Experimental investigation of speech directivity mechanisms

Rémi Blandin, Jingyan Geng, Peter Birkholz

16ème Congrès Français d'Acoustique
11-15 Avril 2022, Marseille

Experimental investigation of speech directivity mechanisms

R. Blandin[a], J. Geng[a], and P. Birkholz[a]

[a]Institute of Acoustics and Speech Communication, TU Dresden, Dresden 01062, Germany

Speech directivity has been often investigated by measuring the sound radiated by human speakers at multiple directions. However, little work has been done to understand the mechanisms influencing the shape of the directivity patterns. This work aims at better understanding the contribution of the head and torso, the lips and the vocal tract geometry. For this purpose, a simplified head and torso replica was used that consists of two spheres that represent the head and the torso. The vocal apparatus was simulated either by a simple loudspeaker or by 3D-printed MRI-based vocal tract replicas corresponding to different vowels and genders with and without lips. The radiation patterns of this head and torso simulator were measured with a turntable and a microphone placed at different angular positions. The different contributions of the torso, the head, the lips and the vocal tract shape were identified in different frequency bands.

1 Introduction

Speech directivity induces a change of amplitude and frequency content of speech sounds with the direction. Studying it is important for applications such as virtual reality or fundamental research. Its properties have been measured by several authors on real subjects in anechoic rooms, see as an example [17, 18, 11, 16, 19]. It was observed that the directionality of speech increases with the frequency and that it depends on the phoneme, the articulation and the anatomy of the subjects.

Several studies relying on theoretical models, measurements on replicas and real subjects have shed some light on the parameters influencing speech directivity.

The simplest and earliest model for predicting speech radiation properties is a vibrating piston [10, 14]. It predicts that larger mouth dimensions induce a more directional radiation. However, this correlates with measurements of real subjects only in the octave bands at 2 kHz and 4 kHz [15, 7, 19], and the predicted patterns are generally much simpler than the ones of real subjects.

Arnela et al. [2] simulated with finite elements the radiation of vocal tract geometries set in a head geometry with different degrees of simplification. They found that, except for the lips, the precise shape of the face has little impact on the radiated sound, and that approximating the face with a sphere while keeping the lips was almost equivalent to using a realistic face. Yoshinaga et al. [20] compared the radiation patterns of a vocal tract replica of the consonant /s/ set in a rectangular baffle with and without lips and found that the lips increase the directionality of the radiated sound.

Brandner et al. [9] compared the radiation patterns of a commercial head and torso simulator (HATS) measured with the head alone and with the head and torso together. This showed that the torso generates diffraction patterns appearing as side lobes in the horizontal plane.

Blandin et al. [5] predicted theoretically and observed by measurements with vocal tract replicas that higher-order modes (HOM) propagating inside the vocal tract at high frequency (from about 3.5 kHz) induce significant changes of directivity patterns within small frequency intervals (order of 100 Hz). This effect was predicted to be stronger for open vowels (e.g. /a/) than for closed vowels (e.g. /u/) [6]. This effect was confirmed experimentally with measurements of real subjects [9, 7].

No substantial differences related to the gender of the subjects were observed [22]. Monson et al. [18] found only that male speech was slightly more directional above 8 kHz.

Despite the better understanding due to the works previously mentioned, the impact of some parameters is still unknown. As an example, the effect of the lips have been observed only for the fricative /s/. On the other hand, some parameters have been investigated separately excluding others, and thus, preventing to observe potential interactions between them. As an example, the lips and the HOM effect have been studied only in geometries included in an infinite baffle without integrating them in a HATS.

The objective of this work was to investigate the influence of the torso, lips and the vocal tract using a HATS which integrates all these parameters simultaneously.

The used HATS consisted of two spheres simulating the head and the torso. Such a design is inspired from the work of Algazi et al. [1] who proposed to approximate the head and torso by two spheres for the computation of the head related transfer functions (HRTF). They found that such a configuration reproduces the major features of the HRTF. Beyond the advantage of the simplicity of the practical implementation, another advantage of such a design is that it can be compared with simple and fast computations of the radiated field [12] (not done in this work). To study the influence of the vocal tract, this simplified HATS was designed so that a simple loudspeaker or various realistic 3D printed vocal tract replicas could be inserted in the head. Vocal tract replicas corresponding to the vowels /a/, /i/ and /u/ of a male and a female subject with and without lips were used. The radiation patterns of these different configurations were measured in an anechoic room.

