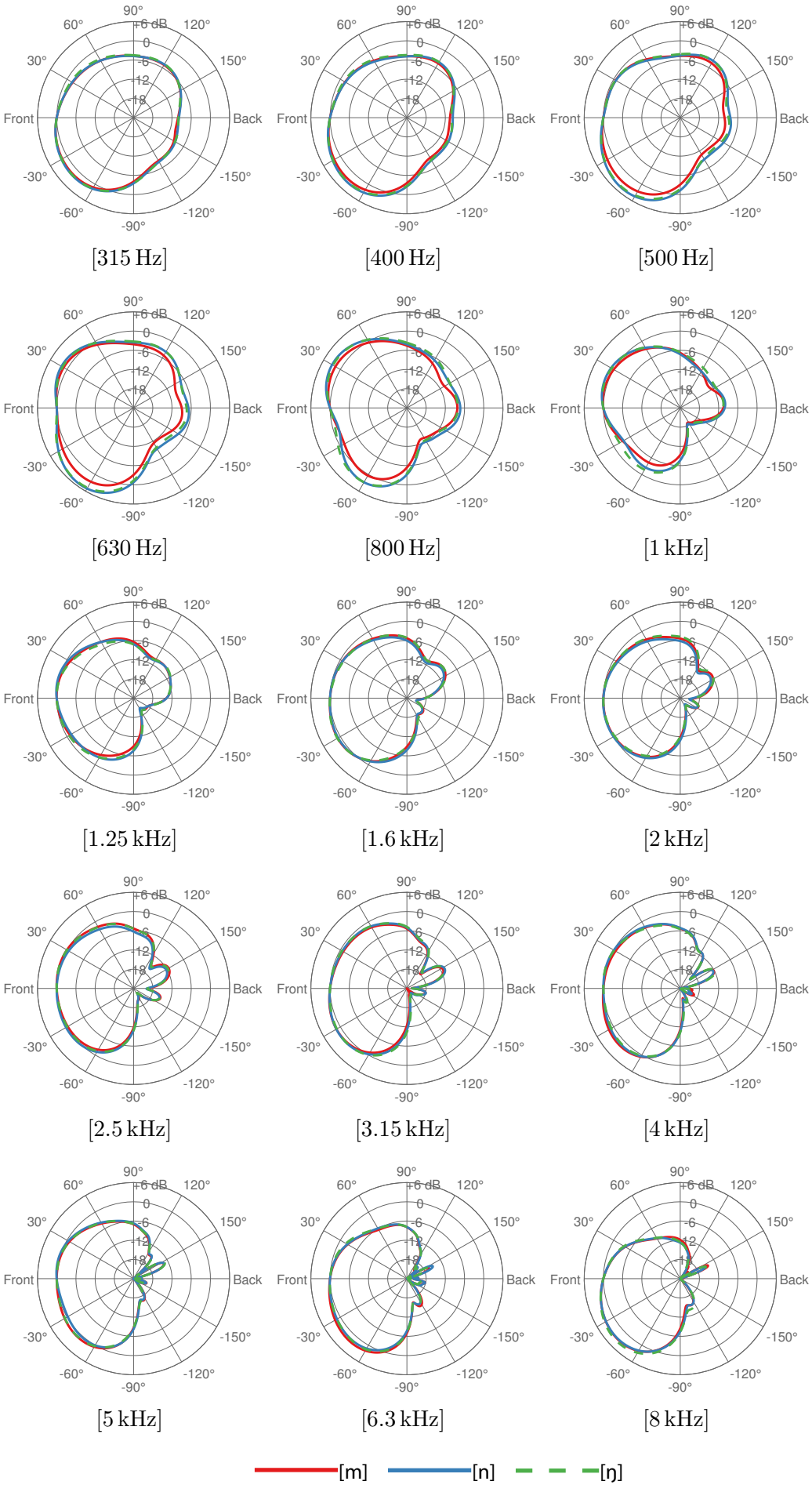

Figure 11: Polar plots of the directivity pattern in the vertical plane, determined for the nasals. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

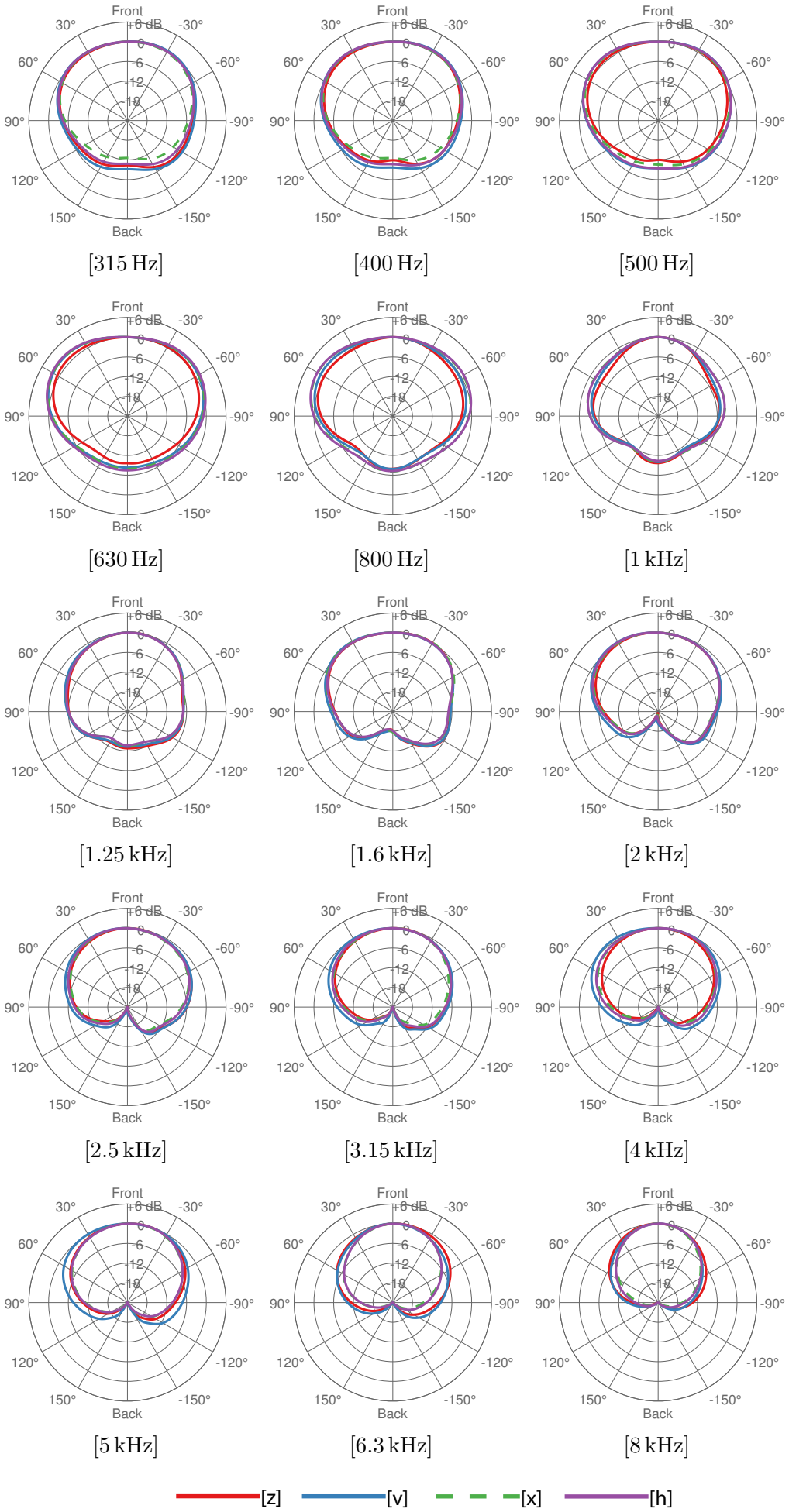

Figure 12: Polar plots of the directivity pattern in the horizontal plane, determined for the fricatives [2]. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

