Comparing the directivity of a mouth simulator and a simple physical model

Georg Fischer, Christian Schneiderwind, Annika Neidhardt
Technische Universität Ilmenau, Institut für Medientechnik, 98684 Ilmenau, Deutschland,
Email: g.fischer@tu-ilmenau.de

Abstract

For increasing refinement of virtual acoustic environments, a detailed modelling of source directivities is desired. To include human speech in virtual scenarios it is of interest to create appropriate representations of the mouth as a sound source.

Physical mouth simulators are available to test telecommunications devices. However, their directivity is invariant and may not represent the directivity of a human mouth accurately. In order to gain flexibility a customizable parametric model is employed. As a starting point a simple physical model is considered.

In this paper the physical properties of this simple model are compared with that of a mouth simulator. It is shown, that the overall directivity of the simple model resembles that of the mouth simulator.

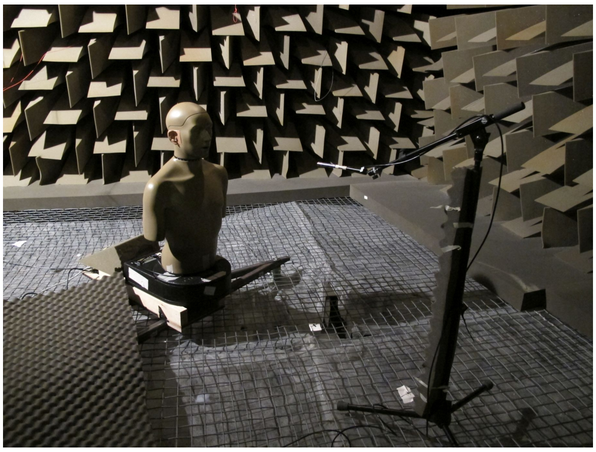

Figure 1: Setup for the directivity measurement of the Kemar 45BC's mouth simulator in the horizontal plane with an angular resolution of 4°. The microphone was placed at a distance of 60 cm from the center of the head.

Introduction

There have been various attempts to determine the directivity of human speakers [1, 2] or singers [3, 4]. These measurements are challenging especially due to the dynamic nature of the mouth directivity among different produced sounds. Furthermore, similar to HRTFs individual anthropometry may affect the directivity of the mouth for different people. The realization of a measurement setup achieving an adequate spatial resolution is another challenge in the determination of a mouth's sound emission characteristics.

Different manufacturers offer head-and-torso-simulators with mouth simulators mainly for testing telecommunication devices. They provide an invariant directivity, but the characteristics of the sound field produced by the mouth farther away from the head were of minor relevance in the development. The mouth simulators may not be a precise representation of the human mouth.
Halkosaari [5] compared the directivity of a Brüel & Kjær HATS 4128 mouth simulator and 13 human speakers by analyzing the sound fields at few measurement point very close to the head. The comparison of the sound field at the microphone positions brought up differences of more than 10 dB between the mouth simulator and the human mouths.

To model the acoustics around a human head, a rigid spherical model was used in previous studies [6, 7]. The localization cues of such a model were investigated e.g. in [8, 9].
For finding a suitable model to represent the sound emission of a mouth, Chalker and Mackerras [10] compared different approaches. The authors conclude that a circular piston in a spherical baffle is the most appropriate among the tested models. Halkosaari [5] included a spherical head model with a circular piston in their study and compared the modeled sound field with the measured values at few selected points close to the head.
Still, a detailed investigation of the directivity of the artificial mouth and the spherical model and the occurring deviations is missing.

The Kemar mouth simulator

Head and torso simulators are common tools to conduct binaural recordings and capture binaural room impulse responses. Selected products are additionally equipped with a mouth simulator that is realized by a loudspeaker behind the mouth opening. The development of these products is mainly motivated by testing applications in the telecommunication technology. These mouth simulators are usually not optimized for measuring OBRIRs. There may be partly even strong deviations between the directivity of a mouth simulator compared to those of a human mouth.
For this physical comparison the directivity of a Kemar 45BC's mouth simulator was measured with an omnidirectional microphone at a distance of 60 cm from the center of the head. An electronic turntable Outline ET-250 3D ensured accurate rotation of the Kemar. The loudspeaker in the mouth has a frequency range of 100 Hz to 24 kHz. Fig. 1 shows the measurement setup.