2 Method

2.1 Measurement setup

The HATS was consisted of two spheres of diameter 21 cm and 51 cm simulating the head and the torso respectively, as can be seen in Fig. 1. A tube made of unplasticised polyvinyl chloride (uPVC) with a diameter of 110 mm connects both spheres and simulates the neck. The spheres were made of polystyrene covered with a 5 mm coating of adhesive and reinforcing mortar (Maxit multi 300) to enhance their reflectivity. A vocal tract replica was placed inside the head. The sound was generated by a sound source connected to its glottal end. The sound source consisted of a loudspeaker enclosed in a casing whose design is

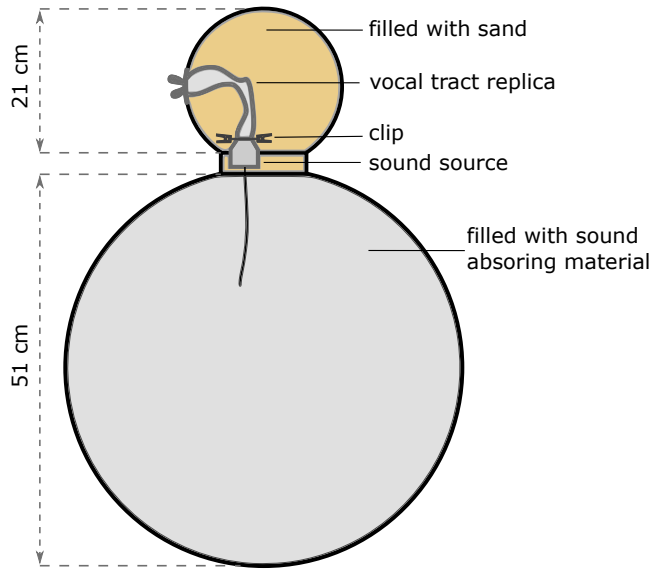

FIGURE 1 – Schematic of the head and torso simulator.

detailed in [13]. It was attached to the vocal tract replica using clips and a soft silicone seal to make the connection airtight. A loudspeaker could also be placed on the surface of the head instead of the vocal tract replica. The head was filled with sand in order to absorb as much as possible the vibrations generated by the sound source and the sound potentially radiated from the walls of the replica and the sound source. To further reduce the potential transmission of sound through other paths than the mouth of the vocal tract replica, the torso was filled with sound absorbing material and modelling clay was used to seal the gaps between the various elements of the HATS.

The vocal tract replicas are shown in Fig. 2. They have been designed from the geometries extracted from magnetic resonance images of real subjects provided in the Dresden Vocal Tract Dataset [1] [4]. The original geometries corresponding to the vowels /a/, /i/ and /u/ of the male and the female subject were modified to integrate them into the HATS and to create two versions with and without lips. Fixations were added to attach them inside the HATS and a spherical cap was added around the mouth opening to integrate them into the head of the HATS. In the case with lips, this flange was positioned at the corner of the lips. For the case without lips, it was positioned halfway between the corner of the lips and the most external part of the lips. This makes the effective length of the vocal tract about the same as for the replicas with the lips [3]. The part of the lips outside of the flange was removed. The remaining gaps after the corner of the lips were filled manually using the sculpture function of Blender [8]. The mesh processing was done with Blender and Meshmixer. The replicas were 3D printed with Ultimaker UM3 using polylactic acid (PLA) with 100% filling.

Figure 3 presents a schematic of the measurement setup. The HATS was placed on a turntable (LinearX LT360) to automatically rotate the HATS to specific angular positions.