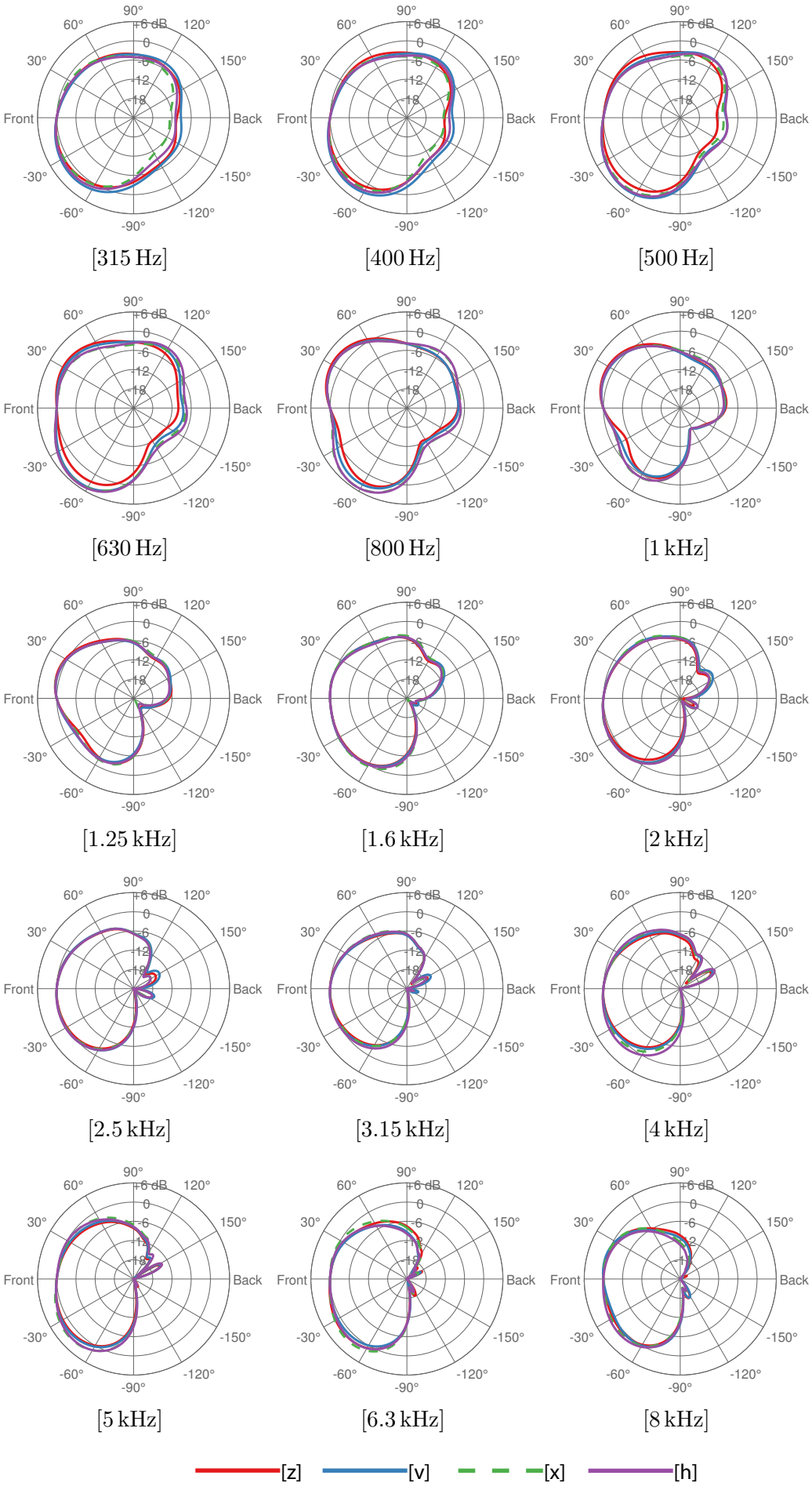

Figure 13: Polar plots of the directivity pattern in the vertical plane, determined for the fricatives [2]. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

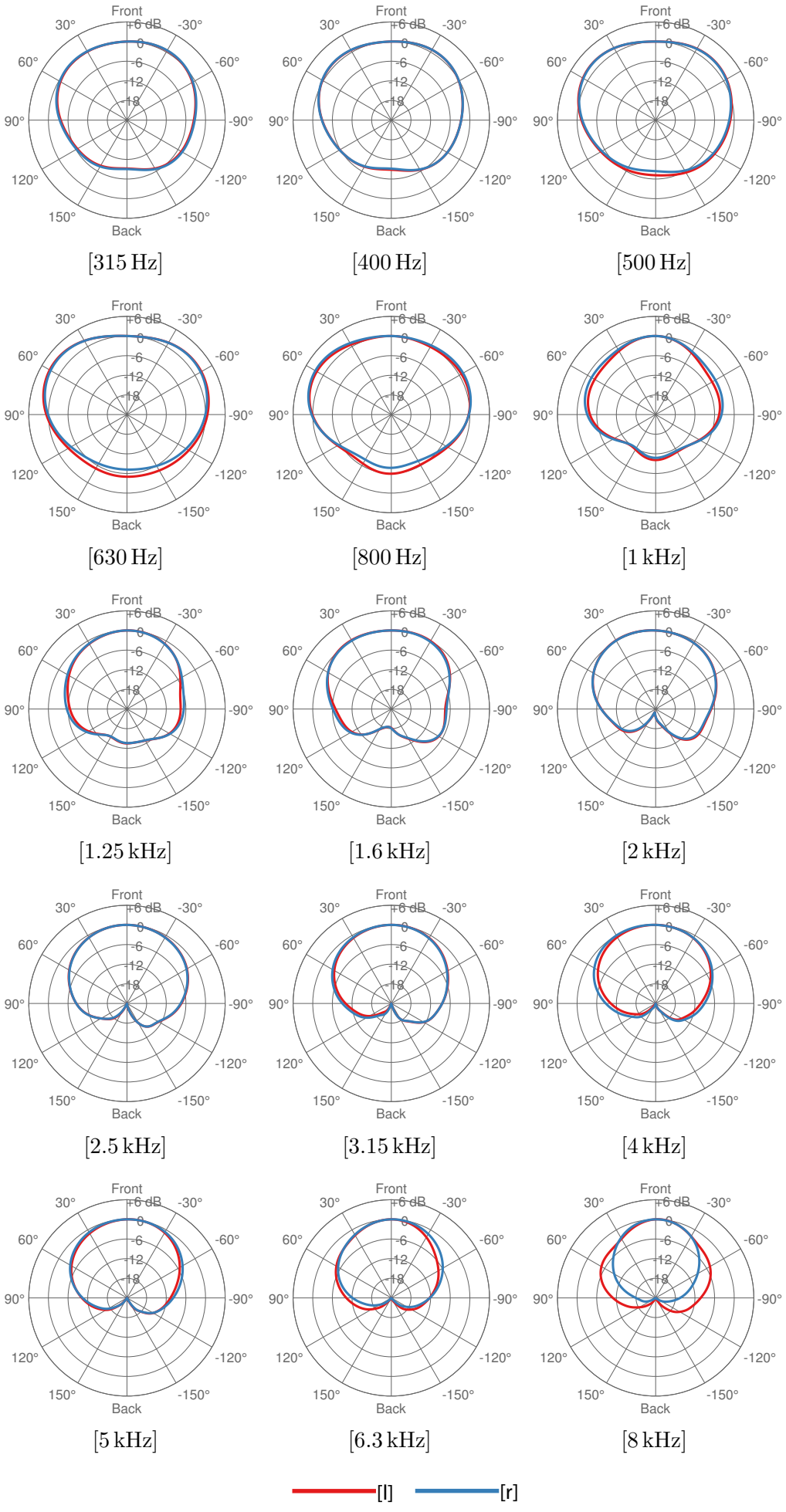

Figure 14: Polar plots of the directivity pattern in the horizontal plane, determined for the voiced alveolars [l] and [r]. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