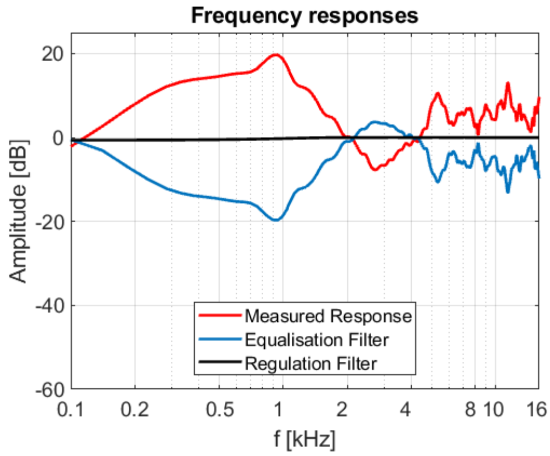

Figure 2: Frequency response of the originally measured mouth simulator transfer function for the direction at 0° (red) and the corresponding equalization filter.

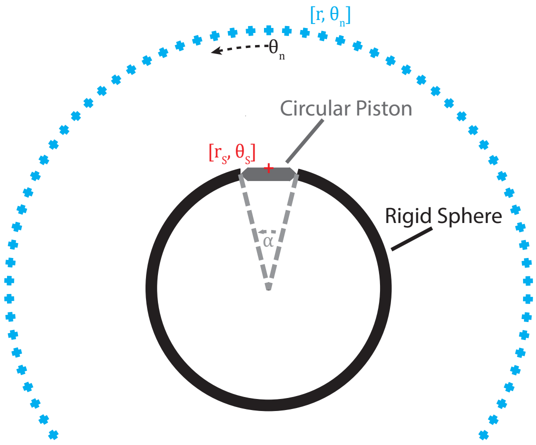

Figure 3: A rigid sphere with a circular piston is considered as physical model to represent the head with a mouth. The source position at $[r_S, \theta_S]$ is marked in red. The sampling points placed on a circular grid with 4° resolution at positions $[r, \theta_n]$ are depicted in blue.

Creation of equalization filter

The manufacturer recommends an equalization based on a measurement with a microphone placed 2.5 cm in front of the mouth [11]. However, for an OBRIR measurement with a reflecting surface at distances between 25-200 cm from the center of the head, it appeared as more suitable to choose a MRP (mouth references point) at 50 cm in front of the mouth (about 60 cm distance to the center of the head).

For the creation of the equalization filter, the impulse response determined in the directivity measurement for the 0° direction in front of the mouth was chosen. An inversion of the frequency response was realized with the least-mean-squares approach according to the description by Schärer et al. [12]. The roll-off frequencies were set to 60 Hz and 20 kHz.

This procedure is not standardized and might require further investigation on its own.

Spherical head model with circular piston

In order to model the directivity pattern of a human mouth the head is regarded as a rigid sphere with a vibrating circular piston representing the mouth. The directivity $D_n(\omega)$ can be calculated as the transfer function from the sound source, i.e. the middle of the circular piston at point $[r_S, \theta_S]$ to various points $[r, \theta_n]$, with $n = 1, 2, ..., N$ on a circular grid around the head as illustrated in fig. 3. Here, θ is the azimuth angle that ranges from 0° to 360°. The transfer function can be calculated by

$$D_n(\omega) = \frac{p_{r_S,\theta_S}(\omega)}{p_{r,\theta_n}(\omega)} \tag{1}$$

Here p_{r_S,θ_S} is the sound pressure generated by the sound source, p_{r,θ_n} is the pressure at the evaluation points $[r, \theta_n]$. According to [13] the pressure at the circular piston can be calculated by

$$p_{r_S,\theta_S} = \frac{i\rho_0 cW}{2} \sum_{n=0}^{\infty} [P_{k-1}(\cos\alpha) - P_{k+1}(\cos\alpha)] \cdot \\ \cdot \frac{h_k^{(2)}(kr_S)}{h_k^{(2)'}(kr_S)} P_k(\cos\theta_S), \tag{2}$$

where i is the imaginary number, ρ_0 is the fluid density, c is the speed of sound and the capital W is the velocity of the piston. P_k are the Legendre polynomials of $k-th$ order and h_k and $h_k^{'}$ are the spherical Hankel functions of second kind and $k-th$ order and their derivative, respectively. The angle α is the half-angle, corresponding to the mouth opening.

The pressure at the surrounding points can be calculated similarly, but by exchanging θ_S with θ_n and replacing r_S in the spherical Hankel function in the numerator with r, i.e. the radius at the evaluation points.

