Journal of Sound and Vibration 432 (2018) 621–632

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

The effect on vowel directivity patterns of higher order propagation modes

Rémi Blandin [a,*], Annemie Van Hirtum [b], Xavier Pelorson [b], Rafael Laboissière [c]

[a] GIPSA-Lab, UMR CNRS 5216, Grenoble Alpes University, 11 rue des Mathématiques (BP46), 38402 Grenoble, France
[b] LEGI, UMR CNRS 5519, Grenoble Alpes University, 38058 Grenoble, France
[c] LPNC, UMR CNRS 5105, Grenoble Alpes University, (BP47), 38040 Grenoble, France

ARTICLE INFO

Article history:
Received 4 October 2017
Revised 15 June 2018
Accepted 22 June 2018
Available online XXX
Handling Editor: R. E. Musafir

Keywords:
Speech
Vocal tract
Waveguide
Higher order mode
Vowel
Directivity

ABSTRACT

Measurements of speech directivity patterns show that it differs according to the phoneme pronounced. In the case of the vowels, these differences can be attributed to variations of the vocal tract shape. The multimodal method, which takes into account higher order propagation modes, is used to simulate directivity patterns (2 kHz–15 kHz) for simplified vocal tract geometries of the vowels [a], [e], [i], [o] and [u]. The directivity patterns of a simplified and a realistic replica of the vowel [a] are measured experimentally. The comparison of the experimental data with the simulations shows a good agreement (average difference of 1.7 dB). It is observed that the amplitude, orientation and number of lobes can change significantly for some small frequency intervals, of the order of 100 Hz; these are shown to be caused by higher order modes. These changes can occur as low as 3 kHz if a wide cavity is present near the mouth exit. A small mouth exit limits the effect to specific frequency intervals, and a narrow channel limits the transmission of the higher order mode effect to the mouth exit. The comparison of the directivity measured on a realistic replica corresponding to the vowel [a] with simulations performed on two simplified geometries shows that a fully asymmetric shape (for which the centers of the cross-sectional contours are not aligned) is qualitatively more realistic than a partially asymmetric shape (when the centers of the cross-sectional contours lie along the same axis).

1. Introduction

The speech directivity consists in the variation of the amplitude of the sound radiated by a speaker with the direction. It is important for various applications including microphone placement [1], telecommunication [2], vocal performance practice [3–5], architectural acoustics [6], auralization and 3D sound synthesis [7–9]. Directivity patterns have been measured experimentally on human subjects [1,3–5,10] as well as modelled physically [2,11,12]. It was observed that they are almost omnidirectional at low frequencies (of the order of 100 Hz) and that they become more complex and directional at higher frequencies (starting at 500 Hz).

Whereas speech directivity has been mostly considered for words and sentences, some studies highlighted differences between the directivity patterns of different phones. Marshall and Meyer [4] observed a difference between the directivity patterns of vowels [a], [o] and [e]. Halkosaari [2] also reported differences between vowels and related it to the difference of mouth aperture size. Likewise, for the fricatives, Monson et al. [5] observed differences between [s], [s], [f] and [θ]. Katz et al. [10]

* Corresponding author.
E-mail address: remi.blandin@gmail.com (R. Blandin).

https://doi.org/10.1016/j.jsv.2018.06.053
0022-460X/

showed differences for both vowels and fricatives in a study specifically dedicated to the radiation patterns of the phones. On the other hand, higher order propagation modes (HOM) are present inside the vocal tract [13,14] above 3 kHz, and it is known that they can influence significantly the radiation of ducts [15–18]. It has been shown on simplified vocal tract geometries corresponding to the vowel [a] that HOM can influence significantly the directivity [19]. Thus, HOM could play an significant role for high-frequency speech directivity (above 3 kHz).

With the progress of technologies and applications, the increase of quality requirements necessitates a better knowledge of the physics underlying high-frequency speech radiation. Indeed, the simple physical models of speech directivity [2,11,12] account only for the external reflections, the diffraction by the head and the diffraction of the plane wave at the mouth exit. So the only parameter specific to a particular phone is the mouth exit cross-sectional area. The effect of the lips shape, the mouth exit cross-sectional shape and the three-dimensional (3D) aspect of the internal acoustic field in high-frequency are not considered. Thus, they cannot be used to predict accurately the directivity pattern of a specific phone for frequencies up to 15 kHz. On the other hand, directivity measurements are usually performed with a minimal frequency resolution of third-octave bands and with a minimal angular resolution of the order of 10°. This may mask phenomena related to the presence of HOM which occur within smaller frequency intervals and angular regions as reported in Ref. [19].

