

A voice directivity measurement system with facial tracking and augmented acoustics

Manuel Brandner¹, Matthias Frank¹, Alois Sontacchi¹

¹ *Institute of Electronic Music and Acoustics, 8010 Graz, Austria, Email: brandner@iem.at*

Introduction

Voice directivity has an influence on the perceived acoustics for both the singer/speaker and the audience [1, 2, 3]. One of the most important aspects of voice directivity in a room is the direct-to-reverberant energy ratio (D/R ratio) at the listening position. The more focused the voice directivity is, the higher the D/R ratio [4, 5]. This is why voice directivity is becoming more and more important in virtual or augmented acoustic reality [6].

Another field of interest is the performance analysis of professional voices as part of the education. Typical analysis tools for voice characteristics include linear prediction. This method gives incomplete or even erroneous information if the voice is analysed at higher pitch [7, 8]. Voice directivity may help to identify general characteristics in addition to other acoustic features.

Previous studies showed that a change in mouth opening has an effect on the directivity characteristics, i.e., the main direction and the beam width [9, 10]. Different characteristics may hold valuable information about specific vocal tract configurations. Acquiring detailed voice directivity data from human speakers or singers is no easy undertaking due to head movements or mouth changes during measurements. Therefore, we propose an acoustic measurement system enhanced by a facial tracking system to build a ground truth for different mouth opening configurations in speaking and singing. Furthermore, to support natural usage of the voice while actually being in an anechoic measurement room, the system provides augmented acoustics to simulate a more comfortable acoustic environment [6].

This contribution explains the systems, its components, and the methods to process the measurement data. Finally, we present some exemplary results for measurements of different vowels and voice qualities sung by two classical singers.

Measurement system

A measurement system for the determination of sound radiation characteristics of the singing voice was set up by using the double circle microphone array (DCMA) [11], optical tracking sensors in order to measure the oral posture and center position of the singer, and a video camera. The video camera allows to validate the measurement results from the tracking system and opens the possibility to calculate the mouth opening from video directly. The measurement software was implemented in Pure Data ¹.

Double circle microphone array

The used microphone array has a radius of 1 m and consists of two circular rings, one placed in the horizontal plane and one in the vertical plane [11]. Each ring can hold up to 32 microphones resulting in an angular spacing of 11.25° and a total number of 62 microphones. In addition, a reference microphone is used, which is located at the exact center of the microphone array if the glissando method is used [9, 12].

Tracking

The mouth opening and absolute position of the singer/speaker inside the measurement array can be captured by a tracking system. The tracking system uses ten cameras, six of them are positioned on the ceiling of the measurement room and four closer in front of the singer/speaker. The closer cameras increase the localization accuracy for the mouth tracking.

Augmented acoustics

For the acoustic analysis, dry signals measured in an anechoic environment are ideal. However, in singing/speaking, room acoustics support the voice, which is a necessity in a longer recording session. Therefore, we use an augmented acoustic system [6] that gives the singer natural room acoustics via transparent headphones [13] without creating reverberation on the microphone signals. The augmented acoustics is fed by the microphone in front of the singer and employs a static, however frequency-dependent directivity [14] to excite the virtual room. The virtual room simulates a shoe-box-like concert hall with a size of roughly 30 m × 24 m × 20 m and reverberation time of 2.2 s.

Methods

This section presents methods to analyse voice directivity and metrics to discuss directivity characteristics.

Tracking data analysis

The tracking data is used to analyse the steadiness of a sustained sung vowel and to calculate the mouth opening area. Therefore, the tracking data of each time frame of the single facial markers around the mouth are put in order by calculating the convex hull. From the resulting polygon, the area is calculated, which gives the effective mouth opening area over time. This area comes with a bias of around 5 to 7 cm² due to the positioning of the markers around the lips, which needs to be considered for further analysis. Furthermore, the height and width of the mouth can be calculated and give valuable information about effects occurring in each single plane (horizontal or vertical). The tracking data during each segment (e.g. sustained vowel) can be used to analyse whether there is an increase or decrease of the mouth area during phonation.

