Applied Acoustics 96 (2015) 101–107

Contents lists available at ScienceDirect

Applied Acoustics

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

Technical Note

Towards sibilant physical speech screening using oral tract volume reconstruction: Some preliminary observations

Y. Fujiso [a], K. Nozaki [b], A. Van Hirtum [a,*]

[a] GIPSA-Lab, UMR CNRS 5216, Grenoble University, France
[b] Osaka University Dental Hospital, Japan

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 18 March 2015
Accepted 20 March 2015

Keywords:
Sibilant fricative
Oral tract reconstruction
Speech/noise screening
Oral health care

ABSTRACT

The current paper considers physical speech screening by supplying air to a rigid replica based on oral tract volume reconstruction of a speaker uttering phoneme /s/ (Reynolds number about 5300). Radiated acoustic pressure spectra were measured for different flow conditions upstream from the reconstructed portion: Reynolds number (2800, 5300, 8900) and turbulence intensity (<4% and >4% by varying the method of air supply). Acoustic spectra obtained with the replica were compared to spectra of phoneme /s/ uttered by the same speaker at different loudness levels. It is found that noise emitted by the replica reproduces the spectral shape (in particular for frequencies up to 8 kHz) and the order of magnitude of spectral features (spectral slopes and dynamic amplitude) of phoneme /s/. Nevertheless, spectral differences (energy discrepancy, peak frequency, negative spectral slope) between phoneme /s/ and sound generated with the replica are observed as a function of Reynolds number as well as a function of upstream turbulence intensity. Therefore, current data suggest that in order to perform sibilant physical speech screening Reynolds number as well as upstream flow conditions need to be taken into account.

1. Introduction

The human oral tract is due to its multiple functions and its crucial role in daily life of interest to many disciplines ranging from physiology to entertainment. A common way to evaluate some aspects related to the oral tract either for clinical, oral health care or research purposes is physical speech screening. Consequently, many research efforts focus on quantifying speech features as a function of oral tract features (e.g. geometrical) [3,10]. Nevertheless, some difficulties are inherently related to speech screening on human subjects such as intra- and inter speaker variability or inherent correlation of flow and geometrical quantities [8]. To avoid these problems, some researchers and developers working in relation to oral health care, make use of the fast development of volume reconstruction and rapid prototyping to study noise production using a reconstructed oral tract portion instead of a human speaker [12].

Sibilant physical speech screening is a concrete example of direct use for speech researchers as well as for clinical researchers e.g. dentistry [13,16]. Therefore, in the following, we consider sibilant fricative physical speech screening based on a partial reconstruction of the oral tract during phoneme /s/ production. The underlying mechanism of sibilant fricative sound production is generally described as noise produced due to the interaction of a turbulent jet, issued from a constriction between the tongue and the hard palate – i.e. the sibilant groove – somewhere in the vocal tract, with a downstream wall or obstacle [3,18]. Consequently, acoustic features of sibilant noise are influenced by both flow and geometrical parameters. Sibilant physical speech screening using a reconstructed geometry allows to study the potential impact of flow parameters on noise production independently from geometrical parameters. Recently, evidence was provided suggesting that different upstream flow conditions affect spectral features used to describe sibilant fricatives such as phoneme /s/, but comparison with spectra of phoneme /s/ uttered by a human speaker was lacking [20]. Such a comparison is needed when aiming sibilant physical speech screening using a replica. Concretely, in the current study spectral features of noise generated using a replica containing a reconstructed oral tract portion are gathered as a function of two flow parameters. Firstly, different Reynolds numbers are assessed to account for the effect of mean velocity on the sound outcome [18,14]. Secondly, the turbulence intensity upstream from the reconstructed portion is altered by varying

* Corresponding author at: GIPSA-Lab, UMR CNRS 5216, Grenoble University, 11 rue des Mathematiques (BP46), 38402 Grenoble, France. Tel.: +33 (0)4 76 57 43 41; fax: +33 (0)4 76 57 47 10.
E-mail address: annemie.vanhirtum@grenoble-inp.fr (A. Van Hirtum).

http://dx.doi.org/10.1016/j.apacoust.2015.03.015
0003-682X/

the method of air supply to the replica in order to change the contribution of aerodynamic noise production expressed by the Reynolds stress tensor [5,20]. Spectral features obtained with the replica are then compared to spectral features of sibilant /s/ uttered by the subject for which the oral tract reconstruction was made. In addition, flow downstream from the replica is visualised in order to qualitatively assess the flow field. It is discussed to which extent a replica based on a reconstructed oral tract can be used for sibilant physical speech screening.

2. Method

All human data are acquired on the same adult male subject (Japanese native speaker, normal sitting position, no health issues and without speech disorder, normal dentition of angle class I of occlusal type) [12].