1. https ://vocaltractlab.de/index.php ?page=dvtd

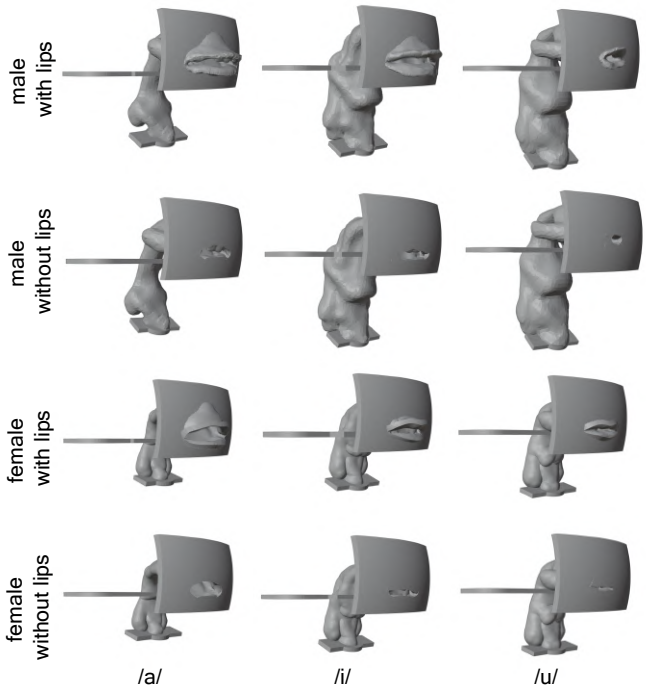

FIGURE 2 – 3D-printed vocal tract replicas.

The sound was recorded with a measurement microphone (MTG MK 250) attached to a circle arc support having 7 different possible positions spaced by 15°. The microphone support could be flipped down, as illustrated by the dashed lines in Fig. 3, to measure position ranging from 0°to 165°in the vertical plane. The microphone was connected to the Klippel Distortion Analyzer 2 which was also used to generate the source signal amplified with an amplifier Samson Servo 120a. The output voltage of the amplifier was measured by the Klippel Distortion Analyzer 2 before being fed to the sound source. The measurement process was controlled with a laptop computer through the Klippel Robotics software. The measurements were done in the anechoic room of the TU Dresden.

2.2 Data processing

The input signal was an exponential sine sweep generating frequencies ranging from 100 Hz to 20 kHz in 1.4 s. The signals were recorded with a sample rate of 48 kHz. In addition to the sound radiated from the HATS, the background noise was also recorded. The spectra of the signals were computed using the software Klippel dB-Lab RnD using a half-Hanning window (67,200 samples).

The spectra obtained with Klippel dB-Lab RnD were exported as text files and smoothed by applying a Gaussian Weighted averaging in Matlab 2021a using the `smoothdata` function. For this purpose, a window length of 100 and 1000 samples was used for the radiated sounds and the background noise respectively. More samples were used for the background noise to obtain a smoother noise threshold. The valleys of the transfer functions of the vocal tract replicas make the amplitude of the radiated sound very