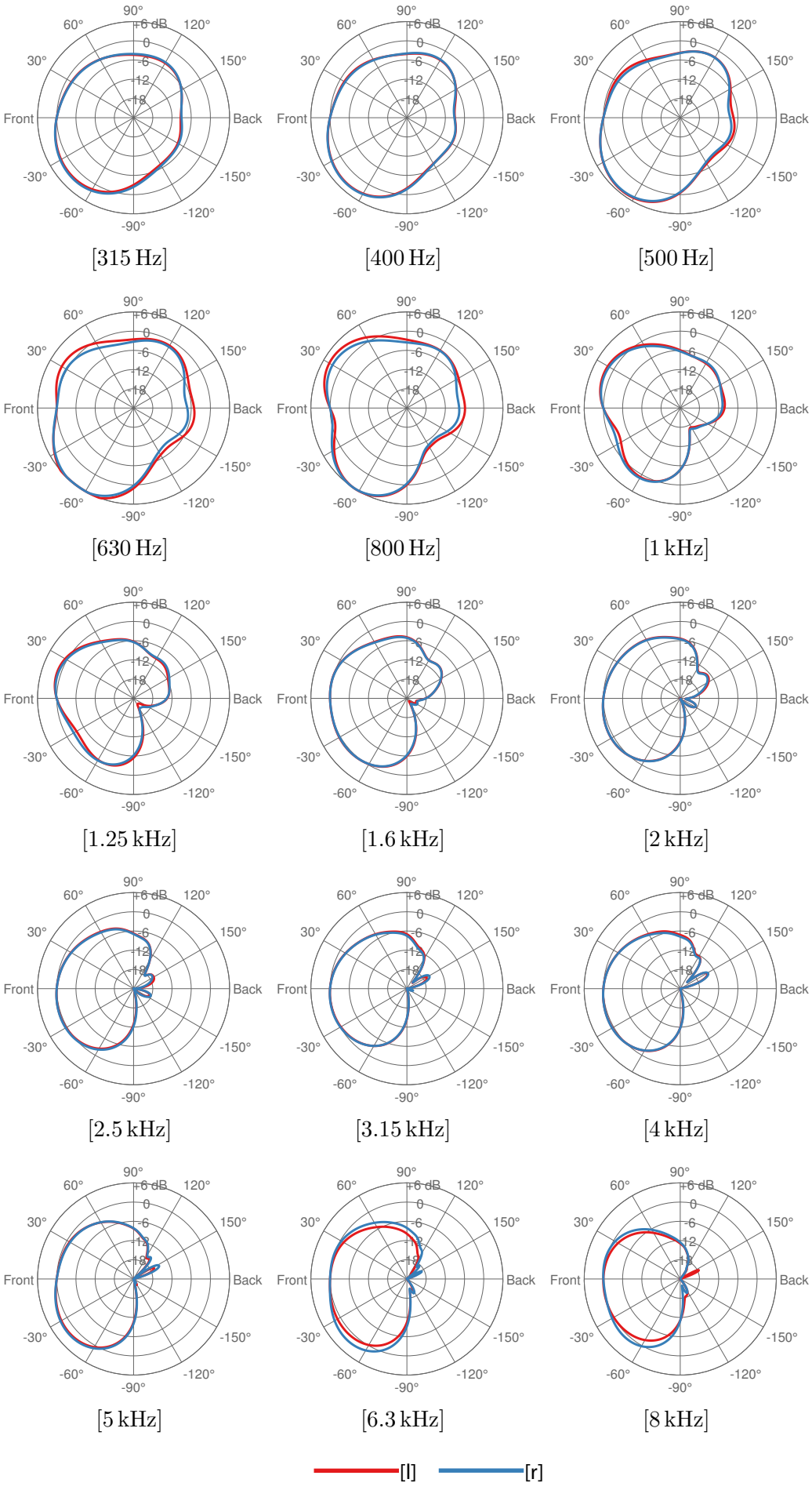

Figure 15: Polar plots of the directivity pattern in the vertical plane, determined for the voiced alveolars [l] and [r]. Shown are the mean values in third-octave bands with center frequencies between 315 Hz and 8 kHz.

Directivity plots

Figures 16 – 21 show the directivity patterns of all phonemes examined in [1, 2]. For this, we determined the directivity pattern per phoneme normalized to the frontal direction and then averaged the directivity patterns over all subjects.

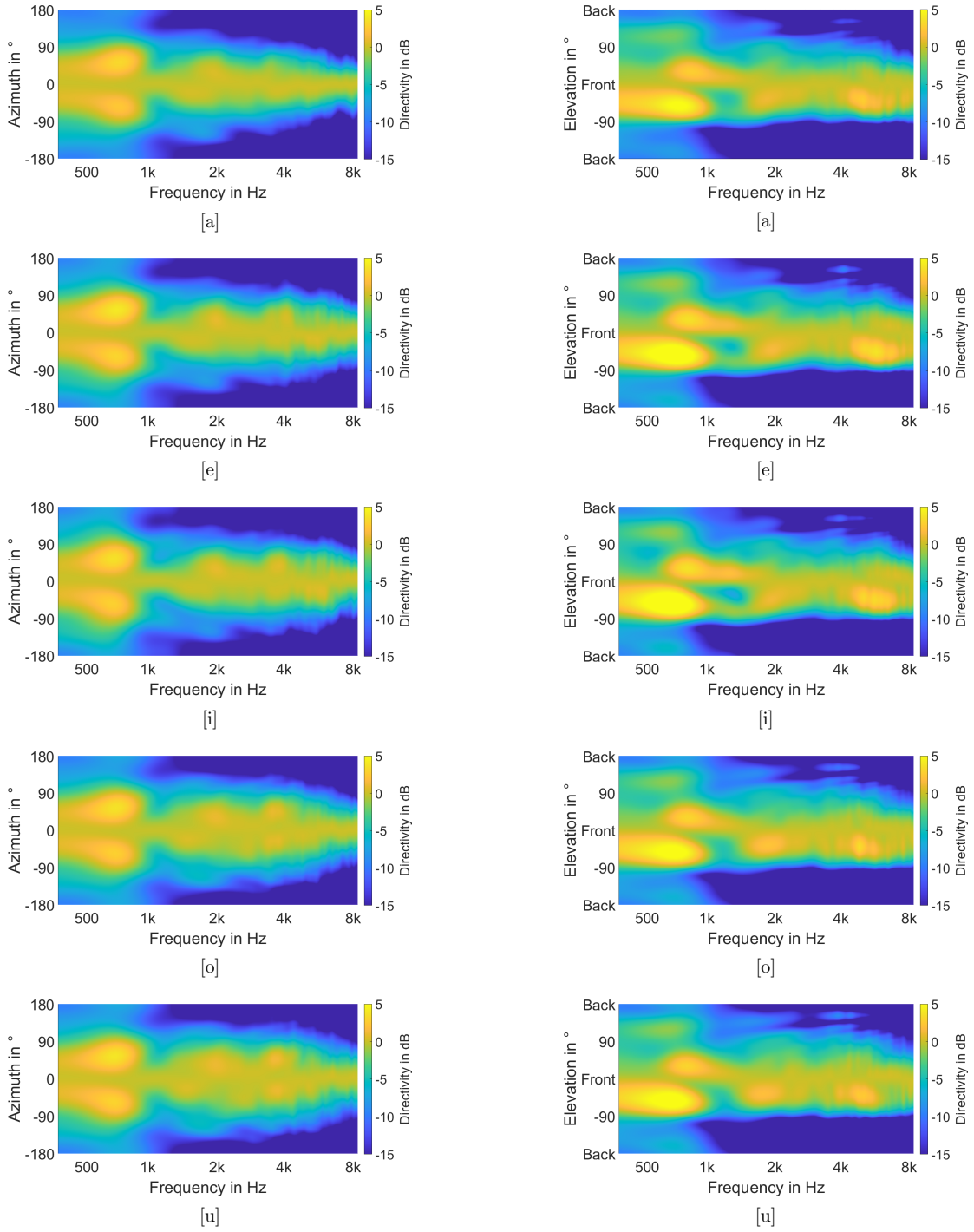

Figure 16: Directivity patterns of the vowels [a], [e], [i], [o], [u]. Horizontal plane (left column) and vertical plane (right column).