2.1. Reconstructed geometry

A cone-beam CT scan (CB MercuRay, 512 slices of 512×512 pixels with accuracy ±0.1 mm) was made while uttering phoneme /s/ during 9.6 s following a 'medium' loudness instruction (with a flow rate about 21 L/min) [12]. The oral cavity volume and its shape were reconstructed using a marching cube method and an optical modelling machine (SOUP 2 600GS, material TSR-829) with spatial accuracy ±0.1 mm [12]. Note that due to surface smoothing the exact accuracy is unknown [1,11]. The minimum area of the constricted passage (sibilant groove) yields 3.5 mm^2. The inlet of the reconstructed portion has an elliptical shape with hydraulic diameter 13 mm. A rigid mechanical vocal tract replica (Fig. 1) for sibilant /s/ was obtained by smoothly connecting an upstream circular duct with outlet diameter $D = 8$ mm to the reconstructed geometry (transparent) with elliptical inlet. The connecting portion was designed to have a length of 30 mm with a divergent angle of about 9° so that the flow remained unstalled [15]. The total length of the mechanical replica was 187 mm which corresponds to the averaged length of the vocal tract of an adult male subject [3]. The constriction degree is defined as $1 - A(x)/A_{max}$ with cross-sectional area $A(x)$ perpendicular to the streamwise x-direction and A_{max} the maximum area of the reconstructed portion (bottom right in Fig. 1). The length of the constricted portion l_c with constriction degree greater than 88% ($1 - A(x)/A_{max} \geqslant 88\%$, bottom right in Fig. 1) yields about 15 mm and the length of the resulting front-cavity l_f (portion downstream from the constricted portion) yields about 20 mm. For these dimensions and sound speed $c = 343$m/s (air at a temperature of 20 °C), the half wavelength constriction resonance ($c/2l_c$) approximates 11 kHz and the quarter wavelength front-cavity resonances ($nc/4l_f$ with $n = 1, 3, \ldots$) occur at uneven multiples of 4 kHz (*i.e.* 12 kHz) [18,14]. The acoustic response of the vocal tract replica in absence of flow is further illustrated in Fig. 2. The acoustic response is obtained assuming stationary plane wave propagation within the cross section area perpendicular to the streamwise x direction (Fig. 1) [14]. Moreover, a flanged outlet of the replica is assumed to model the radiation impedance, a complex wavenumber is considered to account for cisco-thermic losses and finally a harmonic noise source is placed at the inlet [14]. The cross-section area is discretized in uniform sections of 1 mm length. At the junctions a transfer matrix is used to propagate the acoustic pressure and volume velocity from the exit to the entrance [14]. The method is repeated for frequencies within the range of interest, from 1 kHz up to 20 kHz using the same source strength. Note that this acoustic response is a first approximation since based on the largest outlet dimension of the replica of $w = 55$ mm (mouth width), the first cut-on frequency of higher order propagation modes is about 3.5 kHz ($1.84c/\pi w$) [14]. The minimum area A_{min} perpendicular to the x-axis yields 5 mm^2 instead of the effective minimum area of 3.5 mm^2 due to the angle of about 45° between the constricted portion and the streamwise x direction (bottom right Fig. 1). In any case – considering either 5 mm^2 or 3.5 mm^2 as the minimum area – the maximum constriction degree yields $\geqslant 99.5\%$, illustrating the severeness of the constriction.

The largest dimension of the replica at its outlet is $w = 55$ mm (mouth width).

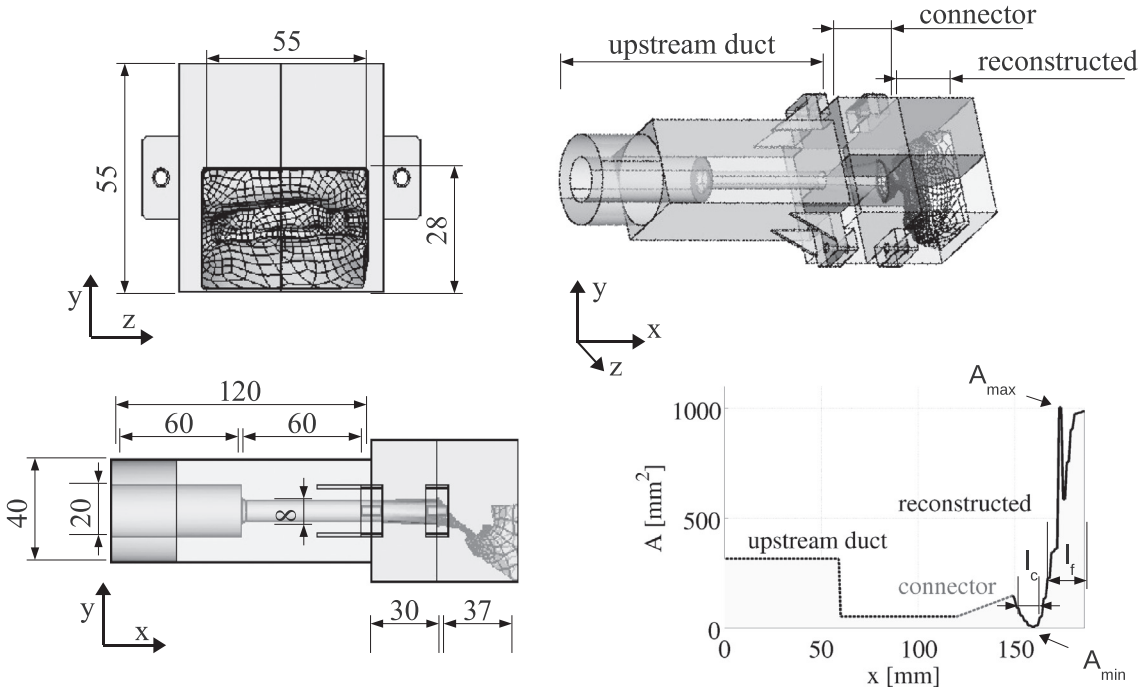


Fig. 1. Schematic overview of the vocal tract replica [mm]: front view (top left), centre side view (bottom left), overall view (top right) and cross-sectional streamwise area A [mm^2] perpendicular to streamwise x direction (bottom right).

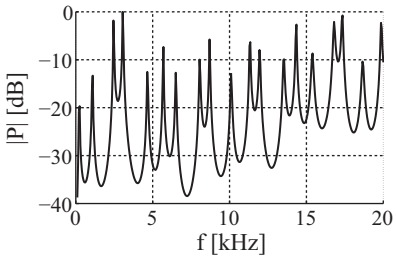


Fig. 2. Acoustic plane wave response of the vocal tract replica for the area cross section given in Fig. 1: Normalised pressure $|P|$ [dB] as a function of frequency f [kHz] at the outlet of a flanged replica.

