Technische Universität Berlin
Fakultät I - Geisteswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

Expose

Numerical simulation of voice direcitvity patterns for different phonemes

Leif Johannsen
Martrikelnummer: 403380

Supervisors: Dr. Paul Luizard
Dr. Fabian Brinckmann

Abstract

This expose gives an overview of the master thesis with the title “Numerical simulation of voice direcitvity patterns for different phonemes”.
The thesis deals with the effect of the mouth shapes of different phonemes on the radiated sound field and thus the directional characteristic of the human voice. The Boundary Element Method (BEM) is used as numerical simulation. Transfer functions between the mouth and multiple points on a sphere surrounding the head are simulated with the MESH2HRTF BEM solver [1]. This allows directional characteristics with high angular resolution to be calculated. The FABIAN artificial head is used as the basic model as a scanned mesh. In combination with different vocal tract configurations and mouth shapes, mesh models can be generated for different phonemes. Radiated sound fields can be calculated from these models. This makes it possible to compare the directional characteristics for different spoken phonemes in fine angular resolution over a wide frequency range.
After the introduction, this expose presents the state of research with other studies in this field and the methods used in this work. This report ends up with the intended time table related to each work package of this study.

Contents

1 Introduction

The human voice plays a major role in interpersonal communication. Even in times of increasing digitalization, this remains the case, since many technologies use the human voice as an input signal (e.g. Internet-based intelligent personal assistants). A key property of the voice is its directionality, as it influences the energy distribution in the room and at the ear of the receiver. By varying the intensity, vocal tract position and mouth shape, the directivity and thus the propagation of the sound field emitted by the speaker also changes. Research on this topic is of great relevance for architectural design, the development of new technologies in the field of telecommunications or the modelling of speech [2, 3, 4]. The measurement of the sound field at specific angles around the head of a test person is a common research topic. The angular resolution is given by the number of microphones used and therefore not very fine (e.g. 7.5 degrees [5]). In addition, the measurement with microphone arrays is complex and the reproducibility can be problematic when measuring with human beings. As a result, there are not enough measurement data in fine angular resolution over the entire audible frequency range.

Research on Head Related Transfer Functions (HRTFs) has shown that it is possible to simulate propagating sound fields by numerical simulation using the boundary element method (BEM) [6]. Thereby, a mesh generated from a surface scan of the head can be generated. BEM provides robust HRTF simulations by acoustic reciprocity. This approach was validated by comparative measurements on the artificial head.

The aim of this master thesis is to apply the procedure for the simulation of HRTFs to the calculation of the radiation pattern of the human voice. Therefore a mesh of the artificial head FABIAN will be extended by providing various mouth shapes for different phonemes. Using the software MESH2HRTF and BEM the propagating sound field for the different mouth shapes will be calculated [1].

Research questions of this thesis are: How do different mouth shapes of different phonemes affect the simulated directional characteristic of the human voice? To what extent can the speech directivity be validly simulated with the simplified assumption of rigid boundaries and conphase moving radiation areas?

2 State of the Art

The investigation of the directivity of human speech is of high relevance for different fields of research such as telecommunications, building and room acoustics, musicology and media applications with speech simulation, to name but a few. For this reason, there are numerous publications dealing with this topic. A possible subdivision here would be into papers that measure the radiated sound field with microphone arrays at different measuring points around a test person and papers that model or simulate the sound field. Furthermore, there are studies comparing both approaches for the validation of used methods. In the following, some relevant work is described in order to shed light on the state of the art in research on the directional characteristic of the human voice.

Various research projects are dedicated to the measurement of the radiated sound field of a test person with microphones equidistantly arranged in an arc of 180 degrees. The angular resolution is determined by the number of microphones used. This measurement setup makes measurements in the horizontal plane possible. In some work, the arc with the measuring microphones can also be changed in elevation so that the sound propagation can also be recorded for the vertical plane. As speech material either vocals, spoken language or individually held phonemes are used. The examined phonemes can be divided into vowels (e.g. /a/, /e/, /i/, /o/, /u/), nasal (e.g. /m/, /n/) and fricative (e.g. /s/, /sh/, /f/, /th/, /ch/) consonants.

