




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

Master thesis

**Numerical simulation
of voice directivity patterns
for different phonemes**

Leif Johannsen

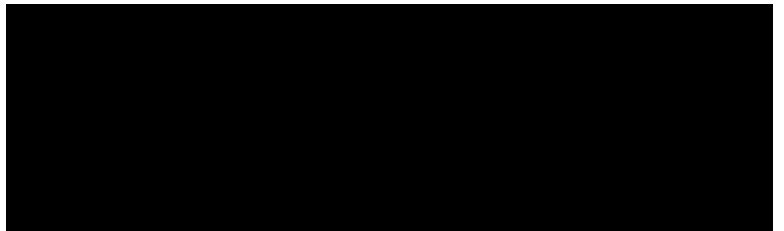
A solid black rectangular box used to redact information, likely a student ID or contact details.

Supervisors: Prof. Dr. Stefan Weinzierl
Dr. Fabian Brinkmann
Dr. Paul Luizard

Dass diese Arbeit in vorliegender Form entstehen konnte, verdanke ich vielen Menschen, die ihren individuellen Teil dazu beitrugen. Besonders danke ich den Dozierenden, die meine Neugier geweckt und mir das nötige Werkzeug an die Hand gegeben haben. Gleichmaßen bedanke mich bei meinen Betreuern Dr. Fabian Brinkmann und Dr. Paul Luizard für die besonders gute Zusammenarbeit, die sowohl durch fachliche Kompetenz als auch Offenheit geprägt war.

Hereby I declare that I wrote this thesis myself with the help of no more than the mentioned literature and auxiliary means.

Ulm, 13.03.2021



Abstract

The master thesis with the title “Numerical simulation of voice directivity patterns for different phonemes” deals with the effect of the mouth shapes related to different phonemes on the radiated sound field and thus directional characteristic of the human voice. The Boundary Element Method (BEM) is used as numerical simulation. This yields directional characteristics with high angular resolution. The FABIAN artificial head is used as the basic model as a scanned three-dimensional mesh. In combination with different vocal tract configurations and mouth shapes, mesh models are generated for different phonemes. Radiated sound fields and directivity patterns are calculated from these models. This makes it possible to compare realistic directional characteristics for different spoken phonemes in fine angular resolution over a wide frequency range. This work provides 3D models of the FABIAN artificial head with torso for eight different phonemes with different vocal tract configurations. The simulated transfer functions are also published.

The simulations of the models without vocal tract show a relatively straight frequency response in front of the mouth opening. For the models with vocal tracts, the formants are clearly visible in the spectra and match the position of the formants from corresponding microphone recordings.

With regard to sound propagation, the simulations show omnidirectional behavior up to 500 Hz and above that an increase in directionality with increasing frequency. Directionality is higher for models with larger mouth opening. The vocal tract has the effect of slightly shifting the radiation patterns. The main direction of radiation is frontal and slightly below the median plane for all models. When comparing the Directivity Indexes (DI) for the different models, the vocal tract causes a higher directivity, which is already visible below 1000 Hz.

Zusammenfassung

Die Masterarbeit mit dem Titel “Numerical simulation of voice directivity patterns for different phonemes” befasst sich mit dem Einfluss der Mundformen verschiedener Phoneme auf das abgestrahlte Schallfeld und damit auf die Richtcharakteristik der menschlichen Stimme. Zur numerischen Simulation wird die Boundary-Element-Methode (BEM) verwendet. Dies ermöglicht die Berechnung von Richtcharakteristiken mit hoher Winkelauflösung. Als Grundmodell dient der FABIAN-Kunstkopf als gescanntes Oberflächennetz. In Kombination mit verschiedenen Vokaltraktkonfigurationen und Mundformen werden Netzmodelle für verschiedene Phoneme erstellt. Mit diesen Modellen lassen sich abgestrahlte Schallfelder und Richtcharakteristiken berechnen. Damit ist es möglich, die Richtcharakteristik für verschiedene gesprochene Phoneme in feiner Winkelauflösung über einen weiten Frequenzbereich realistisch zu vergleichen.

Durch die Arbeit werden 3D Modelle des FABIAN Kunstkopfes mit Torso für acht verschiedene Phoneme mit verschiedenen Vokaltraktkonfigurationen und die simulierten Übertragungsfunktionen bereitgestellt.

Die Simulationen der Modelle ohne Vokaltrakt weisen einen relativ graden Frequenzgang vor der Mundöffnung auf. Bei den Modellen mit Vokaltrakt sind die Formanten in den Spektren gut sichtbar und stimmen in ihrer Position mit den Formanten aus dazugehörigen Mikrofonaufnahmen überein.

Im Hinblick auf die Schallausbreitung zeigen die Simulationen ein omnidirektionales Verhalten bis 500 Hz und darüber eine Zunahme der Direktionalität mit steigender Frequenz. Die Direktionalität ist für Modelle mit größerer Mundöffnung höher. Der Vokaltrakt wirkt sich hierbei als leichte Verschiebung der Abstrahlcharakteristiken aus. Die Hauptabstrahlrichtung ist für alle Modelle frontal und leicht unterhalb der Medianebene. Bei der Gegenüberstellung der Bündelungsmaße für die verschiedenen Modelle bewirkt der Vokaltrakt eine höhere Richtwirkung, die bereits unterhalb von 1000 Hz sichtbar ist.

Contents

1. Introduction	10
2. State of the Art	11
3. Methods	14
3.1. Overview	14
3.2. Mesh generation	14
3.2.1. Overview	14
3.2.2. Adjustment of the FABIAN jaw position	16
3.2.3. Combination of mouth models and FABIAN model	19
3.2.4. Separation and deletion of the vocal tract	22
3.2.5. Model positioning	24
3.3. Mesh postprocessing and re-meshing	24
3.4. BEM computation	30
3.4.1. Sample grids	30
3.4.2. Mesh2HRTF	32
4. Results	35
4.1. General results	35
4.2. Spectral results	37
4.3. Results of directivity	43
4.4. Error estimation	63
5. Discussion	65
A. Digital appendix	67

List of Figures

1.	General pipeline with the necessary work steps	14
2.	Mesh of the FABIAN HATS and mesh of vowel /e/ in Blender	15
3.	Amature positioning for jaw movement	16
4.	Weight map to determine the influence of the amature movement	17
5.	Positioning of the vowel /e/ in relation to the FABIAN mesh	20
6.	Mesh of phonemes /i/ in the opening of the FABIAN mesh	21
7.	Mesh of phonemes /a/ after smoothing	22
8.	Mesh of vowel /e/ with positioned ellipsoid	23
9.	Analysis of the 3D-Print Toolbox for Vowel /ae/	25
10.	Remeshing in Meshmixer	26
11.	Comparison of all remeshed wvt models	29
12.	Comparison of all remeshed wovt models	29
13.	Determining the position of the sampling grid for vowel /e/ with positioning marker	31
14.	Visualization of the sample grids and the mesh for vowel /a/	32
15.	Source definition wovt model vowel /u/	33
16.	Computation time for simulations	36
17.	Spectra of wvt models up to 16000 Hz	38
18.	Spectra of wovt models up to 16000 Hz	40
19.	Formant spectra from recordings and simulations up to 4 kHz	42
20.	Fully spherical sound propagation vowel /i/, 100 Hz to 500 Hz	44
21.	Fully spherical sound propagation vowel /i/, 630 Hz to 1600 Hz	45
22.	Fully spherical sound propagation vowel /i/, 2000 Hz to 5000 Hz	46
23.	Fully spherical sound propagation vowel /i/, 6300 Hz to 12500 Hz	47
24.	Polar diagrams of the vowels /ae/, /e/, /a/, /i/, third-band averaged, 100 Hz to 400 Hz	50
25.	Polar diagrams of the vowels /oe/, /u/, /o/, /y/, third-band averaged, 100 Hz to 400 Hz	51
26.	Polar diagrams of the vowels /ae/, /e/, /a/, /i/, third-band averaged, 500 Hz to 1000 Hz	52
27.	Polar diagrams of the vowels /oe/, /u/, /o/, /y/, third-band averaged, 500 Hz to 1000 Hz	53
28.	Polar diagrams of the vowels /ae/, /e/, /a/, /i/, third-band averaged, 1250 Hz to 2500 Hz	54
29.	Polar diagrams of the vowels /oe/, /u/, /o/, /y/, third-band averaged, 1250 Hz to 2500 Hz	55
30.	Polar diagrams of the vowels /ae/, /e/, /a/, /i/, third-band averaged, 3150 Hz to 6300 Hz	56
31.	Polar diagrams of the vowels /oe/, /u/, /o/, /y/, third-band averaged, 3150 Hz to 6300 Hz	57
32.	Polar diagrams of the vowels /ae/, /e/, /a/, /i/, third-band averaged, 8000 Hz to 12500 Hz	58

33.	Polar diagrams of the vowels /oe/, /u/, /o/, /y/, third-band averaged, 8000 Hz to 10000 Hz	59
34.	Test-retest deviation for wvt simulations /i/ and /oe/	63
35.	Spectral influence of sample grid positioning, wvt model vowel /a/	64

List of Tables

1.	Measured lip spacing of phonemes and required jaw openings	18
2.	Positioning of the phoneme meshes	19
3.	Shift of the models for alignment to the coordinate origin	24
4.	Used remeshing settings for Meshmixer	27
5.	Used remeshing settings for pmp-library	27
6.	Used remeshing software of the different models	28
7.	Positioning of the sample grid at the mouth opening	31
8.	Export setting of Mesh2HRTF with Blender	34
9.	Frequencies without iterative BEM solution for the different models . . .	35
10.	Third octave band averaged sound pressure levels of the wvt models . . .	39
11.	DI of wvt models	61
12.	DI of wovt models	62

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 surrounding space and at the ear of the receiver. By varying parameters such as 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 [1, 2, 3]. 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 in most cases (e.g. 7.5 degrees [4]). In addition, the measurement with microphone arrays is complex and the reproducibility can be problematic when measuring the voice directivity of 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) [5]. Thereby, a mesh generated from a surface scan of the head can be produced. 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, in addition to the related vocal tract configurations. Using the software Mesh2HRTF and BEM the propagating sound field for the different phonemes will be calculated [6].

Research questions of this thesis are: How do different mouth shapes and vocal tract configurations related to different phonemes affect the simulated directional characteristic of the human voice? How great is the effect of the vocal tract on the propagation of the sound field? 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. It would be possible to divide the existing scientific works into two parts: 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, the 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 papers, the arc with the measuring microphones can also be changed in elevation so that the sound propagation can also be recorded for the vertical planes. As speech material either vowels, 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. [4] 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 to 1250 Hz, 2500 to 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 [3]. However, differences in the radiation characteristics of the individual vowels cannot be confirmed in this research. Below 600 Hz the radiation pattern of the voice is almost omnidirectional, the greatest differences in directionality are found at 800 to 1000 Hz, 2500 Hz and 4000 Hz. Since the radiation patterns in high-energy regions of the spectrum have a greater effect on the perceived directionality, both the spectra and the radiation patterns must be evaluated (combined spatial-frequency analysis). The spectrum also shows a shift of energy to higher frequencies with increased intensity.

Kocon et al. [7] took up the question in 2018. Their study investigates the radiation characteristics of the human voice with regard to the individual phonemes. In running speech, the vowel /a/ exhibits the greatest directionality. 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 of sound sources compared to an omnidirectional characteristic can be expressed as Directivity Index (DI) in dB. The DI for an omnidirectional characteristic is

0 dB and higher values describe a greater directivity. The DIs for the examined vowels 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 /θ/ in a direction of 90 degrees to the right. The calculated DIs 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 is challenging to 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. 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 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, vocal tract transfer functions and play a significant role in the frequency range of 5000 Hz to 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 is based on several points of the presented studies. The radiation pattern of the human voice for different phonemes is simulated. For this purpose the FABIAN head-and-torso mesh 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 [6]. Thus, the influence of different mouth shapes on the sound propagation is investigated. The results can be compared with the results of previous studies 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 data generation consists of the three subtasks mesh generation, remeshing and BEM computation. In mesh generation, a head-and-torso mesh is extended by different mouth and vocal tract configurations. 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 the remeshing software Meshmixer [13] and the Polygon Mesh Processing Library (pmp-library) [14] are used. The numerical BEM simulation is performed with Mesh2HRTF [6]. Finally, the processing and comparison of the calculated data is done in Matlab (see figure 1) [15].

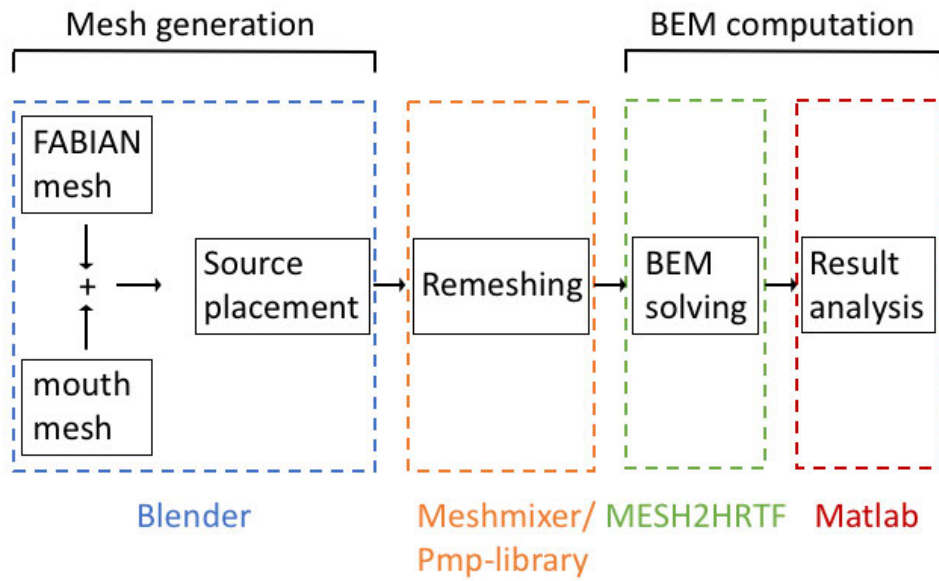


Figure 1: General pipeline to calculate transfer functions. Assignment of the work steps to the software Blender, Meshmixer, pmp-library, Mesh2HRTF, Matlab.

3.2. Mesh generation

3.2.1. Overview

For this thesis, one mesh is required for each of the phonemes examined. This consists of a head-and-torso mesh and the specific mouth and vocal tract configuration of the phoneme. A collection of phonemes can be found in the International Phonetic Alphabet (IPA chart) of the International Phonetic Association [16]. For the generation of the meshes of the mouth shapes existing models are used. For the Finnish vowels /a/, /ae/,

/e/, /i/, /o/, /oe/, /u/, /y/ meshes are already available, which also include the vocal tract [17]. From a study from 2017 a triangle mesh of the FABIAN artificial head-and-torso simulator (HATS) with an edge length of 2 mm for the pinna and 5 mm for head and torso exists as a Standard Triangulation/Tessellation Language (.stl) file [5]. In this master thesis the different mouth and vocal tract meshes are combined with the FABIAN mesh using the rendering software Blender (see figure 2) [12]. In the resulting models a vibrating plane is defined as the sound source for the BEM simulation. All models should be available with vocal tract (wvt) and without vocal tract (wovt) to determine the effect of the vocal tract on the propagation of the sound field. Subsequently, the resulting models have to be remeshed to meet the requirement of six elements per wavelength for the frequency range up to 22000 Hz [18]. For this purpose a remeshing software like Meshmixer is used [13].

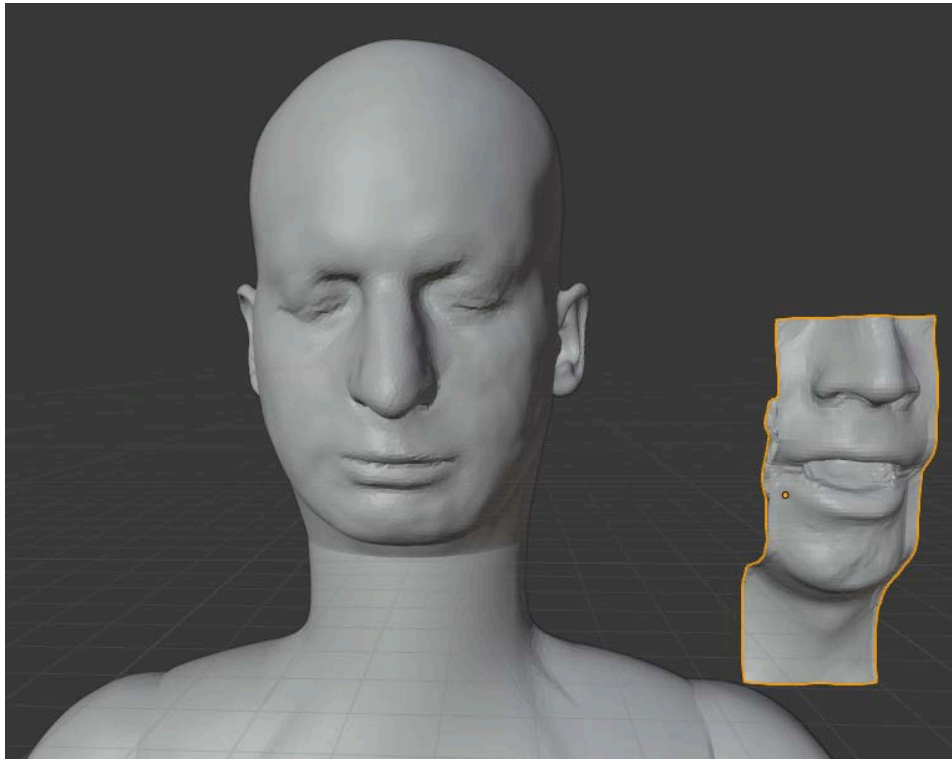


Figure 2: Mesh of the FABIAN HATS with mesh of the mouth and vocal tract mesh for the vowel /e/ (framed in yellow) in the rendering software Blender.

3.2.2. Adjustment of the FABIAN jaw position

In order to connect the models of the individual phonemes with the model of the FABIAN artificial head, its jaw position must be able to be varied. In humans, the temporomandibular joint (TMJ), as the connection of the temporal bone and the mandible, allows the lower jaw to move against the upper jaw. Thus, in this mesh work, the goal is to allow the part of the lower jaw to rotate downward around the TMJ. However, the upper lip should remain in place.

In the first step, the mesh of the FABIAN HATS is imported into Blender as a .stl file. The positioning remains unchanged afterwards. This orientation of the model is shown in Blender as neither moved nor rotated.

An amature is now added in Object Mode to enable the movement of the jaw section. Amatures consist of the three parts head, body and tail and are used in Blender to transfer movements to a mesh. As the movement in rotations refers to the position of the head of the amature, it is positioned at the TMJ of the FABIAN. The Body and tail point towards the chin (see figure 3).

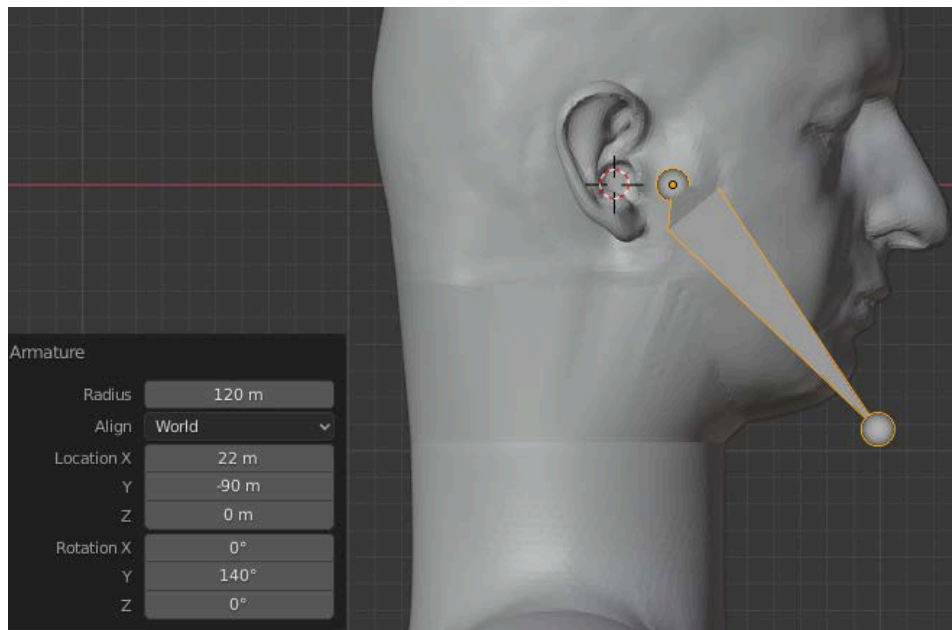


Figure 3: An amature is added at the position of the TMJ to later allow movement of the lower jaw around this position.

