Journal of Sound and Vibration 411 (2017) 247–259

Contents lists available at ScienceDirect

Journal of Sound and Vibration

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

Multimodal modeling and validation of simplified vocal tract acoustics for sibilant /s/

T. Yoshinaga [a], A. Van Hirtum [a,b,*], S. Wada [a]

[a] Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, 560-8531, Osaka, Japan
[b] GIPSA-Lab, UMR CNRS 5216, Grenoble Alpes University, 11 rue des Mathématiques (BP46), 38402, Grenoble, France

A R T I C L E I N F O

Article history:
Received 17 October 2016
Revised 6 September 2017
Accepted 6 September 2017
Available online XXX

Keywords:
Speech production
Multimodal theory
Simplified vocal tract geometry
Sibilant fricatives

A B S T R A C T

To investigate the acoustic characteristics of sibilant /s/, multimodal theory is applied to a simplified vocal tract geometry derived from a CT scan of a single speaker for whom the sound spectrum was gathered. The vocal tract was represented by a concatenation of waveguides with rectangular cross-sections and constant width, and a sound source was placed either at the inlet of the vocal tract or downstream from the constriction representing the sibilant groove. The modeled pressure amplitude was validated experimentally using an acoustic driver or airflow supply at the vocal tract inlet. Results showed that the spectrum predicted with the source at the inlet and including higher-order modes matched the spectrum measured with the acoustic driver at the inlet. Spectra modeled with the source downstream from the constriction captured the first characteristic peak observed for the speaker at 4 kHz. By positioning the source near the upper teeth wall, the higher frequency peak observed for the speaker at 8 kHz was predicted with the inclusion of higher-order modes. At the frequencies of the characteristic peaks, nodes and antinodes of the pressure amplitude were observed in the simplified vocal tract when the source was placed downstream from the constriction. These results indicate that the multimodal approach enables to capture the amplitude and frequency of the peaks in the spectrum as well as the nodes and antinodes of the pressure distribution due to /s/ inside the vocal tract.

1. Introduction

The production mechanisms of sibilant fricatives have been discussed using simplified vocal tract geometries and aeroacoustic theory [1–3]. Stevens [1] argued that the sibilant sound is generated by turbulent flow generated by the constriction in the vocal tract, and modeled the sound generation with a vibrating spoiler in a tube. Shadle [2] assumed that the frequency characteristics of sibilants are determined by the position of the constriction formed by the tongue and the alveolar ridge, and constructed simplified vocal tract replicas to test this hypothesis. The sound generated by the replicas was measured experimentally, and two constriction positions, 15 mm and 25 mm from the lip outlet, reproduced the characteristic spectrum peak of sibilant /s/ (4 kHz), and /ʃ/ (2.5 kHz), respectively. An example of a spectrum of sustained /s/ pronounced by a Japanese male speaker is shown in Fig. 1. In the spectrum, first characteristic peak appeared at 4 kHz and the maximum amplitude was observed at 8 kHz. In previous studies, the first peak of /s/ varied from 4 to 7 kHz, and the high frequency peak can be broad or more peaky depending on the speaker's language or gender [4,5] whereas frequency values depend mainly on the constriction

* Corresponding author. GIPSA-Lab, UMR CNRS 5216, Grenoble Alpes University, 11 rue des Mathématiques (BP46), 38402, Grenoble, France.
E-mail address: annemie.vanhirtum@gipsa-lab.grenoble-inp.fr (A. Van Hirtum).

http://dx.doi.org/10.1016/j.jsv.2017.09.004
0022-460X/

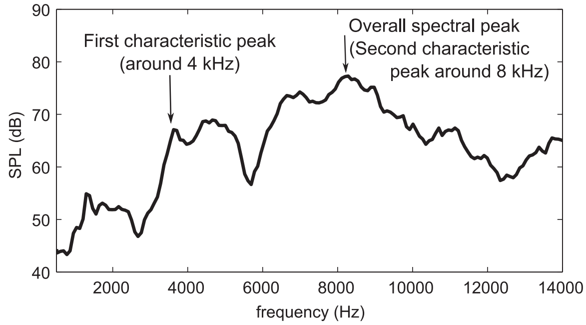

Fig. 1. Example of a spectrum of sustained /s/ pronounced by a male Japanese speaker. There was no vowel context for /s/ and the sound was measured at 30 cm from the lip of the speaker. The simplified geometry is derived from geometrical data measured on this subject (see section 2.2).

position and volume of the front cavity [2].