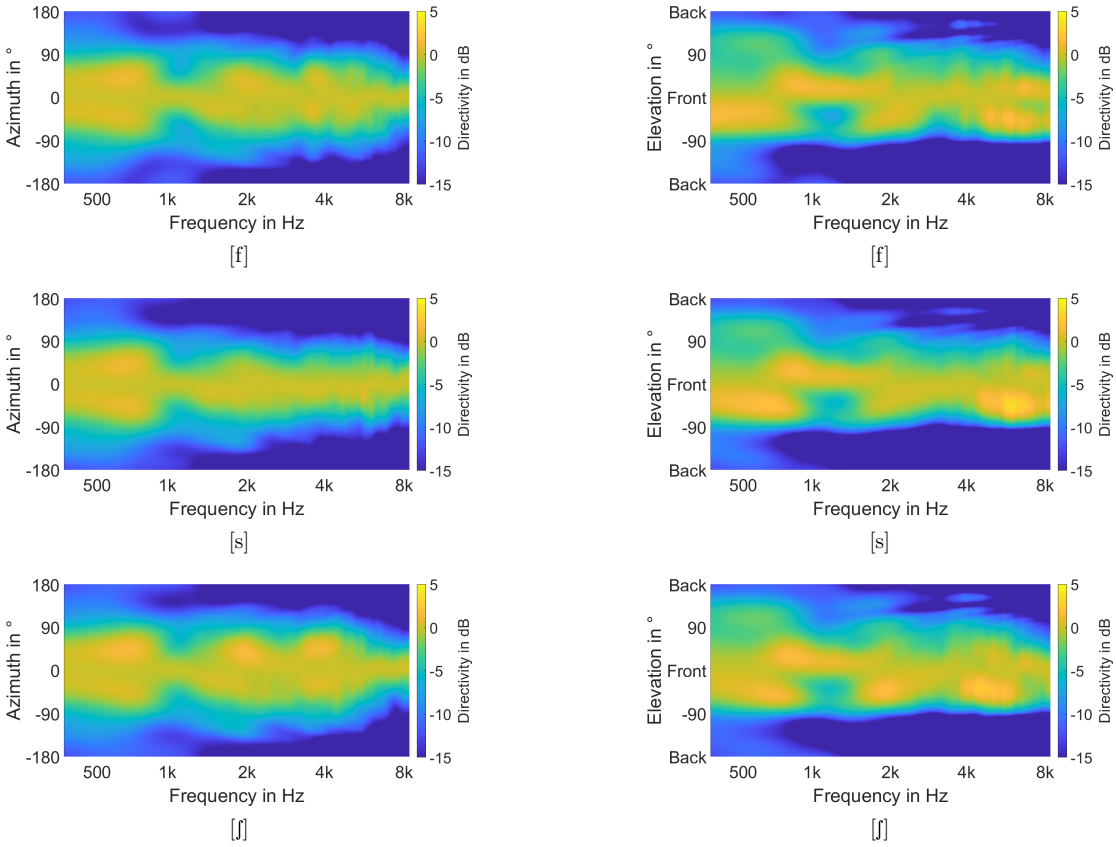

Figure 17: Directivity patterns of the fricatives [f], [s], [ʃ].Horizontal plane (left column) and vertical plane (right column).

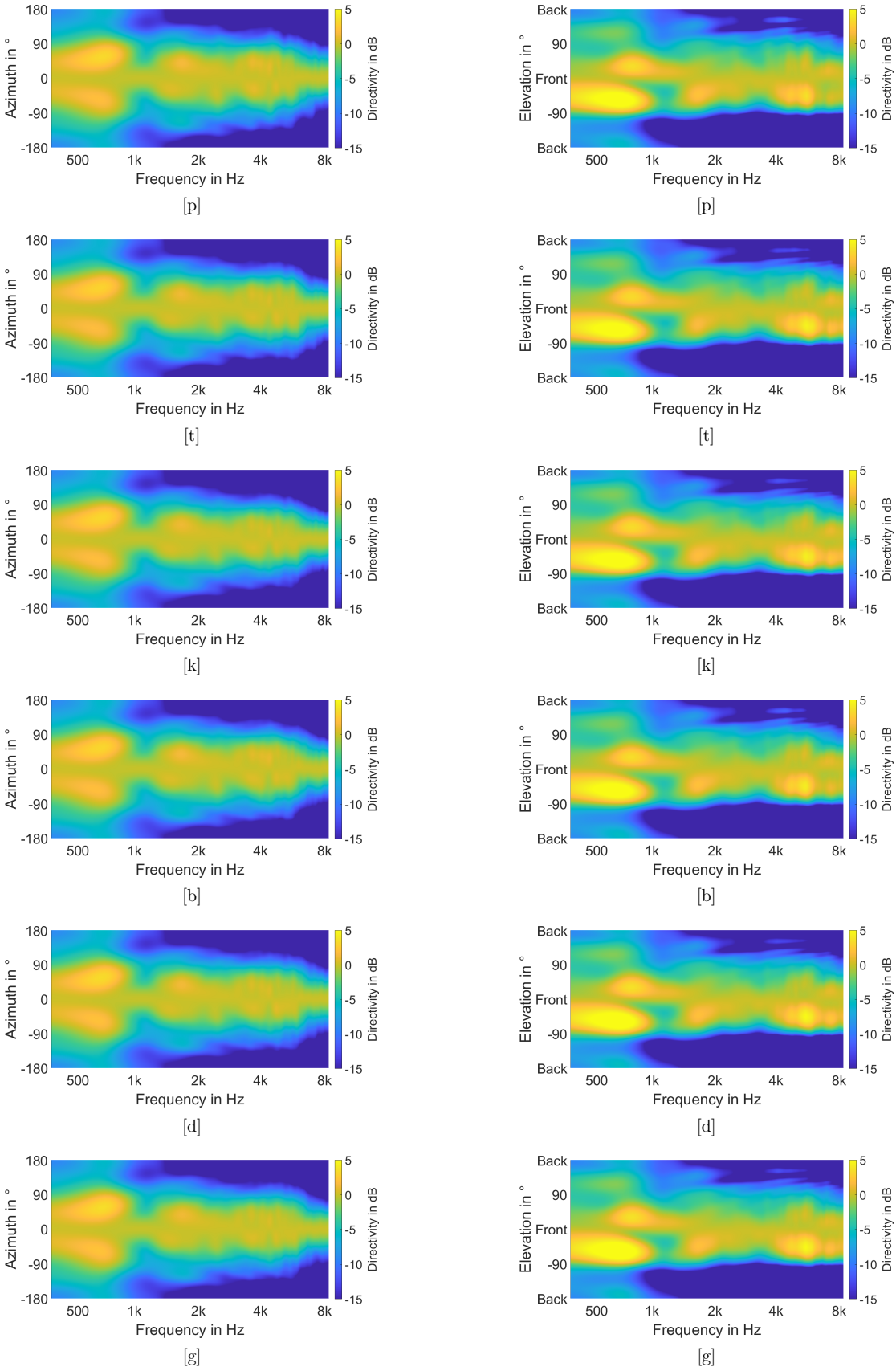

Figure 18: Directivity patterns of the plosives [p], [t], [k], [b], [d], [g]. Horizontal plane (left column) and vertical plane (right column).