In the next step the mesh of the artificial head is assigned to the amature. This is called parenting. In Object Mode the artificial head is selected first and then the amature is selected while shift is pressed. Now the option “amature deform with empty groups” is selected. This adds a new vertex group to the head, called “Bone”. This must be deleted

and the existing group renamed to “Bone”. In Edit Mode all options must be selected in the Modifier tab so that the movements of the amature will later affect the vertices of the mesh in Edit Mode.

In order to determine the influence of the movement of the amature on the mesh of the FABIAN artificial head, a weight map is created in Weight Paint Mode which assigns factors in the interval $[0, 1]$ to the areas of the model. Areas with a low factor are hardly affected by the movement of the amature, while areas with a high factor are greatly affected. To create a symmetrical weight map, the view in Blender is centered on the y-axis and the gradient is used (see figure 4).

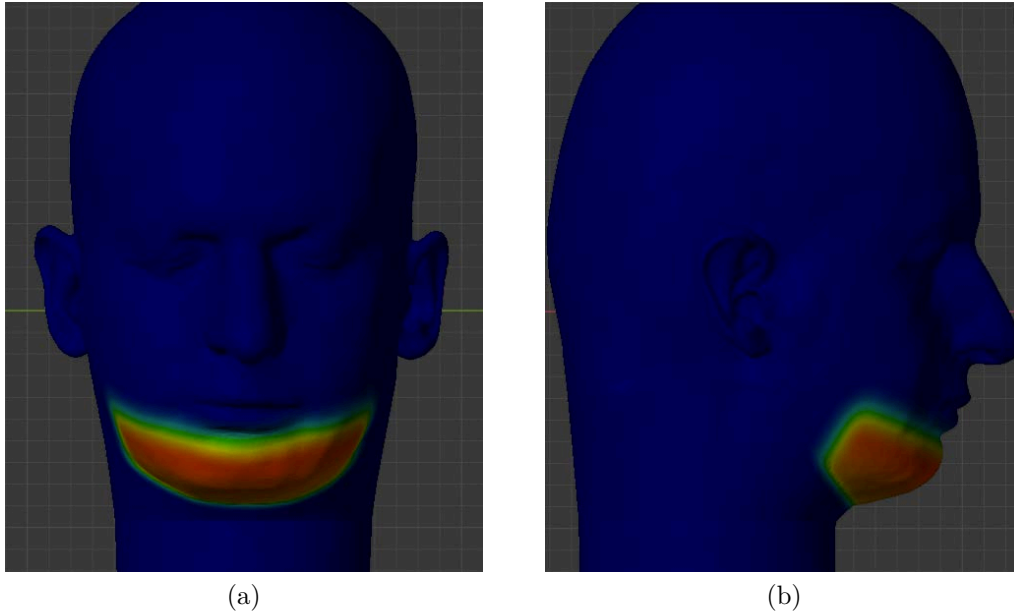


Figure 4: Weight map in Weight Paint Mode to determine the effect of the amature movement on the mesh of the FABIAN artificial head in x-view (a) and y-view (b). Blue areas are not affected by the amature movement, whereas red areas are strongly affected.

The jaw opening of the FABIAN model created in this way can be measured by the edge lengths of the modified faces between the upper and lower lip. The mouth openings of the phonemes are measured as the maximum vertical distance between the upper and lower lip. For the vowels /ae/, /e/, /a/ and /i/ these values can be directly transferred to the required width of the jaw opening of the FABIAN. For the fitting of the remaining models it is not possible to transfer the lip distances to the jaw opening, since in the vowels with a more pointed lip image the vertical distances between the lips are small, but the jaw openings are relatively large. Thus, the jaw positions for the vowels /oe/, /u/, /o/ and /y/ are determined manually, whereby a jaw position is selected so that the upper and lower lips of the FABIAN model are as congruent as possible with the

lips of the respective models of the phonemes (see table 1). The jaw distance has no retroactive effect on the lip distance of the models, but makes it possible to combine the individual mouth and vocal tract models of the phonemes with the FABIAN artificial head.

Vowel	Vertical lip distance	FABIAN jaw opening
/oe/	1.0 cm	1.2 cm
/u/	0.5 cm	1.2 cm
/o/	0.4 cm	0.9 cm
/y/	0.4 cm	1.0 cm
/æ/	2.9 cm	
/e/	1.5 cm	
/a/	1.2 cm	
/i/	1.0 cm	

Table 1: Comparison of the measured lip distances of the individual phonemes and the required FABIAN jaw openings.

3.2.3. Combination of mouth models and FABIAN model

In the previous subchapter (see subchapter 3.2.2) the preparation of the FABIAN mesh was described to enable the combination with the models of the individual phonemes. In this subchapter the steps are described which are necessary to generate an individual FABIAN mesh in combination with each phoneme.

In the first step, in Blender the corresponding phoneme mesh is imported as .stl file to the FABIAN mesh with specific jaw construction. As the models of the phonemes used have a different orientation on the coordinate system and are oriented differently in space, they must be aligned according to the position of the FABIAN mesh. In order to be as congruent as possible with the FABIAN mesh in respect of the lips, they must also be tilted downwards (Rotation X) depending on the mouth image (see figure 5). The phoneme models are not shifted or rotated in the Y-direction. The reorientations result in an individual positioning for each phoneme (see table 2). The angle of inclination (Rotation X) is taken into account in the discussion of the simulated results.

Vowel	Location X	Location Z	Rotation X	Rotation Z
/ae/	-3.3 cm	-4.3 cm	17°	90°
/e/	-2.8 cm	-4.9 cm	13°	90°
/a/	-3.2 cm	-4.7 cm	15°	90°
/i/	-2.6 cm	-5.4 cm	10°	90°
/oe/	-3.1 cm	-5.0 cm	9°	90°
/u/	-3.7 cm	-5.5 cm	8°	90°
/o/	-3.2 cm	-6.0 cm	4°	90°
/y/	-3.2 cm	-6.0 cm	5°	90°

Table 2: Positioning of the phoneme meshes for highest possible congruence with the respective FABIAN mesh with specific jaw position. A shift or rotation in Y-direction is not necessary.

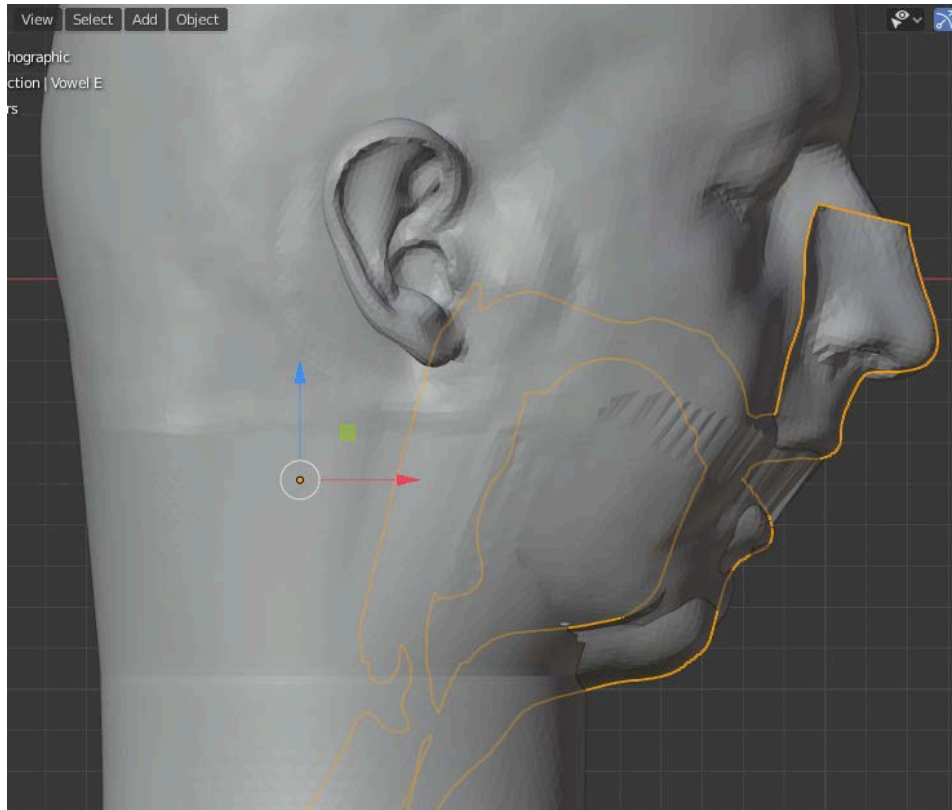


Figure 5: Positioning vowel /e/ with as much congruence as possible with the FABIAN mesh in respect of the lips. The mouth image was rotated by 13 degree in the downward direction (Rotation X).

The next step is to select and remove the unneeded nose and chin areas around the lips for the phoneme model. It is ensured that any remaining vertices left over by this procedure without connection to the mesh are subsequently deleted.

Then, the mouth area can be selected from the Fabian mesh and deleted. Both the trimmed model of the phoneme and the FABIAN mesh can be displayed simultaneously so that the cut-out in the Fabian mesh can be dimensioned correctly. When reworking, care must be taken not to damage the Fabian mesh unintentionally. Such damage is easily done to the back of the FABIAN and can therefore be overlooked. After this step the model of the phoneme sits freely in the opening created in the Fabian mesh. Therefore the two models do not overlap (see figure 6).

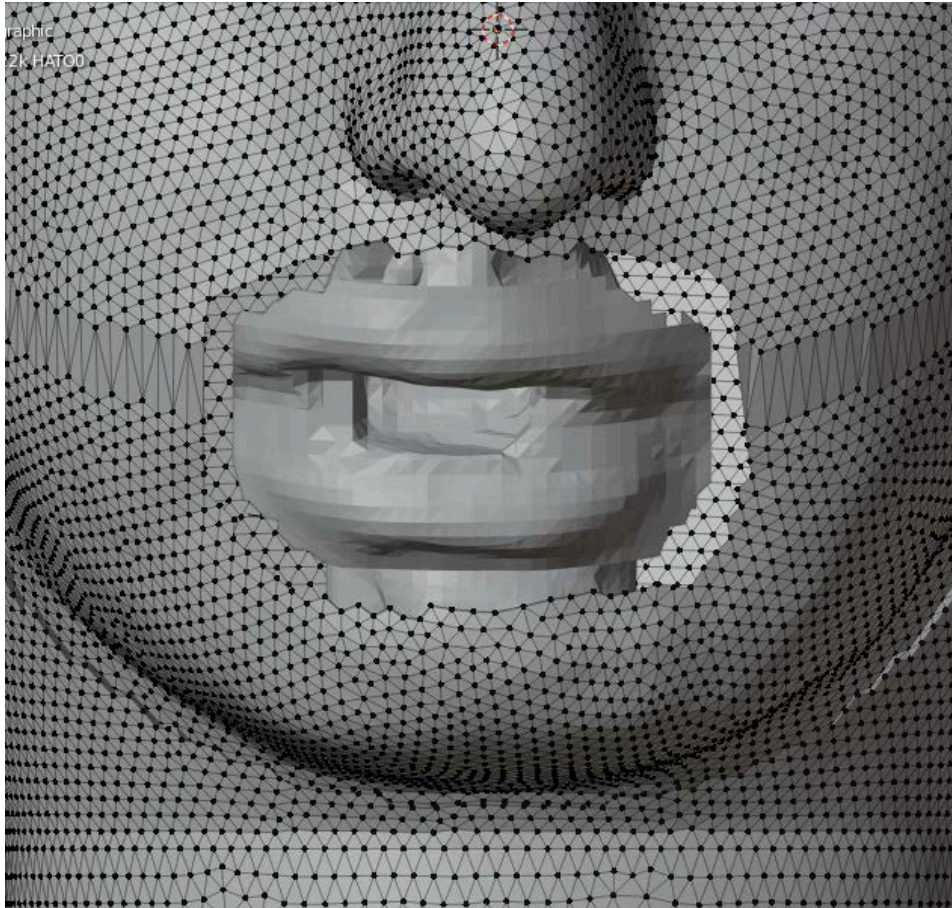


Figure 6: The trimmed model of the phoneme /i/ sits freely on all sides in the cut-out of the FABIAN mesh.

The two models are then combined into one model. To do this, they are both selected in Object Mode and the option join is chosen. Here, the vowel has been given priority to make it easier to align the created models at a later time.

Now, the gap between the lip area and the remaining mesh can be closed. Care is taken to create as few areas as possible that are at a large angle to adjacent areas. Apart from this, smoothing the added areas in a later work step is recommended in any case.

Through the movement of the lower jaw the mesh breaks at the throat of the FABIAN artificial head. This gap must now be closed. To do this, the area around the damaged mesh is selected in the first step and then removed. This increases the distance, but the gap can then be closed at a flatter angle.

When all open areas in the mesh are closed, the edited areas are selected. With the smoothing option, the mesh is smoothed by a factor of 1.0 in 4 iterations so that sharp edges and unevenness almost disappear completely (see figure 7).



Figure 7: After smoothing, almost all sharp edges disappear. Vowel /a/ is shown here as an example.

3.2.4. Separation and deletion of the vocal tract

Blender is used to create versions without vocal tract (wovt) for the different models. By comparing the simulation results with and without the vocal tract for the respective phoneme, the effect of the geometry of the vocal tract on the propagation of the sound field can be evaluated later.

In the first step, a sphere is created as a mesh object. This is placed in the mouth opening and scaled in x-, y- and z-direction and rotated in y direction (see figure 8). It is important, that the entire mouth opening is covered on the outside, on the inside the entire cross-section of the oral cavity must be covered by the formed ellipsoid.

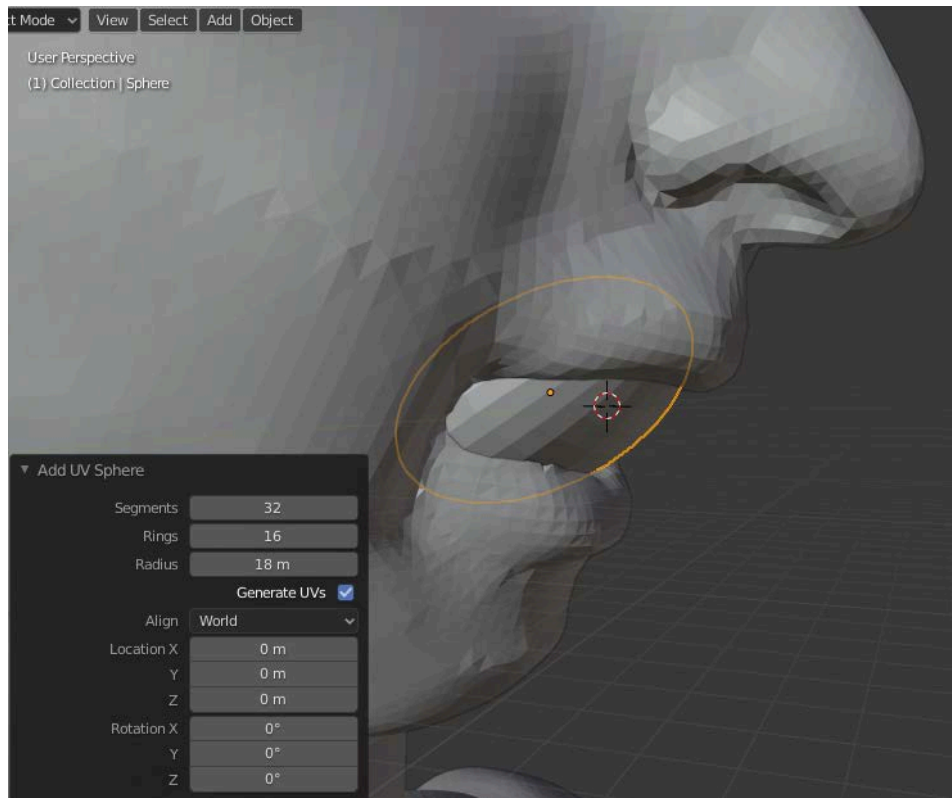


Figure 8: To isolate the vocal tract, an ellipsoid is placed in the mouth opening so that the entire mouth opening is filled and the cross-sectional area of the oral cavity is covered. Later, this creates a slightly convex surface that closes the mouth opening. The mesh of vowel /e/ is shown as an example.

After successfully positioning the ellipsoid, it is duplicated. Next, the Boolean Difference modifier is applied to the head in object mode and the ellipsoid is selected as the target object. The ellipsoid can then be deleted.

Now, a ring of surfaces is selected and deleted on the inside of the model around the cutting edge. This creates a clear separation of the head mesh and the vocal tract. Thus, it can be deleted individually.

To close the mouth opening of the model, the mesh of the head and the duplicated ellipsoid are now joined. After switching to edit mode, the mesh part of the ellipsoid is selected. Now, the selection of the surfaces on the ellipsoid surface in front of the mouth opening can be deselected and the remaining part of the ellipsoid within the head mesh can be deleted. This creates a clean closure of the mouth opening with a slightly convex surface to the outside. No parts of the deleted vocal tract remain inside the mesh.

3.2.5. Model positioning

To make the results of the simulations for the propagation of sound at different phonemes as comparable as possible, the models must be aligned congruently. For this purpose, the model of the vowel /a/ is positioned so that the origin of the coordinates is located centrally in the mouth opening at the height of the corners of the mouth. For this purpose, the model of the phoneme /a/ was shifted in x- and z- direction. Since the models are not axis-symmetrical, a shift in y-direction was omitted (see table 3). Because the FABIAN mesh was not shifted when combining the FABIAN mesh and the individual vocal tract models, the shift of the combined model of phoneme /a/ is transferable to all other models. Thus, all mesh models are congruent at the same place and differ only in the mouth position. Hence, the radiation characteristics of the individual phonemes are easily comparable. In order to simulate the sound in the middle of the mouth opening, an individually positioned sampling grid must be created for each model, which is located at the coordinate origin for vowel /a/ (see section 3.4.1).

Shift X:	-8.4 cm
Shift Y:	0.0 cm
Shift Z:	4.6 cm

Table 3: To position the models in a comparable way, the model of vowel /a/ is aligned so that the coordinate origin is located centrally between the corners of the mouth. This shift is transferred to the other models.

3.3. Mesh postprocessing and re-meshing

To be able to simulate the propagation of sound for the various phonemes using the boundary element method, the models must meet two criteria. Firstly, the meshes must be waterproof so that they are completely closed, and secondly, it must be ensured that the meshes contain six elements per wavelength for the frequency range up to 22050 Hz [18].

To find errors in the mesh, the first step is to add and install the 3D-Print Toolbox in Blender. This is used to check the models for errors such as overlapping surfaces and holes. These are then displayed in a report, can be selected directly in the models and subsequently revised manually (see figure 9). The toolbox also offers the possibility to make the models waterproof.

