On the impact of downward-directed human voice radiation on ground reflections

Christoph Pörschmann1,∗ and Johannes M. Arend2

1 TH Köln, Institute of Computer and Communication Technology, 50968 Cologne, Germany
2 Technische Universität Berlin, Audio Communication Group, 10587 Berlin, Germany

Received 21 November 2023, Accepted 24 January 2024

Abstract — Previous studies on vertical or full-spherical directivity patterns of the human voice showed that the human voice has a slightly downward main radiation direction over a wide frequency range. This paper investigates the phoneme-dependencies of human voice radiation in the vertical plane and analyzes to what extent these characteristics affect the ground reflection and the sound incidence at a listener position. The results show that for most phonemes and below 800 Hz, the ground reflection is stronger than the direct sound component because of the downward-directed main radiation. On the contrary, between 800 Hz and 1.6 kHz, the main radiation direction is upward, probably mainly due to diffraction and reflections from the shoulders and the torso.

Keywords: Human voice directivity, Sound radiation

1 Introduction

Human voice directivity has long been a research topic and has been investigated for more than 200 years [1–3]. However, only a few of these studies determined directivity patterns in the vertical plane, or more generally analyzed full-spherical voice radiation. The first one by Dunn and Farnsworth [4] was published in 1939 and determined the spherical directivity patterns for a spoken sentence at different distances in octave or one-half octave bands from 63 Hz up to 12 kHz. Even though not mentioned in the paper, the data allow the analysis of specific aspects of vertical speech directivity, such as the frequency dependence of the main radiation direction (MRD). Similar patterns of the vertical voice directivity pattern were measured by Marshall and Meyer [5], who already briefly discussed a downward-directed MRD in their study. Chu and Warnock [6] provided means and standard deviations of voice directivity patterns measured for 40 subjects taken at 92 positions on a spherical grid in one-third octave bands. Due to the higher resolution over frequency, the data show better than the previous studies how the MRD of the human voice varies with frequency. While the radiation is slightly downward directed for most frequency bands, at about 1 kHz the MRD is elevated by about 30°. Recently, Leishman et al. [7] published results of measurements on six subjects with a resolution of 5° in the horizontal and vertical planes that further support previous findings on the MRD. However, because the directivity patterns in all these studies were averaged over a full sentence of fluent speech, these datasets do not allow for the analysis of phoneme-dependencies.

∗Corresponding author: christoph.porschmann@th-koeln.de

In Pörschmann and Arend [8, 9], we determined the spherical directivity patterns of 23 phonemes of different groups for 13 subjects using a 32-channel surrounding spherical microphone array and examined both the directivity patterns and the directivity indices. Our results showed statistically significant differences between phoneme groups and between some phonemes within each group. Most importantly in the context of the present study, this comprehensive database allows a phoneme-dependent analysis of the MRD.

Furthermore, only a few studies addressed the perceptual influence of the ground reflection, which is often only slightly lower in intensity than the direct sound, and thus can affect localization, distance perception, and perceived spaciousness [10–15]. However, none of these studies considered natural human voice radiation or a downward-directed MRD, which can lead to a ground reflection reaching a listener at a higher intensity than the direct sound. The present study aims to disentangle these phoneme- and frequency-dependencies of the MRD and its effect on the ground reflection, extending our previous studies on human voice directivity [8, 9].

2 Methods

This work applies the measured datasets from our previous studies [8, 9, 16] that were carried out with 13 subjects in the anechoic chamber of TH Köln using a

1 The datasets are available as supplementary material to these papers and can be accessed at https://doi.org/10.5281/zenodo.7452117 and at https://doi.org/10.5281/zenodo.7834210.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
surrounding spherical microphone array with 32 sampling positions [17, 18]. In the present study, the following articulations were considered:

- **Phonetically balanced sentence**
- **Voiced:** [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].

