Real-Time Calculation of Frequency-Dependent Directivity Indexes in Singing

Manuel Brandner, Alois Sontacchi, and Matthias Frank

Institute of Electronic Music and Acoustics, 8010 Graz, Austria, Email: brandner@iem.at

Introduction

Directivity of speech and singing is of great interest in acoustic modeling [1], performance studies [2, 3] and for architectural design [4]. Voice directivity is frequency-dependent and primarily determined by the physiology of a person. Furthermore, it can be changed to an extend by posture, head inclination, and vocal tract configuration. The most prominent influences on voice directivity can be summarized as:

- shape and size of head and torso (fixed),
- posture, head inclination (changeable),
- vocal tract geometry (changeable),
- spectral emphasis (changeable).

While, a recent database [5] includes the frequency-dependent directivity for each playable tone of some instruments and some of its overtones in order to facilitate maximum authenticity in reproduction, typical models in acoustic simulations do not take the aforementioned effects and properties of speech and singing into account. These models are defined by one global directivity pattern for a large frequency range.

In our study we investigate different methods for calculating the voice directivity of a classical singer. We show the applicability of the short-term Fourier transform analysis (STFT analysis) analyzing singing at steady pitch in comparison to a less time-consuming alternative, namely the glissando method proposed in [6]. We define metrics to evaluate our results and to better investigate the variability of voice directivity over frequency by the example of a classical singer. We use the directivity index for the horizontal and vertical plane (HDI, VDI) as objective metrics [11]. For the real-time calculation, the HDI and VDI values are evaluated at the fundamental frequency and its harmonics to gain more robust information with regard to the SNR (signal-to-noise ratio). The analyzed short-term data is then compared to results calculated by the glissando method.

One professional classical singer with a master's degree in classical voice and international singing experience was asked to participate in our study. We asked the singer to use two different strategies to vocalize the German vowel /a/. One strategy is using a rather small mouth opening and the other a rather large mouth opening achieved by lowering the jaw as discussed in [7]. To analyze the outcome of the strategies we compare the results of each strategy for each measurement method.

Method

Measurement Setup

The measurement of source radiation patterns employed a microphone array consisting of two circular rings, one placed in the horizontal, the other one in the vertical plane, respectively, cf. Fig. 1 and [8]. Each of the rings can hold up to 32 microphones resulting in a maximum number of 62 microphones as both rings intersect in the front and back of the array. The angular spacing of the microphones is 11.25°. To reduce reflections the center-facing side of the rings with a thickness of 21 mm are beveled. The distance from the microphone capsules to the center-facing side is 8 cm to ensure a radius of 1 m from the capsules to the center of the array. The full diameter of the apparatus spans 2.56 m at its widest point. The center of the array can be lifted to any height between 1.3 m and 2 m by adjustable stands. As microphones, we used NTI MA 2230 connected to an Andiamo.MC Directout Technologies microphone preamplifier. An optical tracking system (optitrack flex 13, 6 cameras) was used to validate the position of the head during the recordings to account for possible head movements. Furthermore, the system allows to some extent to guide the test subject by a visual feedback of the current head and actual center position during measurements. In the case of a singer, the mouth is defined as the acoustical center.

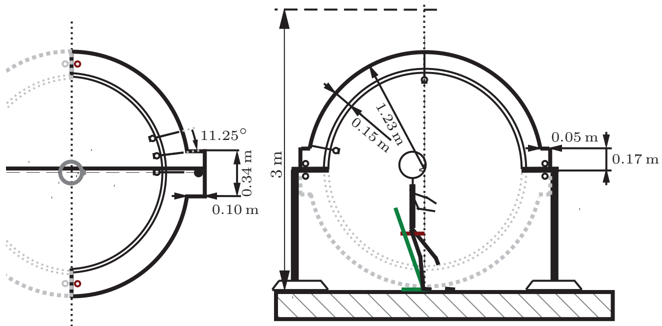


Figure 1: Double Circle Microphone Array. Schematic of the measurement setup with two wooden circular rings.

Short-term Fourier transform analysis

The singer is asked to sustain the German vowel /a/ for as long as possible at steady pitch (note G4) and to attempt to use similar vocal effort for all measurement runs. The microphone signals are analyzed in blocks of equal length and transformed from time domain to the frequency domain. The magnitude spectra are averaged over time for each channel and then third-octave smoothed.

