Influence of speech sound spectrum on the computation of octave band directivity patterns

Rémi Blandin, Brian Monson, Manuel Brandner

INFLUENCE OF SPEECH SOUND SPECTRUM ON THE COMPUTATION OF OCTAVE BAND DIRECTIVITY PATTERNS

Rémi Blandin[1] **Brian Monson**[2] **Manuel Brandner**[3]
[1] Institute of Acoustics and Speech Communication , TU Dresden, Germany
[2] Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, USA
[3] Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria
remi.blandin@tu-dresden.de

ABSTRACT

Speech radiation patterns exhibit angle-dependent variations of the amplitude and spectrum of the radiated sound. Speech directivity is gaining interest for the rendering of speech in three dimensional environments (real or virtual), but it is also related to more fundamental research questions, such as speech intelligibility in the presence of competing speech (the cocktail party problem). Speech directivity is most often quantified by octave-band analysis of speech signals recorded simultaneously with microphone arrays surrounding a talker in an anechoic environment. Due to the variability of the physical mechanisms of speech production, the radiation patterns differ between different speech sounds. However, a part of the observed variability may be due to the band analysis process itself, which is influenced by the spectral differences between the different speech sounds. In order to investigate to what extent and in which frequency range this variability is actually due to differences in directionality, directivity patterns are computed in narrower frequency bands with constant width. The details revealed by this higher spectral resolution also allow one to identify the expected influence of the dimensions of the subjects, the mouth opening and the contribution of the nasal cavity to the sound radiation. Octave band directivity patterns are computed from the high spectral resolution directivity patterns and compared with the commonly computed octave band long term averaged spectra directivity patterns.

1. INTRODUCTION

Speech directivity has been measured by several authors using multiple microphones placed at equidistant and regularly spaced positions around a talker (see as an example [1–4]). For most of these studies the sound levels recorded at the different angular positions have been measured in octave or third octave bands. In some cases, the analysis has targeted specific vowels and consonants, revealing directionality differences [1, 4–6].

The smooth polar diagrams presented in these studies can be misleading by suggesting that the directionality evolves progressively from one averaged pattern to the other as frequency increases. In fact, there is a lack of knowledge concerning the detailed features of human speech and singing directivity. Theoretical investigations of the effect of transverse propagation modes predict that, at least beyond 4-5 kHz, substantial variations in the directionality and shape of the directivity patterns can occur within small frequency intervals [7, 8]. These predictions have recently been confirmed by comparing simulations accounting for transverse propagation modes and the directivity patterns measured on two utterances of the vowel /a/ sung by a classical singer [9]. On the other hand, one can also expect that the mouth is not always the only source of sound radiation. For example, simultaneous radiation from the nose and the mouth might generate interference patterns at some frequencies.

Considering that directivity patterns can have significant variations within a relatively small frequency interval, this information is lost in an octave band pattern, which is essentially the average of multiple potentially different directivity patterns. Thus the overall octave band patterns are more heavily weighted by the patterns corresponding to the highest spectral amplitudes within the octave band. As a consequence, the observed differences in directivity between different speech sounds may be due to differences in the spectrum. On the other hand, differences in directivity may be missed simply because less energy is present at some frequencies within the octave band.

In order to investigate the details of speech directivity and the potential influence of the spectrum on octave band directivity patterns, directivity patterns of subjects recorded with a 13 microphone array in an anechoic room were computed in short time windows with 10.78-Hz spectral resolution. We examined the vowels /a/, /e/, /i/, /o/ and /u/. Octave band directivity patterns were computed following two methods:

- averaging the amplitude of the spectrum within the octave band,
- averaging normalized 10.78-Hz-wide directivity patterns for frequencies within the octave band.

The first method is similar to those employed in most previous studies, whereas the second method isolates the directivity phenomenon from the spectrum of the sound radiated.