The phonetically balanced sentence was recorded to determine an average directivity. For all the phonemes, a suitable voice excitation signal was chosen that had a broadband energy spectrum and thus allowed calculating the directivity pattern continuously over frequency. For the voiced phonemes, the glissando method [19, 20] was applied, which means the subjects sang the phoneme with an increasing pitch over at least one octave. For the unvoiced fricatives, the participants repeated the respective phoneme three times for each measurement. For the unvoiced [x], the subjects continuously articulated the phoneme for at least three seconds. For each measured articulation, we determined relative transfer functions related to the frontal direction by deconvolving the microphone array signals with the signal from an additional frontal reference microphone, yielding a dataset with 32 raw impulse responses per articulation. The measured datasets were spatially upsampled to a Lebedev grid [21] with 2702 sampling points using the SUpDEq (Spatial Upsampling by Directional Equalization) method [22], which is suitable for upsampling voice directivities [23]. The spatially upsampled directivity patterns were stored as impulse responses with a length of 128 coefficients at a sampling rate of 48 kHz. Thus, the datasets from [8, 9, 16] provide a spectrally and spatially dense description of the directivities. While for the sentence only one directivity dataset per person was measured, each phoneme was measured twice per person and the resulting directivity data were averaged.

We analyzed the directivities $D(\phi_0, \theta, f)$ in the frontal vertical plane (also called sagittal plane; azimuth $\phi_0 = 0^\circ$, elevation $\theta = -90^\circ$ to $90^\circ$, where positive angles point upward). This was done by transforming the upsampled datasets to the spherical harmonic (SH) domain with an SH order of $N = 35$ and then resampling them using the inverse SH transform to dense datasets along the vertical plane in steps of $1^\circ$ as evaluated in our previous study [23]. From this, the frequency-dependent value of the MRD of each individual dataset was calculated:

$$\text{MRD}(f) = 20 \log \frac{|D(\phi_0, \text{MRD}, f)|}{|D(\phi_0, \theta_0, f)|},$$  

with $\phi_0, \theta_0$ the frontal direction.

Finally, to analyze the strength of the ground reflection, we determined the ground reflection gain (GRG) as the level difference between the ground reflection and the direct sound incidence at the listener’s position:

$$\text{GRG}(d, f) = 20 \log \left( \sqrt{1 - \eta^2} \frac{d}{|D(\phi_0, \theta', f)|} \right),$$

where $\eta$ denotes the frequency-dependent absorption coefficient of the ground floor. The geometric properties depend on the relative position between speaker and listener, $d$ denotes the distance between the human speaker and the listener, $d'$ the path length of the ground reflection, and $\theta'$ the radiation direction of the ground reflection.

### 3 Results

#### 3.1 Main radiation direction

Figures 1 and 2 show the vertical-plane directivity averaged over all subjects for all examined phonemes as well as the respective MRD. For the phonetically balanced sentence as well as for almost all phonemes, there is a prominent maximum of radiation towards $\theta \approx -45^\circ$ between 600 Hz and 800 Hz and upwards towards $\theta \approx 30^\circ$ between 800 Hz and 1 kHz. As an exception, both can hardly be observed for the voiceless fricatives [ʃ], [s], and [ʃ]. For most phonemes, there are further local maxima in a third region between 1.5 kHz and 1.8 kHz and in a fourth region above 4 kHz, where the radiation is directed downwards to $\theta \approx -30^\circ$. The fourth region is for the vowels, the plosives, and most of the fricatives at about 5 kHz, and for the nasals at about 7 kHz. The black lines in Figures 1 and 2 show the MRD averaged over all subjects, which is below 800 Hz and above 1.6 kHz downwards at $-15^\circ \leq \theta \leq -45^\circ$, confirming the results of previous studies [4–7]. The plots also show that for all phonemes, the MRD rises sharply to $\theta \approx +15^\circ$ at $f \approx 800$ Hz and drops back to $\theta \approx -20^\circ$ at $f \approx 1.6$ kHz.