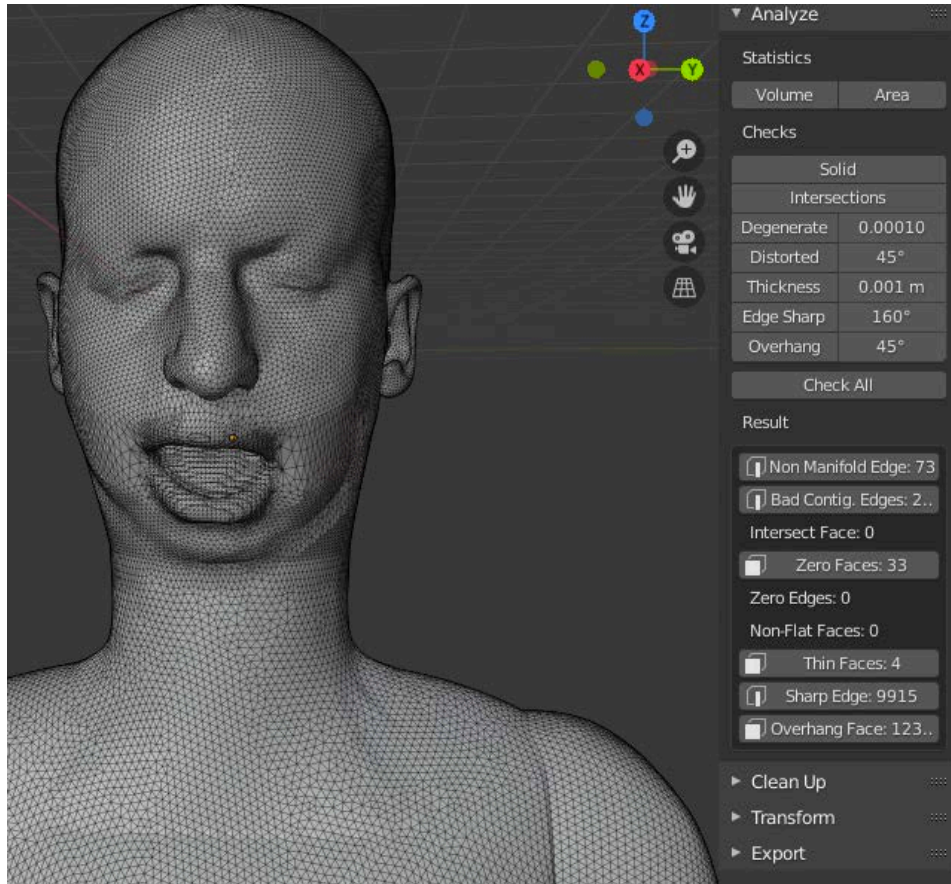


Figure 9: Analysis of the 3D-Print Toolbox shows errors in the mesh of vowel /ae/ and allows manual and automatic correction of the errors. This way the model can be made waterproof.

In the second step, the models are imported into the Meshmixer software [13]. There, they are checked again for holes with the so-called Inspector. If the waterproofness is also confirmed in this step, the models can be re-meshed. In order to keep the computational effort as low as possible, the head with the fully depicted vocal tract is individually re-meshed, while the mesh of the torso remains untouched (see figure 10). The options for the re-meshing are selected so that an edge length of 2 mm is applied to the area of the head with the vocal tract without transition to the torso (see table 4). Since a wrong orientation of the normals leads to unusable simulation results, it is absolutely necessary to pay attention to a correct orientation. In Meshmixer this is clearly visible, because the inside of the models is displayed striped, while the outside is visualized in monochrome grey. In case of a wrong alignment the whole mesh can be selected and the normals flipped. The finished models are exported as .stl files.

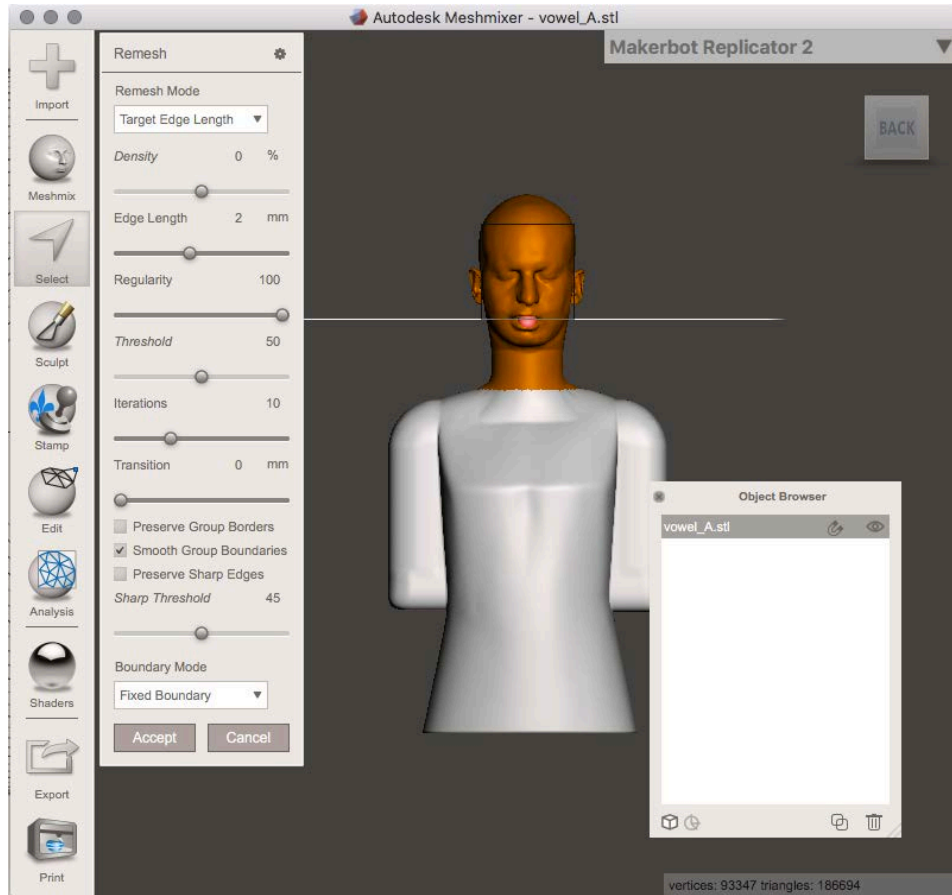


Figure 10: The head with vocal tract is re-meshed separately from the torso for the vowel /a/ with an edge length of 2 mm in the software Meshmixer (a).

The remesh procedure described above is performed for all wovt models. It is also executed for the wvt models with the vowels /ae/, /a/, /i/, /o/ and /y/. For the vowels /e/, /oe/ and /u/ the remesh process with mesh mixer makes the vocal tract so small that the sound cannot propagate through the vocal tract. Since no suitable setting for the remeshing of these models was found with Meshmixer, they are processed with the pmp-library [14] (see table 6). Therefor, the dimensions of the vocal tract in x-, y- and z-direction are determined. Thus, only the vocal tract is remeshed, specifying the minimal and maximal edge length, the maximal deviation of nodes to their previous positioning and the number of iterations (see table 5).

By using the two different software programs, it is possible to find suitable remeshing settings for all eight wvt models (see figure 11) and the corresponding eight wovt models (see figure 12), which allow subsequent simulations of all models.

Parameter	Setting
Remesh Mode	Target Edge Length
Density	0%
Edge Length	2.0 mm
Regularity	100
Threshold	50
Iterations	10
Transition	0.0 mm
Preserve Group Borders	-
Smooth Group Boundaries	x
Preserve Sharp Edges	-
Sharp Threshold	45
Boundary Mode	Fixed Boundary

Table 4: Meshmixer allows you to set many parameters that affect the result of the remesh. A setting is used that allows a subsequent simulation with MESH2HRTF for as many models as possible.

Parameter	Setting
Min. Edge Length	0.5 mm
Max. Edge Length	2.5 mm
Max. Error	0.25 mm
Iterations	10

Table 5: To use the pump-library for remeshing, the desired edge length and the maximum deviation of nodes from their previous position and the number of iterations are given.

Vowel	Meshmixer	pmp-library
/ae/	x	
/ae/ wovt	x	
/e/		x
/e/ wovt	x	
/a/	x	
/a/ wovt	x	
/i/	x	
/i/ wovt	x	
/oe/		x
/oe/ wovt	x	
/u/		x
/u/ wovt	x	
/o/	x	
/o/ wovt	x	
/y/	x	
/y/ wovt	x	

Table 6: The different models are re-meshed either with Meshmixer or the pmp-library. The respective method used is marked by the letter “x”.

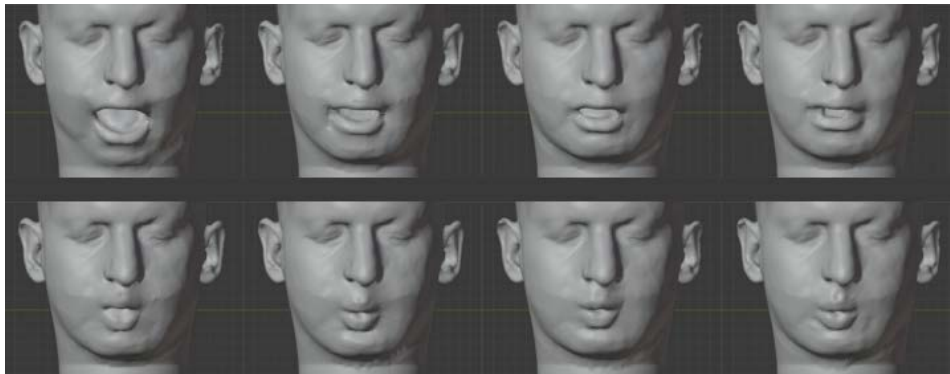


Figure 11: All wvt models of the eight vowels /ae/, /e/, /a/, /i/ (from left to right, top row) and /oe/, /u/, /o/, /y/ (from left to right bottom row) exist as remeshed meshes.

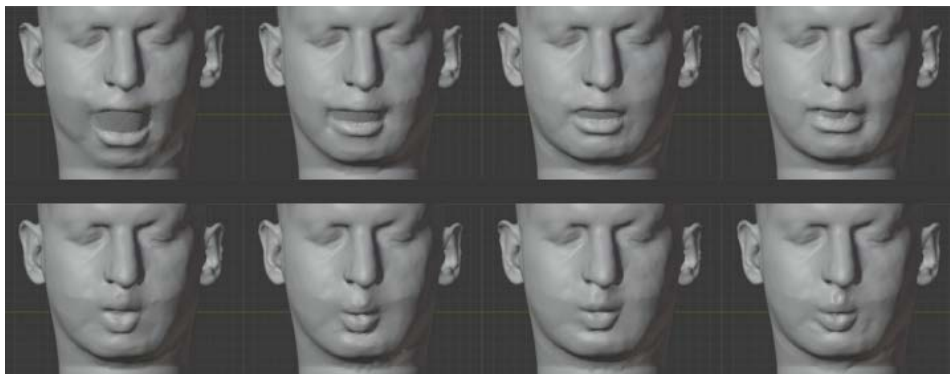


Figure 12: All wovt models of the eight vowels /ae/, /e/, /a/, /i/ (from left to right, top row) and /oe/, /u/, /o/, /y/ (from left to right bottom row) exist as remeshed meshes.

3.4. BEM computation

3.4.1. Sample grids

In this thesis two different sample grids are used. On one hand, the sound pressure level is to be simulated at different positions with high angular resolution and fully spherical around the head. For this purpose a Lebedev grid with 1730 points is created in Matlab with the Sofia toolbox [19], which can be used for Spherical harmonics processing up to order $N_{sh}=35$. The grid is then stored in Mesh2HRTF as .txt files, where one file describes the elements and a second the coordinates of the nodes. When exporting from Blender, Mesh2HRTF loads these files and uses them for the simulations. The sampling points of the used Lebedev grid are distributed at a distance of 1.5 m around the mesh. On the other hand, the sound pressure at the mouth opening is also to be simulated in the simulations. Thereby, the influence of the vocal tract and the formants can be determined. Therefore, a small sample grid is positioned in the mouth opening at the level of the corners of the mouth for the wvt-models. In the wovt-models, this sampling grid is positioned at the same height so that it does not touch the closing surface of the mouth opening, as this would make a simulation impossible. The sample grid consists of 4 nodes, which lie in a horizontally aligned square of edge length 5 mm. The fifth node is located in the middle of the grid and will later be used to evaluate the simulations. The arrangement as well as the positions of the nodes are stored in .txt files and thus made available for Mesh2HRTF. The dimensions are based on the sampling grid used by Arnela et al. [11].

Since the positioning for the individual phoneme models differs, a position marker in Blender is composed of two intersecting planes. By moving this marker, the position of the center of the sample grid is determined (see table 7 and figure 13). Since the model for the vowel /a/ was set as reference when aligning the individual models (see section 3.2.5), the center of the sample grid is located at the coordinate origin for this vowel. The model and the used sample grids can be loaded and visualized from the folders of the simulations created by Mesh2HRTF. This way the correct positioning of the grids can be checked again (see figure 14).

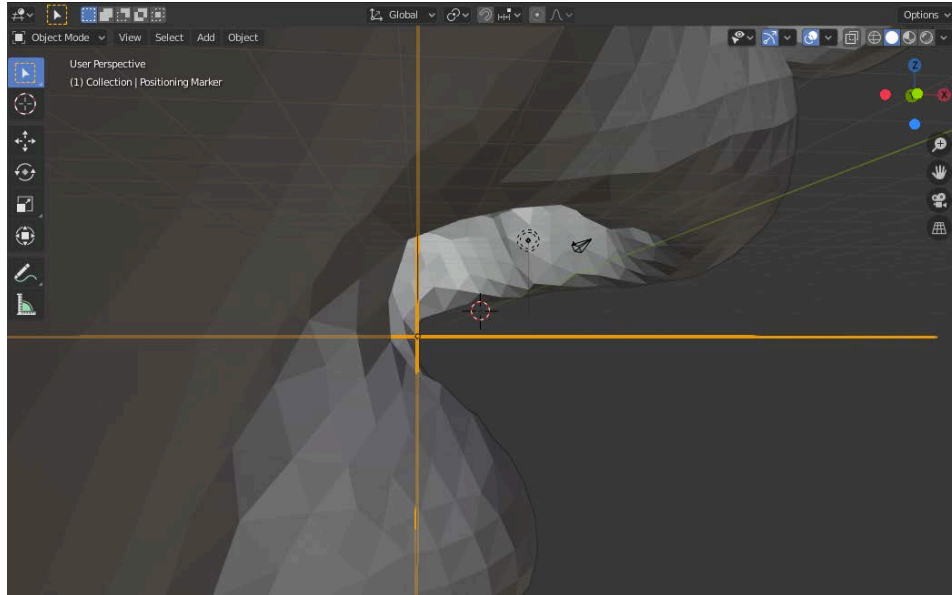


Figure 13: To determine the center of the sample grid in the mouth opening for vowel /e/, a positioning marker is created in Blender by two intersecting planes. This marker is then moved so that it is at the level of the corners of the mouth.

Vowel	Location X	Location Y	Location Z
/ae/	-1.3 cm	0.0 cm	-0.6 cm
/ae/ wovt	0.4 cm	0.0 cm	-0.6 cm
/e/	-0.5 cm	0.0 cm	-0.2 cm
/e/ wovt	0.9 cm	0.0 cm	-0.2 cm
/a/	0.0 cm	0.0 cm	0.0 cm
/a/ wovt	1.1 cm	0.0 cm	0.0 cm
/i/	0.5 cm	0.0 cm	0.0 cm
/i/ wovt	1.3 cm	0.0 cm	0.0 cm
/oe/	0.5 cm	0.0 cm	-0.1 cm
/oe/ wovt	0.8 cm	0.0 cm	-0.1 cm
/u/	0.7 cm	0.0 m	-0.4 cm
/u/ wovt	1.0 cm	0.0 m	-0.4 cm
/o/	0.8 cm	0.0 cm	-0.1 cm
/o/ wovt	1.0 cm	0.0 cm	-0.1 cm
/y/	0.8 cm	0.0 cm	-0.3 cm
/y/ wovt	1.1 cm	0.0 cm	-0.3 cm

Table 7: Individual positioning of the sample grid at the mouth opening. For vowel /a/, the center of this sampling grid is at the coordinate origin, since vowel /a/ is the reference when positioning the meshes.

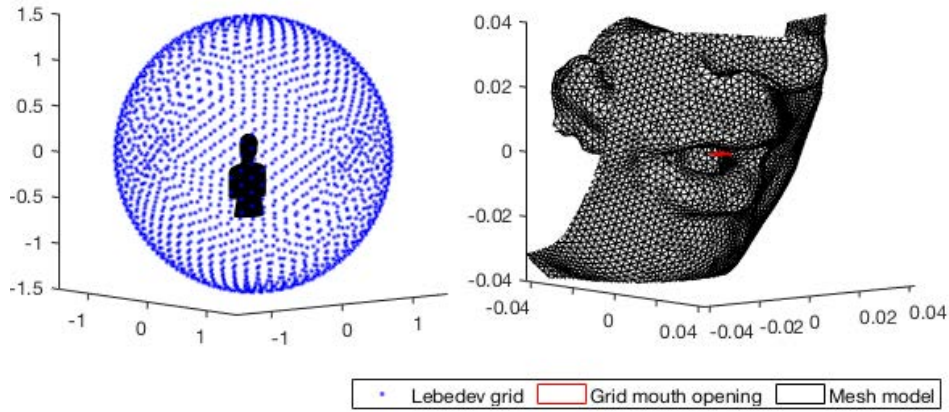


Figure 14: The Lebedev sample grid and the mesh model (left), as well as the sample grid at the mouth opening and the mesh model (right) can be visualized to check the sample grid positioning. This unprecedented example is the vowel /a/.

3.4.2. Mesh2HRTF

The simulation of the propagating sound field is done with the software package Mesh2HRTF. This software, published in 2015 by Ziegelwanger et al. [6], 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). Mesh2HRTF assumes sound hard surfaces unless specified otherwise by the user, e.g. by assigning a velocity boundary condition. 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 [20]). 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 study from 2017 [5] 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.

At this moment Mesh2HRTF is supported by Blender 2.7x. To use Mesh2HRTF it is installed as an add-on in Blender. The software, as well as a detailed manual, can be found on the SOURCEFORGE page [21].

After the successful installation, the head-and-torso models can now be imported into Blender. In the second step, the positions of the sources as vibrating planes are defined.

The first source is defined as the lower surface of the vocal tract in the wvt models. For the simulation results, whether a single element or the entire bottom surface is selected makes a negligible difference. However, the calculation time for a simulation with several elements is shorter in the test run, which is due to the clustering in the numerical simulation. In wovt models, the source elements are specified to represent the area of the mouth opening of the representative model with vocal tract (see figure 15). The second source is positioned in the middle of the right ear canal, because Mesh2HRTF always needs two sources, even if only one is calculated later.

Now, the file structure is exported from Blender, which is needed by NumCalc for simulation. Thereby, the parameters of the simulation are specified (see table 8). As first sample grid the “ARI” grid is selected, because the Lebedev grid is stored in this folder. As second sample grid “Custom” is selected, because in this folder the individually positioned sample grid at the mouth opening is saved with Matlab (see chapter 3.4.1).

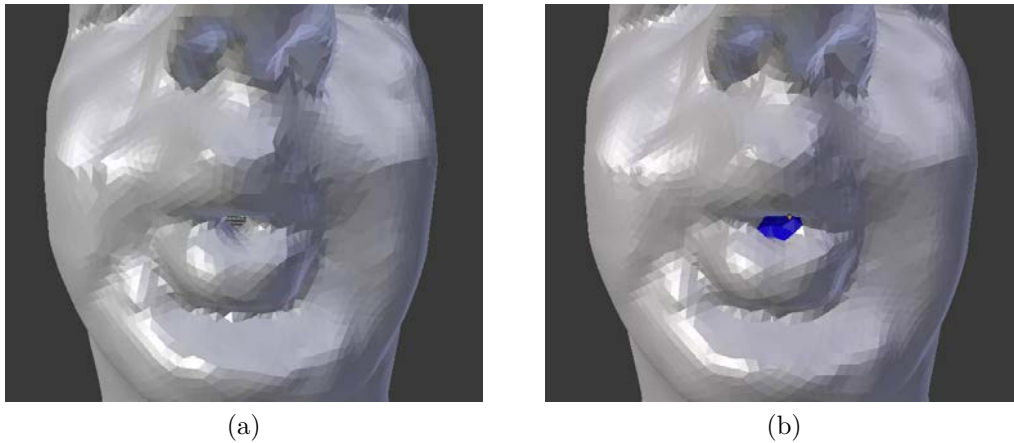


Figure 15: When defining the source elements in the wovt models, care is taken to ensure that the width of the source corresponds to the mouth opening in the corresponding wvt model. Here vowel /u/ is shown as wvt model a), and the wovt model with selected source elements in the same form b).