2.2. Acoustic measurements

Acoustic measurements were done using a calibrated (B&K 4231 calibrator) pressure-field microphone (B&K 4192). The microphone is connected to a pre-amplifier (B&K 2669) and an amplifier (B&K 5935) adding 30 dB. Data were recorded using a data acquisition card (NI PCI-MIO 16 XE) at a sampling frequency of 44.1 kHz. The microphone was located in a quasi-anechoic chamber (Fig. 3) in the horizontal xz-plane at an angle of 37° with respect to the entrance of the chamber [19]. Moreover, the distance between the microphone and the walls yielded at least 800 mm and the distance between the microphone and the exit of the noise generation system was about 940 mm. Experiments were performed at a temperature of 20 ± 1 °C.

Noise was generated at the entrance of the quasi-anechoic chamber in three different ways labelled (I), (II) and (III):

(I) Phoneme /s/. Phoneme /s/ was uttered following three different loudness instructions 'soft', 'medium' and 'loud'. These instructions are commonly used in sibilant fricative speech screening to evaluate the impact of volume flow rate Q_{oral} [8]. The turbulence intensity (T_u, root mean square of streamwise velocity) upstream from the constricted vocal tract portion is unknown. Note that the oral tract geometry is not controlled.

(II) Flow facility (with settling chamber) + replica. Air was supplied to the replica (Fig. 1) using an air compressor (Atlas Copco GA7) followed by a pressure regulator (Norgren type 11-818-987) and a manual valve so that volume flow rate Q_{comp} is controlled (4043 TSI) [20]. The replica is mounted to an upstream settling chamber (0.4 m × 0.4 m × 0.5 m) tapered with acoustic foam (SE50-AL-ML Elastomeres Solutions) to avoid acoustic resonances due to the experimental setup upstream from the replica. The turbulence intensity upstream from the reconstructed vocal tract portion (T_u) is smaller than 4% [20] for all assessed volume flow rates Q_{comp}.

(III) Human blowing + replica. Air was supplied to the replica (Fig. 1) by human blowing following three different effort instructions: 'soft', 'medium' and 'loud'. The turbulence intensity upstream from the reconstructed vocal tract portion (T_u) is greater than 4% for all assessed volume flow rates Q_{blown} [20].

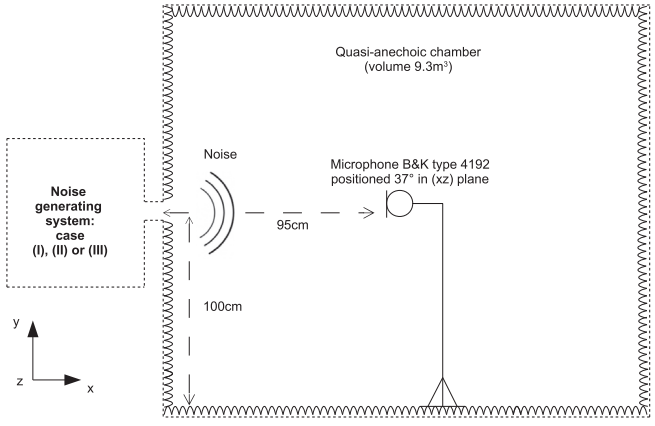


Fig. 3. Overview of experimental setup for acoustic measurements. The microphone is positioned at an angle of 37° with respect to the main airflow direction in the horizontal xz-plane.

The qualitative estimation of the volume flow rate Q_{oral} (I) and Q_{blown} (III) was obtained using the volume flow meter (TSI 4000 series) in combination with the instruction. The upstream turbulence intensity T_u was measured at the centre of the exit of the upstream circular duct using hot-film anemometry (TSI1201-20 and IFA 300) [6,20]. Experimental upstream flow conditions associated with the three different noise generation systems are summarised in Table 1. Reynolds number $Re = 4Q/\pi Dv$ upstream from the reconstructed portion (II) is estimated using the exit diameter of the upstream circular duct of the vocal tract replica $D = 8$ mm (Fig. 1) and kinematic viscosity of air $v = 1.5 \times 10^{-5}$ m^2/s. A qualitative estimation for the Reynolds numbers associated with the different loudness conditions ('soft', 'medium' and 'loud') of noise generation systems (I) and (III) was obtained using again diameter $D = 8$ mm in combination with the measured volume flow rate Q_{oral} (I) and Q_{blown} (III). It is seen that the resulting Reynolds numbers (2800, 5300 and 8900) are the same for all noise generation systems so that they can be used as quantitative labels as was done in Table 1. The same labels are used to indicate the imposed condition of either loudness (I) – flow (II) or blowing effort (III) in the figure legends further on.

Acoustic spectra L_p as a function of frequency f are obtained as the sound pressure level (SPL) of the Welch averaged Power Spectral Density $P(f)$ of the measured acoustic pressure signal [14]:

$$L_p(f) \approx 10 \cdot \log_{10}\left(\frac{|P(f)|}{p_{ref}^2}\right), \quad (1)$$

with $p_{ref} = 2 \times 10^{-5}$ Pa. Concretely, time-averaging of the periodiograms was performed using 20 Hamming windowed energy normalised time segments of fixed duration (7 ms) with 10% overlap resulting in stable spectra with a standard deviation of 5 dB up to 15 dB [8,4]. The sound pressure spectra were parameterized by

Table 1
Summary of upstream flow supply or instruction for different noise generation systems (I, II and III).