In 2006, Katz et al. [5] published the results that the directivity of the voice differs for vowels in the middle frequency range at 1000 Hz and 1600 Hz. The vowel /a/ is more directional than /o/. The lowest directionality in this investigation is found in the vowel /i/. The nasal consonants show a similar radiation behaviour over the whole frequency range. The fricative consonants /f/, /ch/, /s/ differ in their directionality in the middle and high frequency bands, with /f/ having the narrowest lateral projection. The characteristics of the phonemes generally differ in the ranges 630-1250 Hz, 2500-3125 Hz. In a subsequent study, the change in mouth geometry is cited as the reason for the spectral differences for all vowels and thus the frequency-dependent directivity [4]. However, the difference in the radiation characteristics of the individual vowels cannot be confirmed here. Below 600-1000 Hz the radiation pattern of the voice is omnidirectional, the greatest differences in directionality are found at 800-1000 Hz, 2500 Hz, 4000 Hz. To evaluate the perceived directionality, the spectrum should be evaluated as a weighting with the radiation patterns (combined spatial-frequency analysis). The spectrum also shows a shift of energy to higher frequencies with increased intensity.

Kocon et al. [7] will take up the question in 2018 and investigate the radiation characteristics of the human voice with regard to the individual phonemes. In running speech, the vowel /a/ exhibits the greatest directionality. According to the study, the reason for this is that the mouth opening and vocal tract of this vowel has the least change compared to the average mouth and vocal tract position. For individually spoken vowels, the directionality differs for third-octave band weighting above the frequency 1000 Hz with the greatest difference in the frequency band of the center frequency 4000 Hz. The Directivity Indexes (DI) are: /a/ = 3.9 dB, /i/ = 3.3 dB, /e/ = 3.1 dB, /u/ = 2.9 dB, /o/ = 2.8 dB. The results of the study support the findings of Katz et al. from 2006.

Monson et al. [8] underline the relevance of the frequency bands 8000 Hz and 16000 Hz and the so-called High-Frequency Energy (HFE). They come to the conclusion that there is no significant difference in the radiation characteristics of singing and speech. When looking at the individual fricative phonemes /s/, /ʃ/, /f/, /θ/ they differ in the patterns above 2 kHz with the largest difference of 9 dB at 8000 Hz between the phonemes /s/ and /th/ in a direction of 90 degrees. The calculated DI are /s/ = 5.3 dB, /ʃ/ = 4.6 dB, /f/ = 3.2 dB and /θ/ = 2.5 dB. In general, the study shows how decisively the shape of the mouth affects the radiation of the voice and that attention should be paid to the HFE, since all the detected differences are located in this frequency range [9].
The studies described are representative of work where the propagating sound field is measured with a limited number of microphones. This approach gives a good indication of the directivity of the human voice. However, it cannot provide fully spherical data for fine angular resolution.
Besides measuring with microphones, it is also possible to simulate a propagating sound field. This requires a mesh model that can be generated by surface scans. Using numerical methods, differential equations can be solved to calculate sound pressure level and velocity for different points in space. The effects of geometry on simulated results is addressed by the research work of Arnela et al. [10]. The mesh used in this study is a head model without torso, whereas a model with torso allows more valid simulations in this pending master thesis. The effect of simplifying the geometry of the head on the propagation of the sound field of the human voice is investigated with the use of a Finite Element Method (FEM) in the time domain. A Gaussian pulse is used as a stimulus signal, which is impressed on on a planar surface at the lower end of the vocal tract at the position of the vocal folds. The boundary conditions are calculated for the glottis from the sound velocity of the exiting air. For the vocal tract, sound-soft conditions are chosen frequency independent. For the surface of the head Neumann conditions of perfectly rigid boundaries are applied and the environment is modeled with non-reflecting Sommerfeld boundary conditions. The authors conclude that head attachments such as ears and nose are perceptually irrelevant. The exception are the lips, which have an effect on the radiation impedance in a similar manner as for brass instruments that present an exponentially open bell, and vocal tract transition functions and play a significant role in the frequency range of 5000 Hz-10000 Hz. For a modeling above 5000 Hz the lips should therefore be considered.
In a subsequent study, the frequency range in which the lips are relevant for the radiation pattern was examined for the vowels /a/, /i/ and /u/ [11]. For the vowel /a/, the lips are also relevant for the frequency range below 5000 Hz due to the large mouth opening. This is not the case for the vowel /u/. The results for vowel /i/ lie between those for /a/ and /u/. All in all, the modelling of the lips also affects the simulation of the radiation of the vowels in the low frequency range. Unfortunately, the results of these studies do not allow a thorough comparison of the directional characteristic of the individual vowels with the presented work with microphone measurements. However, they can be used to validate the simulation (see chapter 3).