In addition to the mechanical experiment, Shadle [2] theoretically modeled the sound generation using a plane-wave model for 30 mm front cavity and a dipole source at the upstream surface of the obstacle, and discussed about the frequency characteristics of the generated sound. Howe and McGowan [3] assumed that a monopole sound source is generated at the space between the upper and lower teeth, and constructed a one-dimensional theoretical model. The modeled spectrum and overall sound pressure level (SPL) of /s/ were validated against measurements of the same sound made by human speakers. Previous studies have shown that simplified geometries can be used to reproduce spectral features of sibilant fricatives. However, in the studies described above, the potential impact of higher-order acoustic resonance modes related to the three-dimensional (3D) geometry of the vocal tract on the spectral features was not studied.

Blandin et al. [6] applied multimodal theory, which considers propagation of 3D modes, to vocal tract geometries for vowels. The modeled pressure fields and transfer functions were validated against experimental observations and finite-element simulations. The influence of higher-order modes on spectral amplitudes above 4.5 kHz was observed. Because fricatives are characterized by the acoustic energy in this frequency range, this influence indicates that higher-order modes affect perceptually-significant spectral features of fricatives more than those of vowels. Therefore, influence of higher-order modes on pressure fields inside the vocal tract, as well as radiated pressure outside of the vocal tract for sibilant /s/ will be investigated in this paper.

Motoki et al. [7] and Motoki [8] applied multimodal theory to a simplified vocal tract geometry of /ʃ/ to investigate the influence of geometrical changes on the transfer function of the vocal tract. However, the characteristic spectrum peak of /ʃ/ was lacking from the modeled transfer function because the sound source was located at the inlet of the vocal tract. Indeed, since unvoiced sibilant fricatives are generated mainly by the impingement of jet flow on the teeth and lips, the main sound source is located downstream from the constriction [1–3].

Therefore, in this study, multimodal theory is applied to a simplified vocal tract geometry of sibilant /s/ with two different source positions, at the vocal tract inlet and downstream from the constriction. Since the excitation and propagation of higher order acoustic modes is known to depend on the center line curvature of the geometry, the used simplified geometry accounts for the curvature of the vocal tract near the teeth which was omitted in the cited studies [1–3] although some recent studies do propose a simplified curved geometry to study the flow field [9,10]. The outcome of the model is compared with experimental data obtained by imposing a known acoustic source and by supplying airflow at the inlet. In addition, the position of the source downstream from the constriction was varied in order to investigate the effect of the source position on the acoustic characteristics of sibilant /s/. By comparing theoretically modeled spectra with experimental measurements with flow, we examined whether multimodal theory can predict the characteristic peak of sibilant /s/ or other acoustic characteristics, such as antinodes in the vocal tract. Frequencies of up to 14 kHz are measured and compared with multimodal modeling in this paper.

2. Method

2.1. Multimodal theory

Multimodal theory has been developed and implemented by several researchers, *e.g.* Refs. [6–8,11–15]. Vocal tract geometry is simplified as a concatenation of waveguides with constant cross-sections. In the following, z indicates the main propagation direction; cross-sections are situated in the (x, y)-plane, with the x-axis from left to right and the y-axis from inferior to superior.

In the 3D acoustic field, the amplitude of sound pressure $p\,(x, y, z)$ and particle velocity vector $v\,(x, y, z)$ are defined as

$$p(x, y, z) = \jmath\omega\rho\phi(x, y, z), \tag{1}$$

$$\mathbf{v}(x, y, z) = -\nabla\phi(x, y, z), \tag{2}$$