Even though some phoneme dependencies can be observed, it is difficult to compare them directly based on these plots. Therefore, Figure 3 shows the subject-averaged MRD for the different phoneme groups, revealing deviations of the MRD within the vowels, nasals, and especially the plosives. For further comparison, each plot also depicts the MRD for the sentence, which can be regarded as a phonetically averaged value. Generally, above 4 kHz the differences between the phonemes and the standard deviations, which indicate the amount of inter-subject variation increase. Both could be due to the higher spatial complexity of the directional characteristics with peaks and dips at different frequencies and directions. In the case of the nasals, the MRD is slightly shifted towards lower frequencies, so that the sharp increase from a downward to an upward MRD starts at about 700 Hz. The MRD of the voiced [z]
Figure 1. Vertical directivity averaged over all subjects over frequency for the sentence, the vowels and the plosives. The black line shows the MRD averaged over all subjects.
Figure 2. Vertical directivity averaged over all subjects over frequency for the fricatives, nasals and alveolars. The black line shows the MRD averaged over all subjects.
is similar to that of the voiceless [s] and [ʃ] fricatives. However, the standard deviations are much larger for the [z] below 1 kHz than for any of the other phonemes. On the contrary, for the plosives, the standard deviations are small compared to the other phonemes.

3.2 Main radiation direction gain

Figure 4 shows the means and standard deviations of the MRDG for the different phoneme groups, illustrating the strength of the increased sound radiation towards the MRD. The plots indicate frequency- and phoneme-dependent differences in MRDG, which was statistically confirmed by a Greenhouse-Geisser (GG) corrected [24] two-way repeated measures analysis of variance (ANOVA) with the within-subject factors phoneme (23 phonemes) and frequency (19 third-octave frequency-bands from 125 Hz to 8 kHz), revealing a significant main effect of phoneme $F(22,264) = 15.42$, $p_{\text{GG}} < .001$, $\eta_p^2 = .56$, $\epsilon = .25$, and frequency $F(18,2136) = 50.80$, $p_{\text{GG}} < .001$, $\eta_p^2 = .81$, $\epsilon = .15$, as well as a significant phoneme × frequency interaction effect $F(396,4752) = 7.48$, $p_{\text{GG}} < .001$, $\eta_p^2 = .38$, $\epsilon = .02$.

For most phonemes, there is a prominent peak at $f \approx 700$ Hz. However, some phonemes, such as the unvoiced fricatives [f], [s], and [ʃ], do not show this prominent peak in this frequency range, whereas others, such as the voiced alveolars [ɻ] and [ɾ], show a stronger peak when compared visually to the other phonemes. For statistical analysis of the MRDG across phoneme groups, we performed pairwise comparisons between the frequency-dependent means of each phoneme group (i.e., averaged MRDG across the respective phonemes of each group) using nested GG-corrected two-way repeated measures ANOVAs with

![Figure 3. Mean values and standard deviations across subjects of the MRD (1/12-octave smoothed) for the phoneme groups. For comparison, each plot also shows the MRD for the phonetically balanced sentence.](image-url)
the within-subject factors frequency and phoneme group. The initial significance level of 0.05 was further corrected according to Hochberg [25] to prevent alpha-error accumulation. Most interestingly, the ANOVAs yielded significant frequency x phoneme group interaction effects for all five comparisons involving the unvoiced fricatives (all $p_{GG} \leq .003$); for the sake of conciseness, the statistical results of the nested ANOVAs are not reported in greater detail throughout the manuscript). The effect is driven by the deviations in MRDG below 1 kHz for [f], [s], and [ʃ], indicating that the downward radiation is weaker for these unvoiced fricatives than for any of the other phonemes examined in the present study. Furthermore, the analysis revealed significant frequency x phoneme group interaction effects for the five pairwise comparisons involving the voiced alveolars (all $p_{GG} \leq .01$), indicating that the downward radiation below 1 kHz is stronger for these phonemes compared to all other phonemes. The MRDG remains below 2 dB for all phonemes between 1 kHz and 4 kHz. Above 4 kHz, the MRDG increases most for the nasals, reaching values of up to 6 dB. For all other phonemes, the MRDG remains below 2 dB. This is consistent with our observations [9, Fig. 6], where we found for the nasals a strong dip in frontal radiation relative to other directions above 5 kHz. The statistical analysis further confirmed these observations, with significant frequency x phoneme group interaction effects for the five pairwise comparisons involving the nasals (all $p_{GG} < .001$), which is driven by the comparably high MRDG of the nasals in the higher frequency range. Thus, in the frequency range above 4 kHz, the downward radiation seems to be significantly stronger for nasals than for any of the other phonemes analyzed.