¹freely available under <http://puredata.info/>

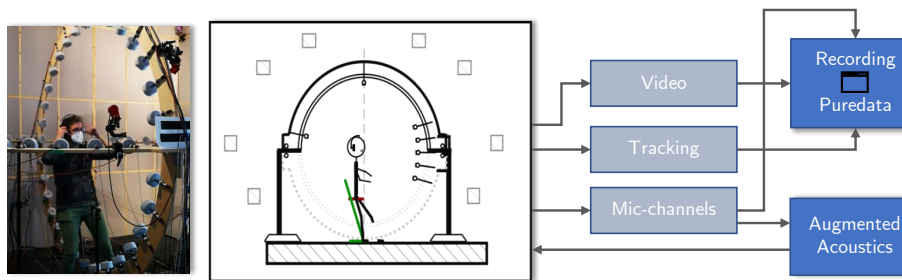


Figure 1: Left: person in the microphone array. Right: schematic representation of the measurement setup.

Calculation of directivity characteristics

Directivity characteristics are computed from frequency data of a sung phoneme at a single pitch calculated by Welch’s method (averaged periodogram method)[15] or sung glissandi by the use of the Glissando method. Analysing glissandi can help to overcome the problem of having energetic gaps between harmonics in the frequency domain data of sung sustained vowels at a single pitch [9, 12].

Averaged periodogram method

The microphone signals at the measurement positions are segmented and transformed via the Fourier Transform into the frequency domain. The estimated frequency responses are averaged over time and smoothed afterwards in, e.g., octave, third, or semitone bands. The advantage of the method is that it can be utilized to calculate directivity patterns from shorter speech or sung segments. Nevertheless, it is not possible to easily guarantee that each frequency bin holds valuable information and that it is above the noise floor. Due to the harmonic structure there are gaps between the harmonics that only hold low energy and possibly do not exceed the noise floor. If too much noise energy is averaged the resulting directivity pattern tends to have an omnidirectional shape with a directivity index (DI) of 0 dB. A fundamental frequency tracking algorithm and a proper noise floor calculation can be utilized to partly overcome this problem.

Glissando method

If the Glissando method is used, a reference microphone is needed. The acoustic signals are transformed via the Fourier Transform into the frequency domain. Then, the complex signals are divided by the complex reference signal. This gives the output to input relationship including the phase of the acoustic paths from the reference microphone to each measurement position. The transfer functions transformed back into the time domain give then impulse responses which can be cut to a minimum length to remove room influences, such as floor reflections. This method offers a high signal-to-noise ratio and due to the sung glissando the transfer functions can be calculated for a broad frequency range. Furthermore, this is a much faster method to acquire data in comparison to the averaged periodogram method and less exhausting for a singer/speaker if the glissandi have been trained in advance.

Directivity index for a single plane

The directivity factor $\gamma_p(\omega) = \frac{P_{on-axis}}{P_{mean}}$ in each plane is defined by the ratio of the on-axis power $P_{on-axis}$ to the average power P_{mean} of all sampling positions on the respective plane. The horizontal and vertical directivity index (HDI, VDI), evaluated at an angular frequency ω are defined in dB as follows:

$$DI(\omega) = 10 \log_{10}(\gamma_p(\omega)). \quad (1)$$

Energy vector

The energy vector \mathbf{r}_E in Eq. (2) can be utilized to describe the direction and the width of the main lobe of an acoustic source. This measure is commonly used in the context of 3D loudspeaker setups, but is as well useful in the description of the characteristics of any arbitrary sound source radiation [16].

$$\mathbf{r}_E = \frac{\sum_{i=1}^L |H(\omega, \phi_i)|^2 \mathbf{m}_i}{\sum_{i=1}^L |H(\omega, \phi_i)|^2} \quad (2)$$

The frequency-dependent magnitudes $H(\omega, \phi_i)$ are multiplied by the vectors $\mathbf{m}_i = [\cos(\phi_i), \sin(\phi_i)]^T$ of each measurement position i , $i = 1, 2, \dots, L$ in each respective plane, and normalized by the sum of the energy, yielding a normalization of the vector between the limits 0 (omnidirectional) to 1 (maximum focus to one direction). The following two metrics are used, here: (i) the main beam width in each plane $\theta_w = 2 \arccos \|\mathbf{r}_E\|$ and (ii) the main direction in the vertical plane $\theta_s = \arctan y_{\mathbf{r}_E} / x_{\mathbf{r}_E}$.