In the study [5] it was shown that this simple model can represent the directivity characteristics of an artificial head model (B&K HATS) with differences of about 5 dB in the frequency range below 1000 Hz at the few tested sampling points.

Comparison of directivity responses

The simulation was based on a spherical model of radius $r_S = 9$ cm, a mouth opening half-angle of $\alpha = 7°$ and a polynomial order of $k = 50$. For comparing the simulated model with the mouth simulator, a gain factor was applied to the measurement data to achieve the same amplitude at 1 kHz at 0 as in the simulated data.

The results of the measurements and simulations can be observed in figures 4 and 6. It is obvious that the measurements contain more irregularities than the simulations. This holds for the frequency responses (fig. 6) but also for the directivities (fig. 4).

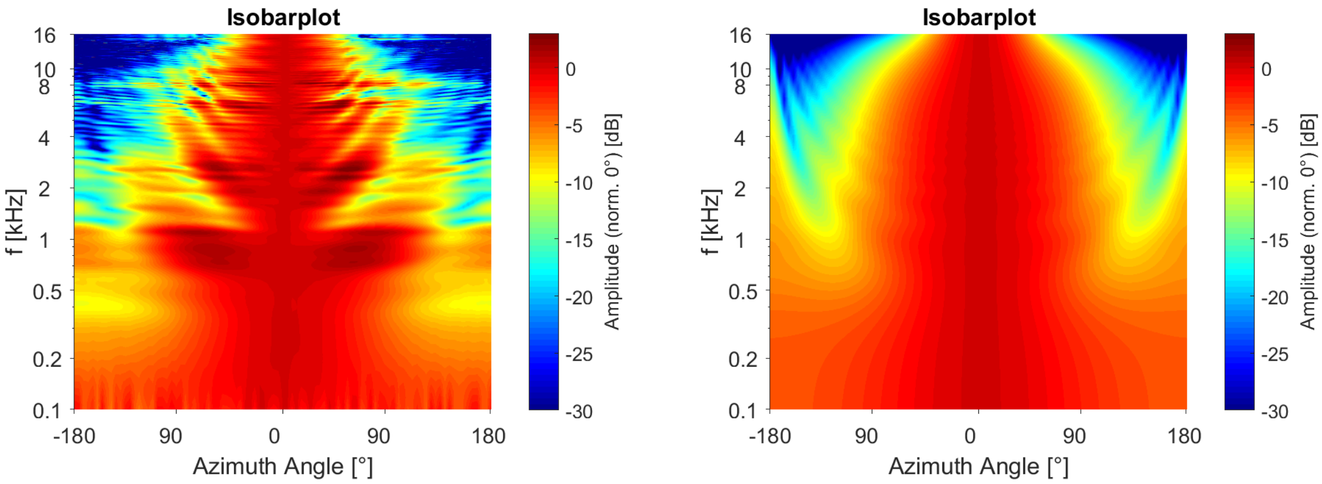

Figure 4: The directivity of a Kemar 45BC's mouth simulator (left) and the spherical head model (right) depicted as isobars with the reference axes at 0° azimuth.