2. METHOD

The data analyzed were sentences pronounced by 15 subjects (8 female and 7 male) recorded at a 44.1-kHz sampling rate with 13 microphones spaced at 15° intervals and equally distant from the head of the subjects in the horizontal plane, from 0° (directly in front of the talker) to 180° (directly behind the talker). Details of the acquisition of the data have been previously published [4]. The segmentation of the vowels /a/, /e/, /i/, /o/ and /u/ performed for a previous analysis of these data [6] was used here.

For the present study, the recordings from each microphone were sliced in windows of 2048 samples overlapping by 90%. The spectrum was computed using a discrete Fourier transform and zero padding (2048 additional samples) so that the frequency resolution was 10.78 Hz. These parameters were the same for both methods detailed hereafter.

2.1 Averaging of the amplitude of the spectrum

Before averaging, unwanted noise was removed from the data. A pure tone at 17.6 kHz was removed. The microphone at 165° had a higher background noise than the other microphones, which created artefacts when the signal to noise ratio was low. Consequently, the data for this microphone were excluded from the averaging process from 5 kHz on. Because this microphone is located behind the head of the subject where the radiated amplitude is low at high frequencies, it did not substantially affect the analysis of the directivity patterns.

An average spectrum for each vowel was generated by averaging across each windows for all the utterances of each vowel by all subjects (see as an example Figs. 2a and 2c for the 500Hz octave band interval). The octave band directivity patterns for each vowel were then computed by calculating the average amplitude of the averaged spectrum over the frequency interval corresponding to each octave band (see Fig. 3). The directivity patterns were then normalized by the maximum amplitude across all angles.

2.2 Averaging of the directivity patterns

Directivity patterns were computed for each frequency and each time window by subtracting the maximal amplitude over the angular positions from the amplitude of the other positions. Thus, the spectral level was normalized at each frequency, allowing each directivity pattern at each frequency to be equally weighted during the averaging process.

However, before computing directivity patterns, it was necessary to define a frequency dependent noise threshold. This is necessary because computing directivity patterns from data with too poor signal to noise ratio would lead to wrong patterns. For example, if a highly directional pattern is observed with a poor signal to noise ratio, only the highest amplitudes would emerge from the noise.

A 9s background noise recording was used to compute a noise threshold for each microphone. Spectra of overlapping windows were computed in the same way as for the other data. The median of the background noise amplitude was computed for each frequency and each microphone in order to get a smooth approximation of the background noise profile. The obtained curve was shifted up in level so as to exceed the maxima of all windows from the noise signal. The 17.6 kHz peak was added to the threshold. Any data having an amplitude lower than this threshold curve were excluded from the analysis.

Directivity patterns were computed only if at least 3 microphones registered amplitude higher than the noise threshold. To visualize the directivity patterns, they were represented in color scale as a function of the frequency and the angular position (see Figs. 1 and 2). This representation is referred to as a directivity map hereafter.

The obtained directivity maps were averaged over time windows for individual utterances of each vowel for each subject (see Figs 1a and 1b), for all utterances of all vowels for each subject (see Figs 1c and 1d), and for all utterances of all subjects for each vowel (see Figs 2b and 2d for the frequency interval corresponding to the 500Hz octave band). Octave band directivity patterns were computed averaging the directivity maps over octave bands (see Fig. 4).

3. RESULTS

3.1 Directivity maps

Figs. 1a and 1b show the directivity maps (calculated by averaging directivity patterns) for single utterances of the vowels /a/ and /i/ pronounced by the same female subject. The evolution of the directivity patterns with increasing frequency show abrupt transitions to different patterns and directionality within small frequency intervals (on the order of 100Hz). For example, in Fig. 1a there is a sudden appearance of a pattern with lower directionality near 8kHz. More complexity can generally be found toward high frequencies. It is noteworthy that the complex variations of the directivity patterns with increasing frequency are also variable in time. They change from utterance to utterance and even possibly within a single utterance (data not shown).