Exemplary Results

In this section we present exemplary results from two classical singers for the methods described in the section above.

Analysis of tracking data and audio signals

In Fig. 2 we show the representation of the time domain audio signal and the voice activity detection (VAD) of sung segments during one recording run. The vowel sequence /a:/, /e:/, /i:/, /o:/, and /u:/ is repeated three times as each sequence is sung with a different voice quality (modal, breathy, and pressed). Furthermore, we show in Fig. 2 the corresponding mouth area calculated from the data stream of the tracking markers around the mouth. During a sustained sung vowel a change of mouth opening can occur, which is indicated by a regression line above each segment. If the mouth opening increases or decreases the straight line is tilted either way.

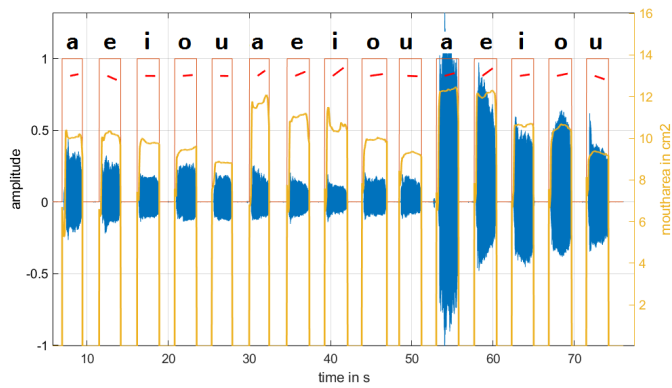


Figure 2: Time domain representation of the audio signal, VAD, and mouth area tracking. Tilted lines over each segment indicate a dynamic change during phonation - mouth area has been increased or decreased by the singer.

Directivity analysis

For a change of vowel or mouth opening we expect a change in directivity or in other words some effect on the directivity characteristics (e.g. directivity index, beam width, beam direction).

Averaged periodogram method

The directivity analysis with the averaged periodogram method of vowels sung by a classical singer at the pitch $a/A3$ of approx. 220 Hz is shown in Fig. 3. Again, the vowel sequence /a:/, /e:/, /i:/, /o:/, and /u:/ sung with three different voice qualities is investigated. In the upper plot of Fig. 3 a decrease of the effective mouth area used by the singer can be seen. Furthermore, a slight decrease of the mouth width and more pronounced decrease of the mouth height can be recognized. In the lower plot of Fig. 3 a decrease of the directivity index and the beamwidth in both planes from vowel /a:/ to /u:/ is visible for all three voice qualities.

Glissando method

A soprano singer was asked to sing the german vowel /a:/ with different provoked mouth openings (similar as in [9]). This is of special interest as the classical singing technique is often associated with a lowered jaw [17] and much larger mouth openings are expected to occur in singing than in conversational speech. In the horizontal plane a larger mouth width provokes a decrease in the beam width starting at around 2 kHz, whereas larger mouth openings along the vertical axis only show a change in beam width starting at around 5 kHz (see Fig 4(a) and (b) - top plots). In the bottom plots in Fig 4(a) and (b) the corresponding main direction of the found beam widths are shown. In the horizontal plane symmetric patterns provoke a focusing of the radiated sound towards the 0° direction for most frequencies. This differs in the vertical plane as frequencies below 1 kHz and above 5 kHz are more focused towards the floor, whereas between 1 kHz and 5 kHz sound is more radiated upwards or to the front. In Fig. 5(a) we show polar patterns normalized to the maximum with differences of around 5.4 dB towards the sides and up to 8.4 dB in