With NumCalc the actual simulation is performed. In this thesis, the different simulations are calculated on the High-Performance-Computing-Cluster (HPC-Cluster) of the TU-Berlin [22]. Therefore, NumCalc is compiled on the cluster. The blender exports with Mesh2HRTF are then copied to the cluster with scp commands and the simulation can be started with Simple Linux Utility for Resource Management (SLURM) commands.

After the simulation is finished, the Matlab file Output2HRTF.m is executed with Matlab. Here, the variable “Reference” is defined as “true”. This compensates the increasing energy of the oscillating source elements to higher frequencies in relation to their surface. Thus, the transfer functions are output in SOFA format and additional information, e.g. the computation time of the simulation, is provided.

Parameter	Setting
Ear	left
Pictures	-
Point Source:	
Source (x)	0
Source (y)	101
Source (z)	0
Reciprocity	x
Constants:	
c (m/s):	343
rho():	1.1839
Object Meshes:	
Unit:	mm
Evaluation Grids:	
Ev.Grid 1:	ARI
Ev.Grid 2:	Custom
NF-Calc.:	-
Frequencies:	
Freq min:	100
Freq max:	16000
Freq step:	x
Num Freq step:	160
Freq.-dep.	-
Method:	ML-FMM BEM
Cluster:	
CPU (first):	1
CPU (last):	1
Num. of used cores:	8
Mesh2HRTF:	
Mesh2HR...	Path to Mesh2HRTF

Table 8: To create simulation files for NumCalc Blender requires different settings in the Mesh2HRTF export. Among other things the frequencies of the simulation and the sample grids used can be specified here.

4. Results

4.1. General results

The applied methodology provides one wvt model and one wovt model for each of the eight vowels and transfer functions for further research (see section A). With BEM numerical simulation the sound propagation in the frequency range from 100 Hz to 16000 Hz is computed for all models. Mesh2HRTF returns logging files “NC.out” for each core of the cluster on which the simulation is performed for different frequencies. In these files it is noted how many interactions were necessary for the calculation and at which frequencies no iterative BEM solution was possible (see table 9). The models differ in these characteristics. The wovt models are more prone to errors in the iterative calculation. These errors occur at high frequencies starting from 14800 Hz (vowel /oe/). In the wvt models, all calculations iterated except vowel /oe/.

Vowel	Frequencies without iterative BEM solution in Hz
/ae/	-
/ae/ wovt	15000, 15200, 15300, 15400, 15600
/e/	-
/e/ wovt	15300, 15400, 15500, 15700, 15900
/a/	-
/a/ wovt	15300, 15500, 15800, 15900
/i/	-
/i/ wovt	15300, 15500, 15800
/oe/	400, 8900, 9000, 9400, 9600, 9900, 10100, 10300
/oe/ wovt	14800, 15300, 15400, 15500, 16000
/u/	-
/u/ wovt	15300, 15400, 15500, 15700, 15900
/o/	-
/o/ wovt	15300, 15400, 15500, 15800, 15900
/y/	-
/y/ wovt	-

Table 9: For each simulation, the output files of Mesh2HRTF contain a list of frequencies for which no iterative calculation is possible. The symbol “-” indicates that all calculations iterate.

When evaluating computation time, simulations for the wvt models range from 193:10 h (vowel /i/) to 215:18 h (vowel /a/). The wovt models can be computed faster with computation times between 103:12 h (vowel /y/) and 158:20 h (vowel /ae/) (see figure 16). A regularity is neither to be recognized with regard to the geometry of the individual models, nor to the remeshing method used (see section 3.3). Only the additional effort in the simulation of the vocal tract becomes clear. On average, the wvt models require 53.3 % more computation time than the wovt models.

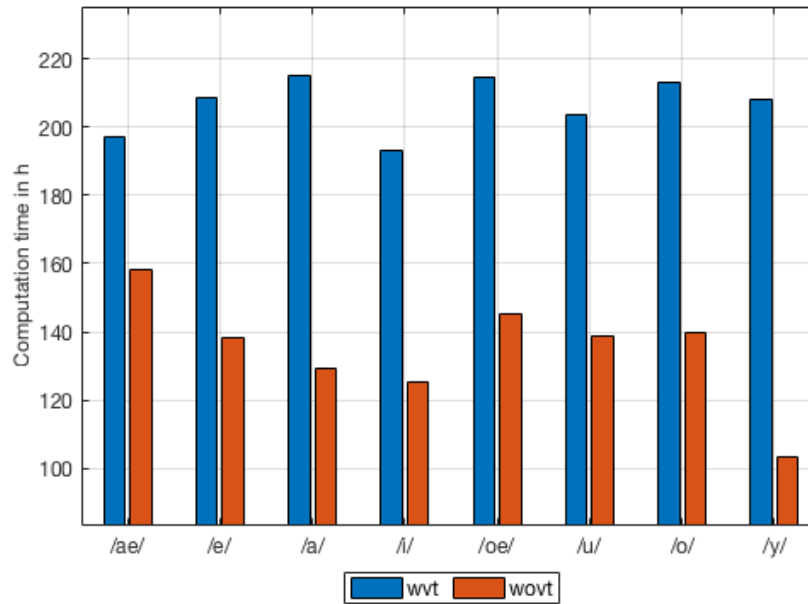


Figure 16: The computation time is shown for all simulations. The wvt models have a significantly higher computation time than the wovt models. Otherwise no regularity is recognizable.

4.2. Spectral results

The spectra of the wvt models are simulated and displayed in the frequency range up to 16000 Hz, or the maximum possible frequency. The sampling point is the center of the sample grid in front of the mouth opening. Each vowel shows an individual distribution of levels in this representation (see figure 17). In the lower frequency range up to 4000 Hz, the positions of the formants are visible. However, they are easier to locate when viewed separately (see figure 19).

Noticeable are vowel /oe/ and vowel /o/. The simulation of /oe/ leads to relatively high sound pressure levels compared to the other vowels. This happens especially at frequencies above 8000 Hz and is due to errors in the BEM simulation. This vowel also required another method of remeshing and is generally more prone to errors (see section 3.3). Vowel /o/ shows a strong decrease of the simulated sound pressure levels in the spectrum between about 2500 Hz and 4000 Hz. Compared to work with microphone recordings with singers, the spectra of the different vowels in this work differ more clearly from each other [3]).

To give a better overview of the sound level distribution in the spectra of the wvt models, the third-octave band averaged sound pressure levels are presented in tabular form (see table 10). It can be observed that the sound pressure levels in the bands first increase, then, depending on the vowel, are maximum in one one-third octave band and decrease to higher frequency bands afterwards. So for each vowel, there is a dominant one-third octave band with the highest average sound pressure level. This is relevant in combination with directivity of sound propagation (see section 4.3), since the directionality of bands with high sound pressure levels is more important for perception [3].

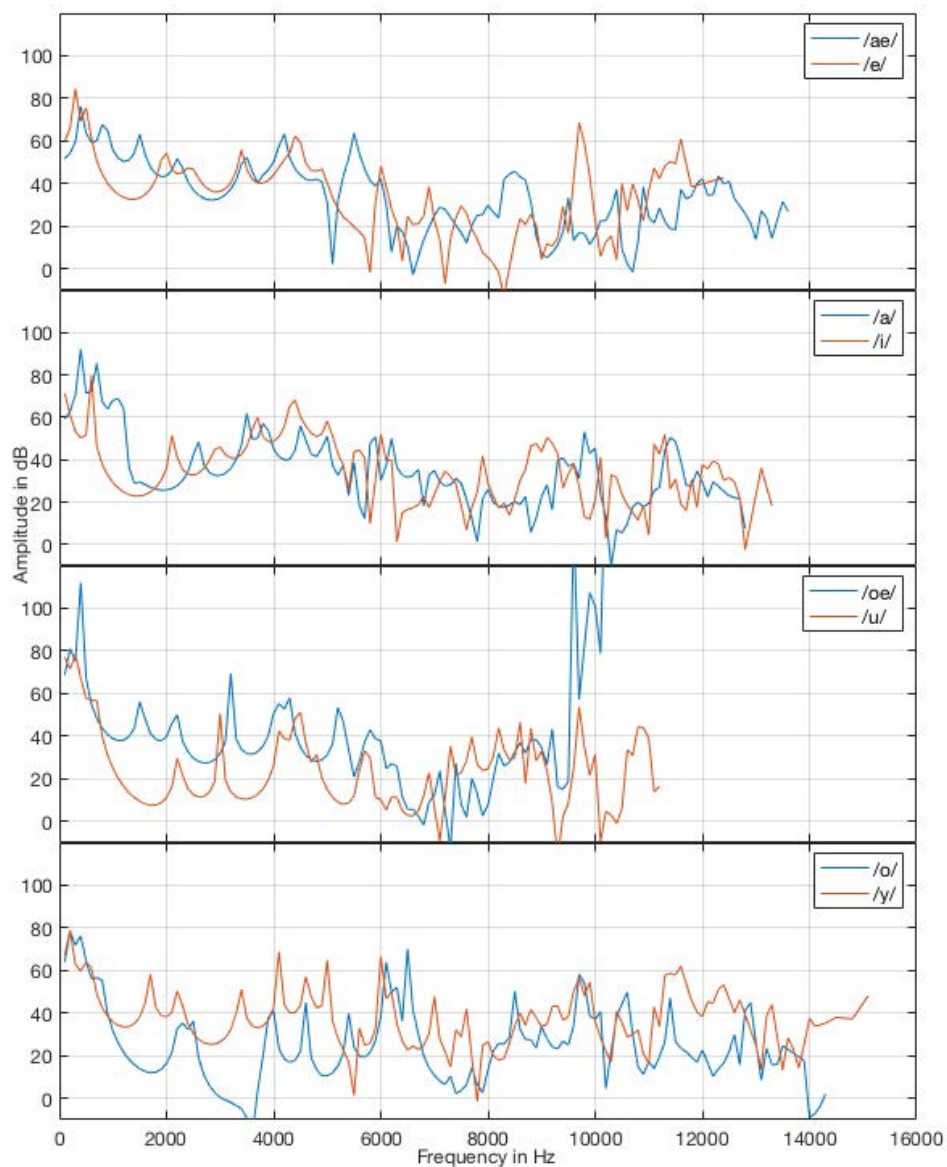


Figure 17: For the wvt models the spectra are calculated at the center of the sample grid in front of the mouth opening. They show the energy rich regions of the formants and the distribution of the sound pressure levels up to 16000 Hz, or the maximum calculable frequency.

3rd-oct. f	Vowel wvt	/ae/	/e/	/a/	/i/	/oe/	/u/	/o/	/y/
	100 Hz	51.7	60.2	59.3	71.4	68.8	77.2	64.1	67.0
	200 Hz	54.2	65.8	62.6	61.3	81.1	71.5	78.4	78.9
	315 Hz	59.6	84.6	70.5	53.3	75.2	78.2	72.0	62.9
	400 Hz	76.3	69.5	92.2	50.3	112.0	67.4	76.2	59.8
	500 Hz	63.7	75.7	71.3	51.6	67.0	57.6	64.5	63.9
	630 Hz	59.8	57.5	81.2	73.9	52.2	56.8	56.4	57.0
	800 Hz	67.9	44.5	67.3	37.9	43.8	39.8	55.4	42.6
	1000 Hz	59.4	37.7	67.2	29.9	39.3	26.5	34.4	36.5
	1250 Hz	52.1	33.0	55.1	23.7	40.9	14.0	18.9	34.2
	1600 Hz	57.2	35.4	28.3	23.8	50.6	8.4	12.9	51.0
	2000 Hz	46.5	48.8	26.4	41.6	43.8	19.2	23.0	43.5
	2500 Hz	39.3	43.8	40.0	35.4	31.4	15.3	29.5	34.1
	3150 Hz	43.5	45.7	48.0	43.7	53.5	34.5	-1.7	39.2
	4000 Hz	52.6	50.6	48.9	58.8	48.9	36.5	28.5	51.8
	5000 Hz	48.7	44.8	44.5	51.6	38.8	33.1	29.5	48.8
	6300 Hz	33.1	31.8	39.4	34.9	30.0	17.9	51.9	46.6
	8000 Hz	34.7	18.6	22.3	34.0	28.1	34.5	29.4	31.9
	10000 Hz	23.8	46.2	35.5	36.9	(174.8)	34.6	40.8	42.4
	12500 Hz	33.6	50.4	37.8	35.6	-	-	28.9	47.5
	16000 Hz	-	-	-	-	-	-	2.1	41.4

Table 10: The table shows the third-octave band averaged sound pressure levels in dB of the wvt models in the center of the sample grid in front of the mouth opening. This allows an overview of the sound level distribution for the individual vowels. The maximum sound pressure level of the third-octave bands is printed bold for each vowel.

For the wovt models, the spectra in the center of the sample grid in front of the mouth opening are calculated and compared. The spectra of the vowels /ae/, /e/, /a/, /i/, /oe/ fall slightly from a frequency of 6000 Hz and thus have a low-pass behaviour with low slope. The spectra of the vowels /u/, /o/, /y/ increase steadily. At high frequencies above 15000 Hz the spectra of the vowels /e/, /a/, /i/, /o/ with very high amplitude values indicate erroneous results. However, we can speak of relatively straight spectral progressions of the simulations (see figure 18).

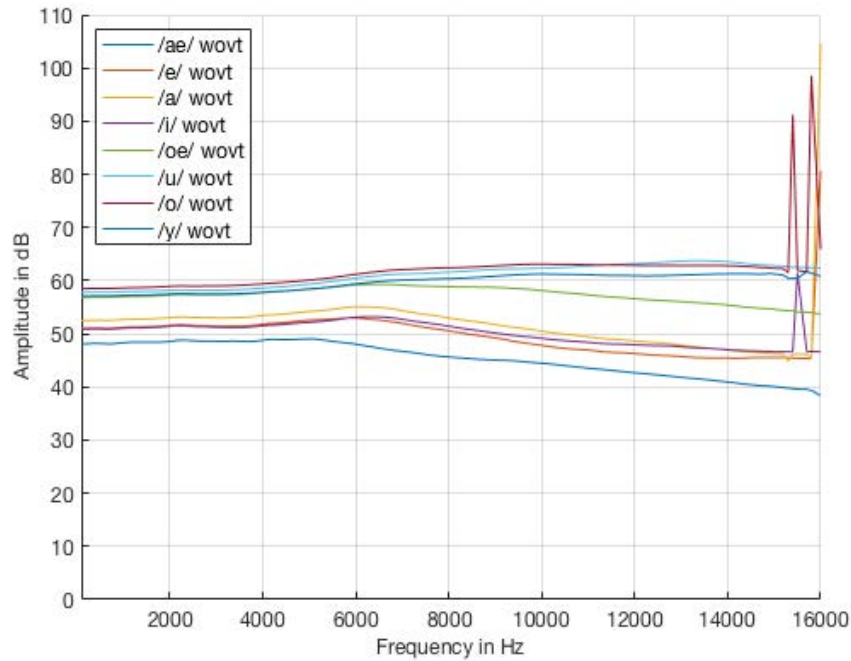


Figure 18: The spectra of wovt models generally have a relatively flat course. For the vowels /ae/, /e/, /a/, /i/, /oe/ the spectra fall from a frequency of 6000 Hz, for the vowels of other vowels the spectrum increases steadily. At high frequencies above 15000 Hz one sees erroneous values with very high amplitudes.

For each vocal tract model there is a sound recording of the test person, which was made during the MRI scan. These signals are cropped to minimize noise (see section A). With Matlab the power spectral density (PSD) estimate of these signals with a window length of 512 samples and an overlap of 256 samples is calculated and compared to the simulated formant spectra of the center of the sample grid at the mouth opening in the frequency range up to 4000 Hz. The respective signals are normalized to a maximum amplitude of 0 dB in the frequency range up to 4000 Hz. Thus, the position of the formants in the recordings can be compared with the formant spectra of the simulations (see figure 19). For all vowels, the calculated formant spectra from the simulations have a similarity to the spectra of the respective recordings. For the vowels /e/ and /y/, the spectral similarity is greatest.

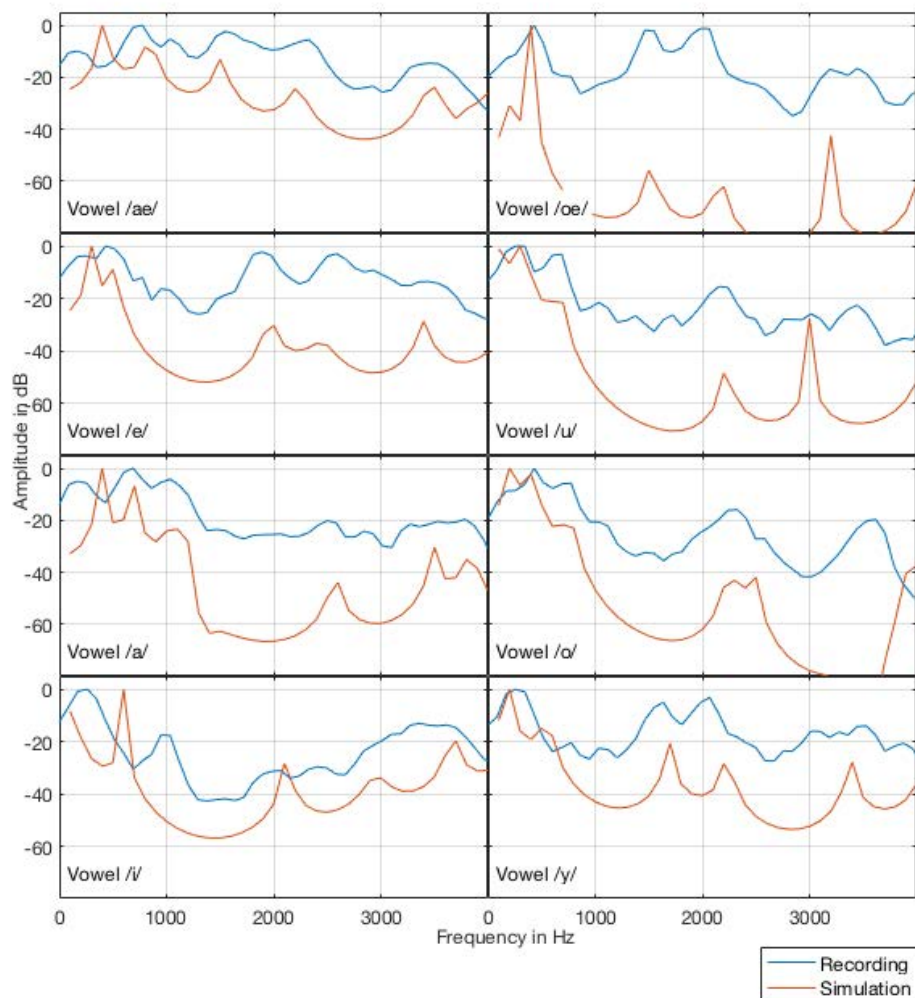


Figure 19: The formant spectra of the vowels from recordings and simulations are comparable. All signals were normalized to a maximum amplitude of 0dB in the frequency range up to 4000 Hz.

4.3. Results of directivity

With the Matlab Toolbox AKtools [23], the full-spherical results of the simulations are displayed as third-octave band averaged. The average complex sound pressure level L_{mean} in dB is calculated from the sound pressure values p of the third octave band as:

$$L_{\text{mean}} = \text{dB}(\text{mean}(\text{abs}(p))) \quad (1)$$

The results of the wvt model, the wovt model and the difference between the two are compared. This kind of visualization allows to visualize the influence of the vocal tract on the propagation of the sound on the entire sphere surface around the head of the model. In general, it can be seen that the radiation pattern at frequencies below 500 Hz is rather omnidirectional and the directionality is more noticeable at higher frequencies. In all models the pressure maximum is frontal at an azimuth angle of 0 degrees and slightly below the horizontal plane. This is mainly due to reflection of the torso and external geometry, since the wovt models have the pressure maximum at a comparable position. The influence of the vocal tract seems to be rather small on the sound propagation, since the results of the wvt models and the wovt models are very similar. The difference between the two simulations also looks similar in shape, indicating a slight shift in the beam shape. The results of the simulations for Vowel I are shown as an example. For all vowels, the comparison of the simulations of the wvt and wovt models is performed (see section A).