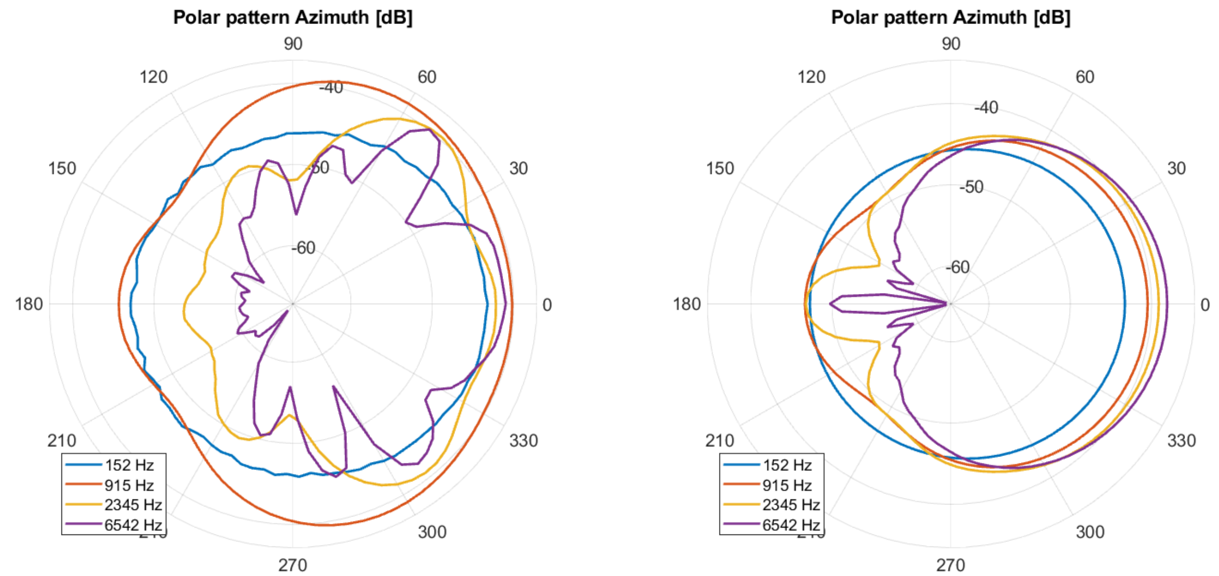

Figure 5: The polar patterns of a Kemar 45BC's mouth simulator (left) and the spherical head model (right) frequencies 152 Hz (blue), 915 Hz (red), 2345 Hz (yellow), 6542 Hz (purple).

In the frequency responses of the measured data comb-like filter structures around 1 kHz to 8 kHz are visible. They can be explained by reflections from the measurement setup, but might also arise from the body and head structure of the Kemar mouth simulator. The amplitude drop below 200 Hz is caused by the frequency response of the mouth simulator. The simulated responses are flat over the full range of the spectrum.

For side and backward directions, the energy of the higher frequencies in the measured data decreases due to the shape of the head at around 4 kHz to 5 kHz. Similar to the measurement, the directivity of the simulation shows continuously decreasing energy in the high frequencies for the side and backward directions. From the polar plot in figure 5 for the mouth simulator some sidelobes can be observed that are stronger than the response in the frontal directions. These sidelobes arise for 1 kHz, 2 kHz to 3 kHz and 7 kHz at directions of 60° to 90°.

Figure 7 visualizes the deviations for all directions. Due to the normalization of the measurement data mentioned above, a deviation of 0 dB can be observed at 1000 kHz and 0°. The sidelobes in the measurements can be seen in the negative dB values. The strongest deviations of up to around 50 dB arise at high frequencies in the back of the head, where the measurements show strong dips in the frequency responses. Overall the mean deviations lie in the region of 5 dB to 12 dB.

Discussion

In this paper, the directivities of a Kemar 45BC mouth simulator and a spherical head model with a circular piston are compared. It was shown that the overall directivity pattern of the spherical head model represents that of the mouth simulator with mean deviations of up to 12 dB in the back of the head. However, the spherical head model produces very smooth directional frequency

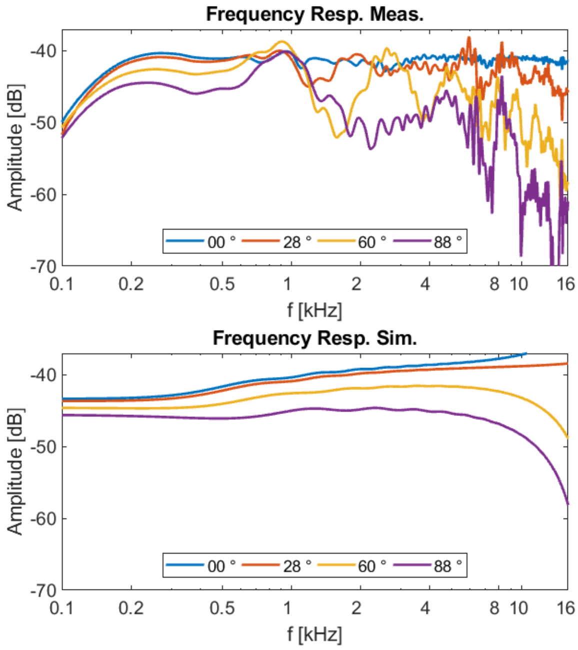

Figure 6: The frequency responses of a Kemar 45BC's mouth simulator (top) and the spherical head model (bottom) for 4 different directions.