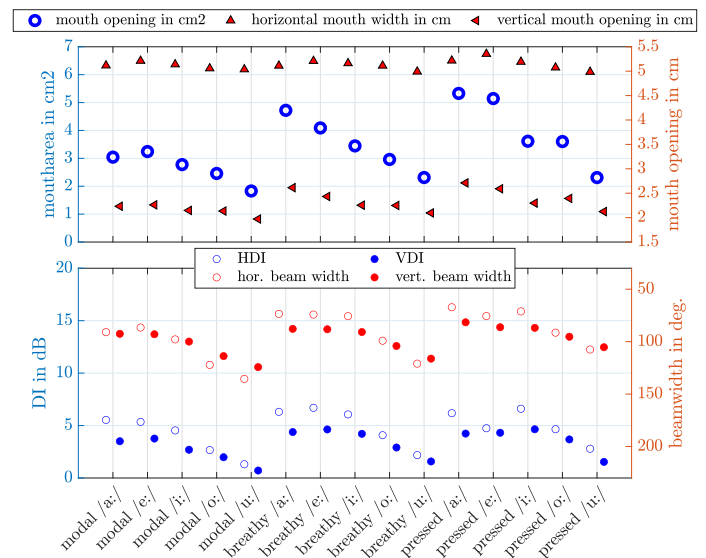


Figure 3: Top: averaged mouth opening area for each segment and the corresponding mouth width and height in cm. Bottom: horizontal, vertical, and averaged directivity index and for the corresponding plane the beamwidth of the energy vector and its average calculated from one-third-octave smoothed data.

Fig. 5(b) in the vertical plane for the forward upwards direction. In Tab. 1 we list the additional information from the tracking system which provides insight on the quality of the measurement results. While the effective mouth opening areas differ between the provoked mouth openings, the head inclination angle changes no more than 4.1° .

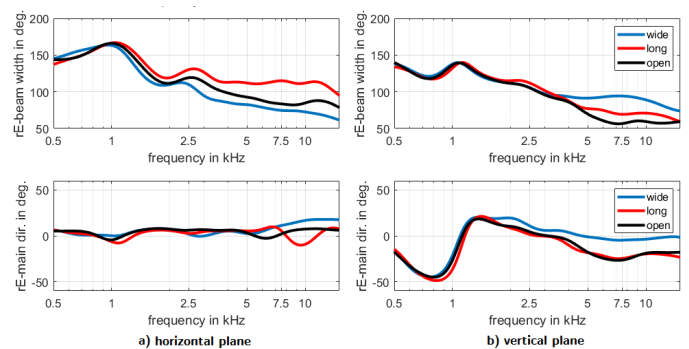


Figure 4: Glissandi analysis. In the top plots (a) horizontal plane and (b) vertical plane the differences between provoked mouth openings are displayed. In the bottom plots we provide the corresponding main direction of the displayed beam widths in the top plots.

Conclusion

We presented an enhanced measurement system to measure voice directivity and discussed strategies and methods to investigate directivity characteristics in detail by exemplary measurement results from two classical singers. The results show that differences between vowels and between different mouth openings can be made visible and substantiated by the use of a tracking system to verify the center position and validate the effective mouth

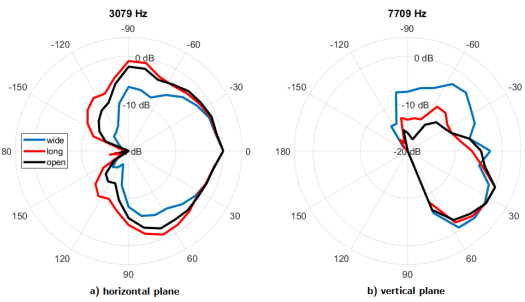


Figure 5: Polar patterns for the three mouth openings (a) in the horizontal plane at 3 kHz and (b) in the vertical plane at 7.7 kHz.

mouth		wide	long	open
Tracking				
Δ Area in cm ²		+0.52	+5.06	+10.66
Δ Head inclination in °		+0.40	+1.79	+4.10