This master thesis ties in with several points of the presented studies. The radiation pattern of the human voice for different phonemes is simulated. For this purpose the mesh of the FABIAN artificial head with torso is used, which is extended in Blender by different mouth shapes for different phonemes [12]. With MESH2HRTF and the numerical method BEM the emitted sound field is subsequently calculated in very high angular resolution and fully spherical over the entire frequency range up to 22000 Hz [1]. Thus the influence of different mouth shapes on the sound propagation is investigated. The results can be compared with the results of research work with microphone measurements. Furthermore, data of the propagating sound field of the different phonemes will be provided in the form of impulse responses for high angular resolution. These can be used for musical acoustics, musicology, and room acoustic simulation in future research or applications.

3 Methods

3.1 Overview

The research work consists of the three subtasks mesh generation, BEM computation and validation and analysis. In mesh generation, a mesh of the head, preferably with a torso, has to be extended by different mouth images. This is done in the software Blender [12]. Subsequently, it is necessary to remesh the mesh in order to meet the requirements for numerical simulation. For this step a remeshing software like Meshmixer is used [13]. With MESH2HRTF the numerical BEM simulation is performed [1]. The processing and comparison of the calculated data is finally done in Matlab (see figure 1) [14]. To validate the procedure, results for vowel /a/ are compared with data from Arnela et al. [11]. Afterwards the radiation patterns for different phonemes are simulated and compared to each other.

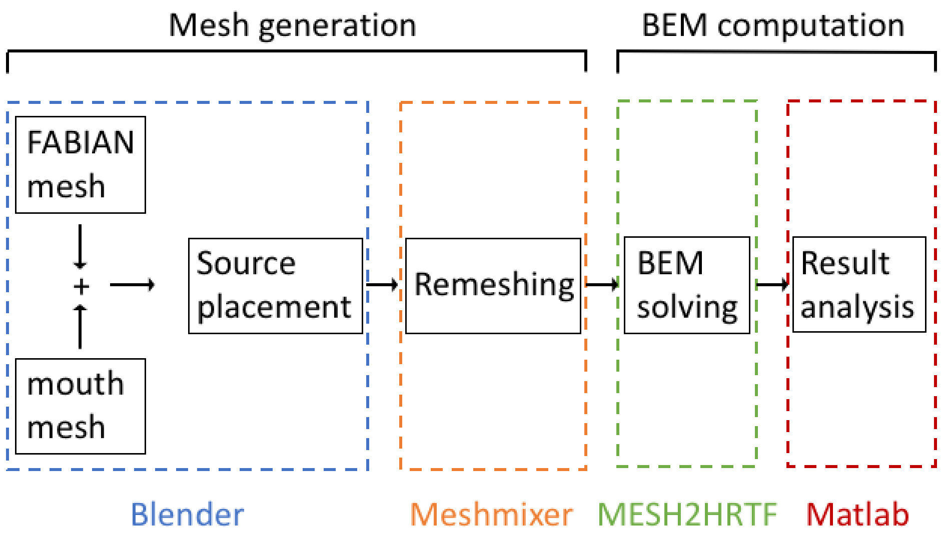

Figure 1: Construction of the general pipeline to calculate the effect of a mouth shape on the radiation characteristics of the voice. Assignment of the work steps to the software Blender, Meshmixer, MESH2HRTF, Matlab.

3.2 Mesh generation

For the work, one mesh is required for each of the phonemes examined. This consists of a head-torso mesh and the specific viseme of the phoneme. A collection of phonemes can be found in the International Phonetic Alphabet (IPA chart) of the International Phonetic Association [15]. For the generation of the viseme meshes existing models are used as far as possible. For the phonemes /a/, /ae/, /e/, /i/, /o/, /oe/, /u/, /y/ meshes are already available, which also include the vocal tract [16]. From a work from 2017 a triangle mesh of the FABIAN artificial head with torso with an edge length of 2 mm for the pinna and 5 mm for head and torso exists as a Standard Triangulation/Tesselation Language (.stl) file [6]. In this master thesis the different visemes with the FABIAN mesh are combined with the rendering software Blender (see figure 2) [12]. In the resulting models an vibrating plane in the oral cavity is defined as the sound source for the BEM simulation. Subsequently, the resulting models have to be remeshed to meet the requirement of 6 elements per wavelength for the frequency range up to 22000 Hz [17]. For this purpose a remeshing software like Meshmixer can be used [13].