The radiation characteristics can be compared using the frequency-dependent directivity index. The calculation is done in quasi real-time (averaging only over a limited number of blocks) with the 2D Polar Pattern and Spectrum Analyzer [8] where the number of Fourier coefficients and the hopsize is adjustable. The examplary data reported in this document is calculated with a window and FFT size (Fast Fourier transform) of 2048 samples at a sampling frequency of 44100 Hz. We use a 4-term Blackman-Harris window and therefore an overlap of 66.1% between consecutive frames [9]. The power spectra $|H(\omega,n)|^2$, where n denotes the discrete time instants, are averaged by using a first-order recursive filter (Eq. 1). The filter coefficient λ, which can be seen as a forgetting factor, is set to 0.5 (smoothing time $\approx$ 100 ms).

$$\hat{H}(\omega,n) = \lambda|H(\omega,n)|^2 + (1-\lambda)|H(\omega,n-1)|^2 \qquad (1)$$

The harmonic structure of the spectrum in singing does not provide valuable information (amplitude above the noise floor) at each frequency bin and at all time. There are gaps between the overtones that only hold low energy and possibly fall below the noise floor from time to time. If too much energy of the noise floor is averaged, the directivity changes towards an omni-directional pattern in the analysis. This happens as measurement noise provokes a directivity index of around 0 dB. To overcome this misleading effect a pitch tracker is used to calculate the HDI and VDI solely at the fundamental frequency and its harmonics.

Glissando method

The glissando method (vocal sweep method, [6]) allows to calculate impulse responses directly from directivity measurements for vocalized phonemes. To capture the source signal a reference microphone is positioned in front of the singer as close as possible (< 3 cm). The performer is asked to sing a glissando (vocal sweep) starting at a low pitch (G4) and ending at a higher pitch at least one octave above. The impulse responses are calculated by deconvolution of the measured signals at each microphone by the reference signal in the frequency domain and then cut to a length of 512 samples and windowed in the time domain. This simultaneous measurements allow a reduction of both measurement time and positioning errors.

Visualization using circular harmonics

The visualization of the polar patterns uses an interpolation scheme which is applied in the circular harmonics domain (cf. [8]). Further information about circular and spherical harmonic decomposition can be found in [10].

Horizontal and vertical directivity index

The directivity factor $\gamma_p(\omega)$ is calculated at each frequency for the horizontal and vertical plane, respectively. It is defined by the ratio of the on-axis power to the mean power of all sampling positions, where L denotes the total number of measurement points within the corresponding plane.

For both planes we denote the angle ϕ_i for each measurement position independent of the orientation of the plane with ϕ_0 as the on-axis direction.

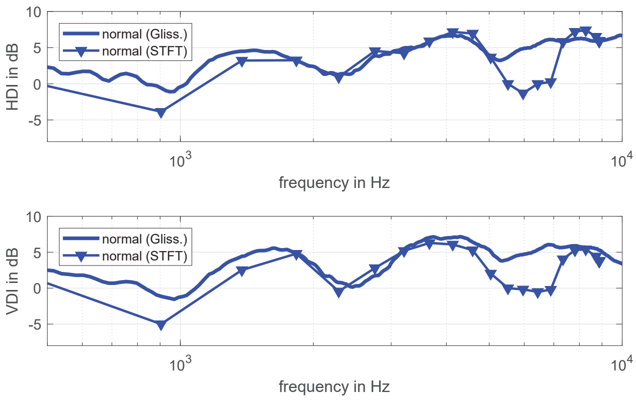
$$\gamma_p(\omega) = \frac{|H(\omega,\phi_0)|^2}{\frac{1}{L}\sum_{i=0}^{L-1}|H(\omega,\phi_i)|^2} \qquad (2)$$

The directivity index at a frequency ω for each plane is then defined in dB as follows