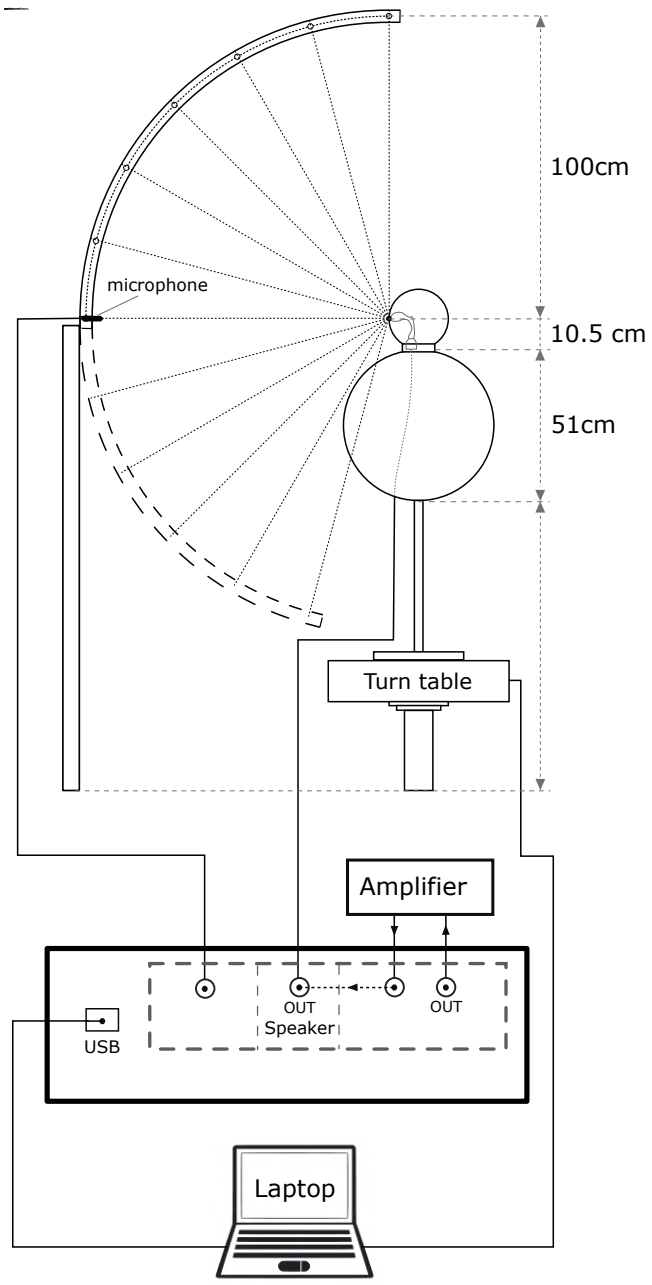

FIGURE 3 – Schematic of the measurement setup.

low, and the signal to noise ratio can be too poor to exploit the data. Thus, the data having an amplitude lower than the background noise were discarded.

The spectral information was discarded to keep only the directivity information by subtracting the maximal amplitude over the position (in dB) to the amplitude of all the positions for each frequency. Directivity maps were plotted as images showing the normalized amplitude as a function of the frequency and the angular position in the vertical and the horizontal planes (see Fig. 4). A directivity index was computed as the ratio of the maximal amplitude over the average amplitude over the positions for each frequency in the horizontal plane (horizontal directivity index, HDI) and in the vertical plane (vertical directivity index, VDI). To get a more synthetic view of the directivity properties, these indexes were averaged over third octave bands and over the different conditions to compare (e.g. all the measurements with and without lips).

3 Results

The directivity patterns of the vocal tract geometries corresponding to the vowels /a/, /i/ and /u/ of a male and a female subject with and without lips have been measured. Examples of directivity maps illustrating some effects further discussed hereafter are presented in Fig. 4. The directivity indexes averaged on all the measurements corresponding to specific configurations are presented in Fig. 5 : with and without torso, with and without lips, the phonemes /a/, /i/ and /u/ and the male and female geometries.

The torso generates a diffraction pattern which is visible as two lobes diverging to the sides with increasing frequency and repeats several times (three times in most of the cases. It is visible in the horizontal plane in Figs. 4a, 4c 4g, 4h and 4i, and not visible in the configuration with only the head (but not the torso) shown in Fig. 4b. In the vertical plane, it is visible as lobes shifting downward with increasing frequency, as can be seen in Figs. 4d and 4f, and is not seen in the head-alone configuration shown in Fig. 4e. The torso diffraction slightly increases the directionality (of the order of 1 dB maximum) in the horizontal plane, except between 0.8-2.4 kHz in which it is decreased, as illustrated in Fig. 5a.

The lips tend to increase the upward radiation between 4-9 kHz and downward radiation above 9 kHz, as can be seen when comparing Fig. 4d, with lips, with Fig. 4f, without lips. Lips slightly increase the directionality at high frequency from about 5 kHz for the HDI up to about 2 dB and from about 10 kHz for the VDI up to about 1 dB (see Fig. 5b).