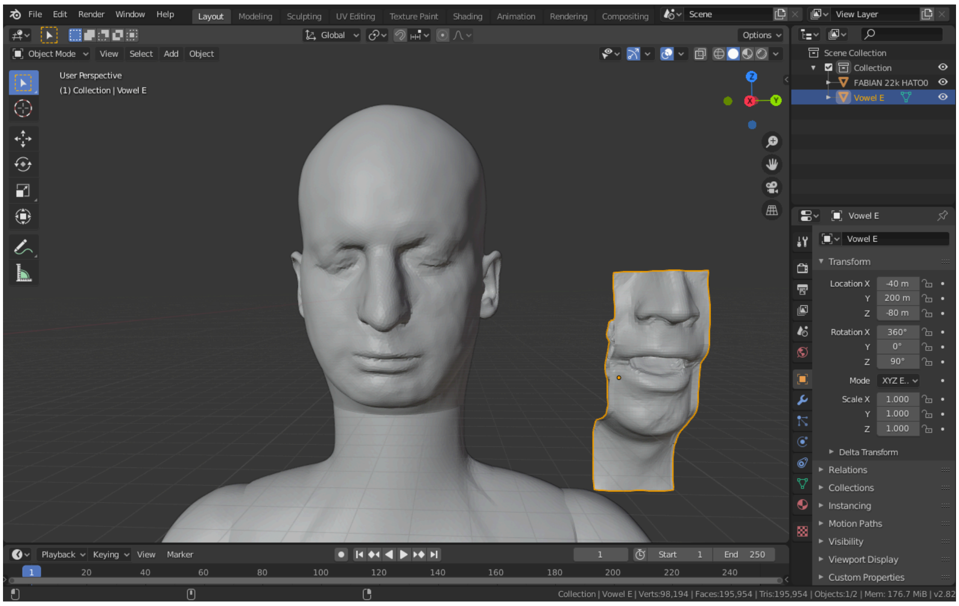

Figure 2: Mesh of the FABIAN artificial head and torso with mesh of the mouth position for the vowel /e/ (framed in yellow) in the rendering software Blender.

3.3 BEM computation

The simulation of the propagating sound field is done with the software package MESH2HRTF. This software, published in 2015 by Ziegelwanger et al. [1], solves the differential equation of the Helmholtz equation and provides sound pressure level and velocity for arbitrary points in space. The algorithm uses a 3-dimensional Burton-Miller collocation Boundary Element Method (BEM) coupled with a Multi-Level Fast Multipole Method (ML-FMM). It applies rigid boundaries with Dirichlet conditions for sound-soft surfaces and Neumann conditions for sound-hard surfaces. As input data MESH2HRTF reads the geometry data of the mesh and provides the results of the numerical simulation as Spatially Oriented Format for Acoustics (SOFA). This format is

standardized and allows a good processing of the data in MATLAB.
The same approach has been successfully applied in the past in research on Head Related Transfer Function (HRTF) simulation. In a work from 2017 [6] a mesh model of the FABIAN artificial head with different resolutions is generated by scans and HRTFs are calculated by numerical simulation with BEM and Fast Multipole Method (FMM). Measurements of the HRTFs validate this procedure.
To calculate the sound field, a vibrating plane is defined as the source. The appropriate position of this plane is determined in advance, whereby 3 different positions in the oral cavity are tested in relation to the reference surface of the mouth opening.

3.4 Validation and analysis

To validate the procedure of the entire simulation, a simulation with 3 different positions of the source plane in the oral cavity is carried out in advance for the vowel /a/ and compared with measurements from the research work of Arnela et al in 2016 [11]. In this thesis the sound propagation for the vowels /a/, /i/ and /u/ was simulated with the Finite Element Method (FEM) for a head model with lips, vocal tract and glottis (see chapter 2). A realistic geometry of the vocal tract is used. In contrast, this master thesis is limited to geometry of the oral cavity. To overcome this difference, the spectrum of the source is adjusted to match the signal delivered by Arnela in the plane of the lips. After this step, the radiation patterns of the different phonemes are simulated and differences can be investigated.

4 Time Schedule

The target time frame for the Master's thesis is 6 months from April to September. The individual work steps (see chapter 3, section 3.1) can be assigned to the categories research, mesh generation, pipeline programming, simulation, evaluation of the collected data and writing the master thesis (see figure 3). The category "Literature research" takes the largest share with 4 months and starts immediately. Thus a good knowledge base is built up and new information is acquired during the other categories. When organising the time for these categories, care is taken to ensure that work that builds on each other is dealt with one after the other and that not too many areas are covered at the same time. The timetable set up in this way serves as an orientation and is not necessarily mandatory.