Noise generation system[c]		Reynolds number[a,b] Re		
		2800	5300[d]	8900
I	loudness[e] of phoneme /s/ (Q_{oral})	'soft' T_u unknown	'medium'	'loud'
II	Flow facility (Q_{comp}) + replica	10 L/min $T_u < 4\%$	24 L/min	45 L/min
III	Blowing effort[e] (Q_{blown}) + replica	'soft' $T_u > 4\%$	'medium'	'loud'

[a] Based on upstream duct outlet diameter $D = 8$ mm and rounded to the nearest multiple of 100.
[b] Corresponding to about 10,000, 19,000 and 21,000 based on the minimum area of the reconstructed portion (Fig. 1).
[c] Human data are gathered on one single subject.
[d] Condition used for oral tract reconstruction (Section 2.1).
[e] Loudness/blowing effort instruction: 'soft' (≈11 L/min), 'medium' (≈21 L/min) and 'loud' (≈45 L/min).

considering dynamic amplitude A_d, spectral peak frequency f_m and linear regression slopes S_1 ($f_{min} \leqslant f \leqslant f_m$, with f_{min} the frequency associated with minimum spectral amplitude, positive slope) and S_2 ($f_m \leqslant f \leqslant 20$ kHz, negative slope) in accordance with fricative spectral characterisation [8]. Dynamic amplitude A_d is commonly reported [8] to depend on volume flow rate (loudness level), peak frequency f_m relates to the geometry (front-cavity resonance) and spectral slopes depend on volume flow rate (loudness level) as well as geometry (front-cavity resonance and hence peak frequency) [8]. In addition, recently [20] evidence is provided that varying upstream flow conditions might affect negative spectral slope S_2 and dynamic amplitude A_d as well, so that the relationship between spectral features and Reynolds number might be altered by varying upstream flow conditions such as turbulence intensity.

Besides spectral parameters, the acoustic energy within each octave band with centre frequency f_c was considered [4]. The acoustic energy was normalised by the total acoustic energy among all bands in order to compensate for a variation of Reynolds number. Note that the octave band with centre frequency $f_c = 16$ kHz is limited to 22,000 Hz instead of 22,627 Hz to avoid aliasing given the acoustic sampling frequency of 44.1 kHz. The difference of normalised acoustic energy $\xi(f_c)$ is quantified with respect to noise generation system II (flow facility + replica) since this is the most reproducible generation system (volume flow rate and geometry are controlled/known). The difference $\xi(f_c)$ is expressed in dB.

2.3. Flow visualisation

The flow facility of noise generation system (II) (compressed air supply with settling chamber upstream from the replica) was used to perform flow visualisation of the near field of the replica exit. Neutrally buoyant white smoke was injected in the settling chamber by means of a fog machine (Kool Light, FOG-1200E, 1200W, maximum volume flow rate 300 m^3/min). The effective Schmidt number of the smoke was of order 10^6. Two dimensional illumination was applied with a two-dimensional laser light beam. The laser light sheet was generated by a class IIIb laser light source (Laserglow Technologies, LRS-0532-TFM-00200-10, 234.2 mW, wavelength of 532 nm and with spectral line-width <0.1 nm) to which a 10° cylindrical lens was added. The resulting laser sheet had a thickness <3 mm. To record the illuminated smoke pattern, a colour camera (Casio, EXILIM Pro EX-F1, 6.0 million effective pixels and 12X optical zoom) was positioned perpendicular to the laser sheet. The back wall of the test chamber was painted in black to ensure a good contrast with the smoke pattern. Movies were recorded at 300 fps so that the time interval between consecutive images was 0.0033 s which ensures a good freezing of the flow development. The exposure time used to gather each image was constant and did not depend on the assessed Reynolds number.

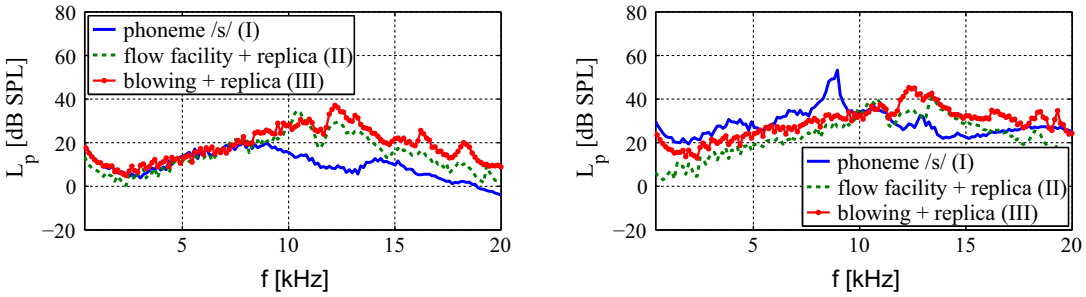
3. Results

3.1. Acoustic spectra

In general, generated sounds are perceived by a human listener as noisy as expected for a natural human phoneme /s/ utterance. A more quantitative analysis is obtained by analysing acoustic spectra. Spectra of the radiated acoustic pressure are illustrated in Fig. 4 for $Re \approx 5300$ and $Re \approx 8900$. All acoustic spectra, *i.e.* for all noise generation systems (I, II and III) and Reynolds numbers (2800, 5300 and 8900), are dominated by frequencies greater than 3.5 kHz, which corresponds to the order of magnitude of the first cut-on frequency of higher order propagation modes (Section 2.2). Spectra observed for $Re \approx 5300$ (Fig. 4(a)) match well for all three noise generation systems (I, II and III) up to 8 kHz. This is reassuring since $Re \approx 5300$ corresponds to the condition for which the oral cavity reconstruction was obtained (Section 2.1). Beyond 8 kHz spectral differences occur, which are more important when the geometry is not controlled (I versus II/III: different shape and amplitude) than when the method of air supply is varied for a same geometry (II versus III: overall similar shape and different amplitude). Increased amplitude in the high frequency noise portion might be due to the increase in upstream turbulence intensity causing an increase in aerodynamic noise (Reynolds tensor) when comparing air supplied using a flow facility ($T_u < 4\%$) and air supplied by human blowing ($T_u > 4\%$) (II versus III in Table 1). Comparison of spectra for noise generation systems II and III for $Re \approx 8900$ (Fig. 4(b)) reveals the same tendencies as observed for $R \approx 5300$ (overall similar shape and different amplitude) whereas spectra for phoneme /s/ (I) differ in shape as well as amplitude from those obtained with the replica (II and III). Moreover, amplitude and shape of phoneme /s/ spectra (I) for $Re \approx 8900$ and $Re \approx 5300$ differ in general whereas the impact of Reynolds number on spectra obtained with the replica is limited to amplitude and the spectral shape is less affected.