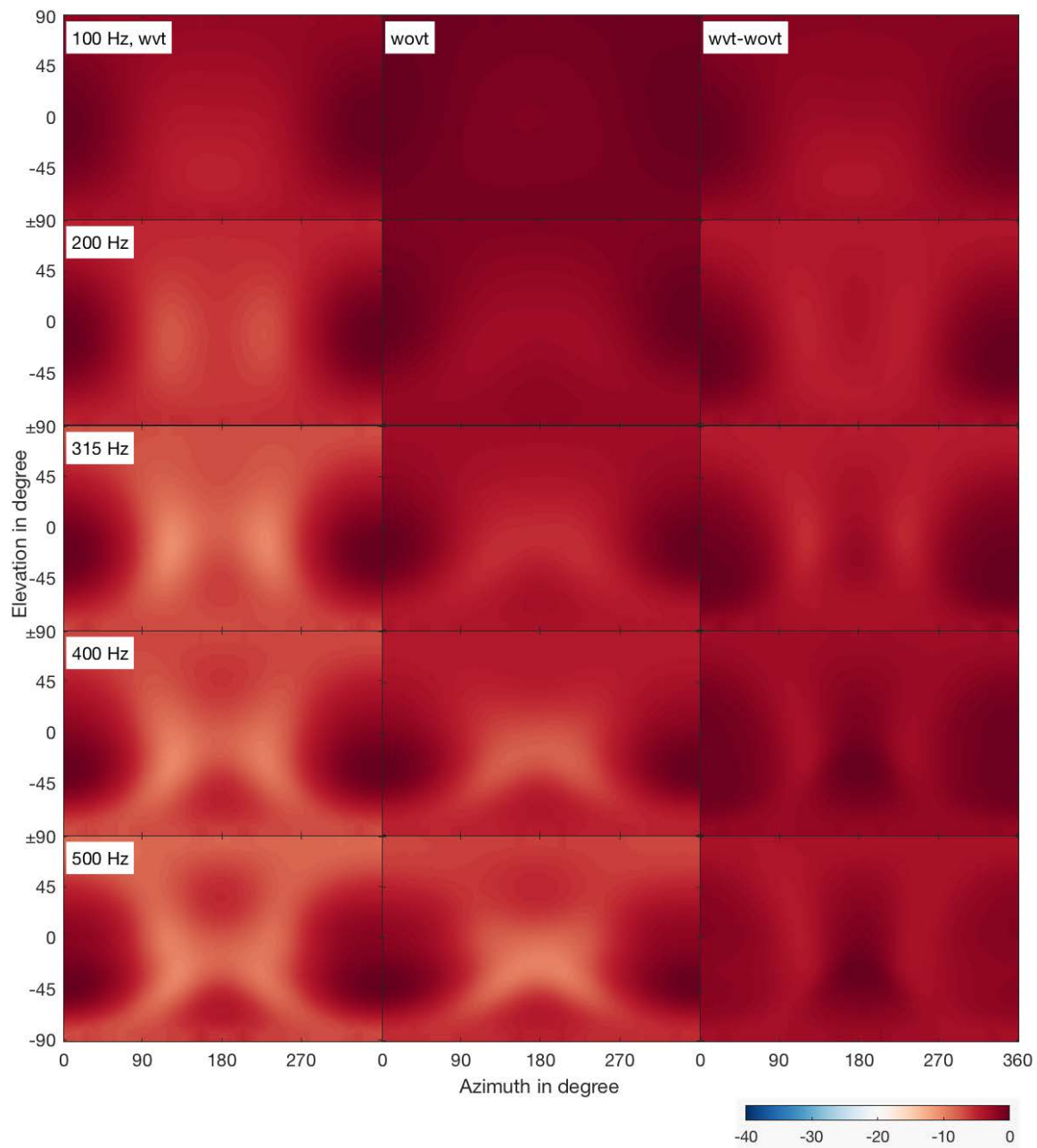


Figure 20: The sound propagation of vowel /i/ is displayed full-spherical third-band averaged for the center frequencies 100 Hz, 200 Hz, 315 Hz, 400 Hz, 500 Hz. The columns show the results of the wvt models, the wovt models and the difference of both normalized to 0 dB.

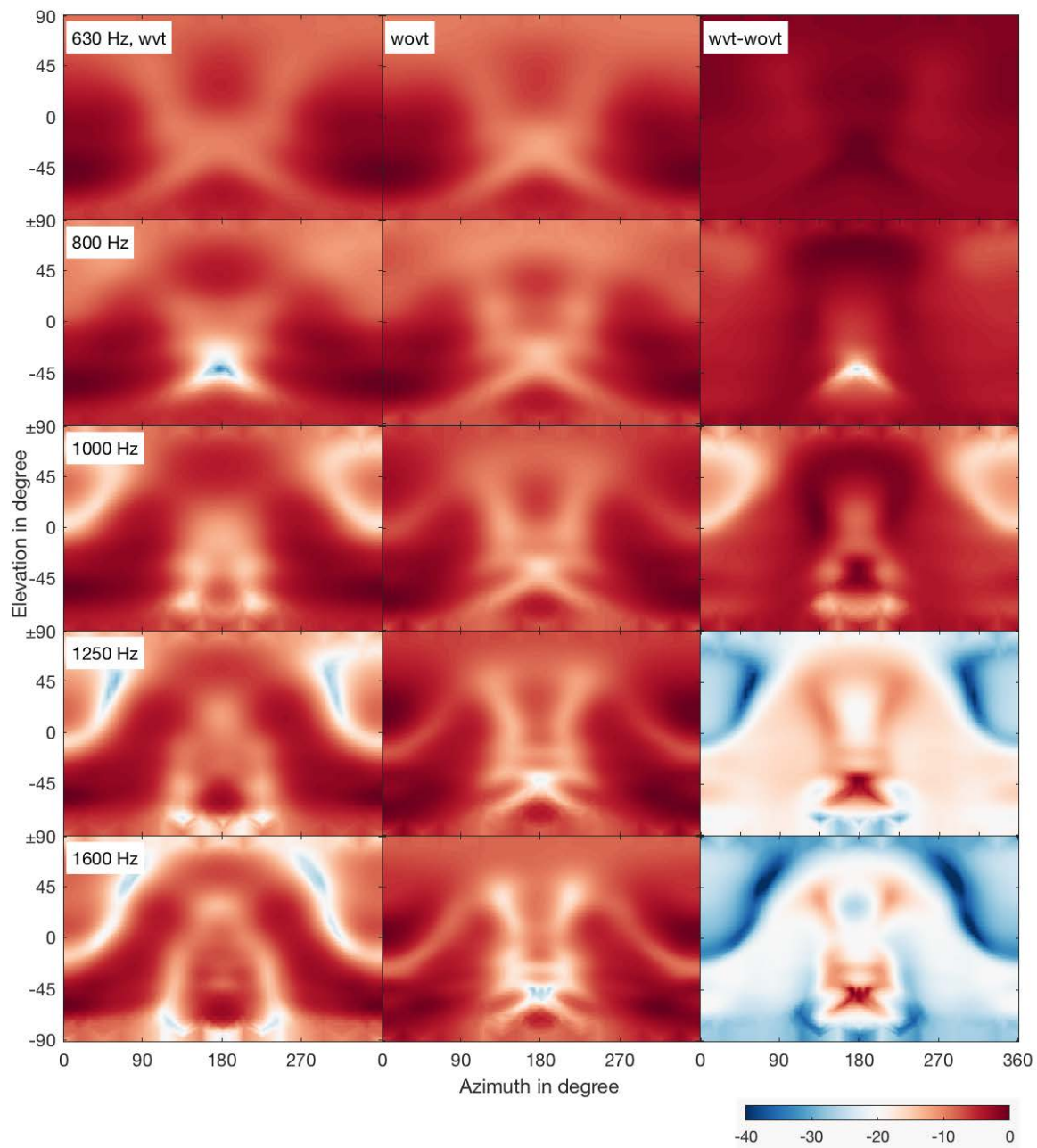


Figure 21: The sound propagation of vowel /i/ is displayed full-spherical third-band averaged for the center frequencies 630 Hz, 800 Hz, 1000 Hz, 1250 Hz, 1600 Hz. The columns show the results of the wvt models, the wovt models and the difference of both normalized to 0 dB.

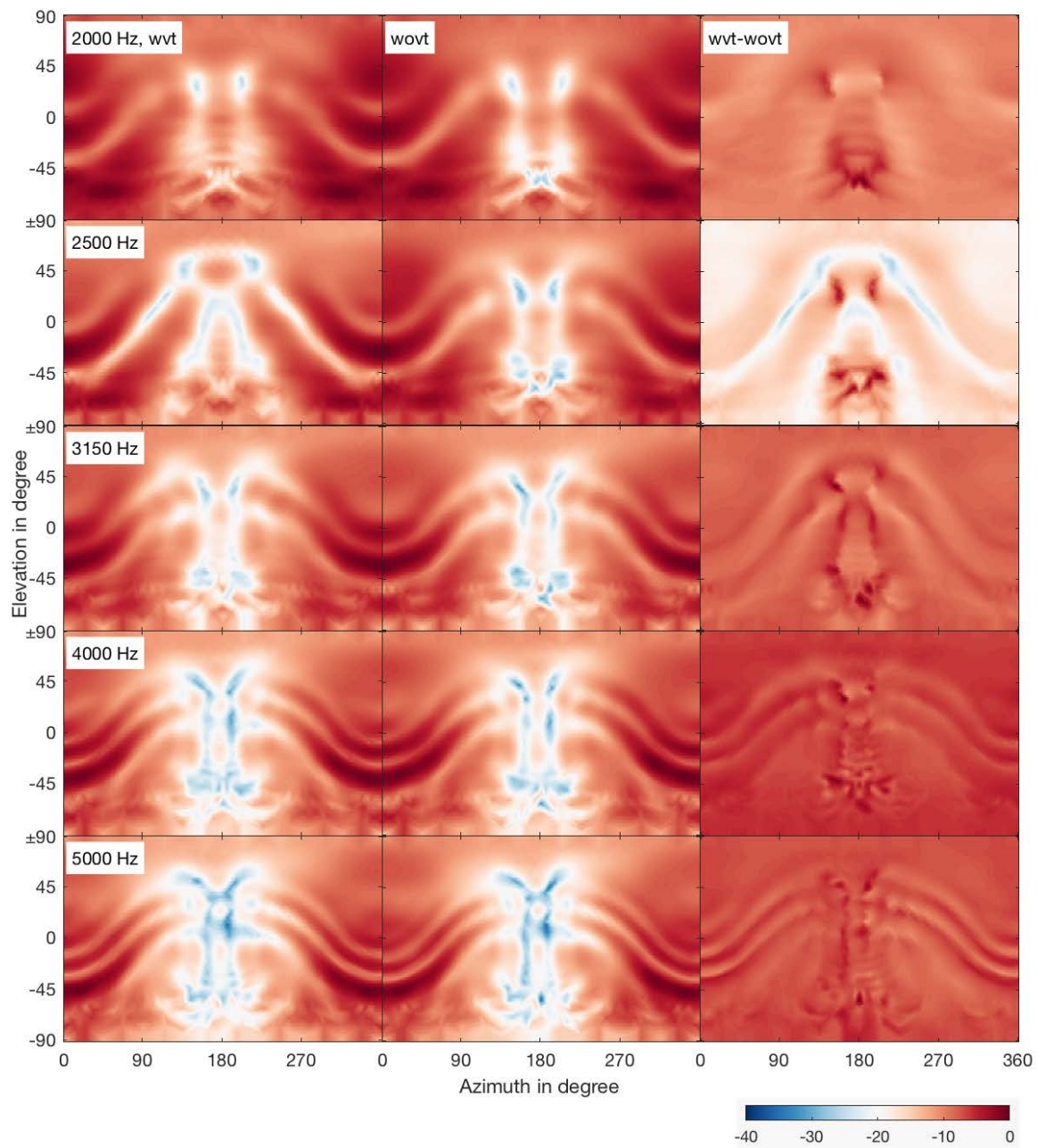


Figure 22: The sound propagation of vowel /i/ is displayed full-spherical third-band averaged for the center frequencies 2000 Hz, 2500 Hz, 3150 Hz, 4000 Hz, 5000 Hz. The columns show the results of the wvt models, the wovt models and the difference of both normalized to 0 dB.

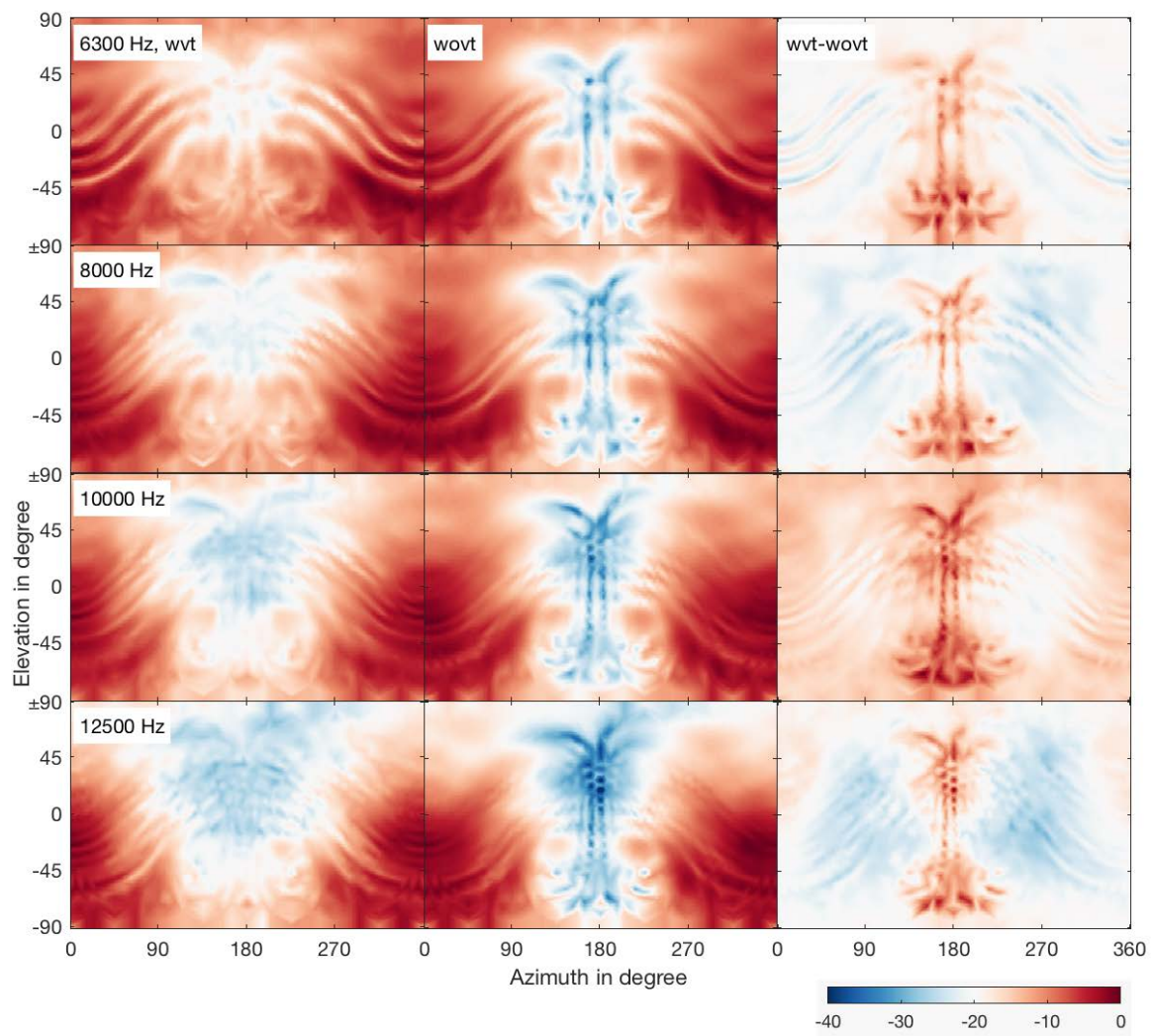


Figure 23: The sound propagation of vowel /i/ is displayed full-spherical third-band averaged for the center frequencies 6300 Hz, 8000 Hz, 10000 Hz, 12500 Hz. The columns show the results of the wvt models, the wovt models and the difference of both normalized to 0 dB.

In polar diagrams, the radiation characteristics of the individual vowels can be easily displayed and compared. The three main planes transversal plane (0 degrees represents front), frontal plane (0 degrees represents top) and median plane (90 degrees represents front) are displayed. The values are third-octave band averaged, spline-interpolated and the maximum level normalized to 0 dB. In direct comparison, four vowels each are shown, sorted by the size of the mouth opening. Thus the vowels /ae/, /e/, /a/, /i/ are compared and the vowels /oe/, /u/, /o/, /y/. The representation of the sound propagation by polar diagrams is preferred to the representation in balloon diagrams, because a comparison of the different models can be realized better this way. The results show the wvt models. Comparable plots of the wovt models can be found in the digital appendix (see section A).

In the frequency range up to 200 Hz, all models have a relatively direction-independent characteristic in the transverse and frontal planes. The mean attenuation at these frequencies across all vowels at these two planes is -2.5 dB. On the median plane, the vowels with larger mouth openings /ae/, /e/, /a/, /i/ form a directionality of approximately 95 degrees even at 100 Hz and 200 Hz, defined by a slope width of -3 dB. Among each other these directional effects are very similar (see figure 24). The vowels with smaller mouth openings /oe/, /o/, /y/ show an almost omnidirectional behaviour on all levels in this frequency range. Vowel /u/ shows a deviating behavior with a supercardioid characteristic at 100 Hz on the transverse plane. However, this might be due to errors in the calculation, since there is no reasonable explanation for this. At the other levels and also at 200 Hz, vowel /u/ shows a slight directionality comparable to the vowels of the larger mouth openings (see figure 25).

Between 500 Hz and 1000 Hz, the vowels become more directional forward on the transverse plane up to 630 Hz. At 800 Hz and 1000 Hz, on the other hand, they exhibit a broader radiation pattern. This behavior is particularly pronounced for the vowels /i/, /oe/ and /y/. Vowel /u/ forms a characteristic at 800 Hz on the transversal and frontal plane which resembles a dipole. Vowel /o/ shows a similar radiation pattern at 1000 Hz on the transverse plane. On the frontal plane, all vowels show a downward shift of the radiation direction. On the median plane, the vowels become more directional with larger mouth opening in the direction of oblique frontal bottom. The individual vowels do not differ greatly. In the vowels with smaller mouth opening, /oe/ and /y/ are more directional than /u/ and /o/ (see figures 26, 27).

In the frequency range from 1250 Hz to 2500 Hz the polar diagrams still look relatively smooth. The vowels /e/, /i/, /oe/, /y/ show at 1250 Hz on the transverse plane an attenuation at about 30, 150, 210 and 330 degrees. In addition, the characteristics are very wide to the sides. At 1600 Hz the representation of the sound propagation of vowel /i/ looks almost backwards. Vowel /a/ shows a dipole characteristic at 2000 Hz on the transverse plane and the frontal plane. Vowel /i/ has a comparable tendency at 2500 Hz. The other vowels show a slight downward directionality on the frontal plane in the entire frequency range. On the median plane, all vowels show a more diffuse sound propagation, with the coarse directionality still pointing forward and downward. No systematic relation to the size of the mouth opening is discernible (see figures 28, 29).

Between 3150 Hz and 6300 Hz, the polar diagrams of sound propagation for all vowels

look quite similar, pointing forward on the transverse plane, downward on the frontal plane, and forward and downward on the sagittal plane at about 130 degrees. The vowels with larger mouth opening /ae/, /e/, /a/, /i/ are more strongly directed than the other vowels. At 6300 Hz the sound propagation of vowel /u/ is most non-directional (see figures 30, 31).

Above 8000 Hz all vowels show a similar behavior, with directionality increasing. Again, directionality is higher for vowels with a larger mouth opening. The diffuse radiation pattern for vowel /u/ on all levels at 10000 Hz indicates errors in the simulation (see figures 32, 33). No values could be obtained for higher frequencies in the wovt models.