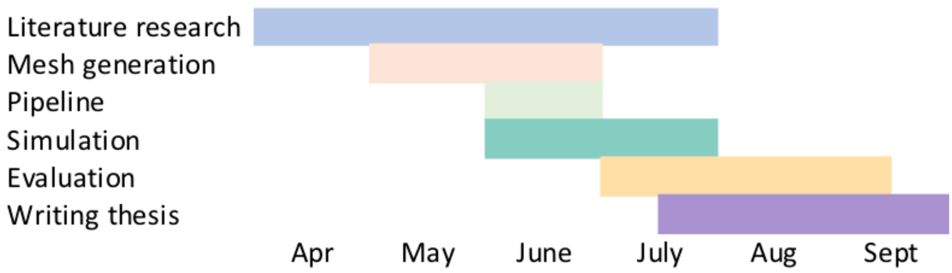

Figure 3: Intended time schedule with the work packages Literature research, Mesh generation, Pipeline, Simulation, Evaluation and Writing thesis with time frame from the beginning of April to the end of September.

References

[1] Harald Ziegelwanger, Wolfgang Kreuzer, and Piotr Majdak. "Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions". In: *22nd International Congress on Sound and Vibration*. 2015.

[2] Wing Tin Chu and ACC Warnock. "Detailed directivity of sound fields around human talkers". In: (2002).

[3] Teemu Halkosaari, Markus Vaalgamaa, and Matti Karjalainen. "Directivity of artificial and human speech". In: *Journal of the Audio Engineering Society* 53.7/8 (2005), pp. 620–631.

[4] Brian Katz and Christophe d'Alessandro. "Directivity measurements of the singing voice". In: 2007.

[5] B Katz, Fabien Prezat, and Christophe d'Alessandro. "Human voice phoneme directivity pattern measurements". In: *4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan*. 2006, p. 3359.

[6] Fabian Brinkmann et al. "A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations". In: *Journal of the Audio Engineering Society* 65.10 (2017). Available Open Access publishedVersion at https://depositonce.tu-berlin.de/handle/11303/9779, pp. 841–848. DOI: 10.17743/jaes.2017.0033. URL: https://doi.org/10.17743/jaes.2017.0033.

[7] Paulina Kocon and Brian B Monson. "Horizontal directivity patterns differ between vowels extracted from running speech". In: *The Journal of the Acoustical Society of America* 144.1 (2018), EL7–EL12.

[8] Brian B Monson, Andrew J Lotto, and Sten Ternström. "Detection of high-frequency energy changes in sustained vowels produced by singers". In: *The Journal of the Acoustical Society of America* 129.4 (2011), pp. 2263–2268.

[9] Brian B Monson, Eric J Hunter, and Brad H Story. "Horizontal directivity of low- and high-frequency energy in speech and singing". In: *The Journal of the Acoustical Society of America* 132.1 (2012), pp. 433–441.

[10] Marc Arnela, Oriol Guasch, and Francesc Alías. "Effects of head geometry simplifications on acoustic radiation of vowel sounds based on time-domain finite-element simulations". In: *The Journal of the Acoustical Society of America* 134.4 (2013), pp. 2946–2954.

[11] Marc Arnela et al. "Influence of lips on the production of vowels based on finite element simulations and experiments". In: *The Journal of the Acoustical Society of America* 139.5 (2016), pp. 2852–2859.

[12] *Blender Software*. https://www.blender.org. Viewed on 13.04.2020. URL: https://www.blender.org.

[13] *Autodesk Meshmixer*. http://www.meshmixer.com. Viewed on 16.04.2020. URL: http://www.meshmixer.com.

[14] *MathWorks MATLAB*. https://de.mathworks.com/products/matlab.html. Viewed on 16.04.2020. URL: https://de.mathworks.com/products/matlab.html.

[15] *International Phonetic Association*. https://www.internationalphoneticassociation.org. Viewed on 16.04.2020. URL: https://www.internationalphoneticassociation.org.

[16] *Speech & Math*. http://speech.math.aalto.fi/data.html. Viewed on 16.04.2020. URL: http://speech.math.aalto.fi/data.html.

[17] Robert D Ciskowski and Carlos Alberto Brebbia. *Boundary element methods in acoustics*. Springer, 1991.