The objective of this work is to study the influence of the HOM, without other directivity mechanisms, on the directivity patterns for the vowels [a], [e], [i], [o] and [u]. Simulated directivity patterns of simplified vowel geometries are compared to each other and to measurements performed on mechanical replicas. The aim is also to provide a method which allows one to account for HOM while relying on simplified geometries and reasonable computational complexity in order to make systematic studies possible. For this purpose, the multimodal method (MM) [20–23] is used to predict the particle velocity distribution at the mouth exit at a relatively low computational cost compared to finite element or finite difference methods [14]. The radiated acoustic pressure is then calculated with the Rayleigh-Sommerfeld integral accounting for this distribution. In order to observe the directivity patterns accurately, the frequency range and the angular resolution of the measurements and the simulations have been set to 2–15 kHz and 2° respectively. Simplified geometries are designed from area functions provided by Story [24]. The design is made so that as much HOM as possible are elicited. Elliptical cross-sections are used as a simple way to take into account the width and height of the lip opening provided by Fromkin [25]. Eccentric junctions (*i.e.* the centers of consecutive contours are not aligned) are used in order to elicit as much as possible the HOM as it prevents cancelation of HOM due to symmetry [14,19].

The reliability of the simulation method is first assessed comparing simulations and experiments on a vowel [a] simplified geometry. Then, MM is applied to five simplified vowel geometries corresponding to [a], [e], [i], [o] and [u], in order to study the effect of HOM on these different vowels. Finally, the effectiveness to reproduce the effect of the HOM of the simplified geometries is evaluated for the vowel [a] by comparing the simulated directivity patterns with that measured on a realistic replica designed from magnetic resonance images (MRI) [26]. The vowel [a] is used since it has the largest mouth exit cross-section area and hence the effect of HOM is expected to be significant [19].

The paper is structured as follows: first the geometries, the experimental setup, the simulation method and the data analysis are detailed in section 2. Then, the results obtained from the comparison between simulation and experiment, the simulation for the vowels [a], [e], [i], [o] and [u] and the comparison between the simplified geometries and the realistic one are presented in section 3. Finally, the agreement between the simulations and the experiments, the relationship between the observed directivity patterns and the shape of the geometries, and the similarities between the simplified geometries and the realistic one are discussed in section 4.

2. Methods

This section presents the methods used to acquire and analyse the experimental and simulated data presented in section 3. These data consist of the acoustic pressure radiated by vocal tract geometries at different angular positions (spatial sampling each 2°) between 2 kHz and 15 kHz (frequency sampling every 10 Hz). The design of the geometrical approximations and the experimental setup are detailed in sections 2.1 and 2.2 respectively. An improvement of the implementation of the MM with respect to the previous works [14,19] is presented in section 2.3. Finally, the data analysis method is detailed in section 2.4.

2.1. Geometries

In order to perform simulations and experiments, different geometries approximating the shape of the vocal tract for the vowels [a], [e], [i], [o] and [u] are used. They are designed from area functions provided by Story [24] and from a 3D geometry extracted from MRI provided by Aalto et al. [26].

A 3D view of the simplified geometries is provided in Fig. 1. They are designed in order to exhibit HOM with cut-on frequencies of the same order of magnitude as for realistic geometries. They are constituted of 44 constant length tubes whose cross-sectional area is provided by area functions from Story [24]. Since the mouth exit cross-sectional shape is prominent for radiation problems, it is schematized as an ellips with a width to height ratio of the same order of magnitude as real subjects. In this purpose, the average ratios of 4 subjects presented in Fig. 5 of Fromkin [25] are used. Since no information about the internal cross-sectional shape is provided by Story [24], this ratio is kept constant for all the tubes for a given vowel. For each tube, elliptical shapes having the same area as the one provided by Story [24] are generated with the ratio averaged on Fromkin data [25]. The width and height of the mouth opening, the corresponding ratio and the length of the tubes are presented in