In summary, all vowels at low frequencies have an almost omnidirectional sound propagation on all levels. At higher frequencies, the directionality for all vowels increases. Vowels with a larger mouth opening have a higher directionality than vowels with a smaller mouth opening. On the median plane, the sound propagation is slightly tilted downwards, which is partly due to the tilting of the vowel tract and partly and to a greater extent to reflections from the torso. This has also been observed in measurements with microphones in other research projects. Otherwise, propagations can be seen at some frequencies for different vowels, which rather indicate errors in the simulation. However, a regularity is not recognizable, because different models are affected.

The results found can be compared with findings from other research. In research using microphone measurements of the sound propagation of the human voice, Katz et al. 2006 and 2007 find that the radiation patterns of vowels differ in the mid-frequency range between 1000 Hz and 1600 Hz and are otherwise the same [4], [3]. Below 1000 Hz, all vowels have omnidirectional radiation patterns due to the relationship of wavelength and head geometry, and above 1600 Hz they become more directional at higher frequencies. This is also evident in the present work, except that the range with different characteristics here is between 630 Hz and 2000 Hz. The individual vowels do not differ very much from each other in the works. Also, the attenuation at 0 degrees on the transversal plane, can be seen at 1000 Hz or 1250 Hz in the polar diagrams in both studies. This phenomenon is also evident in the work of Marshall et al [24].

In this thesis, it is also recognized that the maxima of the sound pressure levels are directed downward by 20 degrees on the median plane. This is reported in a very early paper by Dunn et al as well [25]. and is attributed to reflections from the torso. Furthermore, in the present work, the direction of sound propagation is below the transverse plane, but rather at -45 degrees.

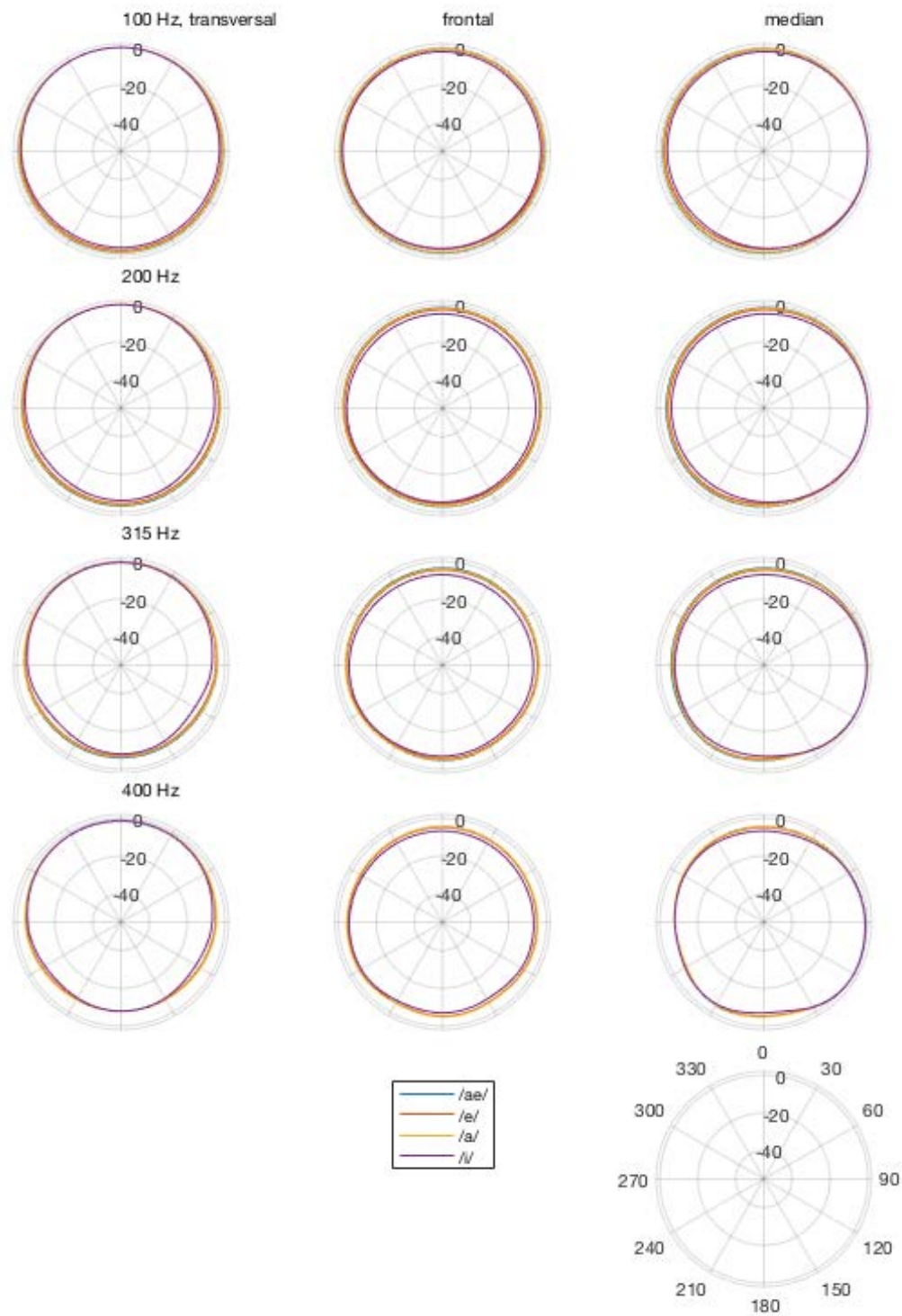


Figure 24: Polar diagrams of the radiation pattern of the vowels /ae/, /e/, /a/, /i/, third-band averaged in the frequency range from 100 Hz to 400 Hz, peak normalized to 0 dB.

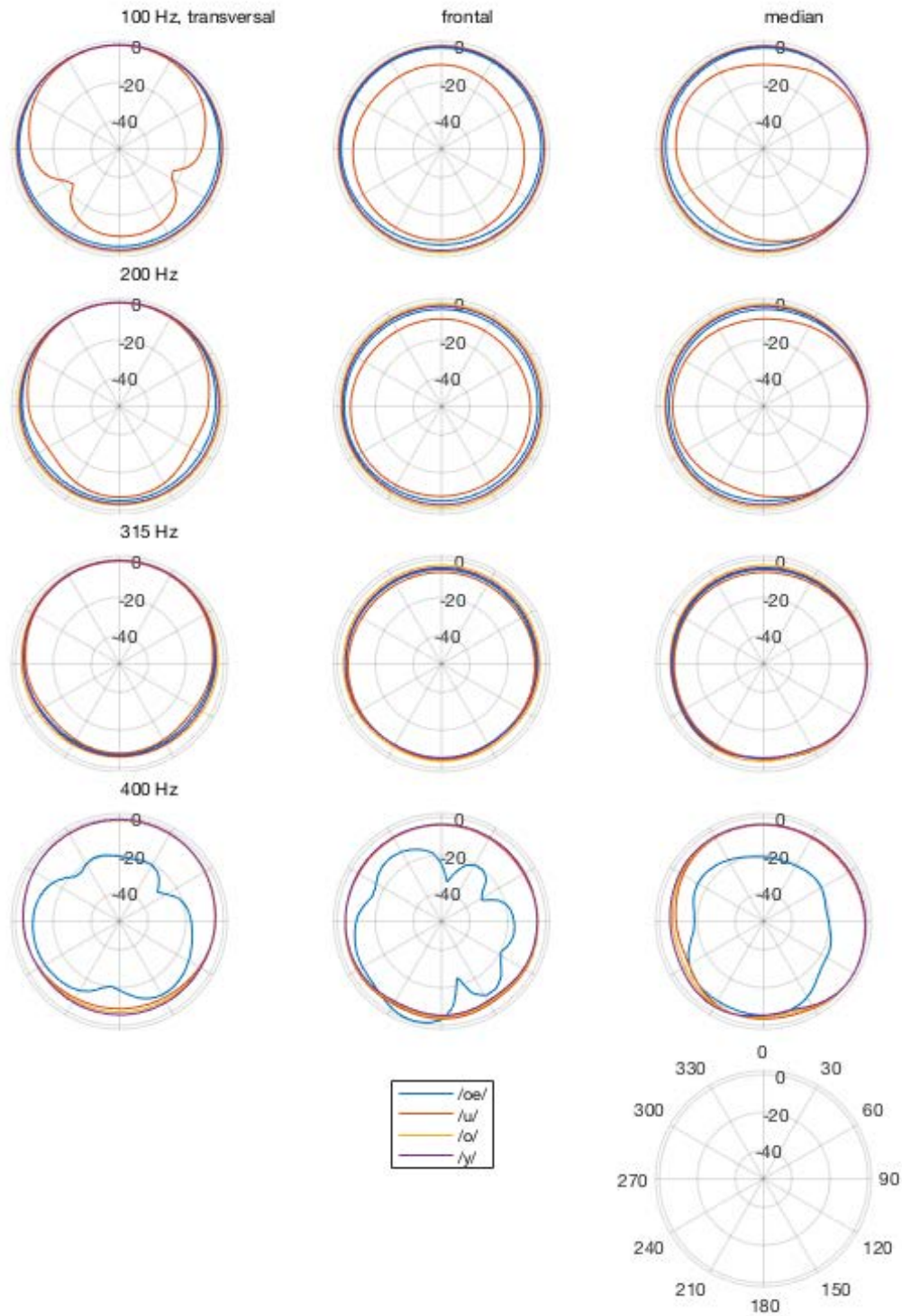


Figure 25: Polar diagrams of the radiation pattern of the vowels /oe/, /u/, /o/, /y/, third-band averaged in the frequency range from 100 Hz to 400 Hz, peak normalized to 0 dB.

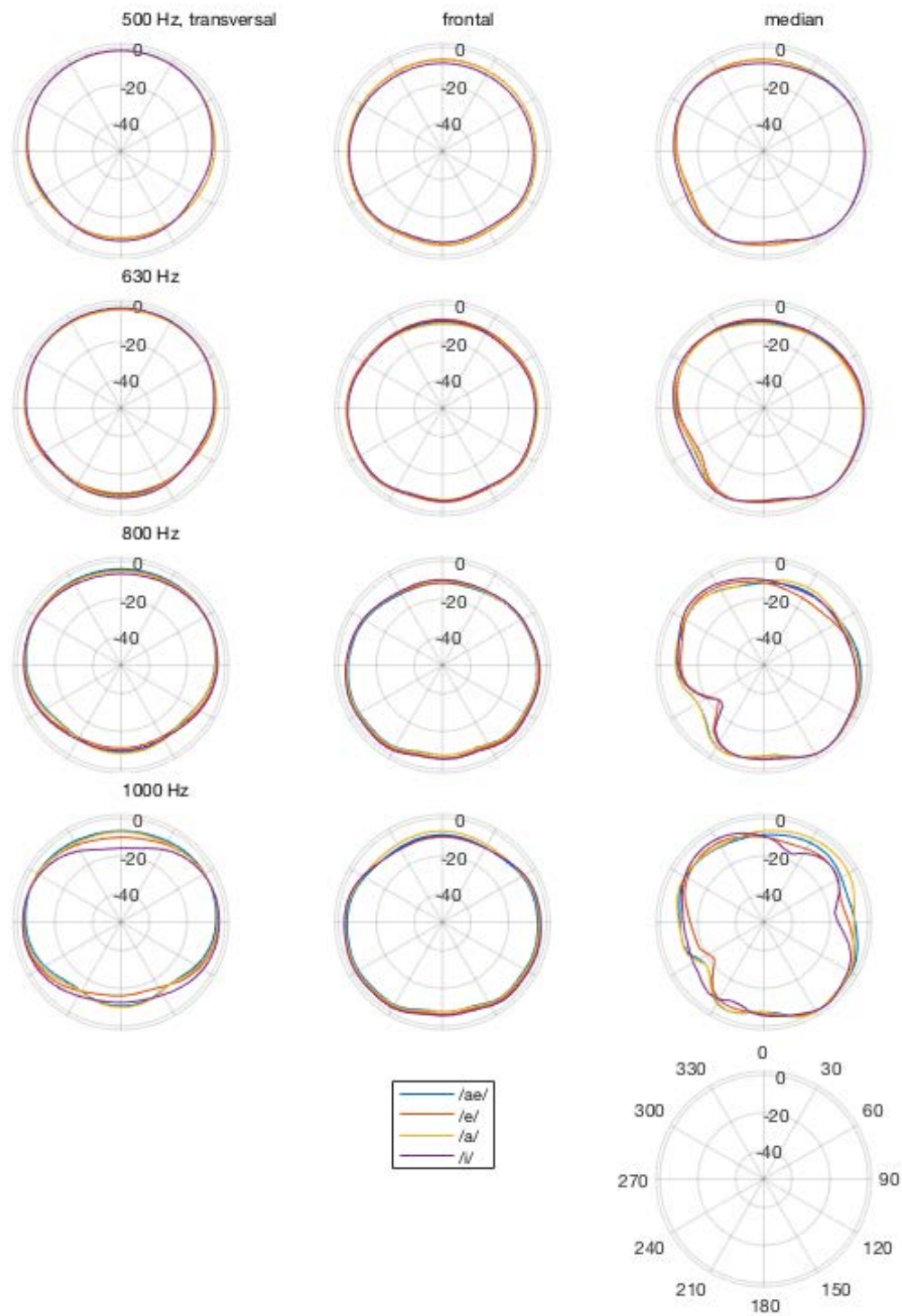


Figure 26: Polar diagrams of the radiation pattern of the vowels /ae/, /e/, /a/, /i/, third-band averaged in the frequency range from 500 Hz to 1000 Hz, peak normalized to 0 dB.

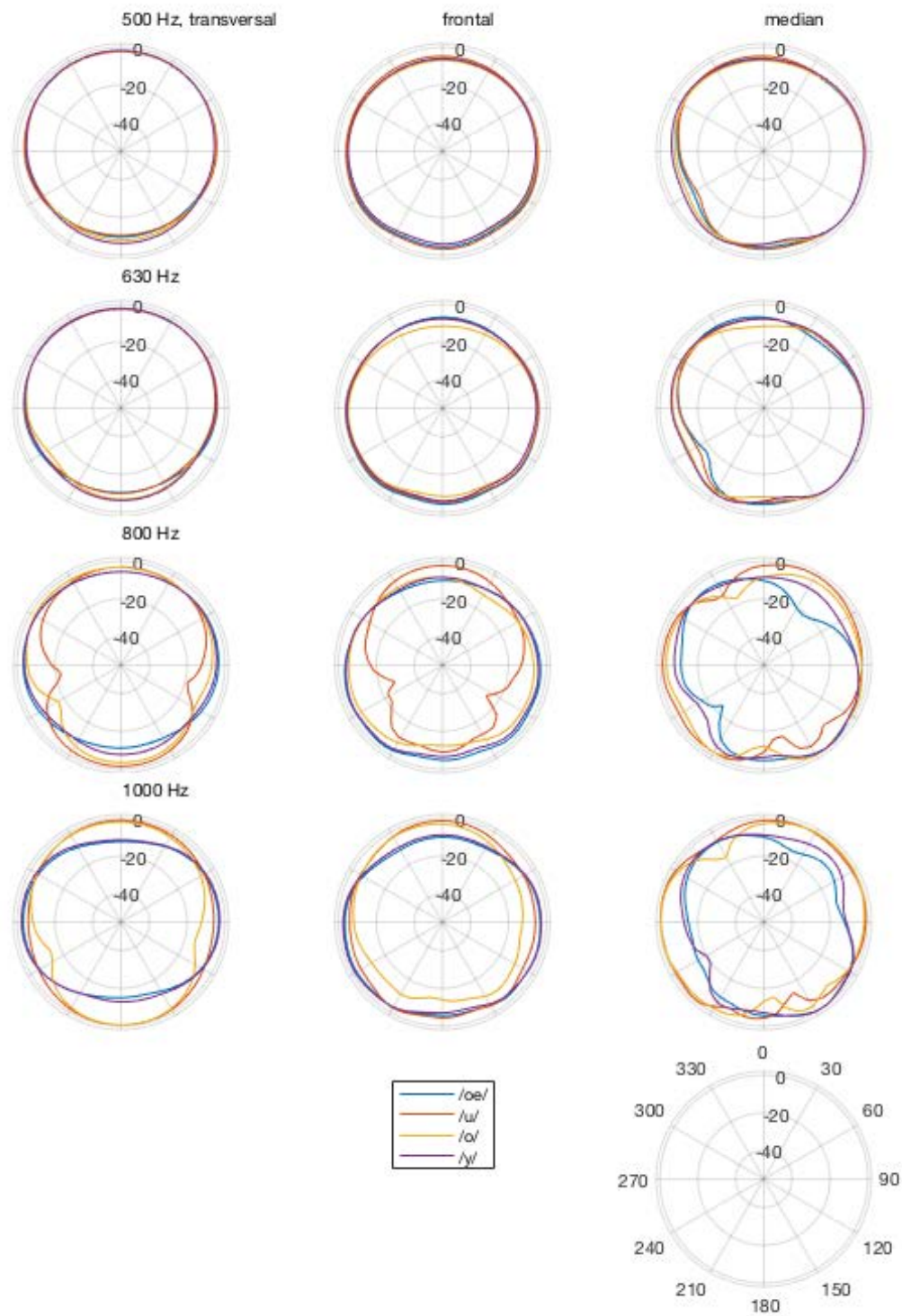


Figure 27: Polar diagrams of the radiation pattern of the vowels /oe/, /u/, /o/, /y/, third-band averaged in the frequency range from 500 Hz to 1000 Hz, peak normalized to 0 dB.

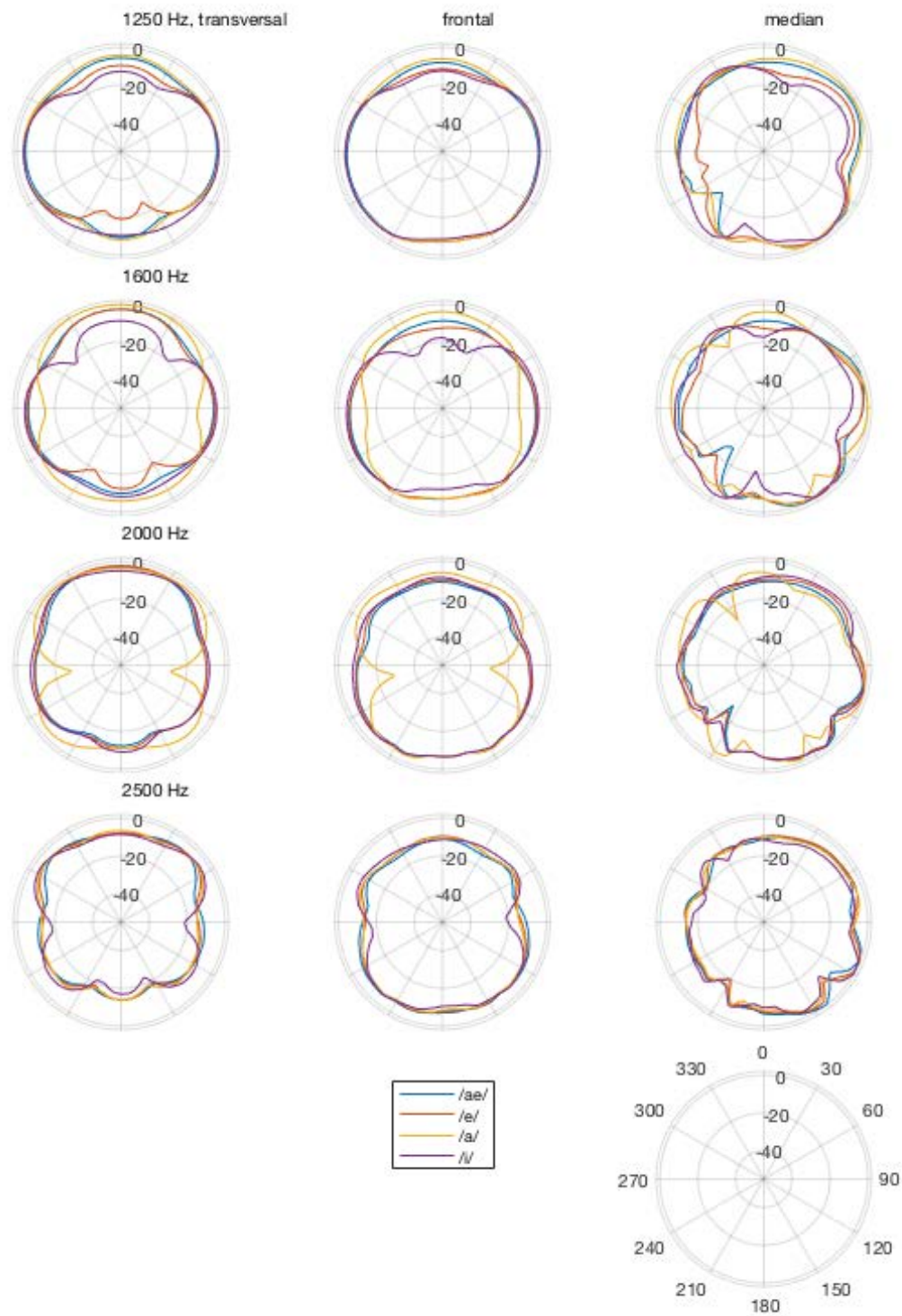


Figure 28: Polar diagrams of the radiation pattern of the vowels /ae/, /e/, /a/, /i/, third-band averaged in the frequency range from 1250 Hz to 2500 Hz, peak normalized to 0 dB.