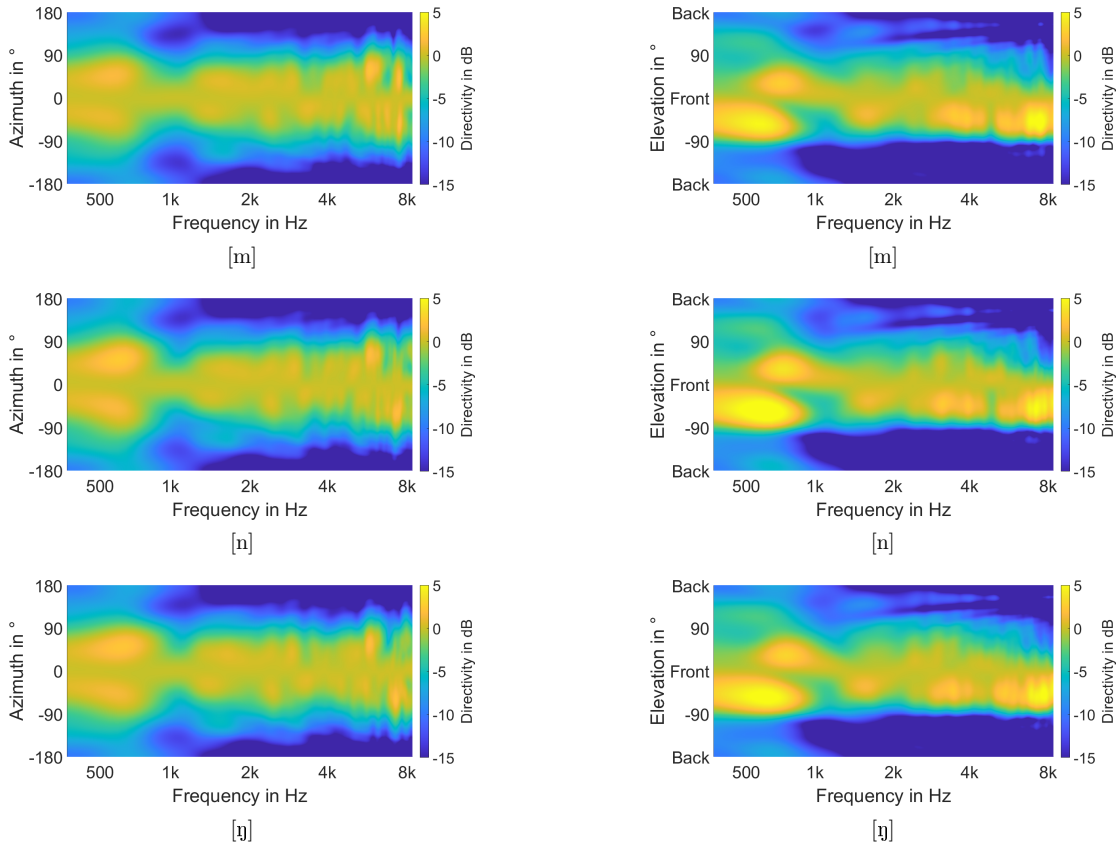

Figure 19: Directivity patterns of the nasals [m], [n] and [ŋ] . Horizontal plane (left column) and vertical plane (right column).

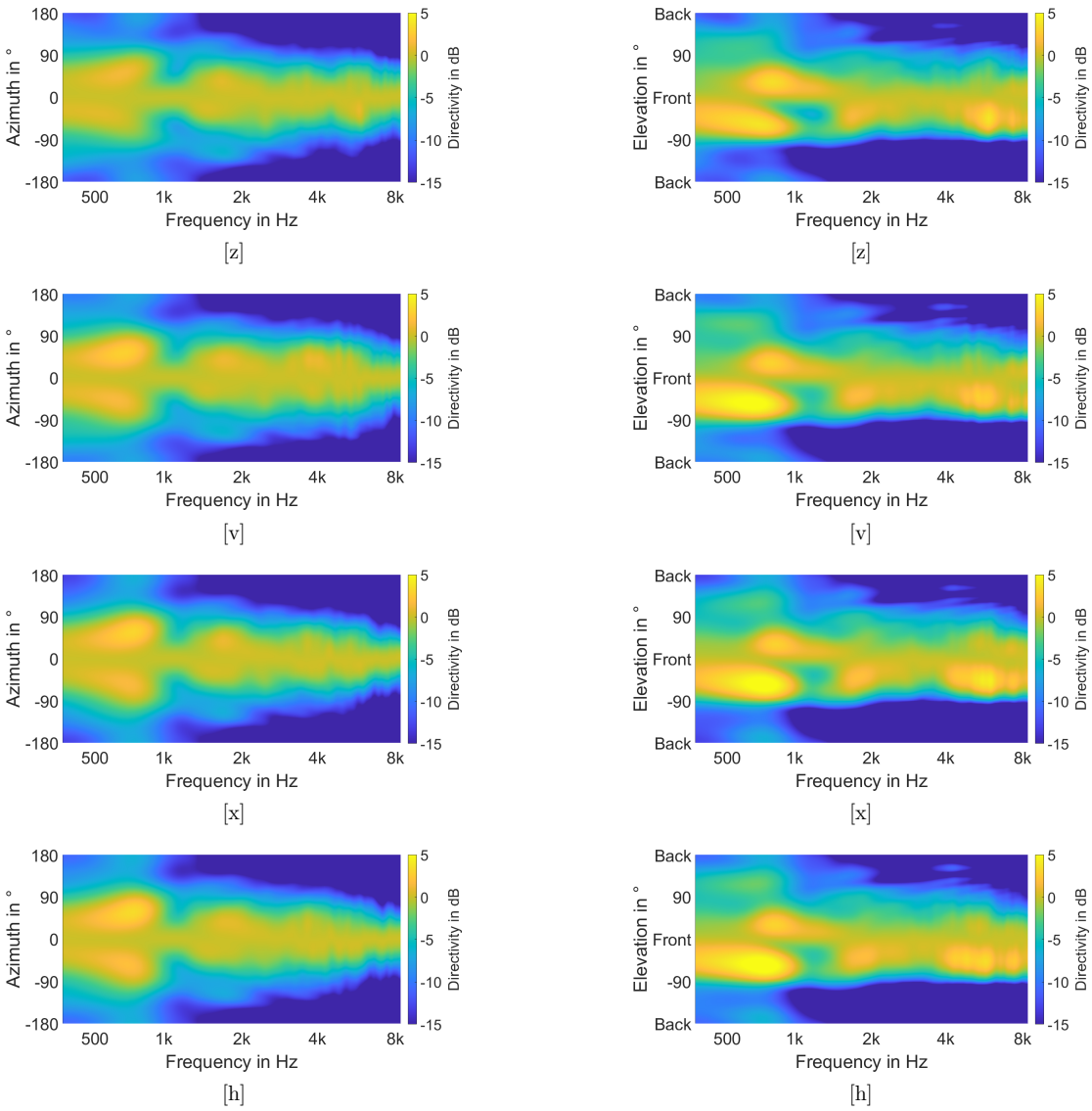

Figure 20: Directivity patterns of the fricatives [z], [v], [x] and [h]. Horizontal plane (left column) and vertical plane (right column).