Directivity maps averaged on all utterances of individual subjects are presented in Figs. 1c and 1d, for a female and a male subject respectively. One can see that the complexity is substantially reduced compared to the unique utterances (Figs. 1a and 1b). A global increase of directionality can be seen up to about 10 kHz. Above 10 kHz, the directionality appears to slightly decrease.

All directivity maps in Fig. 1 reveal a similar pattern at frequencies below 6 kHz, consisting of 3 or 4 lobes diverging toward the side as frequency increases. Slight changes in the number of lobes and their angular position are observed between different subjects and between different utterances within a given subject. In some utterances, other patterns featuring substantial changes within small frequency intervals appear superimposed to this lobe pattern, as seen in Fig. 1b in the 0-2 kHz interval.

In Figs. 2a and 2c the average spectrum is represented

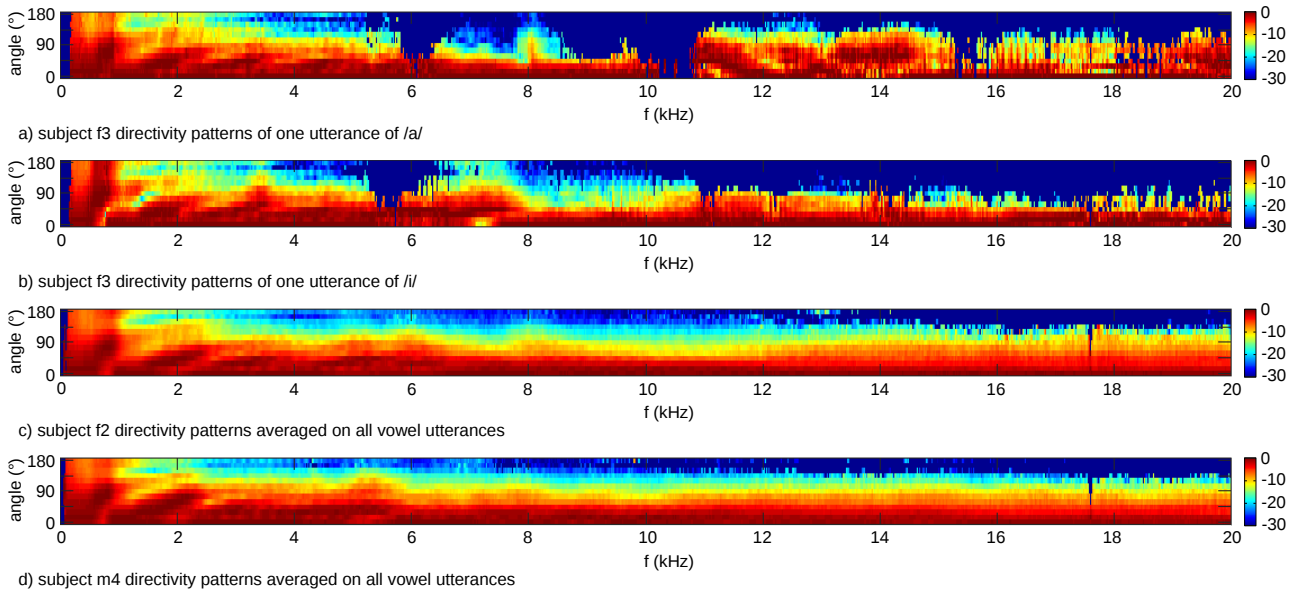

Figure 1. Directivity maps (calculated by averaging directivity patterns) obtained with a 10.78 Hz discretization represented in color scale as a function of the frequency and the angular position. (a) a single utterance of the vowel /a/ by a female subject, (b) a single utterance of the vowel /i/ by the same female subject, (c) directivity patterns averaged on all utterances of all vowels of one female subject and (d) directivity patterns averaged on all utterances of all vowels of one male subject. The color indicates the sound level relative to the maximal level over the 13 positions for each frequency. The deep blue color corresponds to data under a noise threshold.