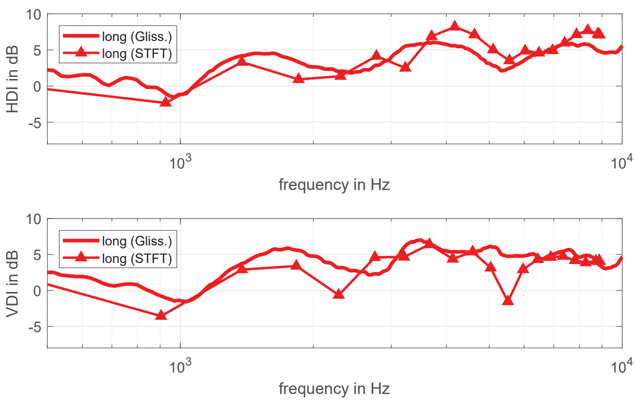
$$DI(\omega) = 10\log_{10}(\gamma_p(\omega)). \qquad (3)$$

Results

Within this section we present a comparison of the results for the two proposed methods to calculate the frequency-dependent directivity index. For each method the frequency data used for calculating the directivity indexes is third-octave smoothed. The classical singer was asked to use two different strategies to produce a similar German vowel /a/. Our suggested strategies are to use a rather small mouth opening "normal", more similar to speech, and in comparison a vertical larger mouth opening "long" achieved by lowering the jaw.



(a) Comparison of HDI/VDI values for "normal" condition



(b) Comparison of HDI/VDI values for "long" condition

Figure 2: Comparison of HDI/VDI values calculated with the STFT analysis (only evaluated at △ or ▽) and glissando method for two strategies for the vowel /a/. The results of both methods show good agreement, besides minor deviations.

Comparison of the proposed methods

In Fig. 2 we compare the results of the glissando method with the results of the STFT analysis. In most of the frequency regions the results agree very well. Although, the HDI/VDI values differ for the "normal" mouth opening at frequencies below 2 kHz, as seen in Fig. 2a. It is especially shown that around 1 kHz larger differences occur between the methods. This can be explained by the different frequency resolutions used for the two measurement methods and the spectral sparsity of a sung note (STFT). The dip around 1 kHz, which is provoked by the shadowing and reflection properties of the torso, occurs within a bandwidth of around 200 Hz. As the metrics are calculated from data smoothed over third octaves and the methods use different frequency resolutions this dip is less pronounced for the glissando method (Δf=86 Hz) than for the STFT analysis (Δf=21.5 Hz). In Fig. 3 we show the influence of the smoothing bandwidth on the results of the glissando method for the "normal" mouth opening. The decrease of the directivity index around 1 kHz, as seen in the results by STFT analysis, is can be made visible if the smoothing bandwidth is reduced to sixth-octave bandwidth.

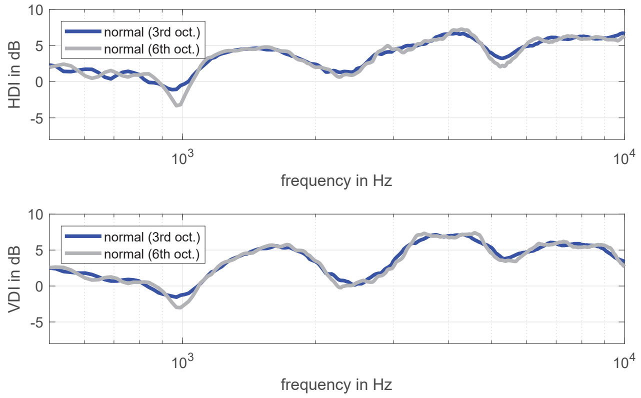


Figure 3: Comparison of HDI/VDI values for the glissando method with third-octave and sixth-octave band smoothing (light gray) for "normal" condition.

For the larger mouth opening "long" much more deviations between the methods are visible (Fig. 2b). These deviations can occur because the exact same mouth opening cannot be guaranteed for both runs and due to differences resulting from averaging; and again due to the spectral sparsity of a sung note. Furthermore, for larger mouth openings higher-order modes are more likely to be radiated from the mouth in comparison to small mouth openings [1].

At higher frequencies the largest deviations are shown around 5 to 7 kHz for the "normal" condition and at 5.5 kHz for the "long" condition which occur due to a low SNR (cf. Fig. 4). Therefore, a threshold for low SNR should be used in the future to easily identify the validity of the measurement data.