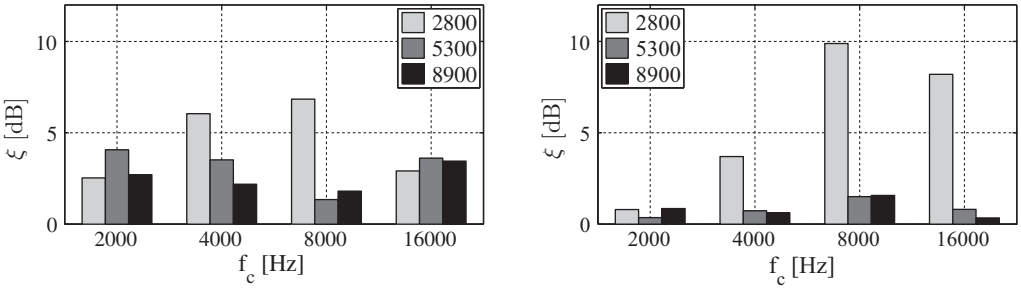
The given qualitative spectral characterisation is confirmed when considering the normalised acoustic energy discrepancy ($\xi(f_c)$ for $f_c \geqslant 2$ kHz) shown in Fig. 5. For $Re \approx 5300$ and $Re \approx 8900$, the overall energy discrepancy is small ($\xi(f_c) < 2$ dB) when the replica is used and the method of air supply is varied (III versus II in Fig. 5(b)). Again for $Re \approx 5300$ and $Re \approx 8900$, the overall energy discrepancy increases ($\xi(f_c) < 4$ dB) when comparing noise generated with the replica and phoneme /s/ (I versus II in Fig. 5(a)). For $Re \approx 2800$ on the other hand, varying the method of air supply to the replica increases the spectral discrepancy significantly ($\xi(f_c \geqslant 4$ kHz$) \geqslant 4$ dB), which emphasizes the impact of upstream flow conditions (such as turbulence intensity T_u) on the acoustic outcome and in particular on the high frequency portion of the spectra. For spectra obtained for phoneme /s/ at $Re \approx 2800$ (I versus II in Fig. 5(a)) the overall energy discrepancy is important ($\xi(f_c) > 2$ dB). Nevertheless, compared to spectra obtained for higher Reynolds numbers $Re \geqslant 5300$ the increase in energy discrepancy is observed only for $4 \leqslant f_c \leqslant 8$ kHz and is less pronounced ($\xi(f_c = 4)$ kHz <4 dB). This suggests that a change in geometry (affecting e.g. peak frequency position) contributes more to the spectral discrepancy than changes in upstream turbulence intensity. Although, since the upstream turbulence intensity T_u for phoneme /s/ (I) is unknown (Table 1) no absolute conclusion can be made.

Spectral features outlined in Section 2.2 are quantified: dynamic amplitude A_d, spectral peak frequency f_m and linear regression slopes $S_{1,2}$. Fig. 6(a) illustrates dynamic amplitude A_d and linear regression slope S_2 who were observed to be correlated to the loudness level (volume flow rate) in human sibilant fricative production [8]. Fig. 6(b) shows linear regression slope S_1 and spectral peak frequency f_m who were observed to be correlated to the geometry (front cavity resonance) in human sibilant fricative production [8]. The impact of Reynolds number on spectral features A_d, f_m and $S_{1,2}$ is summarised in Table 2. For all noise generation systems (I, II and III) and Reynolds numbers (2800, 5300 and 8900) spectral parameters yield values within the range reported for sustained /s/ phonemes [8]. Dynamic amplitude A_d increases with Reynolds number Re (Fig. 6(a)) for all noise generation systems (I, II and III) [8]. Nevertheless, the increase is more important for phoneme /s/ ($\leqslant$110% for I) than when the replica with constant geometry is used ($\leqslant$10% for II and III). This is due to the sharp spectral peak observed for phoneme /s/ (Fig. 4(b)). For the highest assessed Reynolds number $Re \approx 8900$, the order of magnitude of



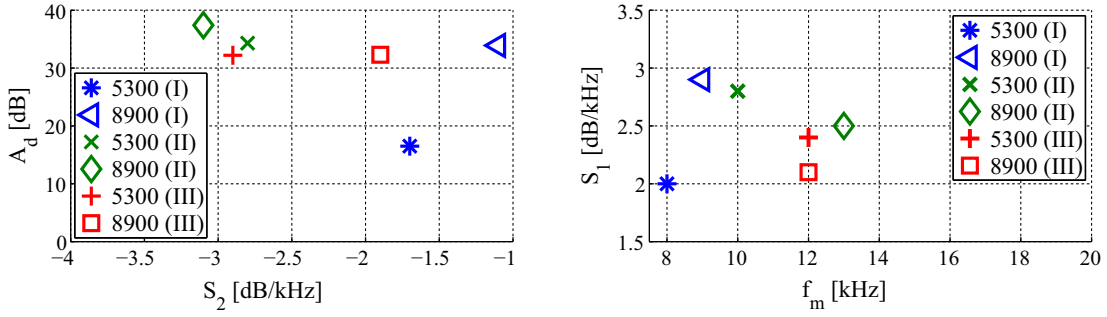
(a) $Re \approx 5300$ ('medium') (b) $Re \approx 8900$ ('loud')

Fig. 4. Illustration of acoustic spectra $L_p(f)$ for all noise generation systems (Table 1) [full line (I), dashed line (II) and dotted full line (III)]: (a) $Re \approx 5300$ ('medium') and (b) $Re \approx 8900$ ('loud').