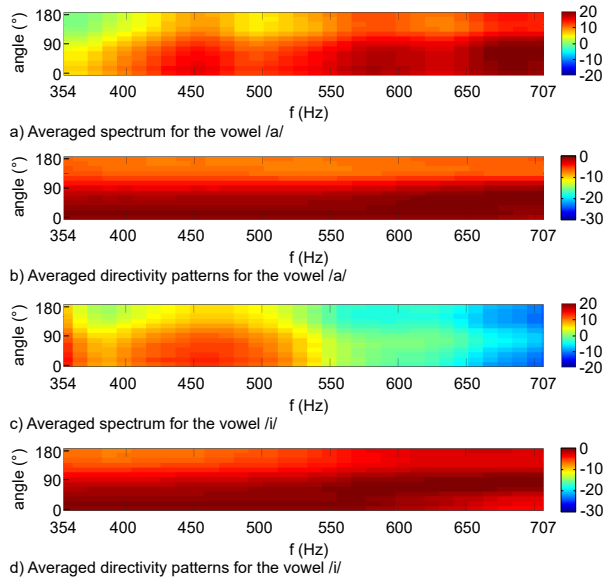

Figure 2. Amplitude of the spectrum and directivity patterns averaged on several utterances of the vowel /a/ and /i/ pronounced by 15 female and male subjects in the frequency interval corresponding to the 500Hz octave band. (a) averaged amplitude spectrum for the vowel /a/, (b) averaged directivity patterns for the vowel /a/, (c) averaged amplitude spectrum for the vowel /i/, and (d) averaged directivity patterns for the vowel /i/.

in color scale over the frequency range corresponding to the 500Hz octave band (354-707Hz) for the vowels /a/ and /i/. The directivity maps corresponding to the same frequency range and vowels are presented in Figs. 2b and 2d.

The averaged spectrum amplitude and the directivity patterns of the vowels /a/ and /i/ are presented in Fig. 2 for the frequency range corresponding to the 500Hz octave band (354-707Hz). The acoustic energy appears more evenly distributed for the vowel /a/ (Fig. 2a) than for the vowel /i/ (Fig. 2b). The energy of the vowel /i/ is mainly present in the first half of the frequency band (354-550Hz), and substantially lower amplitudes are found between 650Hz and 707Hz. The directivity patterns of the vowel /a/ (Fig. 2b) are rather similar all over the 500Hz frequency band, with a slight sideward shift of the maximum of amplitude from 600Hz on. The directivity patterns of /i/ (Fig. 2d) show more variations: a more pronounced lobe oriented toward 90° can be seen from 600Hz on.

3.2 Octave band directivity patterns

The directivity patterns obtained from averaged spectra and averaged directivity patterns are presented in Figs. 3 and 4, respectively. 0° corresponds to the front and 180° to the back of the subjects. Both methods generate globally similar patterns: the shape of the patterns is very similar and the directionality increases with the frequency of the octave bands, except for the 16kHz band.

However, there are substantial differences in the 8kHz and 16kHz band: the patterns obtained from averaged spectra (Fig. 3) are more directional. Another noticeable difference is the greater variability of the patterns across