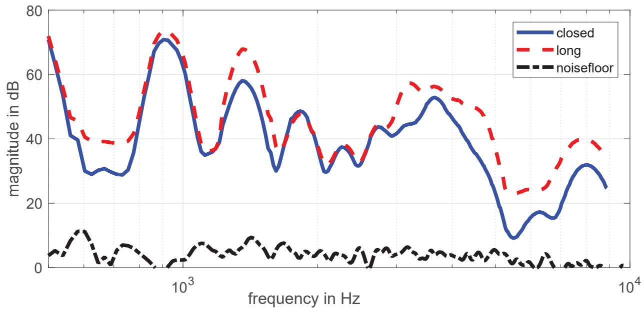


Figure 4: Comparison of the on-axis response third-octave smoothed) for the two strategies for the vowel /a/ from STFT analysis.

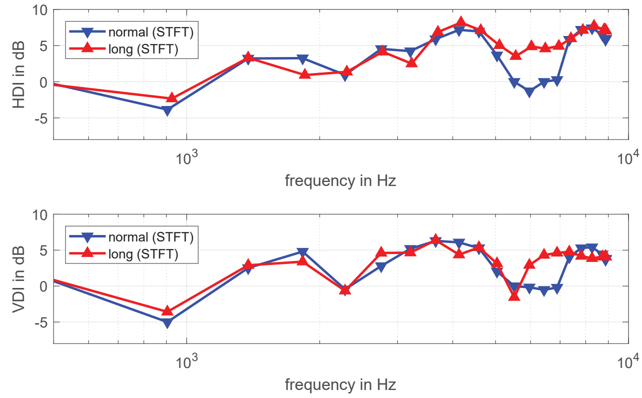


Figure 5: Comparison of HDI/VDI values for the two strategies for the vowel /a/ with the STFT analysis (from third-octave smoothed data).

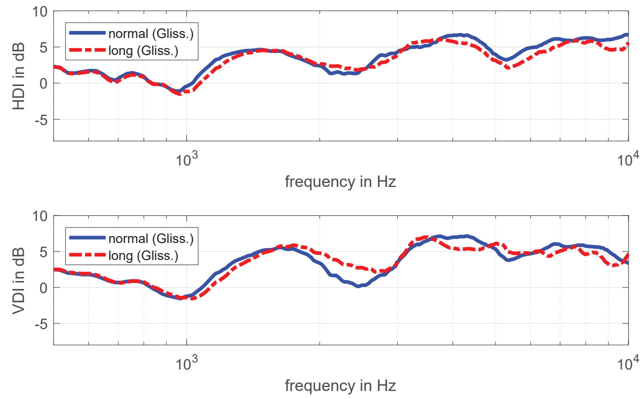


Figure 6: Comparison of HDI/VDI values for two mouth openings for the vowel /a/ with the glissando method.

Comparison of the singing strategies

This section compares the results of each method with regard to the singing strategy (mouth opening) used. The largest deviations are about 2 dB and can be seen at several frequencies, if we exclude the frequency region of the STFT analysis where we attested bad SNR. Nevertheless, the explained better SNR for the "long" condition within the frequency region of 5 to 7 kHz implies higher spectral energy in comparison to the "normal" singing condition. Therefore, the "long" condition yields higher directivity if we incorporate the level in our analysis, see Fig. 5.

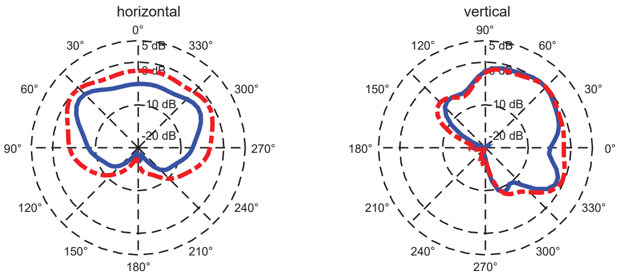


Figure 7: Comparison of the polar patterns at 2.5 kHz for two mouth openings: (i) "normal" (solid) and (ii) "long" (dash-dotted) for the vowel /a/ calculated with the glissando method (third-octave smoothed).