Figure 4. Mean values and standard deviations across subjects of the MRDG (1/12-octave smoothed) for the phoneme groups. For comparison, each plot also shows the MRDG for the phonetically balanced sentence.
Figure 5. Subject-averaged GRG over frequency for the phonetically balanced sentence, the vowels and the plosives. The black lines indicate where the direct sound and the ground reflection reach the listener with equal strength (GRG = 0 dB).
Figure 6. Subject-averaged GRG over frequency for the fricatives, nasals and alveolars. The black lines indicate where the direct sound and the ground reflection reach the listener with equal strength (GRG = 0 dB).
3.3 Ground reflection gain

To estimate how the MRD affects the ground reflection, we further determined the GRG. A downward orientated MRD amplifies the ground reflection compared to the horizontally-directed direct sound component. As a result, the ground reflection can be even stronger than the direct sound component, resulting in GRG > 0 dB. The GRG is maximal at distances where the radiation direction of the ground reflection is the same as the MRD and thus has a strong frequency dependence that is additionally affected by the distance between the human speaker and the listener. In the following, we assume an ideal reflecting ground floor \((a = 0)\) and a height of the human speaker’s mouth and the listener’s ears of 1.5 m, which represent typical values for a natural communication scenario with two persons facing each other. We consider a varying distance between the human speaker and the listener.

Figures 5 and 6 show the resulting subject-averaged GRG for the phonetically balanced sentence and each of the phonemes. In each plot, the black line indicates where the direct sound and the ground reflection are equally strong (GRG = 0 dB).

Below 800 Hz and for most phonemes, the subject-averaged GRG is higher than 0 dB for distances above 3 m. For the sentence, the GRG is maximal at a distance of 4 m at 750 Hz with a GRG of 2.6 dB. The phonemes lead to a maximal GRG of 4.6 dB for an [l] at 700 Hz at a distance of 3.85 m and an [l] at 800 Hz at a distance of 3.2 m. On the contrary, for frequencies between 800 Hz and 1.6 kHz, the GRG is for all phonemes below 0 dB. Thus, we found a similar structure of the GRG for most phonemes. Exceptions are the voiceless fricatives [f], [s], [ʃ], and the [z], for which maximal values of the GRG are less than 1 dB below 800 Hz. For the [s] the [ʃ], and the [z], the highest GRG of more than 1 dB can be found in a region between 4 kHz and 7 kHz. These similarities between the [z] and the [s], which are unvoiced and voiced counterparts, are worth noting. For the plosives [t], [k], [b], [d], and [g] there is a second frequency range with an increased GRG at about 5.5 kHz, reaching values of up to 2.7 dB for the [d]. The nasals show similar behavior but at slightly higher frequencies of about 7.5 kHz, reaching maximal values of 3.2 dB for the [n] at a distance of 4 m.

The ground reflection and thus the GRG are directly affected by the absorption coefficient \(a\) of the ground floor, according to equation (3). The influence of \(a\) is visualized for the phonetically balanced sentence in Figure 7. For \(a = 0.2\), there is still a region below 1 kHz at distances above 3 m where the GRG is above 0 dB and another one above 4 kHz. The region becomes much smaller for \(a = 0.4\) only in an area around about 700 Hz and 5 m distance, and completely diminishes for \(a = 0.6\).