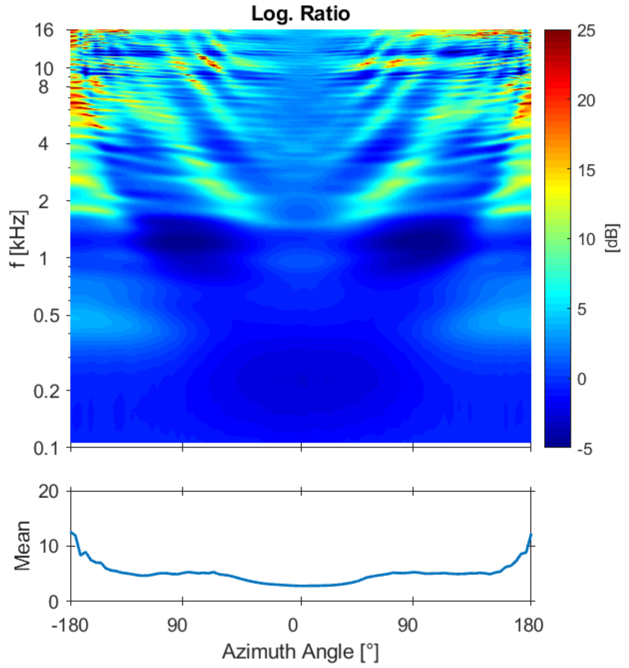

Figure 7: The deviations between the direction dependent responses of a Kemar 45BC's mouth simulator and the spherical head model. This plot visualizes the ratio between both approaches. The depicted data is ranging from around -9 dB to 50 dB, but for better visualization it was decided to clip the data.

responses in contrast to the mouth simulator, because the influence of reflections from the torso and was neglected in this model. Halkosaari et al. [5] suggested the integration of an infinite baffle to model reflections by the body using the mirror image source method.
For the spherical head model further room for improvement is identified. The mouth is positioned at 0° elevation, which does not exactly correspond to the usual position of the mouth. Placing it a few degrees lower may achieve a slightly better approximation. Furthermore, the human head would be better resembled by an ellipsoidal shape, rather than a sphere.
As discussed in the introduction, it should be kept in mind, that the mouth simulator may not be an adequate representation for the directivity of human speech.
In the future, to account for the different anthropometry of different people, an adaptation of head size and mouth openings may be required. For the creation of a dynamic model, various types of oral sounds like whispering, singing, shouting or different consonants and vowels should be considered. In addition, potential psychoacoustic tolerances for deviation in the directivity have to be quantified.

References

[1] W.T. Chu and A.C.C. Warnock. Detailed directivity of sound fields around human talkers. Technical report, Institute for Research in, 2002.

[2] H. K. Dunn and D. W. Farnsworth. Exploration of pressure field around the human head during speech. *J. Acoust. Soc. Am. 10, pp. 184-199*, 1939.

[3] d'Alessandro C. Katz, B.F.G. Directivity measurements of the singing voice. In *19th Int. Congress on Acoustics, Madrid, Spain, September*, 2007.

[4] M. Kob and H. Jers. Directivity measurement of a singer. Technical report, 1999.

[5] T. Halkosaari, M. Vaalgamaa, and M. Karjalainen. Directivity of artificial and human speech. *J. Audio Eng. Soc., Vol. 53, No. 7/8, Jul/Aug, pp. 620-631*, 2005.

[6] D.S. Brungart and W.M. Rabinowitz. Auditory localization of nearby sources. Head-related transfer functions. *J. Acoust. Soc. Am., 106 (3), pp. 1465-1479, Sept.*, 1999.

[7] V.R. Algazi, C. Avendano, and R.O.. Duda. Estimation of a Spherical-Head Model from Anthropometry. *J. Audio Eng. Soc., Vol. 49, Issue 6, pp. 472-479, June*, 2001.

[8] R.O. Duda and W.L. Martens. Range-dependence of the hrtf for a spherical head. In *Workshop on Appl. of Sig. Proc. to Audio and Acoustics, New Paltz, USA*, 1997.

[9] S. Bögelein, F. Brinkmann, D. Ackermann, and S. Weinzierl. Localization cues of a spherical head model. In *44th Annual Acoustics Conference (DAGA), Munich, Germany*, 2018.

[10] D.A. Chalker and D. Mackerras. Models for representing the acoustic radiation impedance of the mouth. *IEEE Transactions on Acoustics, Speech, and Signal Proc., Vol. ASSP-33, No. 6, Dec.*, 1985.

[11] G.R.A.S. Sound and Vibration. *Brochure Kemar 45BC*, 2013.

[12] Z. Schärer and A. Lindau. Evaluation of equalization methods for binaural signals. In *126th Convention of the Audio Eng. Soc.*, 2009.

[13] Earl G. Williams. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.