Table 1: Mouth area and head inclination of the three investigated mouth openings listed as delta values. To calculate the delta values we use as a reference the mouth area for the vowel /a:/ used in speech by the singer.

opening area. Furthermore, the tracking data allows us to analyse the steadiness of a singer during recordings. We also enhanced the measurement setup by including augmented acoustics via transparent headphones, which provides natural room acoustics during the recording session. This opens up the possibility to investigate the influence of different room acoustics on the singing strategy of a singer, for example, on the vowel intelligibility. The microphone array offers to encode the actual directivity pattern of the singer into the augmented acoustics which is a goal in the future.

Acknowledgments

Special thanks to Thomas Musil for his help in implementing the measurement software in Pure Data.

References

- [1] Schärer Kalkandjiev, Z. and Weinzierl, S.: The Influence of Room Acoustics on Solo Music Performance: An Empirical Case Study. *Acta Acustica united with Acustica* 99/3 (2013), 433–441
- [2] Fischinger, T., Frieler, K., and Jukka L.: Influence of virtual room acoustics on choir singing. *Psychomusicology: Music, Mind, and Brain* 25/3 (2015), 208–218
- [3] Postma, B., Demontis, H., and Katz, B.: Subjective Evaluation of Dynamic Voice Directivity for Auralizations. *Acta Acustica united with Acustica* 103 (2017), 181–184
- [4] Frank, M. and Brandner, M.: Perceptual Evaluation of Spatial Resolution in Directivity Patterns 2: coincident source/listener positions. 5th International Conference on Spatial Audio (2019), 131–135
- [5] Frank, M. and Brandner, M.: Perceptual Evaluation of Spatial Resolution in Directivity Patterns.

Fortschritte der Akustik - DAGA (2019)

- [6] Frank, M., Rudrich, D., and Brandner, M.: Augmented Practice-Room - augmented acoustics in music education. *Fortschritte der Akustik - DAGA*, 2020
- [7] Arroabarren, I. and Carlosena, A.: Inverse filtering in singing voice: a critical analysis. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14 (2006), 1422–1431, <https://doi.org/10.1109/TSA.2005.858013>
- [8] Bereuter, P., Kraxberger, F., Brandner, M., and Sontacchi, A.: Design of a vowel and voice quality indication tool based on synthesized vocal signals. *Journal of the Audio Engineering Society* (2021)
- [9] 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 (2020), 1169–1180
- [10] Pörschmann, C. and Ahrend, J.: Investigating phoneme-dependencies of spherical voice directivity patterns. *The Journal of the Acoustical Society of America* 149/6 (2021), 4553–4564. Available: <https://doi.org/10.1121/10.0005401>
- [11] 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)
- [12] Malte, K. and Jers, H.: Directivity measurement of a singer. *The Journal of the Acoustical Society of America* 105/2 (1999), 1003–1003
- [13] Meyer-Kahlen, N., Rudrich, D., Brandner, M., Wirler, S., Windtner, S., and Frank, M.: DIY Modifications for Acoustically Transparent Headphones. *AES 148th Convention*, e-Brief 61 (2020)
- [14] Weinzierl, S., Vorländer, M., Behler, G., Brinkmann, F., von Coler, H., Detzner, E., Krämer, J., Lindau, A., Pollow, M., Schulz, F., et al.: A database of anechoic microphone array measurements of musical instruments, 2017, <https://doi.org/10.14279/depositonce-5861.2>
- [15] Welch P. D.: The use of fast Fourier transforms for the estimation of power spectra: A method based on time averaging over short modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, vol. 15 (1967), 70–73
- [16] Gerzon, M. A.: General metatheory of auditory localisation. *Audio Engineering Society Convention* 92, (1992)
- [17] Nair, A., Nair, G. and Reishofer, G.: The Low Mandible Maneuver and Its Resonant Implications for Elite Singers. *Journal of Voice*, vol. 30/1 (2016), 128.e13–128.e32. Available: <http://dx.doi.org/10.1016/j.jvoice.2015.03.010>