At the listener’s position, the direct sound and the ground reflection are superimposed, causing constructive or destructive interference depending on the phase characteristics. Generally, this interference results in comb filter effects with a distance-depending structure of peaks and dips over frequency.

4 Discussion

This study examined frequency dependencies of the MRD for human voice radiation and investigated to what extent the radiation towards the MRD is stronger than in the frontal direction. Our results confirm those of other studies in the literature, all of which, however, are based on datasets averaged over a complete sentence and thus cannot analyze phoneme dependencies as in the present...
study. Dunn and Farnsworth [4] performed measurements in octave bands for \( f < 500 \) Hz and one-half octave bands for \( f \geq 500 \) Hz at a distance of \( 1 \) m and a vertical resolution of \( 45^\circ \). They found that in the one-half octave bands between \( 500 \) Hz and \( 1 \) kHz, the main direction is downward for \( \theta = -45^\circ \) and between \( 1 \) kHz and \( 1.4 \) kHz upward for \( \theta = +45^\circ \), which is consistent with our results for the phono-
tically balanced sentence. It is worth noting that data of that study shows a similar trend for near-field measure-
ments as well. Marshall and Meyer [5] determined the vertical voice directivity in octave bands with center fre-
cuencies between \( 125 \) Hz and \( 8 \) kHz and found the MRD to be about \( \theta = -20^\circ \). Because this study provides only data in octave bands, the abrupt changes in MRD around \( 1 \) kHz could not be resolved. However, in the \( 1 \) kHz octave band, the strength of the radiation is similar for \( \theta = -20^\circ \) and up to \( +45^\circ \). Chu and Warnock [6], who performed mea-
surements in one-third octave bands, found downward MRDs (\( -42^\circ \leq \theta \leq -13^\circ \)) for the bands with center fre-
cuencies between \( 315 \) Hz and \( 800 \) Hz and an upward MRD (\( \theta = 31^\circ \)) at the frequency bands with center fre-
cuencies of \( 1 \) kHz and \( 1.25 \) kHz, which is also in line with our results. Similarly, the results of Moreno and Pretzschner [26] show MRDs upwards and downwards in these fre-
cuency bands, but with slightly different absolute values, probably due to the varying sampling grid. 

The MRDG as depicted in Figure 4 shows a prominent peak for most of the phonemes at \( f \approx 700 \) Hz. Similarly, the spherical directivity index has a characteristic dip in this frequency range [8, 9], as both are direct counterparts. Since the directivity index is calculated related to the frontal direction, it decreases when the radiation into other directions increases. Accordingly, the dip in directivity index is less prominent for the unvoiced fricatives [\( f \), [s], and [\( f \)], and so is the MRDG for these phonemes (cf. [8] and Fig. 4).

Although the general behavior of the MRD can be determined from data in the literature, it has not yet been studied in detail, and no causes have been identified for the increase in MRD in 

MRD at \( f \approx 700 \) Hz. Since our results show that it is similar for a large number of phonemes, it is unlikely to be caused by properties of specific phonemes, such as different mouth opening sizes and radiation through the mouth and nostrils in the nasals. In this context, Birkholz et al. [27] recently simulated the frontal radiation of a modelled human mouth and torso, which showed a sharp dip at \( f \approx 800 \) Hz and a rise to a (local) maximum at \( f \approx 1.3 \) kHz. Their comparative simulations of a spherical cap in a spherical baffle (without any kind of torso) did not show these peaks and dips, indicating that diffraction and reflection at the shoulder and torso could be a cause of these effects. 

In a recent study, Blandin et al. [28] examined torso diffraction and reflections in more detail. They showed that when the torso is present, interference patterns occur containing repeated maxima and minima in both the horizontal and vertical planes. For the frequency range below \( 2 \) kHz, where we found the prominent pattern of downward and upward directed MRDs, the study by Blandin et al. showed a significant effect of the torso on the directivity index.