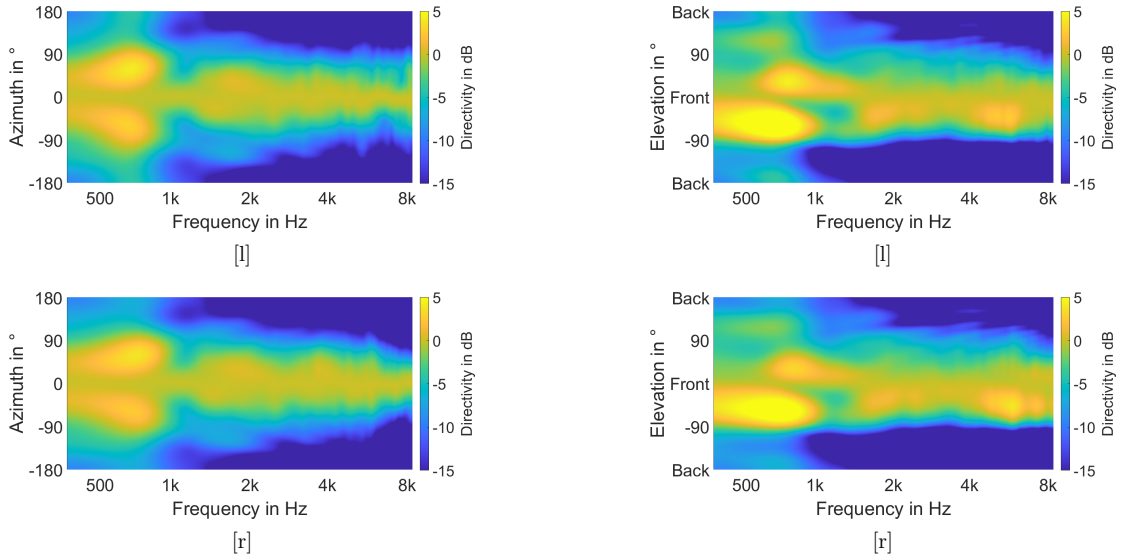

Figure 21: Directivity patterns of the voiced alveolars [l] and [r]. Horizontal plane (left column) and vertical plane (right column).

Spectral differences

In [1, 2], we calculated the spectral deviations ΔG_{sp} of all phonemes from an averaged directivity pattern. For this, we first normalized the radiation power for each measured dataset and then determined the directivity pattern D_{ph} per phoneme averaged over all subjects. The spectral deviations per direction to an averaged directivity D_{av} were calculated in dB as

$$\Delta G_{sp}(\theta, \phi, f) = 20lg \frac{\mid D_{ph}(\theta, \phi, f) \mid}{\mid D_{av}(\theta, \phi, f) \mid}, \tag{1}$$

To be consistent in both publications, we used the same values for D_{av} from [1], which was determined by averaging over the 8 phonemes (5 vowels and 3 fricatives) investigated in that study. In [1, 2] we showed ΔG_{sp} in a range from -2.5 dB to +2.5 dB. Even though this range resolves the spectral differences quite well in a wide angular and frequency area it does not allow studying frequency ranges and spatial regions in which the directivity pattern is affected most by the respective phoneme. Thus, Figures 22 – 27 show ΔG_{sp} in the horizontal and vertical plane in an increased range from –10 dB to +10 dB.

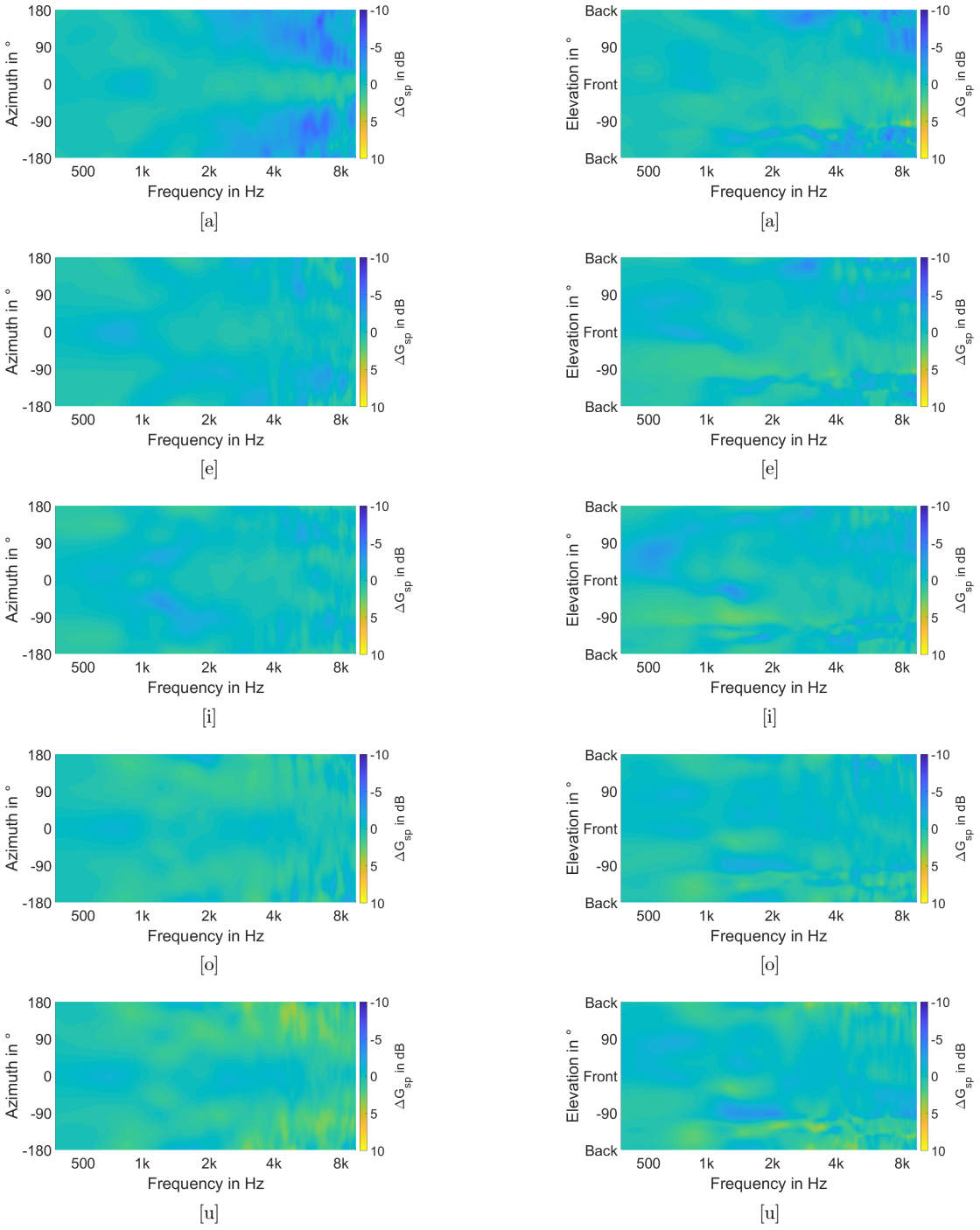

Figure 22: Spectral deviations ΔG_{sp} in a range from –10 dB to +10 dB of the vowels [a], [e], [i], [o], [u]. Horizontal plane (left column) and vertical plane (right column).