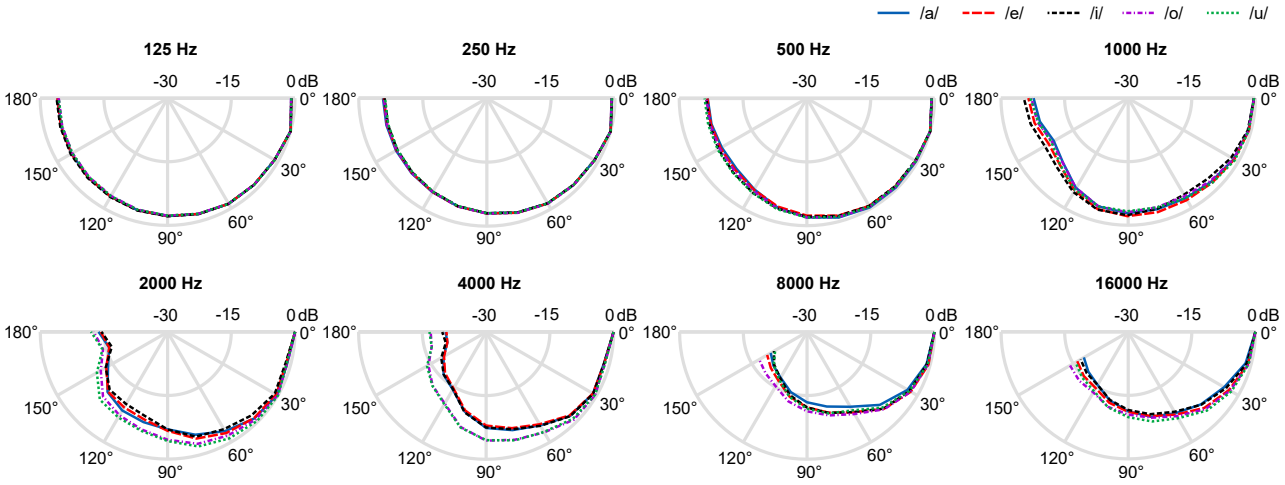

Figure 3. Directivity patterns obtained averaging the sound amplitude over octave bands and over the utterances of 15 male and female subjects for the vowels /a/, /e/, /i/, /o/ and /u/. The radius of the curves indicates sound level relative to the maximal level over the 13 positions.

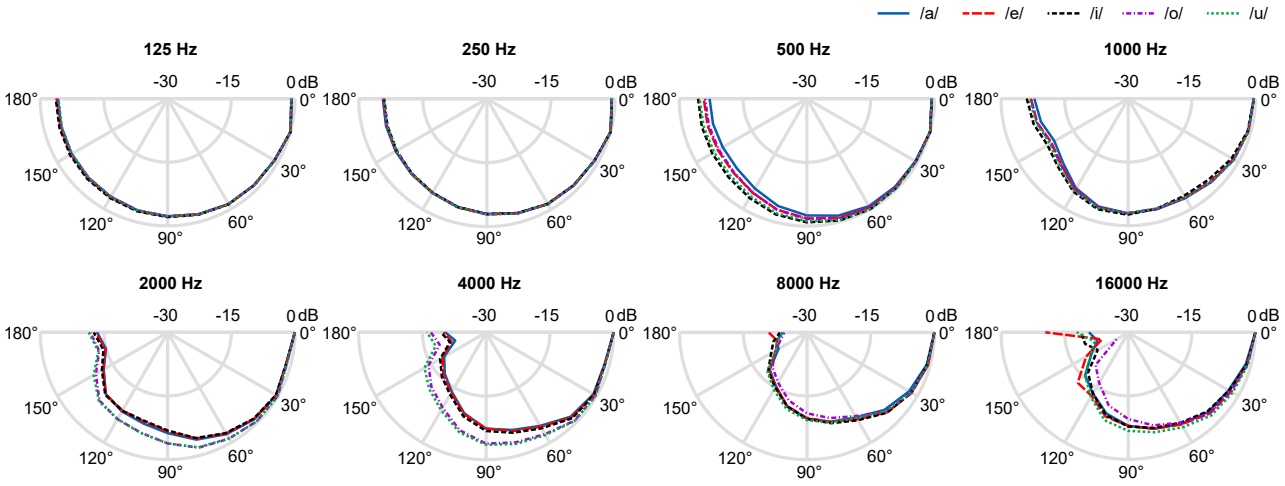

Figure 4. Directivity patterns obtained averaging the directivity patterns computed every 10.78 Hz over octave bands and over the utterances of 15 male and female subjects for the vowels /a/, /e/, /i/, /o/ and /u/. The radius of the curves indicates sound level relative to the maximal level over the 13 positions.