As the glissando method is related to a classic sweep measurement it inherits a better SNR than an averaging method like the STFT analysis. In Fig. 6 we see larger deviations in the vertical plane for the results of the glissando method with an increase of directivity especially for the frequency region around 2.5 kHz. In Fig. 7 the differences at 2.5 kHz are shown in polar plots, which are normalized to the maximum of both patterns in each plane. This rather large variation has been also reported in [3] and seems to be linked to the size of the mouth.

If we consider the radiation of a small piston in an infinite baffle as a simplified model, we expect for the change of mouth opening from "normal" to "long": (i) a slight decrease of directivity in the horizontal plane due to a decrease of the mouth width and (ii) an increase of directivity in the vertical plane as the height is increased because of lowering the jaw.

However, this cannot be attested from our data, as differences do not occur intuitively when using the common metrics (HDI/VDI). If we study the radiation characteristics more thoroughly from frequency versus angle representations, we can see that for the larger mouth opening more energy is radiated towards the floor as the frequency increases. This cannot be represented by the classical directivity index.

Conclusion

We show that similar results can be achieved if the directivity index is calculated from short-time averages in comparison to the glissando method or long-term averages (cf. [2]). Although, the analysis needs to consider the sparsity of the spectrum (low frequencies) and the noise floor (high frequencies) which are the reason for larger deviations between the investigated methods. The real-time calculation is especially useful for a fast view on the data already during or after the measurement. Furthermore, it is a practical way when investigating a large number of conditions, e.g. several different phonemes.

Our results show that the directivity index in the horizontal and vertical plane is increasing with frequency except for some drops around 1 kHz, 2.5 kHz, and 6 kHz. We show subtle variations in the directivity index dependent on the used singing strategy (mouth opening). Our results are in good agreement with results found in literature. Future studies should evaluate the directivity of different phonemes with other metrics than the HDI/VDI and investigate the qualitative performance of the impulse responses achieved by the glissando method in auralization.

References

[1] Blandin, R.: Influence of Higher Order Acoustical Propagation Modes on Variable Section Waveguide Directivity : Application to Vowel [A]. Acta Acoustica united with Acustica 102 (2016) 1-12.

[2] Cabrera, D.: Vocal Directivity Measurements of Eight Opera Singers, Proceedings of the 18th International Congress on Acoustics (2004) 505-506.

[3] Katz, B., and D'Alessandro, C.: Directivity Measurements of the Singing Voice, 19th International Congress on Acoustics (2007) 45-50.

[4] Chu, W. T. and Warnock, A. C.: Detailed Directivity of Sound Fields Around Human Talkers, Technical Report for the Institute for Research in Construction (National Research Council of Canada) (2002) 1–47.

[5] Weinzierl, S., et al.: A Database of Anechoic Microphone Array Measurements of Musical Instruments (2017), URL: http://dx.doi.org/10.14279/depositonce-5861.2

[6] Kob, M. and Jers, H.: Directivity measurement of a singer. Collected Papers from the Joint Meeting "Berlin 99" (1999), ISBN:3-9804568-5-4

[7] Nair, A. and Nair, G., and Reishofer, G.: The Low Mandible Maneuver and Its Resonential Implications for Elite Singers, Journal of Voice 30 (2016) 128.e13-128.e32.

[8] Brandner, M., Frank, M., and Rudrich, D.: DirPat - Database and Viewer of 2D/3D Directivity Patterns of Sound Sources and Receivers. Audio Engineering Society Convention 144 (2018), URL: http://www.aes.org/e-lib/browse.cfm?elib=19538

[9] Heinzel, G., Ruediger, A., and Schilling, R.: Spectrum and Spectral Density Estimation by the Discrete Fourier transform (DFT), including a comprehensive List of Window Functions and some new Flat-Top Windows, URL: http://edoc.mpg.de/display.epl?mode=doc&id=395068

[10] Zotter, F.: Analysis and Synthesis of Sound-Radiation with Spherical Arrays. University of Music and Performing Arts, Graz (2009), URL: https://iem.kug.ac.at/fileadmin/media/iem/projects/2010/zotter.pdf

[11] Tylka, J., Sridhar, R., and Choueiri, E.: A Database of Loudspeaker Polar Radiation Measurements. Audio Engineering Society Convention 139 (2015), URL: http://www.aes.org/e-lib/browse.cfm?elib=17906