The phoneme influences the effect of HOM on directivity, which induces significant changes of directivity in relatively small frequency intervals (order of 100 Hz) [5, 6]. This effect is more visible for /a/, and from lower frequency (from 3.8 kHz for male /a/) than for /i/ and /u/ (see Figs. 4a, 4g, 4h and 4i). It is almost not observable for /u/, however, the torso diffraction pattern is more pronounced and visible at higher frequencies for /u/. The torso diffraction pattern

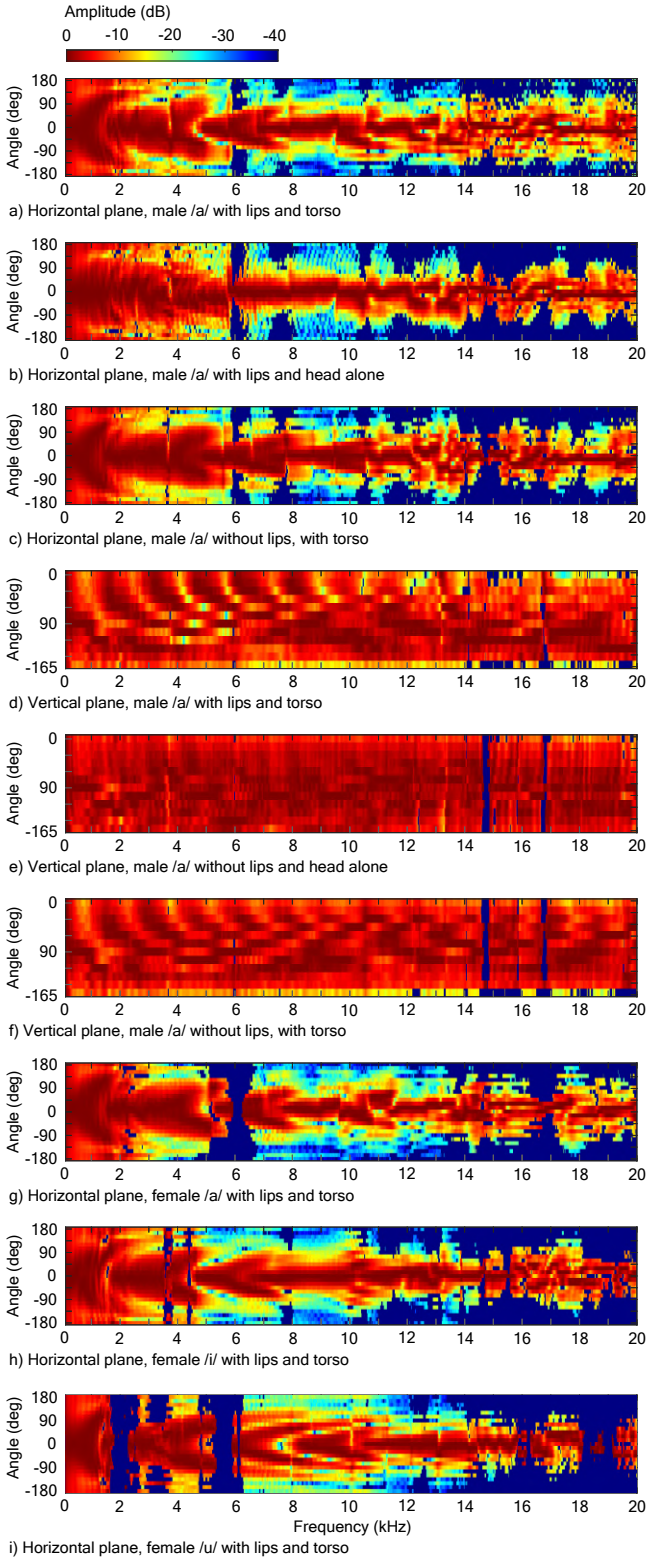

FIGURE 4 – Directivity maps of the head and torso replica in various configurations : with and without torso, with and without lips, in the horizontal or vertical planes and female and male vocal tract geometries corresponding to the vowels /a/, /i/ and /u/.