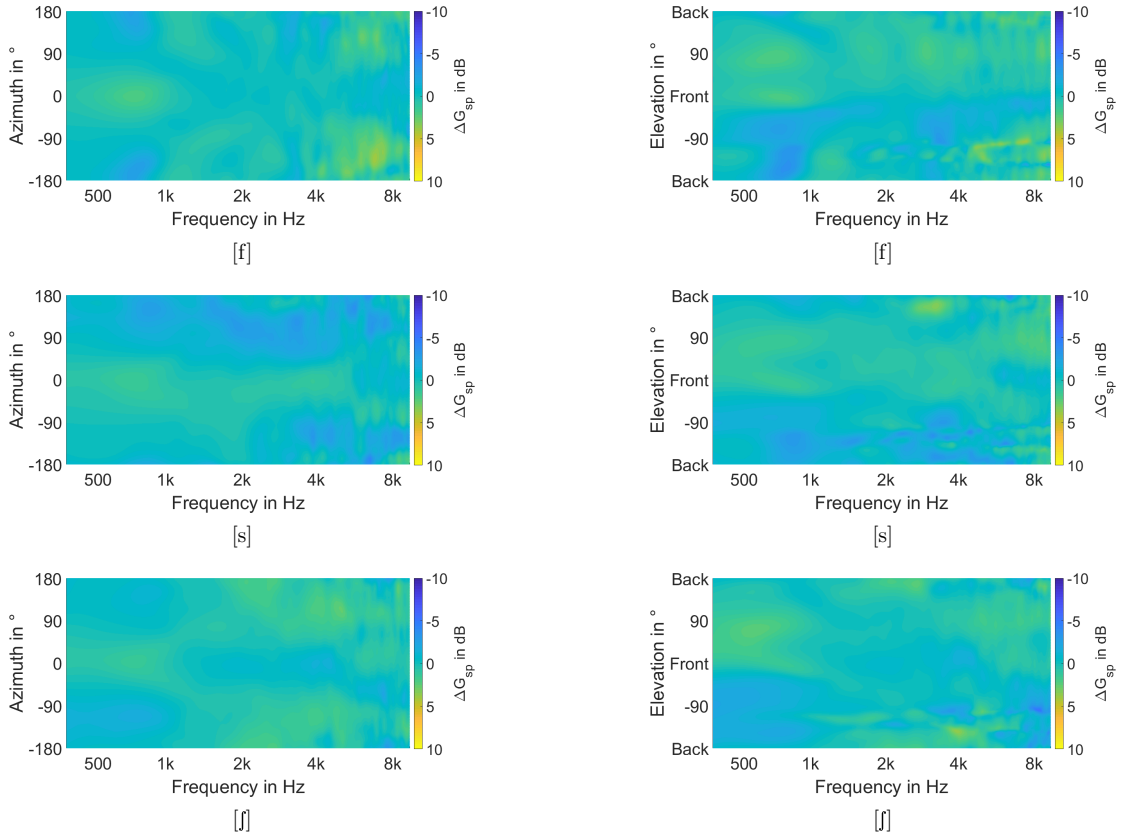

Figure 23: Spectral deviations ΔG_{sp} in a range from –10 dB to +10 dB of the fricatives [f], [s], [ʃ].Horizontal plane (left column) and vertical plane (right column).

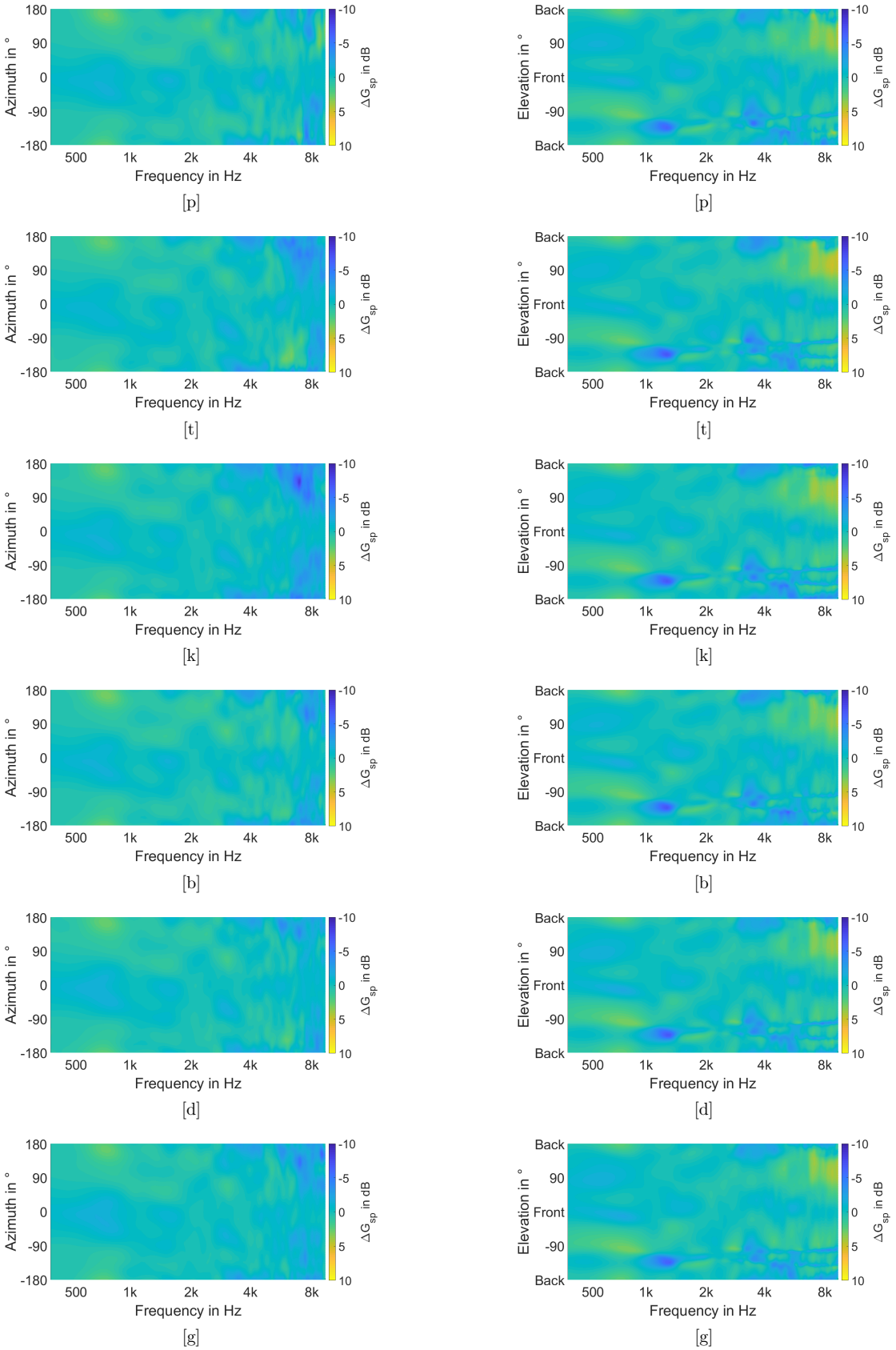

Figure 24: Spectral deviations ΔG_{sp} in a range from –10 dB to +10 dB of the plosives [p], [t], [k], [b], [d], [g]. Horizontal plane (left column) and vertical plane (right column).

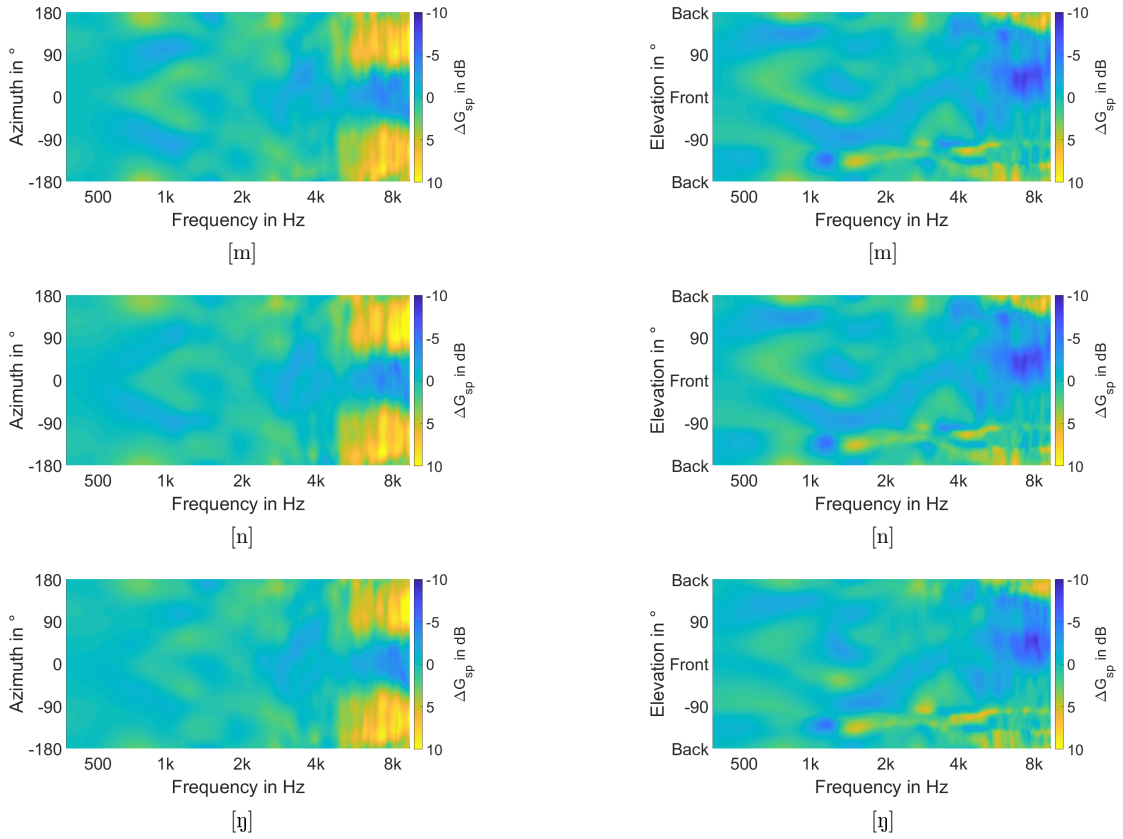

Figure 25: Spectral deviations ΔG_{sp} in a range from –10 dB to +10 dB of the nasals [m], [n] and [ŋ] . Horizontal plane (left column) and vertical plane (right column).