(a) I versus II (b) III versus II

Fig. 5. Difference in acoustic energy as a function of octave band with centre frequency f_c and Reynolds number Re (grey shaded legend) – 2800 ('soft'), 5300 ('medium') and 8900 ('loud') – normalised by the total energy with noise generation system II taken as a reference ($\xi(f_c)$): (a) phoneme /s/ utterance (I versus II). (b) Variation of method of air supply (III versus II).



(a) $Re \approx 5300$ ($\star$, $\times$, $+$), $Re \approx 8900$ ($\triangleleft$, $\Diamond$, $\square$) (b) $Re \approx 5300$ ($\star$, $\times$, $+$), $Re \approx 8900$ ($\triangleleft$, $\Diamond$, $\square$)

Fig. 6. Illustration of spectral parameters for all noise generation systems (Table 1) and $Re \approx 5300$ ('medium') [$\star$ (I), $\times$ (II), + (III)] and $Re \approx 8900$ ('loud') [$\triangleleft$ (I), $\Diamond$ (II), $\square$ (III)]: (a) dynamic amplitude A_d as a function of negative spectral slope S_2 and (b) positive spectral slope S_1 as a function of peak frequency f_m.

Table 2
Summary impact (increase ↗, decrease ↘, no influence →) of increasing Reynolds Re number of spectral parameters for all noise generation systems given in Table 1 (I, II and III).

	Increasing Re		
	I	II	III
A_d	↗ (110%)	↗ (10%)	↗ (2%)
S_1	↗ (30%)	↘ (12%)	↘ (12%)
f_m	↗ (12%)	↗ (25%)	→ (0%)
S_2	↗ (55%)	↘ (12%)	↗ (50%)

the dynamic amplitude ($A_d = 34 \pm 2$ dB) is the same for all noise generation systems (I, II and III). For phoneme /s/ (I), spectral slopes increase with Reynolds number (>30% for S_1 and $> 55\%$ for S_2) as observed for sustained phonemes /s/ [8]. This increase of spectral slopes with Reynolds numbers is not observed when air is supplied to a fixed geometry (II and III) (see e.g. S_1 in Table 2). Spectral peaks f_m for the replica (II and III) occur in the range 10 kHz up to 13 kHz (Fig. 6(b)), which is of the same order of magnitude as the constriction and front-cavity resonances of the reconstructed portion of the replica (Section 2.1). In the case of phoneme /s/, the peak frequency ($f_m < 10$ kHz) is lower than observed with the replica (Fig. 6(b)) indicating a change in geometry compared to the reconstructed oral tract geometry. Moreover, the sharp spectral peak (about 9 kHz in Fig. 4(b)) observed for phoneme /s/ at $Re \approx 8900$ suggests that whistling (Helmholtz resonance) might interfere with fricative sibilant production. For

phoneme /s/ (I), spectral slopes increase with Reynolds number (>30% for S_1 and > 55% for S_2) as observed for sustained phonemes /s/ [8]. This increase of spectral slopes with Reynolds number is not always observed when air is supplied to a fixed geometry (II and III) (see $S_{1,2}$ in Table 2).

The current comparison of spectral features for phoneme /s/ and for noise generated with a replica suggests that the relationship between loudness level and vocal tract geometry is not dual (between Reynolds number and geometry), but more complex since at least a triple relationship between Reynolds number, turbulence intensity and geometry needs to be considered. Since turbulence intensity and more in general upstream flow conditions is currently not measured for human speakers, physical sibilant speech screening using replicas with reconstructed portions seem more appropriate for studies aiming understanding. At the same time, current data suggest that the order of magnitude of spectral features of phoneme /s/ can be reproduced using such a replica. Therefore, physical sibilant speech screening can be a useful tool for oral health care applications as well.

3.2. Flow visualisation

Flow visualisation was assessed for several Reynolds numbers ($2800 \leqslant Re \leqslant 8900$). Instantaneous flow images in the near field downstream from the replica are illustrated in Fig. 7. Observations did not depend on Reynolds number. Flow in the horizontal xz-plane (Fig. 7(a) and (b) taken from a position with positive y-coordinate) is characterised by vortex formation and the interaction of large coherent structures. The jet issuing downstream from the replica is either symmetrical with respect to the xy-plane (Fig. 7(a)) or asymmetrical with respect to the xy-plane (Fig. 7(b)) resulting in a sideways deflected jet. Visualisation of the longitudinal xy-plane shows that the jet is asymmetrical with respect to the horizontal xz-plane (Fig. 7(c)) with an angle determined by the replica's geometry (Fig. 1). In addition, jet bifurcation within the xy-plane (Fig. 7(b)) is observed to occur as well as flow recirculation. Consequently, the flow field immediately downstream the replica is complex and highly three-dimensional. This is further illustrated by visualisation of the flow field in the transversal yz-plane at 1 mm (Fig. 7(e)–(h)) and 2 cm (Fig. 7(i)–(l)) downstream from the replica exit. Indeed, jet deflection (Fig. 7(h)), recirculation and axis switching (Fig. 7(j) compared to Fig. 7(k)) are observed and contribute to a complex, rapidly changing and three-dimensional flow field. Experimentally gathered [18,21,4], numerically simulated [17,2] or modelled [7,9,22] flow field characteristics immediately downstream from simplified geometries of the oral tract during sibilant fricative production do not reflect the complexity of the visualised flow field using the replica obtained with the reconstructed oral tract portion. At current stage, it is not clear to what extent such a complex flow field is realistic for a sustained phoneme /s/ uttered by a human speaker. Moreover, the interaction of the vortex structures with the replica outlet might provide an unrealistic contribution to the radiated sound field [5,4].