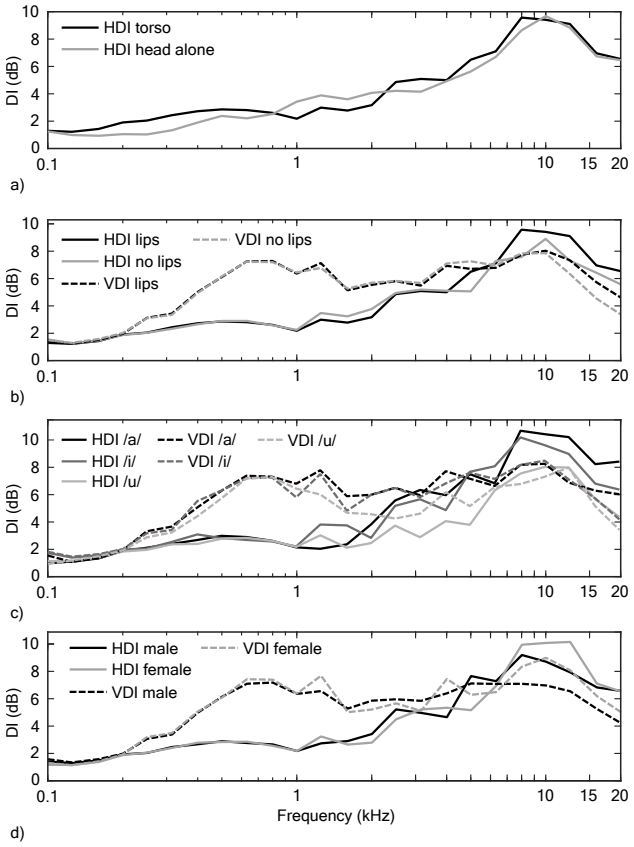

FIGURE 5 – Horizontal and vertical directivity indexes (HDI and VDI) of the head and torso replica averaged on different configurations : a) with and without torso, b) with and without lips, c) for the vowels /a/, /i/ and /u/ and d) male or female.

is less visible for /a/. This effect is not observed on the radiation patterns of the loudspeaker (data not shown).

In the horizontal plane, /a/ is slightly more directional than /i/ and substantially more directional (up to about 3 dB) than /u/ from 2 kHz, except between 5-7 kHz where /i/ is more directional than /a/ (see Fig. 5c). In the vertical plane /a/ and /i/ are substantially more directional (up to about 2 dB) than /u/ above 1 kHz, but even though there are substantial differences between /a/ and /i/, no clear trend can be observed, one being more directional than the other at different frequencies.

The female speaker geometries have a slightly more pronounced downward radiation in the vertical plane from 11 kHz. It is also more directional at high frequencies (see Fig. 5d) from about 6 kHz for the HDI (up to about 2 dB) and from about 8 kHz for the VDI (up to about 1 dB).

Interactions between the effect of the torso and the lips are observed. The torso pattern is affected by the presence of the lips : the frequency of the lobes is changed and the lobes are more pronounced when the lips are present. See Fig. 4a with lips and Fig. 4c without lips for the horizontal plane, and Fig. 4d with lips and Fig. 4f without lips in the vertical plane.

Interactions between the effect of the torso and the phoneme are observed. The torso diffraction pattern is more or less masked by the HOM effect. This becomes gradually more pronounced for /u/, /i/ and /a/. The torso diffraction pattern shape (frequency of the lobes) is affected by the phoneme. It is more pronounced for /u/ (see Fig. 4i).

There is an interaction between the effect of the lips and the phoneme. The effect of HOM is slightly less visible when the lips are present (see Figs 4a and 4c). In the case of the female /i/ a lobe moving downward and then upward is visible in the vertical plane above 11 kHz when the lips are present (data not shown).

There is an interaction between the effect of the phoneme and the gender. The effect of HOM is visible at different frequencies for the female and male geometries : from 3.8 kHz for male /a/ (see Fig. 4a) and from about 7 kHz for female /a/ (see Fig. 4g). The HOM effect is observed for female /u/ (see Fig. 4i) and not for male /u/. There is also globally less HOM effect for female /i/ (data not shown).

4 Discussion

The obtained torso diffraction pattern is similar to the observation reported by Brandner et al. [9]. Thus, even though the HATS used in this study has a simplified geometry, it can reproduce the major features of a realistic torso shape. The pattern observed was also observed in measurements on real subjects in the horizontal plane [7].