The downward-directed MRD strongly affects the ground reflection intensity. The strongest effect can be observed for distances between \( 3 \) m and \( 7 \) m. Below \( 800 \) Hz, the ground reflection is stronger than the direct sound for most phonemes, with a peak of \( 4.6 \) dB for an \( [i] \) and an \( [I] \). Even though this effect is weaker for the other phonemes, the behavior is similar for all phonemes except a few unvoiced fricatives. Towards higher frequencies, there are phoneme-dependent maxima, e.g., for most of the plosives.

Bech [11, 12] determined the perceptual thresholds of 17 early reflections in a simulated small room with a volume of \( 113 \) m\(^3\). The results showed that the signal level of four early reflections exceeded the human threshold of detection which included the ground reflection that was attenuated by \( 1.36 \) dB related to the direct sound. Bech suggested that the ground reflection in particular affects the perceived spaciousness. A few other studies indicate an effect of the ground reflection on localization. Guski [10], studied the influence of reflections from the ground, ceiling, and side walls on localization and showed that adding a reflection from the ground significantly reduces the vertical localization error of a speech signal. Gourévitch and Brette [13] analyzed the influence of ground reflections on binaural cues utilizing numerical models. The authors hypothesized that the ground reflection may potentially provide additional cues for distance and elevation estimation. Wendt et al. [15] investigated aspects of vertical sum localization due to the ground reflection. While the authors did not find a significant influence of the ground reflection for broadband noise, for speech signals the ground reflection contributed to the perception of height. Finally, the ground reflection can affect distance perception. In this context, the results of Ebelt et al. [14] indicate an influence of the ground reflection on distance estimation if the subject is familiar with the environment. None of these studies took into account the directivity of the human voice, which has an MRD that is directed downward over a wide frequency range. Accordingly, for the human voice, the effects of the ground reflection could be even stronger than the ones determined in these studies. As the ground reflection is present in many typical conversa-
tion scenarios, future studies focusing on the perceptual relevance of increased downward radiation of human voice directivity need to be carried out.
In our study, we assumed a non-tilted head with a frontal head orientation. However, it has been shown, e.g., by Munhall et al. [30], that people intuitively tend to make small head movements during conversations and to tilt their head while speaking, which is to some extent to support what is being said. We plan to investigate aspects of non-frontal head orientation on voice radiation in future studies.

5 Conclusion

We investigated the main radiation direction (MRD) of the human voice for a variety of phonemes, showing a characteristic pattern of MRD over frequency. We found statistically significant differences, for example, between the unvoiced fricatives [f], [s], and [ʃ], and the other phonemes. For all phonemes, the MRD is slightly downward in a range of −15° to −45° over a wide frequency range, but between 800 Hz and 1.6 kHz the MRD is upward to about 15°. The downward radiation strongly affects the ground reflection, with amplitudes that are higher than those of the direct sound component over a wide frequency range. As a consequence, for distances greater than 3 m between speaker and listener, the reflection from an ideal reflecting ground floor reaches the listener with a higher amplitude than the direct sound at frequencies below 800 Hz. Accordingly, perceptual aspects of the ground reflection's magnitude and the perceptual relevance of considering phoneme dependencies need to be addressed in future research.

Funding

The research presented in this paper has been partly funded by the Federal Ministry of Education and Research in Germany, support code BMBF 03FH14XS/NarDasS, and by the German Research Foundation under Grant No. DFG WE 4057/21-1.

Conflict of interest

The authors have no conflicts of interest to disclose.

Data availability statement

This work applies measured datasets from our previous studies [8, 9, 16], which are publicly available as research data associated in Zenodo, under the references [31, 32].

References


Cite this article as: Pörschmann C. & Arend JM, 2024. On the impact of downward-directed human voice radiation on ground reflections. Acta Acustica, 8, 12.