4. Conclusion and discussion

The current study is preliminary since it is based on a single speaker and a single rigid replica. Nevertheless, the following observations are made based on (1) the comparison between

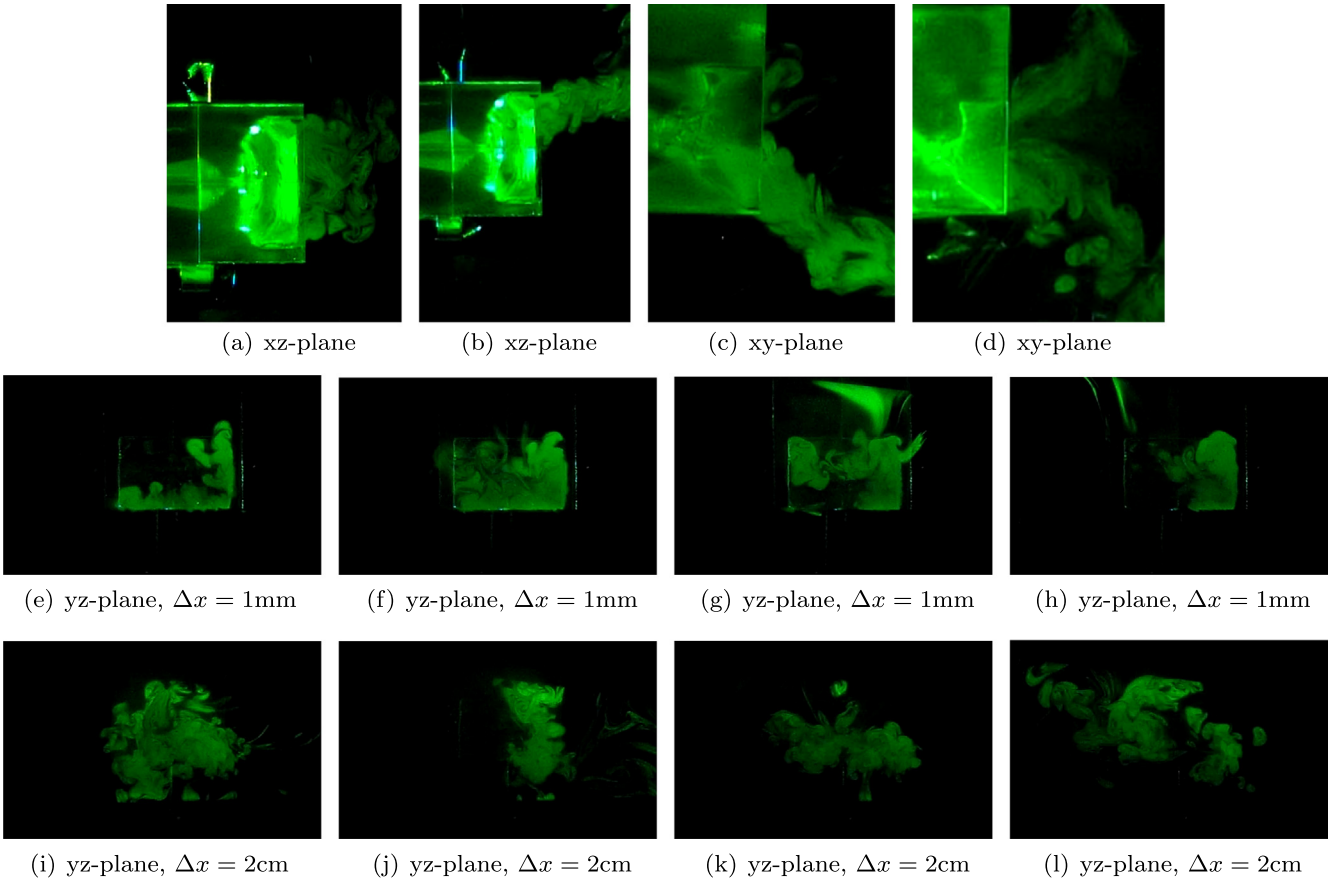


Fig. 7. Illustration of flow visualisation images at $Re \approx 5300$ ('medium'): (a and b) horizontal xz-plane taken from a position with positive y-coordinate. (c and d) longitudinal xy-plane, (e–h) transversal yz-plane at 1 mm downstream from the replica exit ($\Delta x = 1$ mm). (i–l) Transversal yz-plane at 2 cm downstream from the replica exit ($\Delta x = 2$ cm).

radiated acoustic spectra of noise generated by supplying air to the replica and phoneme /s/ utterances and (2) flow visualisation immediately downstream from the replica exit:

- For all assessed Reynolds numbers and for all noise generation systems (either with the replica or from a human speaker) the acoustic energy is situated at frequencies above the cut-on frequency of the first non-plane mode (>3.5 kHz). Consequently, higher order acoustic propagation modes potentially affect the radiated sound field and need to be taken into account when aiming sibilant sound screening either using replicas or using human speakers.
- Spectral similarity between phoneme /s/ and noise generated with the replica is most obvious for loudness instruction ($Re \approx 5300$) for which the oral volume reconstruction is realised. The spectral match for frequencies up to 8 kHz illustrates that such a replica can at least partially reproduce acoustic spectra of the same speaker and hence might be useful for speech screening applications, modelling or experimental studies in the frequency range below 8 kHz. Additional replicas based on volume data obtained for a 'soft' and 'loud' loudness instruction of the same speaker are of interest in order to understand how acoustic radiated sound spectra are affected by a potential change in oral tract shape.
- Noise generated with the replica is characterised by spectral features (dynamic amplitude, spectral slopes and peak frequency) of the same order of magnitude as observed for phoneme /s/. This supports and encourages the use of volume oral tract reconstruction for speech screening applications. Nevertheless, the peak frequency of phoneme /s/ is lower than the one observed with the replica for all assessed Reynolds numbers. The decrease is probably due to intra-speaker variability of the oral tract geometry between different phoneme /s/ utterances. Moreover, current data show that spectra of phoneme /s/ can be influenced by additional noise sources such as whistling observed for the highest assessed Reynolds number ($Re \approx 8900$). Moreover, the tendency of spectral features as a function of Reynolds number is different for noise produced with the replica as for phoneme /s/. This illustrates the interest to study flow parameters in combination with a constant replica geometry.
- Spectral features depend on flow conditions upstream (Reynolds number as well as turbulence intensity) from the reconstructed portion of the replica. The upstream turbulence intensity (method of air supply) influences the energy discrepancy observed between the replica and phoneme /s/. The influence of upstream turbulence intensity is more pronounced for the lowest assessed Reynolds number ($Re \approx 2800$). The Reynolds dependence of spectral features (peak frequency and negative spectral slope) is affected by the turbulence intensity (method of air supply) as well. Therefore, current data provide evidence that in order to perform physical speech screening it is not sufficient to define a constant geometry based on volume reconstruction and vary Reynolds number. In addition, a detailed study of upstream flow conditions in relation to sibilant sound production and radiated spectra using a replica with constant geometry is needed. Such a study can also contribute to criteria characterising upstream flow conditions relevant to human sibilant fricative production since such a direct measurement on a human subject is extremely difficult to achieve.
- The current study focused on the impact of turbulence intensity on the spectral discrepancy between noise generated with a replica and phoneme /s/. Flow visualisation immediately downstream from the replica suggests that future studies need to account for different exit conditions (geometry, rigidity) as well in order to consider the impact of the visualised highly three-dimensional flow field on noise production and radiation as well as its relevance for phoneme /s/ uttered by human subjects.

Acknowledgment

This work was partly supported by EU-FET Grant (EUNISON 308874).

References

[1] Araki K, Itsuo F, Maki K, Okano T, Sakamaki K. Clinical evaluation of a conebeam X-ray ct system (cb mercury) for dentomaxillofacial region. Medix 2003;38(1):4–7.
[2] Cisonni J, Nozaki K, Van Hirtum A, Grandchamp X, Wada S. Numerical simulation of the influence of the orifice aperture on the flow around a teeth-shaped obstacle. Fluid Dyn Res 2013;45(1):1–19.
[3] Fant G. The acoustic theory of speech production. The Hague: Mouton; 1960.
[4] Fujiso Y. Etude aéroacoustique d'un canal avec obstacle(s) – Application à la production de fricatives. PhD thesis, University of Grenoble; 2014.
[5] Goldstein M. Aeroacoustics. New York: McGraw-Hill; 1976.
[6] Grandchamp X, Van Hirtum A, Pelorson X. Hot film/wire calibration for low to moderate flow velocities. Meas Sci Technol 2010;21:115402.
[7] Howe MS, McGowan RS. Aeroacoustics of [s]. Proc Roy Soc A 2005;461(2056):1005–28.
[8] Jesus LMT, Shadle CH. A parametric study of the spectral characteristics of European Portuguese fricatives. J Phonetics 2002;30(3):437–64.
[9] Krane M. Aeroacoustic production of low-frequency unvoiced speech sounds. J Acoust Soc Am 2005;118(1):410–27.
[10] Narayanan SS, Alwan AA, Haker K. An articulatory study of fricative consonants using magnetic resonance imaging. J Acoust Soc Am 1995;98:1325–47.
[11] Nozaki K. Numerical simulation of sibilant [s] using the real geometry of a human vocal tract. In: High performance computing on vector systems 2010. Springer; 2010. p. 137–48.
[12] Nozaki K, Yoshinaga T, Wada S. Sibilant /s/ simulator based on computed tomography images and dental casts. J Dent Res 2014;93:207–11.
[13] Palmer JM. Analysis of speech in prosthodontic practice. J Prosthet Dent 1974;31(5):605–14.
[14] Pierce A. Acoustics. An introduction to its physical principles and applications. New York: Acoustical Society of America; 1991.
[15] Pope S. Turbulent flows. Cambridge University Press; 2005.
[16] Pound S. Let /s/ be your guide. J Prosthet Dent 1977;38(5):482–9.
[17] Ramsay G, Shadle C. The influence of geometry on the initiation of turbulence in the vocal tract during the production of fricatives. In: Proc 7th int seminar on speech production (ISSP7). Ubatuba, Brazil; December 2006. p. 8.
[18] Shadle CH. The acoustics of fricative consonants. PhD thesis, Massachusetts Institute of Technology; 1985.
[19] Van Hirtum A, Fujiso Y. Insulation room for aero-acoustic experiments at moderate Reynolds and low Mach numbers. Appl Acoust 2012;73(1):72–7.
[20] Van Hirtum A, Fujiso Y, Nozaki K. The role of initial flow conditions for sibilant fricative production. J Acoust Soc Am 2014;136:2922–5.
[21] Van Hirtum A, Grandchamp X, Cisonni J. Reynolds number dependence of near field vortex motion downstream from an asymmetrical nozzle. Mech Res Commun 2012;44:47–50.
[22] Van Hirtum A, Pelorson X, Estienne O, Bailliet H. Experimental validation of flow models for a rigid vocal tract replica. J Acoust Soc Am 2011;130:2128.