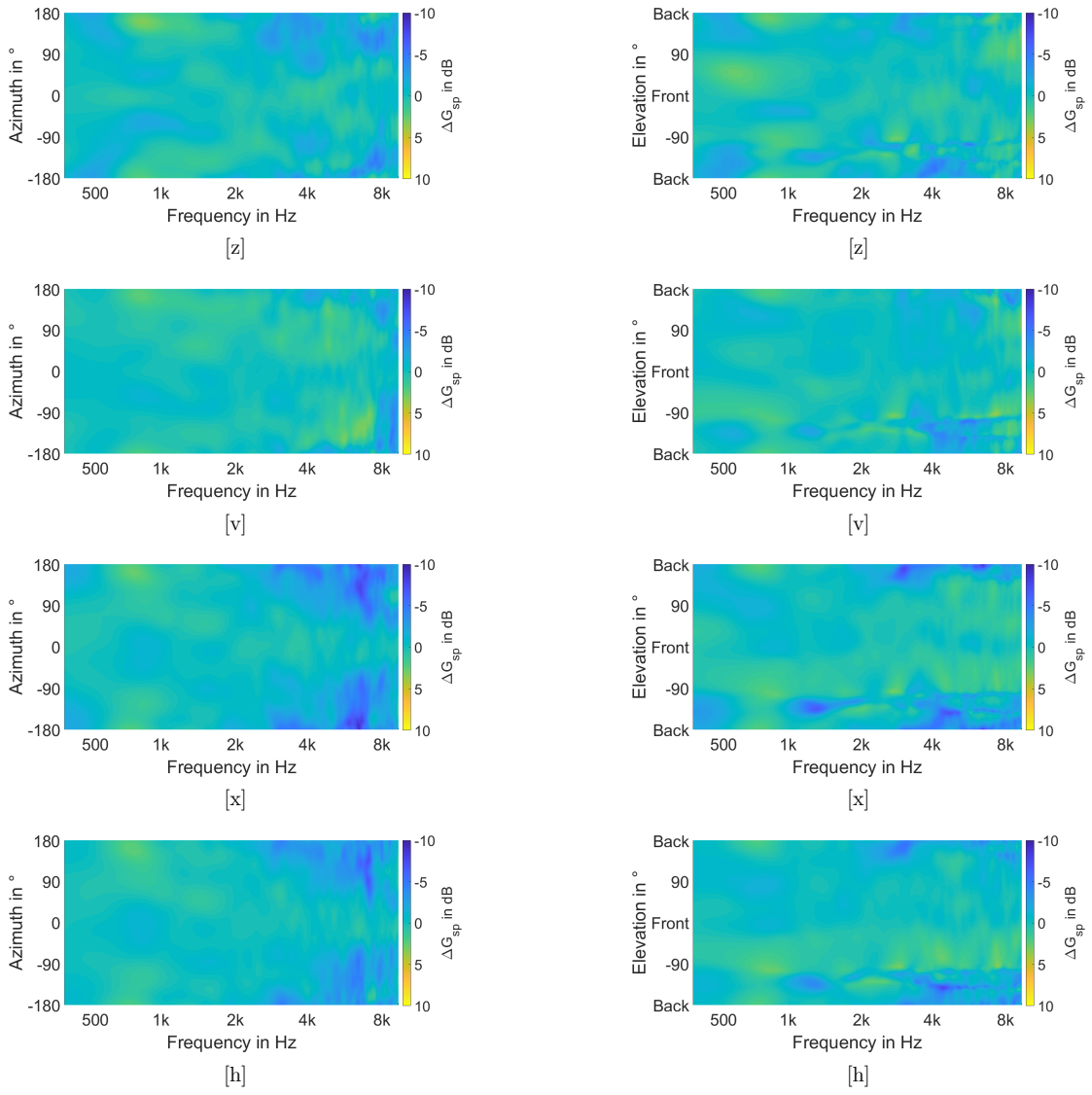

Figure 26: Spectral deviations ΔG_{sp} in a range from –10 dB to +10 dB of the fricatives [z], [v], [x] and [h]. Horizontal plane (left column) and vertical plane (right column).

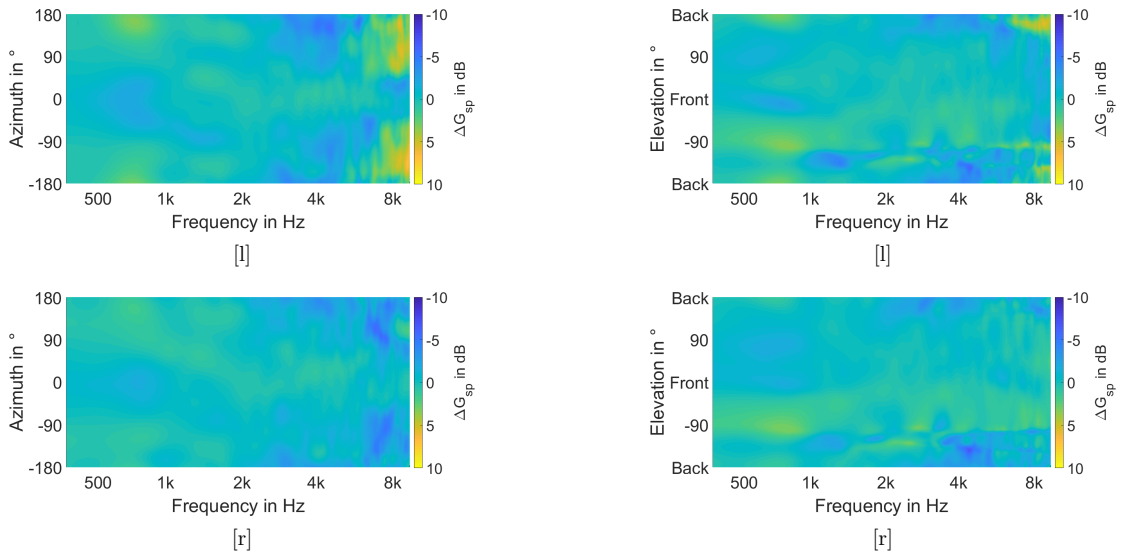

Figure 27: Spectral deviations ΔG_{sp} in a range from –10 dB to +10 dB of the voiced alveolars [l] and [r]. Horizontal plane (left column) and vertical plane (right column).

References

[1] 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.

[2] ——, "Investigating phoneme-dependencies of spherical voice directivity patterns II: Various groups of phonemes," *The Journal of the Acoustical Society of America*, vol. 153, no. 1, pp. 179–190, 2023.

[3] 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," *Journal of the Audio Engineering Society*, vol. 70, no. 7/8, pp. 565–584, Jul. 2022.

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

[5] C. Pörschmann, J. M. Arend, and F. Brinkmann, "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 6, pp. 1060 – 1071, 2019.

[6] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model from Anthropometry," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472 – 479, 2001.