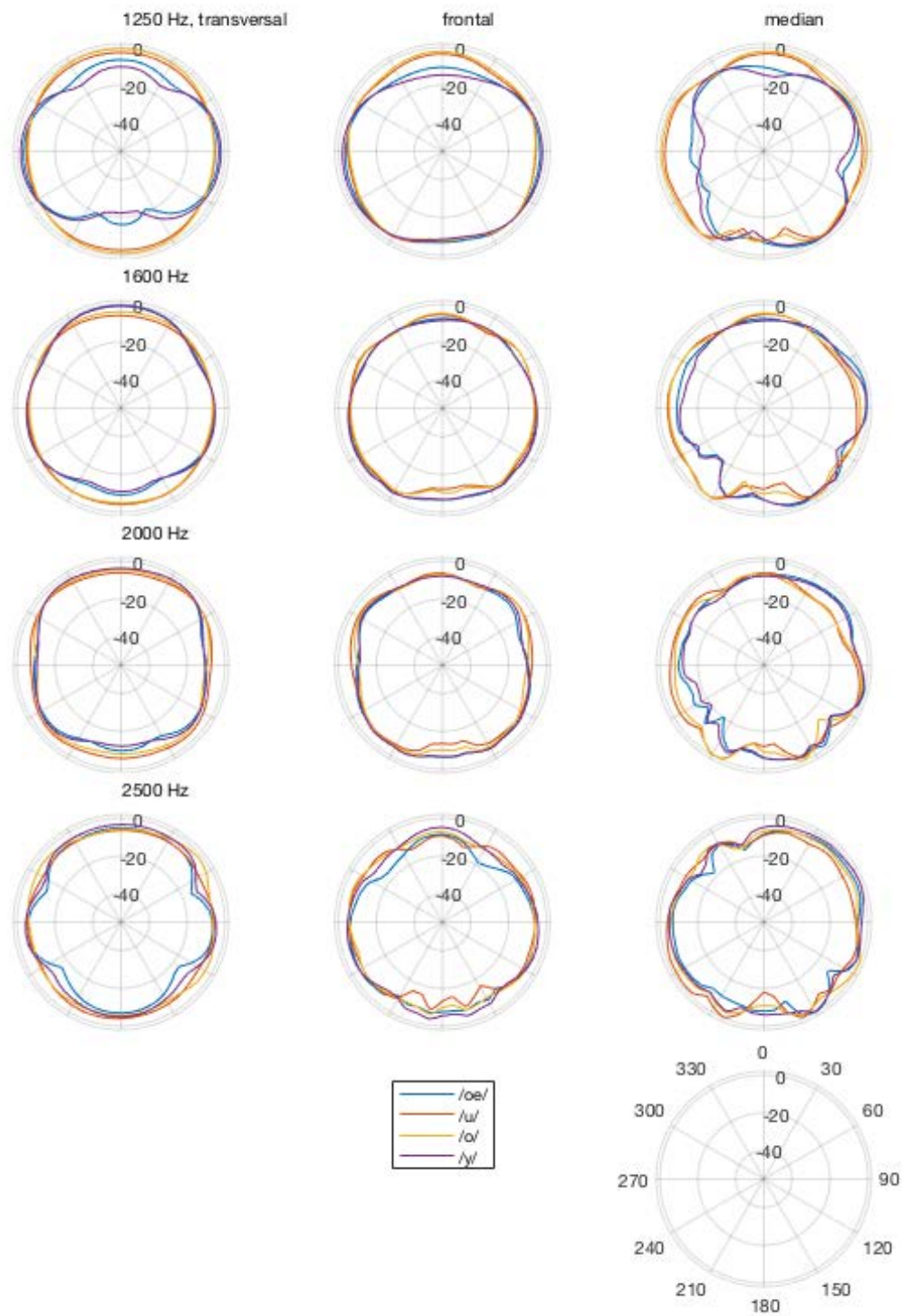


Figure 29: Polar diagrams of the radiation pattern of the vowels /oe/, /u/, /o/, /y/, third-band averaged in the frequency range from 1250 Hz to 2500 Hz, peak normalized to 0 dB.

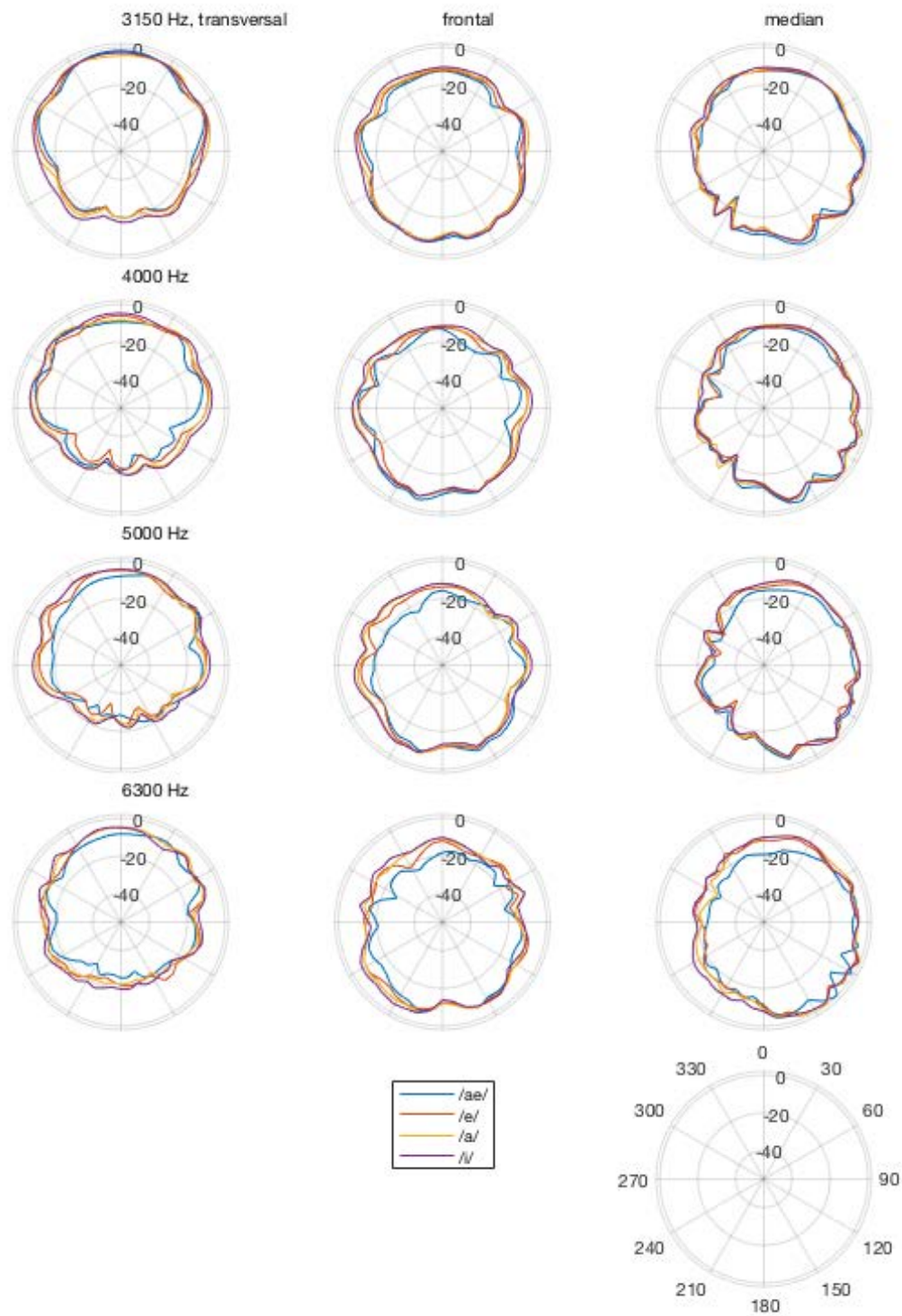


Figure 30: Polar diagrams of the radiation pattern of the vowels /ae/, /e/, /a/, /i/, third-band averaged in the frequency range from 3150 Hz to 6300 Hz, peak normalized to 0 dB.

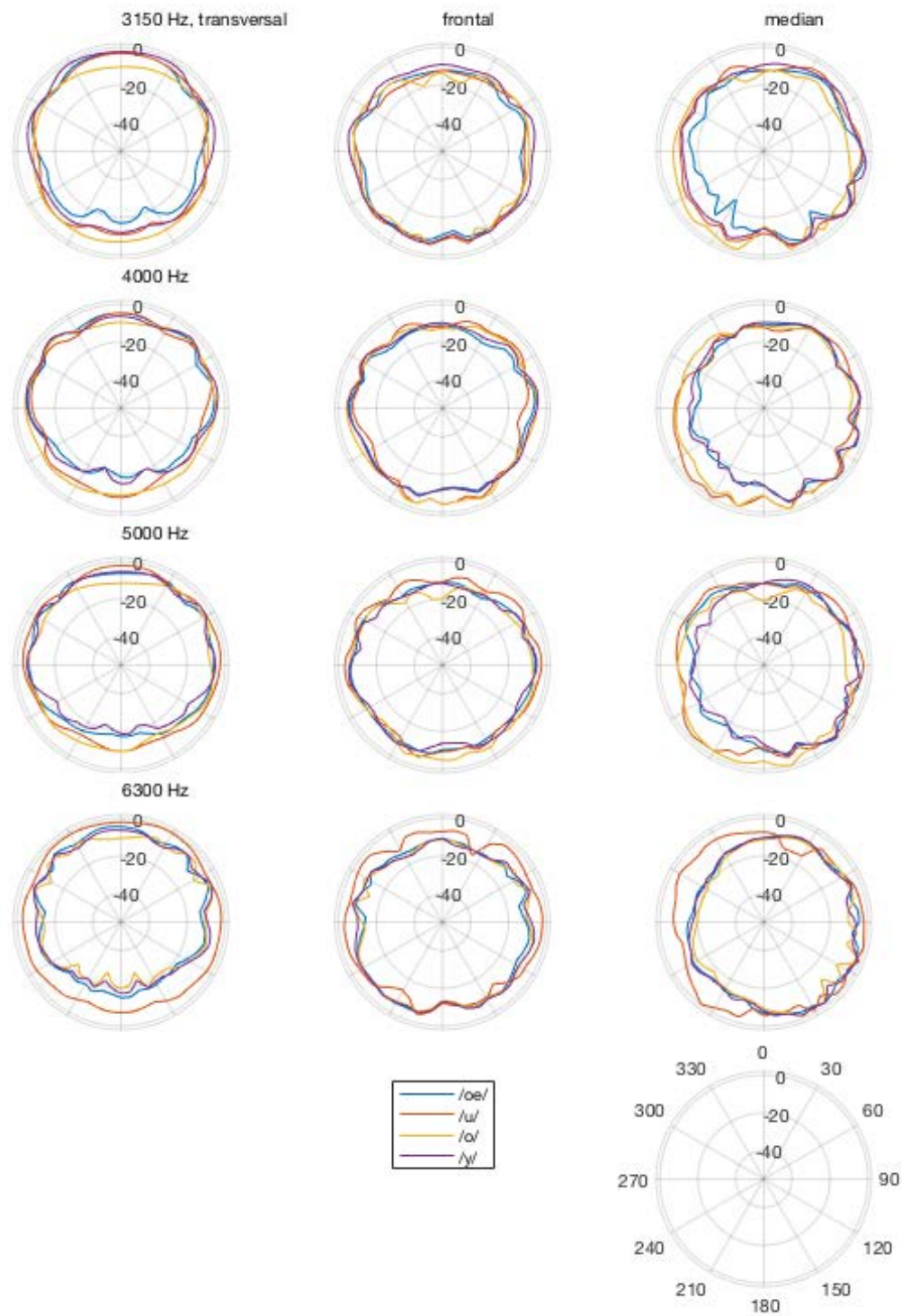


Figure 31: Polar diagrams of the radiation pattern of the vowels /oe/, /u/, /o/, /y/, third-band averaged in the frequency range from 3150 Hz to 6300 Hz, peak normalized to 0 dB.

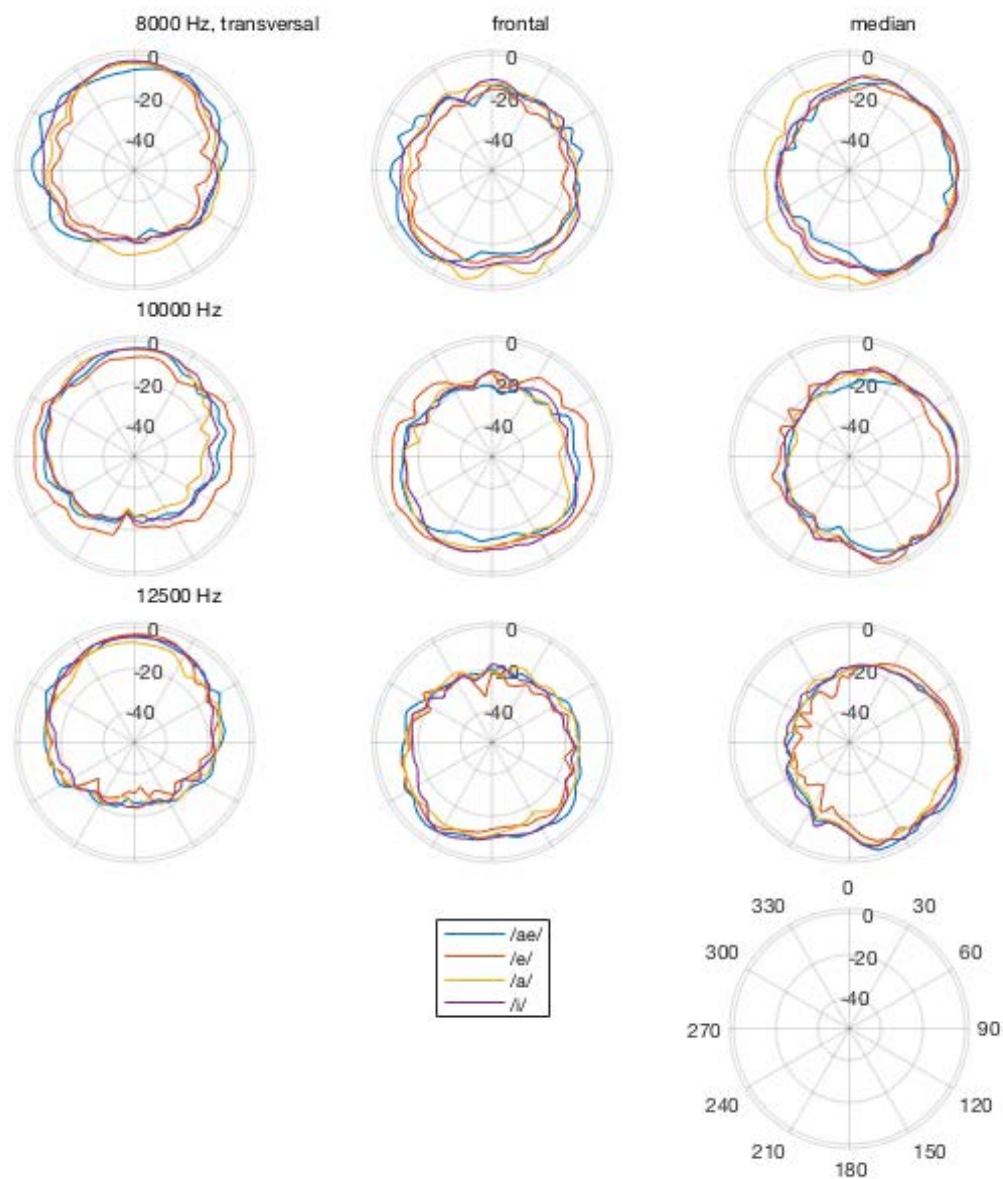


Figure 32: Polar diagrams of the radiation pattern of the vowels /ae/, /e/, /a/, /i/, third-band averaged in the frequency range from 8000 Hz to 12500 Hz, peak normalized to 0 dB.

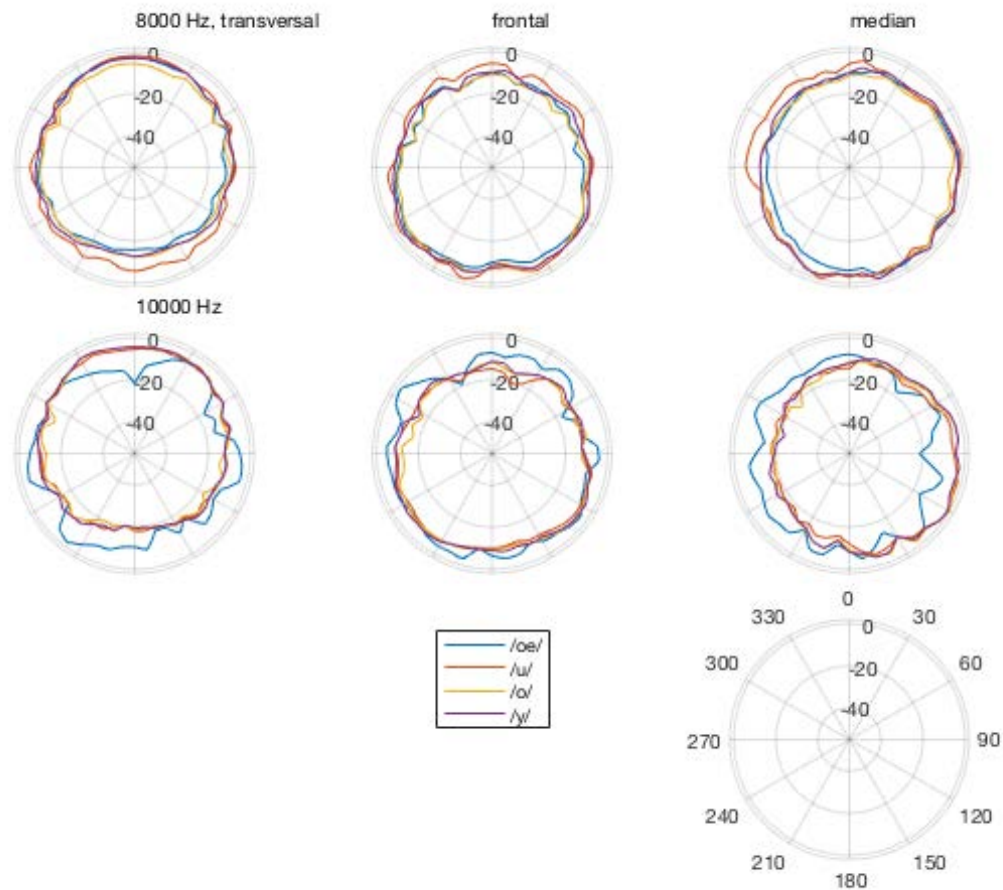


Figure 33: Polar diagrams of the radiation pattern of the vowels /oe/, /u/, /o/, /y/, third-band averaged in the frequency range from 8000 Hz to 10000 Hz, peak normalized to 0 dB.