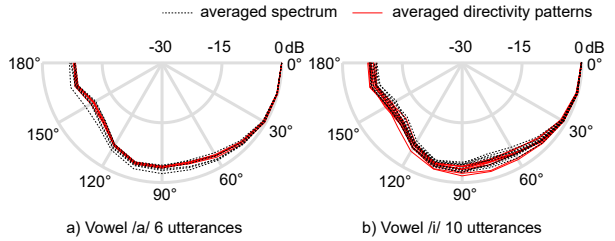

Figure 5. Directivity patterns computed for multiple individual utterances of the same male subject from averaged spectrum (dashed black lines) and from averaged directivity patterns (full red lines) for (a) the vowel /a/ and (b) the vowel /i/.

vowels in the 500Hz band for the patterns obtained from averaged directivity patterns (Fig. 4). This variability appears to be greater than that observed in the 1kHz band. In the 2kHz band, the vowels are more clearly separated in two groups in the octave band patterns from averaged directivity patterns.

For both octave band patterns computation methods, in the 125Hz and 250Hz band exactly the same patterns are obtained for all the vowels. Differences appear in the 500Hz band, in which the vowels are, in order of increasing directionality, /i/, /u/, /e/, /o/ and /a/. Less differences are observed in the 1kHz band, but one can note that the vowel /i/ has a slightly different pattern from the others, with a slightly lower amplitude in the 30°-90° region and the highest amplitude in the 90°-180° region. In the 2kHz and 4kHz bands the vowels are separated in two groups: /a/, /e/ and /i/ more directional and /u/ and /o/ less directional. In the 8 kHz and 16kHz bands there is no more separation in two groups and less differences between the vowels than in the 2kHz and 4kHz bands.

Fig. 5 show the patterns obtained for multiple utterances of the vowels /a/ and /i/ pronounced by the same male subject. In the case of the vowel /a/, less utterance to utterance variations are observed when using averaged directivity patterns. In the case of the vowel /i/, the variability of the patterns is similar for both methods.

4. DISCUSSION

4.1 Comparison of two methods of computation of octave band directivity patterns

The main difference between octave band directivity patterns computed from averaged spectra (Fig. 3) and from averaged 10.78Hz directivity patterns (Fig. 4) is the increased directionality obtained with the first method in the 8kHz and 16kHz octave bands. This can be explained by the expected effect of unequal weighting of patterns at individual frequencies, as pointed out in the introduction. Because these bands cover the most extended frequency ranges, they are the most likely to be influenced by this problem. In the 16kHz band, more acoustic energy is present in the lower frequencies of the band at which the directivity patterns are more directional than in the higher frequencies (see Figs. 1c and 1d). Thus, with these patterns being over-represented, the resulting octave band pattern is more directional than the one obtained from averaged 10.78Hz directivity patterns. In the case of the 8kHz band, the situation is reversed: the directionality increases with frequency in the band and there is less acoustic energy at lower frequencies of the band. The less directional patterns being less weighted, the overall pattern is also more directional than the one obtained from averaged 10.78Hz directivity patterns.

Another difference is observed in the 500Hz octave band: the directivity patterns obtained for each vowel by averaging the spectrum are almost identical, whereas substantial differences across vowels are found when averaging the 10.78Hz directivity patterns. This can be explained by an uneven distribution of the acoustic energy over the frequency band. For the vowel /i/, the less directional pattern compared to the other vowels is due to the presence of a pronounced lobe orientated toward 90° in the upper part of the band (see Fig. 2d). Its orientation, which differs from the other patterns of the band, makes the averaged overall pattern less directional. However, the formant structure of the vowel /i/ is such that there is less energy in the upper portion of the band (see Fig. 2c). Thus, the weight of this 90° lobe is small in the averaging process, and this difference of directivity compared to the other vowels is underestimated. On the other hand, the acoustic energy of the vowel /a/ is more evenly distributed in this band (see Fig. 2a), but the 90° lobe is less pronounced than for the vowel /i/ (see Fig. 2b). Thus, averaging the spectrum to compute the octave band directivity pattern tends to underestimate the differences in directionality of the vowels in the 500Hz octave band.