The torso diffraction pattern is influenced by other parameters :

- The presence of the lips, which tend to make it more pronounced. The reason for this effect is not clearly understood.
- The phoneme can change its shape and can make it more or less pronounced. This could be attributed to the differences of mouth opening dimensions, the mouth position on the head, which can slightly vary between the replicas and the differences in lips shape. Phoneme can also mask more or less the torso pattern with the effect of HOM.
- The gender can also change the shape of the torso diffraction pattern and make it more or less pronounced. The torso dimension differences are likely to induce gender related differences. However, this cannot be explored with our experiment since only one torso size was used. Differences in the frequency of the lobes of this pattern which may be related to gender-specific torso dimension variation are reported by Brandner et al. [9]. For our experiment, the differences observed can probably be attributed to the same cause as the differences related to the phonemes : differences in mouth size, lip shape and vocal tract geometry. It is not possible to generalize these results as more subjects would be necessary to show a trend of the gender-specific vocal tract shape difference.

The increase of directionality induced by the presence of the lips is in agreement with the observation of Yoshinaga et al. [20]. Thus, this effect which was observed only for the phoneme /s/ can be generalized to the vowels /a/, /i/ and /u/. This influence of the lips appears from about 5 kHz, thus it can be assumed that the effect of the torso is predominant up to 5 kHz.

The variation of directionality of the different phonemes is in agreement with the piston model [10, 14] in the range of the octave bands 2-5 kHz, and above 7 kHz in the horizontal plane. This is in agreement with the observations on real subjects [15, 7, 19].

The effect of HOM observed on the different phonemes is in agreement with the predictions from simulations performed by Blandin et al. [6]. It is less visible when the mouth opening is smaller, and the frequency from which it can be observed is in agreement with these simulations. The lips seem to affect the HOM effect, making it less pronounced.

It is difficult to conclude anything concerning the effect of gender from this experiment since not enough subjects were simulated and the torso dimensions were kept the same for both genders. Thus, the differences observed could be simply attributed to differences of individual anatomy or articulation at the moment of the capture of the geometries with MRI. The female subject has a more pronounced articulation than the male subject (lower jaw position). This could explain the stronger directionality of the female subject at high frequency (from about 6 kHz) as well as the more pronounced downward radiation in the vertical plane from about 11 kHz. This is in agreement with the observation of Brandner et al. [9] that a low jaw position had the same effect. The male /a/ has inter-dental spaces which are not present in the female /a/. These side cavities lower

the cutoff frequency of the HOM and can, thus, explain that HOM effect appears at lower frequency for the male subject.

5 Conclusion

The simplified HATS reproduces properly the major features of the torso diffraction pattern. This pattern is influenced by the lips, the mouth dimensions and the vocal tract shape. The increase of directionality above 5 kHz induced by the presence of the lips was confirmed for the vowels /a/, /i/ and /u/ of two different subjects. It was confirmed that the directionality of different phonemes can be related to the mouth dimensions in the octave bands 2 kHz and 4 kHz. The effect of HOM was observed and found to be in agreement with theoretical predictions : /a/ exhibit more HOM effect, /i/ less and /u/ almost none. Given the number of subjects modelled, it was not possible to conclude anything concerning the influence of the gender. However, the comparison of both subjects highlights the effect of differences of articulation and anatomy. A lower jaw position would increase the directionality and induce more downward radiation at high frequency. The presence of inter-dental side cavities enhances the effect of HOM.

6 Aknowledgement

This study was supported by the German Research Foundation (DFG) with the grant no. BI 1639/7-1.

Références

[1] V.R. Algazi, R.O. Duda, R. Duraiswami, N.A. Gumerov and Z. Tang, Approximating the head-related transfer function using simple geometric models of the head and torso, *The Journal of the Acoustical Society of America*, **112(5)**, 2053-2064 (2002).

[2] 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, *The Journal of the Acoustical Society of America*, **134(4)**, 2946-2954 (2013).

[3] P. Birkholz and E. Venus, Considering lip geometry in one-dimensional tube models of the vocal tract, *International Seminar on Speech Production*, 78-86 (2018).