To give an overview of the directionality of sound propagation of the individual models, the Directivity Index (DI) is calculated as a one-third octave band average [26]. This is done fully spherically over the entire sphere surface of the Lebedev grid used (see section 3.4.1). First, the maximum sound pressure level \tilde{P}_{\max} of the sampling points from 30 to 60° below the horizontal plane with an azimuth angle of $\pm 30^\circ$ is evaluated for each third octave band. This window is selected because other research also determines the main radiation direction there [27], [25] and the polar diagrams also point to this direction (see figures 24 to 33). The Directivity Factor (DRF) is calculated as:

$$\Gamma = \frac{\tilde{P}(\varphi, \vartheta)}{\tilde{P}_{\max}} \quad (2)$$

For the calculation of the DI the area weights AW of the Lebedev grid are used as weighting in the numerical integration. The DI thus calculates to:

$$DI = \frac{1}{\sum AW \Gamma^2} \quad (3)$$

The mean DI is determined via the DRF of the third-octave band averaged values from 100 Hz to 10000 Hz, since results can be calculated for all simulations in this range. This procedure prevents high DI at 12500 Hz or 16000 Hz from raising the mean DI of a simulation, thus making comparison with other simulations impossible.

When comparing the calculated DI of the wvt models and the wovt models, the structure of the models should be outlined once again in order to better classify the results. The wvt models include the MRI scans of the vocal tract. The wovt models are closed at the mouth opening but have the same source width and jaw positioning as the corresponding wvt models. Thus, the results of the wovt models can be used to estimate the influence of jaw position and source width on the directionality of sound propagation. The comparison with the wvt models shows the influence of the vocal tract.

In all simulations the directionality increases towards higher frequencies, which is due to the acoustic shadow of the head. Furthermore, the directionality decreases as the mouth opening decreases. This is especially visible in the wovt models, where the directionality of the individual vowels decreases in the same order as the mouth opening. Vowel /ae/ has the largest mean DI and vowel /y/ the smallest. The influence of the vocal tract varies depending on the vowel. For all the vowels the directionality is slightly higher in the wvt models than in the wovt models. At frequencies below 1000 Hz the DI of the different wovt models do not differ greatly. Here, however, the vocal tract has an effect, since the values of the different wvt models differ below 1000 Hz.

In general, the results are comparable to the findings of Katz et al. who measured a greater directionality at vowel /a/ compared to vowel /o/ when using microphone arrays [3]. The fact that the vowel /i/ has the smallest directionality cannot be confirmed by the simulations in this thesis. The simulations in this thesis are about Finnish vowel tract configurations and vowel /y/ is possibly comparable to the English vowel /i/. This would support the measured results of Katz et al. The results of Kokkon et al. are similar to those from the simulations [7] (see section 2). Again, vowels with a larger mouth opening show a higher directionality than vowels with a smaller mouth opening.

3rd-oct. f	Vowel wvt	/ae/	/e/	/a/	/i/	/oe/	/u/	/o/	/y/
	100 Hz	0.6	0.9	0.7	2.0	1.4	5.5	0.4	0.7
	200 Hz	1.3	2.0	1.6	3.5	2.6	4.9	0.9	1.6
	315 Hz	3.0	3.6	3.3	4.7	3.8	4.5	2.6	3.3
	400 Hz	4.0	4.2	4.0	4.9	4.4	3.9	3.6	3.6
	500 Hz	4.6	4.5	4.6	4.9	4.4	4.0	4.6	4.2
	630 Hz	5.1	4.9	5.3	5.0	4.7	4.8	5.2	4.7
	800 Hz	5.8	5.6	5.7	5.4	5.3	-0.4	4.9	5.2
	1000 Hz	5.8	5.9	5.2	5.6	5.9	-0.7	-0.1	5.2
	1250 Hz	5.4	6.3	4.4	6.3	5.9	3.5	2.4	5.6
	1600 Hz	5.6	6.2	1.5	7.0	4.4	4.4	4.1	4.4
	2000 Hz	7.0	5.8	2.8	5.9	4.8	2.8	2.1	3.0
	2500 Hz	7.9	6.5	6.3	7.1	6.2	5.1	3.9	4.4
	3150 Hz	9.0	9.2	8.5	8.2	8.5	6.5	4.8	6.3
	4000 Hz	10.2	10.0	9.0	8.7	8.2	5.5	4.9	7.6
	5000 Hz	11.0	10.4	9.7	9.1	7.8	5.4	3.0	8.4
	6300 Hz	10.7	9.2	8.5	8.4	7.4	4.5	7.1	6.7
	8000 Hz	9.1	10.6	7.0	8.8	7.6	3.8	6.3	6.0
	10000 Hz	10.8	7.7	9.4	8.0	3.3	7.2	6.4	6.5
	12500 Hz	9.6	10.5	11.9	9.6		3.8	7.2	8.1
	16000 Hz							1.6	10.4
	mean 100 Hz - 10 kHz	7.5	7.1	6.3	6.7	5.8	4.5	4.2	5.3

Table 11: The DI in dB is calculated for the wvt models in third-octave bands. The average DI of the one-third octave bands with center frequencies from 100 Hz to 10000 Hz is derived from the average DRF of these bands.

Vowel wovt 3rd-octave f		/ae/	/e/	/a/	/i/	/oe/	/u/	/o/	/y/
100 Hz		0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
200 Hz		0.9	0.9	0.9	1.0	0.9	0.9	0.9	0.8
315 Hz		2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.5
400 Hz		3.9	3.9	3.9	3.9	3.8	3.8	3.8	3.8
500 Hz		4.5	4.5	4.5	4.5	4.4	4.4	4.4	4.4
630 Hz		5.0	5.1	5.1	5.1	5.1	5.1	5.1	5.0
800 Hz		5.7	5.6	5.6	5.6	5.6	5.6	5.6	5.5
1000 Hz		5.6	5.2	5.1	5.0	5.1	5.1	5.0	5.0
1250 Hz		5.1	4.6	4.4	4.4	4.4	4.3	4.2	4.2
1600 Hz		5.5	4.8	4.5	4.5	4.5	4.5	4.2	4.3
2000 Hz		6.4	5.4	4.9	4.7	5.0	5.0	4.7	4.7
2500 Hz		7.5	6.2	6.0	6.0	5.7	5.5	5.8	5.6
3150 Hz		8.0	8.7	8.5	8.2	8.3	7.9	8.0	7.8
4000 Hz		8.9	8.8	8.5	8.1	8.1	7.8	7.7	7.4
5000 Hz		9.6	9.5	9.3	8.7	8.7	8.4	8.4	8.1
6300 Hz		9.7	8.6	8.6	8.2	7.8	7.0	7.3	7.1
8000 Hz		11.2	9.9	8.8	8.2	7.5	6.3	7.1	6.7
10000 Hz		12.1	10.4	9.3	7.5	7.2	7.8	6.8	6.5
12500 Hz		13.4	12.3	11.2	8.4	8.2	8.3	8.1	7.7
16000 Hz		8.5	0.6	7.6	5.8	8.7	7.6	5.0	9.2
mean 100 Hz - 10 kHz		7.4	6.7	6.4	6.0	5.9	5.6	5.6	5.5

Table 12: The DI in dB is calculated for the wovt models in third-octave bands. The average DI of the one-third octave bands with center frequencies from 100 Hz to 10000 Hz is derived from the average DRF of these bands.

4.4. Error estimation

The results of the simulations are tested for reproducibility. For this purpose, the simulations for vowel /i/ and for vowel /oe/ are performed twice, each as a wvt model. The selection was determined in such a way, because for vowel /i/ all BEM simulations of frequencies up to 16000 Hz iterate, whereas this was not the case for vowel /oe/. Vowel /oe/ also looks error-prone in the other results. The models, as well as the export settings with associated source positioning remain the same. For the test-retest deviation, all one-third octave band mean values of the Lebedev Grid are now compared and the largest difference per one-third octave band is output (see figure 34). For vowel /i/ the deviation is 0 dB over the whole simulation. For the error-prone vowel /oe/ the test-retest deviation is also 0 dB for all frequencies up to 10000 Hz. The maximum deviation of the third-octave band averaged sound pressure levels at 10000 Hz for the vowel /oe/ is 87.9 dB. This underlines the error-proneness at high frequencies and shows at the same time that the test-retest deviation of the simulations is otherwise very small.

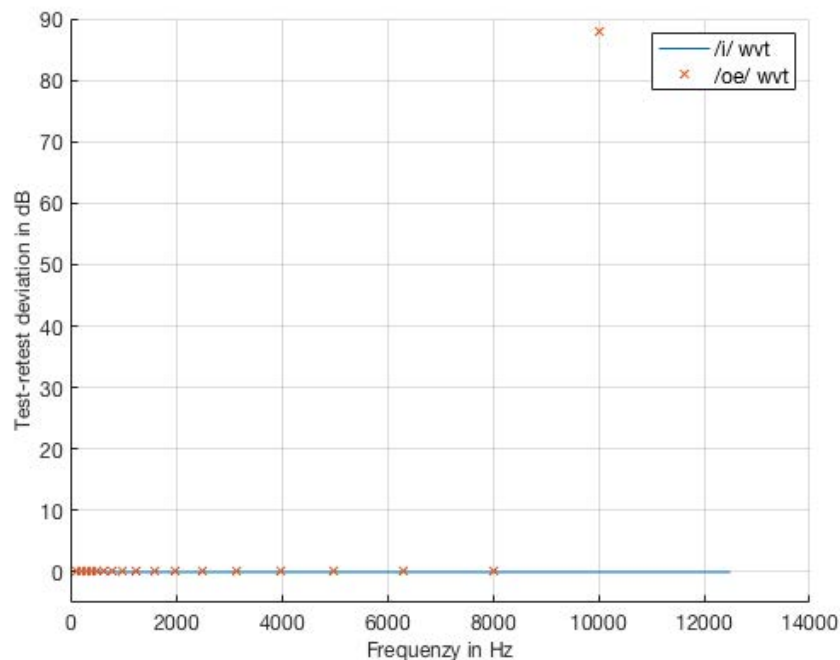


Figure 34: The test-retest deviation is performed for the wvt simulations /i/ and /oe/. The deviation of the third-octave band averaged sound pressure level values of the whole Lebedev-grid is calculated.

Furthermore, the influence of the positioning of the Sample Grid in front of the mouth opening was tested. The wvt model for vowel /a/ was used to compare the sound pressure levels of all five points of the sample grid in front of the mouth opening. The highest

difference is output for each frequency. This procedure was chosen because an incorrect positioning of the sample grid should also be in the range of the distance between the individual sample points on the grid.

The Sample Grid in front of the mouth opening is especially important for the simulation of the spectrum. The main focus is on the position of the formants. In the frequency range up to 4000 Hz the maximum deviations are between 1.7 dB and 5.8 dB. The maximum deviation in this frequency range is 3600 Hz. When considering the whole frequency range, the deviations are between 1.2 dB and 23.8 dB. Here, the differences are greatest at high frequencies above 10000 Hz (see figure 35). Thus, the positioning of the sample grid in front of the mouth opening does not have a strong effect when viewing the formant spectra, but it does have an effect on the values of the sound pressure levels in the higher frequency range.

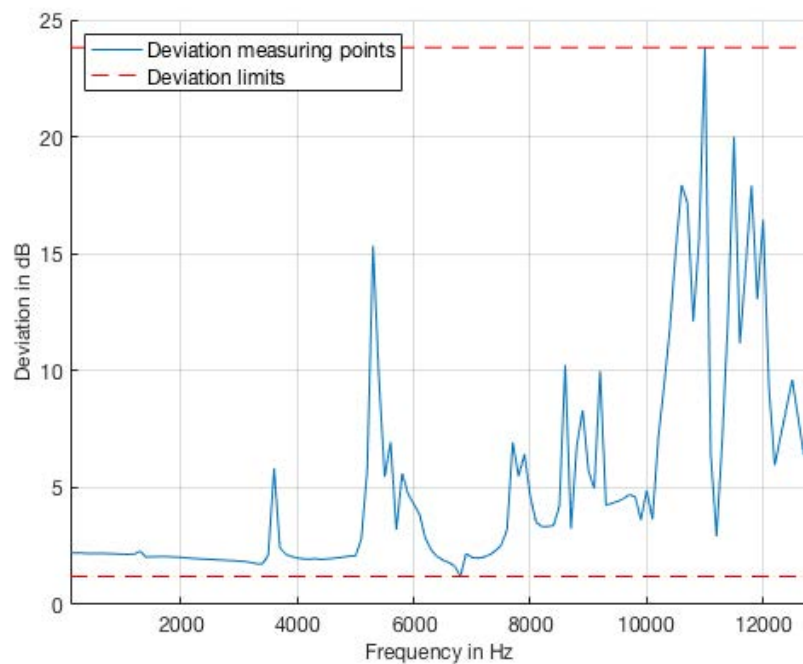


Figure 35: To investigate the spectral influence of the positioning of the sample grid in front of the mouth opening, the maximum difference of the calculated sound pressure levels of the different sample points on the grid is calculated for the wvt model of vowel /a/.

5. Discussion

The present work shows how, using BEM with the simplification of rigid boundaries and conphase moving radiation areas, sound propagation can be simulated for different phonemes. The results are fully spherical with a high angular resolution in contrast to microphone measurements.

For the eight Finnish vowels /ae/, /e/, /a/, /i/, /oe/, /u/, /o/, /y/ models can be created by combining in Blender MRI scans of the vocal tract with mouth image with surface scans of the FABIAN artificial head. These are the wvt models. Furthermore, for each vowel a model with a separated vocal tract is created as wovt model.

The remeshing is done with Meshmixer and the pmp-library. These models will also be provided for future work.

The simulation of sound propagation is calculated in the frequency range up to 16000 Hz, although the solutions by BEM do not iterate at some frequencies for some models.

When considering the simulated spectrum at the mouth opening, the wovt models show equal behavior of all vowels and a relatively smooth frequency response. In the wvt models, on the other hand, the frequency responses are highly individual. In the range up to 4000 Hz, the position of the formants can be determined. The localization of the formants shows similarities to the formants of the audio recordings during the MRI scans. Thus, BEM can largely simulate sound propagation in the vocal tract even with the simplification of rigid boundaries with respect to formant structure.

Looking at the sound propagation, the wvt and the wovt models can be compared. Below 500 Hz, the directionality is omnidirectional for all models. Above 500 Hz, the vocal tract has an effect on sound propagation, but the results tend to indicate a shift and stretching of the characteristics. The basic shapes remain the same.

Using polar plots, the sound propagation of the wvt models can be illustrated on the transverse, frontal and median planes. Thus, a comparison of the different vowels is possible. Here, the vowels exhibit different characteristics in the frequency range from 630 Hz to 2000 Hz. Below this frequency range, all vowels exhibit omnidirectional radiation patterns and above it, directionality that faces forward and is tilted downward. The results compare well with other research using microphone measurements.

In addition, the calculation of the DI shows an increase in directionality for all models toward higher frequencies and a decrease in directionality as the mouth opening is reduced. This is seen for both the wvt models and the wovt models. When looking at the DI, an influence of the vocal tract on the directionality in frequency bands below 1250 Hz is also visible. The results are consistent with other research.

With regard to the research questions of this thesis, it can be summarized that the different mouth openings of the different phonemes affect the radiation characteristics in the sense that phonemes with larger mouth openings show a higher directionality than phonemes with smaller mouth openings. In addition, the vocal tract has an effect on sound propagation. However, these effects are more likely to be seen in the fact that the direction of sound propagation and the extent of the radiation pattern are somewhat affected. Despite this, the radiation patterns look quite similar in the comparison with and without vocal tract despite this.

Even with the simplification of the rigid boundaries and conphase moving radiation areas, the sound propagation through the vocal tract for the human voice can be simulated well with BEM. In another study it is already shown that only a part of the test subjects prefers a static or dynamic directivity of a loudspeaker to an omnidirectional radiation pattern [28]. In view of the results and the computational effort in this thesis, it should be considered whether the calculation of sound propagation in the vocal tract can be omitted in future applications. Here, it would be a compromise to consider the source expansion as well. Otherwise, the individual spectrum of the speech material of a speaker directs the directionality of the basic geometry.

A. Digital appendix

References

- [1] W Chu and A Warnock. “Detailed directivity of sound fields around human talkers”. In: (2002).
- [2] T Halkosaari, M Vaalgamaa, and M Karjalainen. “Directivity of artificial and human speech”. In: *Journal of the Audio Engineering Society* 53.7/8 (2005), pp. 620–631.
- [3] B Katz and C d’Alessandro. “Directivity measurements of the singing voice”. In: 2007.
- [4] B Katz, F Prezati, and C 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.
- [5] F Brinkmann, A Lindau, S Weinzierl, S van de Par, M Müller-Trapet, R Opdam, and Michael Vorländer. “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 published Version 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>.
- [6] H Ziegelwanger, W Kreuzer, and P 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.
- [7] P Kocon and 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] B Monson, A Lotto, and S 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] B Monson, E Hunter, and B 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] M Arnela, O Guasch, and F 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] M Arnela, R Blandin, S Dabbaghchian, O Guasch, F Alías, X Pelorson, A Van Hirtum, and O Engwall. “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] D Sieger and M Botsch. *The Polygon Mesh Processing Library*. <http://www.pmp-library.org>. Viewed on 10.11.2020. URL: <http://www.pmp-library.org>.
- [15] *MathWorks MATLAB*. <https://de.mathworks.com/products/matlab.html>. Viewed on 16.04.2020. URL: <https://de.mathworks.com/products/matlab.html>.
- [16] *International Phonetic Association*. <https://www.internationalphoneticassociation.org>. Viewed on 16.04.2020. URL: <https://www.internationalphoneticassociation.org>.
- [17] *Speech & Math*. <http://speech.math.aalto.fi/data.html>. Viewed on 16.04.2020. URL: <http://speech.math.aalto.fi/data.html>.
- [18] R Ciskowski and C Brebbia. *Boundary element methods in acoustics*. Springer, 1991.
- [19] B Bernschütz, C Pörschmann, S Spors, and S Weinzierl. “SOFiA sound field analysis toolbox”. In: *Proceedings of the International Conference on Spatial Audio (ICSA)*. 2011, pp. 7–15.
- [20] P Majdak, Y Iwaya, T Carpentier, R Nicol, M Parmentier, A Roginska, Y Suzuki, K Watanabe, H Wierstorf, H Ziegelwanger, and M Noisternig. “Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions”. In: *Audio Engineering Society Convention 134*. Audio Engineering Society. 2013.
- [21] *Mesh2HRTF on SOURCEFORGE*. <https://sourceforge.net/projects/mesh2hrtf/>. Viewed on 20.10.2020. URL: <https://sourceforge.net/projects/mesh2hrtf/>.
- [22] *High-Performance-Computing-Cluster (HPC-Cluster) TU-Berlin*. <https://hpc.tu-berlin.de/doku.php?id=hpc>. Viewed on 20.10.2020. URL: <https://hpc.tu-berlin.de/doku.php?id=hpc>.
- [23] F Brinkmann and S Weinzierl. “Aktools – an open software toolbox for signal acquisition, processing, and inspection in acoustics”. In: *Audio Engineering Society Convention 142*. Audio Engineering Society. 2017.
- [24] A Marshall and J Meyer. “The directivity and auditory impressions of singers”. In: *Acta Acustica united with Acustica* 58.3 (1985), pp. 130–140.
- [25] H Dunn and D Farnsworth. “Exploration of pressure field around the human head during speech”. In: *The Journal of the Acoustical Society of America* 10.3 (1939), pp. 184–199.
- [26] *Akustische Wellen. Felder: DEGA-Empfehlung 101, Deutsche Gesellschaft für Akustik*. 2006.

- [27] C Pörschmann and J Arend. “A Method for Spatial Upsampling of Voice Directivity by Directional Equalization”. In: *Journal of the Audio Engineering Society* 68.9 (2020), pp. 649–663.
- [28] J Ehret, J Stienen, C Brozdowski, A Bönsch, I Mittelberg, M Vorländer, and T Kuhlen. “Evaluating the Influence of Phoneme-Dependent Dynamic Speaker Directivity of Embodied Conversational Agents’ Speech”. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 2020, pp. 1–8.