At the scale of single utterances, greater variability in the octave band directivity patterns obtained from averaged spectra (see Fig. 5a) can also be attributed to variations in the distribution of the acoustic energy over the bands. However, the acoustic energy distribution also affects the averaging of 10.78Hz directivity pattern, because there are frequencies at which the amplitude is smaller than the noise threshold. On the other hand, the directivity patterns themselves vary from utterance to utterance, especially at high frequencies. Similar variability can be observed with this method with utterances of other speech sounds (see Fig. 5b). Thus, using multiple utterances to fill acoustic energy gaps when averaging out the variability of directivity patterns is important.

The use of a noise threshold allowed us to obtain information from the two back microphones (165° and 180°), which receive the smallest amplitudes. However, due to the smaller amount of data, the averaged patterns obtained in this angular region may be less accurate. This may be why the shape of the pattern obtained for /e/ departs from the patterns obtained for the other vowels in the 16kHz octave band (see Fig. 4).

4.2 Potential mechanisms inducing vowel directivity pattern variations

The complex variations of the directivity patterns obtained from averaged 10.78Hz directivity patterns (Figs. 1, 2 and 4) can be explained for high frequencies (from 4-5 kHz on) by the propagation of transverse propagation modes inside the vocal tract. In fact, it has been shown theoretically and experimentally that this phenomenon can generate complex variation of the directivity patterns with frequency [8, 9].

However, transverse propagation modes do not explain the abrupt transitions observed at low frequencies, observed in Fig. 1b in the 0-2 kHz interval. This may be due to the nasalisation of the vowels which sometimes occurs when the communication with the nasal cavity is open. In such a case, sound would be radiated simultaneously by the nose and the mouth and produce interference patterns which might be responsible for these abrupt transitions. This needs to be confirmed by proper modelling of the phenomenon.

Some key parameters in determining the directionality of speech are the dimensions of the mouth opening, specifically its width. From the plane piston radiation model, one expects that the wider is the mouth, the more directional the sound is in the horizontal plane [10, 11]. Directionality is expected to increase with the frequency.

However, this tendency is not always observed on the averaged directivity maps of the different subjects (see Figs. 1c and 1d): the directionality slightly decreases from about 10kHz on. On the other hand, the differences of mouth opening would be expected to have a stronger impact at high frequencies. Again, this is not always observed, as less differences between vowels are observed in the 8 kHz and 16 kHz octave bands than in the 2 kHz and 4 kHz octave bands. The transverse propagation modes may be the explanation of this disagreement with the simple plane piston model. Their more frequent occurrence at high frequencies, their tendency to generate less directional patterns and their variability from utterance to utterance would result in an average decrease of directionality. However, this needs to be confirmed by proper modelling. It should be noted that this is an average tendency and that at the scale of single utterances the patterns can be very different from the averaged patterns, as illustrated in Figs. 1a and 1b. The relevance of this variability over time and frequency of the high frequency directivity patterns for the perception of human speech and singing is an open question.

On the other hand, the plane piston model explains very well the division of the directivity patterns in two groups in the 2 kHz and 4kHz octave bands. In fact, the same two groups, /a/, /e/ and /i/ which are more directional and /o/ and /u/ which are less directional, are found when sorting the vowels by mouth width. The mouth widths measured by Fromkin [12] for different vowels show that /a/, /e/ and /i/ corresponds to similar mouth width (about 40mm) which are greater than the ones of /o/ (about 20mm) and /u/ (about 15mm). One can even notice that the vowel /o/ is slightly more directional than the vowel /u/ in the 4 kHz octave band in Fig. 4, which is in agreement with the dimensions provided by Fromkin.

However, the plane piston model fails again to predict the variations of directionality in the 500Hz and 1kHz octave bands:

- more differences are observed in the 500Hz band than in the 1kHz band, whereas one would expect the opposite,
- the differences of directionality are no more correlated to the mouth width, /i/ being the less directional whereas it has a larger width than /o/ and /u/.