[4] P. Birkholz, S. Stone, P. Häsner, R. Blandin and M- Fleischer, Printable 3D vocal tract shapes from MRI data and their acoustic and aerodynamic properties, *Scientific data*, **7(1)**, 1-16 (2020).

[5] R. Blandin, A. Van Hirtum, X. Pelorson and R. Laboissière, Influence of higher order acoustical propagation modes on variable section waveguide directivity : Application to vowel [α], *Acta Acustica united with Acustica*, **102(5)**, 918-929 (2016).

[6] R. Blandin, A. Van Hirtum, X. Pelorson and R. Laboissière, The effect on vowel directivity patterns of higher order propagation modes, *Journal of Sound and Vibration*, **432**, 621-632 (2018).

[7] R. Blandin, B.B. Monson and M. Brandner, Influence of speech sound spectrum on the computation of octave band directivity patterns, *Proceeding of Forum Acusticum 2020*, 2027-2033 (2020).

[8] Blender Online Community, Blender - a 3D modelling and rendering package, Stichting Blender Foundation, Amsterdam, (2018). Available at : http ://www.blender.org.

[9] M. Brandner, R. Blandin, M. Frank and A. Sontacchi, A pilot study on the influence of mouth configuration and torso on singing voice directivity, *The Journal of the Acoustical Society of America*, **148(3)**, 1169-1180 (2020).

[10] J.L. Flanagan, Analog measurements of sound radiation from the mouth, *The Journal of the Acoustical Society of America*, **32(12)**, 1613-1620 (1960).

[11] M. Frič and I. Podzimková, Comparison of sound radiation between classical and pop singers, *Biomedical Signal Processing and Control*, **66**, 102426 (2021).

[12] N.A. Gumerov, R. Duraiswami and Z. Tang, Numerical study of the influence of the torso on the HRTF, *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, **2**, II–1965 (2002).

[13] P. Häsner, A. Prescher and P. Birkholz, Effect of wavy trachea walls on the oscillation onset pressure of silicone vocal folds, *The Journal of the Acoustical Society of America*, **149(1)**, 466-475 (2021).

[14] J. Huopaniemi, K. Kettunen and J. Rahkonen, Measurement and modeling techniques for directional sound radiation from the mouth, *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 183-186 (1999).

[15] P. Kocon and B.B. Monson, Horizontal directivity patterns differ between vowels extracted from running speech, *The Journal of the Acoustical Society of America*, **144(1)**, EL7-EL12 (2018).

[16] T. Leishman, S.D. Bellows, C.M. Pincock and J.K. Whiting, High-resolution spherical directivity of live speech from a multiple-capture transfer function method, *The Journal of the Acoustical Society of America*, **149(3)**, 1507–1523 (2021).

[17] A.H. Marshall and J. Meyer, The directivity and auditory impressions of singers, *Acta Acustica united with Acustica*, **58(3)**, 130-140 (1985).

[18] B.B. Monson, E.J. Hunter and B.H. Story, Horizontal directivity of low-and high-frequency energy in speech and singing, *The Journal of the Acoustical Society of America*, **132(1)**, 433-441 (2012).

[19] C. Pörschmann and J.M. Arend, Investigating phoneme-dependencies of spherical voice directivity patterns, *The Journal of the Acoustical Society of America*, **149(6)**, 4553–4564 (2021).

[20] T. Yoshinaga, A. Van Hirtum, K. Nozaki and S. Wada, Influence of the lip horn on acoustic pressure distribution pattern of sibilant/s/, *Acta Acustica united with Acustica*, **104(1)**, 145-152 (2018).

[21] T. Halkosaari, M. Vaalgamaa, and M. Karjalainen, Directivity of Artificial and Human Speech, *Journal of the audio engineering society*, **53(7/8)**, 620-631 (2005).

[22] W. Chu and A. Warnock, Detailed Directivity of Sound Fields Around Human Talkers, *Technical Report, Institute for Research in Construction* (National Research Council of Canada, Ottawa ON, Canada), 1-47 (2002).