In fact the plane piston model would not predict strong variations of directionality with the width, and finding significant variation in the 500Hz octave band is surprising. The reason of this disagreement is not clearly understood. A possible explanation could be that the nasalisation of the vowels plays a role in the directivity of the vowels in these octave bands. Nasalisation may induce complex variations of directivity patterns with strong minima, such as the ones observed in Fig 1b. This could result in the 90° lobe observed in the upper part of the 500Hz band (see Fig. 2) when averaging over several utterances. Some vowels may tend to be more nasalized than others and, thus have different patterns in these octave bands. Other alternative explanations could be the differences in protrusion, or radiation from other parts of the head such as the cheeks or the larynx. A proper modelling of the potentially implied phenomena as well as a tracking of the nasalization is needed to clarify these various hypothesis.

The lobe pattern between 0 and 4 kHz is most likely due to the diffraction by the torso. In fact, similar lobes are predicted for the head related transfer function, which are very close to the reciprocal situation of speech production [13]. The inter- and intra- subject variations of this pattern can be explained by differences in the dimensions, shape of the head and the torso of the subject as well as their posture.

5. CONCLUSION

The computation of directivity patterns at multiple frequencies with a small frequency spacing (10.78Hz) reveal that speech directivity has complex variations on small frequency scales (on the order of 100Hz). These variations are the consequences of different phenomena implied in speech and singing radiation. More physical modelling is required to identify them. The process of averaging spectrum to build octave band directivity patterns appears to overestimate directionality in the 8kHz and 16kHz bands, and to underestimate the variability related to vowels in the 500Hz. This is the expected consequence of an uneven weighting of directivity patterns resulting from the uneven distribution of the acoustic energy over the speech spectrum. Thus, the process of directly averaging high frequency resolution directivity patterns appears to be more accurate and reliable. However, it requires the definition

of noise thresholds. In this purpose, it is important to perform a recording of background noise free from perturbation from the subjects (breathing) for the measurement of directivity.

6. REFERENCES

[1] A. Marshall and J. Meyer, "The directivity and auditory impressions of singers," *Acta Acustica united with Acustica*, vol. 58, no. 3, pp. 130–140, 1985.

[2] W. Chu and A. Warnock, "Detailed directivity of sound fields around human talkers," 2002.

[3] D. Cabrera, P. Davis, and A. Connolly, "Long-term horizontal vocal directivity of opera singers: Effects of singing projection and acoustic environment," *Journal of Voice*, vol. 25, no. 6, pp. e291–e303, 2011.

[4] B. Monson, E. Hunter, and B. Story, "Horizontal directivity of low-and high-frequency energy in speech and singing," *The Journal of the Acoustical Society of America*, vol. 132, no. 1, pp. 433–441, 2012.

[5] B. Katz and C. d'Alessandro, "Directivity measurements of the singing voice," 2007.

[6] P. Kocon and B. B. Monson, "Horizontal directivity patterns differ between vowels extracted from running speech," *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. EL7–EL12, 2018.

[7] 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*, vol. 102, no. 5, pp. 918–929, 2016.

[8] 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*, vol. 432, pp. 621–632, 2018.

[9] R. Blandin and M. Brandner, "Influence of the vocal tract on voice directivity," 2019.

[10] J. Flanagan, "Analog measurements of sound radiation from the mouth," *The Journal of the Acoustical Society of America*, vol. 32, no. 12, pp. 1613–1620, 1960.

[11] J. Huopaniemi, K. Kettunen, and J. Rahkonen, "Measurement and modeling techniques for directional sound radiation from the mouth," in *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. WASPAA'99 (Cat. No. 99TH8452)*, pp. 183–186, IEEE, 1999.

[12] V. Fromkin, "Lip positions in american english vowels," *Language and speech*, vol. 7, no. 4, pp. 215–225, 1964.

[13] V. Algazi, R. Duda, R. Duraiswami, N. 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*, vol. 112, no. 5, pp. 2053–2064, 2002.