Quantal aspects of non anterior sibilant fricatives: a simulation study

Martine Toda[1], **Shinji Maeda**[2]

[1]LPP – Laboratoire de Phonétique et Phonologie
Université Paris III ; CNRS UMR 7018
19, rue des Bernardins -- 75005 Paris, France

[2]École Nationale Supérieure des Télécommunications
CNRS, LTCI UNR 5141
46, rue Barrault, Paris -- 75634 Cedex 13, France

`martinetoda@yahoo.co.jp` `maeda@tsi.enst.fr`

Abstract. *The quantal theory (Stevens, 1972) argues that there exist stable and unstable regions in the acoustics when the articulatory parameters are varied in a continuous way. Such a relation has been proved to be involved in the /s/-/ʃ/ contrast in English (Perkell et al., 1979). However, the quantal aspects within non anterior sibilant phonemes such as /ɕ/-/ʂ/ in Chinese are not well established. This paper reports an acoustic modeling experiment of sibilant-like configurations where the front cavity and tongue constriction lengths are systematically varied. The results suggest the articulatory space to be parceled into several 'quantal' regions delimited by unstable regions due to (1) the jump of the lowest prominent spectral peak towards higher frequencies when the first front cavity resonance coincides with a palatal channel free zero and is weakened (when a pressure source is assumed to be located at the exit of the constriction); and (2) the interaction of the prominent peak resonance with back cavity resonances at their crossing points, with changes in formant affiliation. The relatively stable regions delimited by those discontinuities are assumed to provide robust prototypes for non-anterior sibilants, although subject-dependent factors can cause some modifications in quantal patterns.*

1. Introduction

The quantal theory (Stevens, 1972) assumes that plateau-like regions bounded by regions that are relatively sensitive to articulatory perturbations constitute the acoustic (and perceptual) correlates of distinctive features. Such a relation is supported by Perkell *et al.* (1979)'s study on the English sibilants /s/ and /ʃ/ ('š'). These authors showed that /s/ involves a tongue tip-incisor contact, whereas /ʃ/ involves little or no contact, which is assumed to give rise to an abrupt increase of the front cavity volume in the s-ʃ continuum responsible for the abrupt change observed in the spectral characteristics of frication noise.

Some of the word's languages possess more than two distinctive places of articulation for coronals. However, the quantal property of those sounds is not well established. Chinese is such a language, and possess two contrastive [- anterior] sibilants /ɕ/ and /ʂ/ in addition to /s/. Ladefoged and Wu (1984) observe that, intriguingly, the Chinese so-called retroflex does not involved a curling-back of the tongue and thus seems to be of a different nature from Dravidian retroflex fricatives, which, in turn, seem to vary in the degree of retroflexion depending on languages (ex: Hindi vs. Telugu -- Ladefoged and Bhaskararao, 1983). As noted by Ladefoged and Wu, we could then wonder into how many categories the 'places of articulation' (specified by place of constriction and tongue shape) should be grouped together. According to Ladefoged and Wu, the [language-specific] phonological rules account for such categories. However, the wide panel of possible articulations for non-anterior coronal fricatives reported in these studies does not disprove the hypothesis that the articulatory space is not uniformly occupied by these fricatives.

The purpose of this study is to examine, through a systematic acoustic simulation, the quantal aspects (or lack of such) of non-anterior fricative sibilants. In fricative consonants, the vocal-tract comprises a narrow constriction required in the generation of turbulent noise source. Thus, the simplest model would consist of a back cavity, a constriction, and a front cavity. In addition, the location of tongue constriction and the shape of the tongue define the 'place of articulation' of natural sibilant fricatives. In the simulation, a backward shift of the point of constriction could be modeled as a lengthening of the front cavity, whereas an apical, laminal or palatalized articulation could be modeled by varying the length of constriction. Such simple models are able to give satisfying agreement with the real spectra of French /ʃ/ variants produced with various shapes of the tongue (Toda, 2006). In that study, it is suggested from the examination of MRI data that there exist (at least) two different ways of producing /ʃ/, which appeared to be acoustically equivalent. Therefore, such three-tube model will be used in this systematic study.

Contrary to vowels, different kinds of excitation source are assumed to arise in sibilant fricatives (e.g. Narayanan and Alwan, 2000). Narayanan and Alwan have shown that the combination of one monopole source (located at constriction) and one or two dipole sources (located respectively inside the front cavity and at the teeth) permit to fit well the spectra of sibilant fricatives. Thus vocal-tract transfer functions are calculated here with a monopole flow and dipole pressure sources located at various points.

2. Method

The DOS-interface program VTF_fric, a frequency domain version of the acoustic simulation of the vocal-tract system (Maeda, 1982), has been used for the calculation of the transfer characteristics of vocal-tract models between 0 and 8 kHz. In our experiment, the following fixed conditions were used:

Table 1. Constant parameter values used with VTF_fric.

source type	output type	wall condition	radiation load at lips	glottal opening	subglottal system
(a) pressure ('dental' and 'palatal' locations) (b) flow ('palatal' location)	radiated sound pressure	yielding	RL-circuit	1 cm^2	none

Our sibilant fricative model consists of a back cavity (BC), a constriction ('palatal channel'; PC), and a front cavity (FC), as shown in Figure 1. The model has only two parameters, the length of PC and that of FC. The cross-sectional areas of cavities are fixed as indicated in Table 2. In simulation, the length of PC and FC are varied, respectively, from 5 to 65 mm and from 10 to 45 mm in 5 mm steps. The length of BC is determined as the total tract length (fixed at 170 mm) minus the sum of PC and FC length.

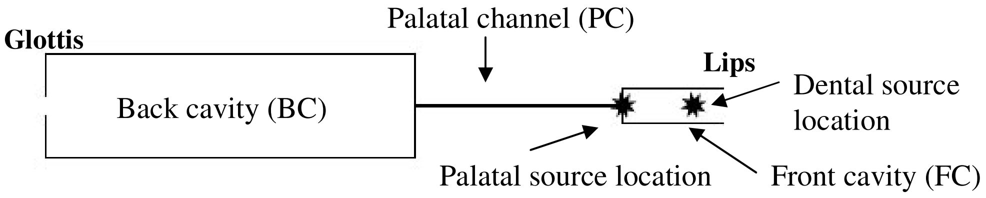


Figure 1. Schematic representation of the vocal-tract model.

The area functions of the sibilant fricative model were generated by using a Matlab script. The values of fixed model variables were set as shown in Table 2.

Table 2. Fixed variables in the vocal-tract model.

length of each section	total length of model	source location	section area of back cavity	section area of palatal channel	section area of front cavity
1 mm	170 mm	(a) 10 mm behind the lip end of the model ('dental source'); (b) at the exit of the palatal channel ('palatal source')	6 cm^2	0.5 cm^2	2 cm^2

The transfer functions obtained by the simulation are characterized by the frequency of their lowest prominent peak (LPP; see Figure 4). The LPP is defined as the frequency of the lowest peak that meets two conditions: (1) its level is at least 20 dB higher than the deapest dip in the lower frequencies (excluding the one at 0 Hz); and (2) no peak of higher intensity exists within the next 500 Hz. The second condition is intended to avoid detecting a BC resonance as LPP when it is close to a PC or FC resonance of greater energy. The LPP has been chosen because it fitted the best the subjective impression of height when compared to other descriptive measures such as the center of gravity (mean of frequency weighted by intensity) or the most prominent peak.

In order to find out the discontinuities in the articulatory-to-acoustic mapping, the average difference of LPP frequencies was calculated between vertical and horizontal neighbors. Figure 2 illustrates how the average difference of LPP is calculated.

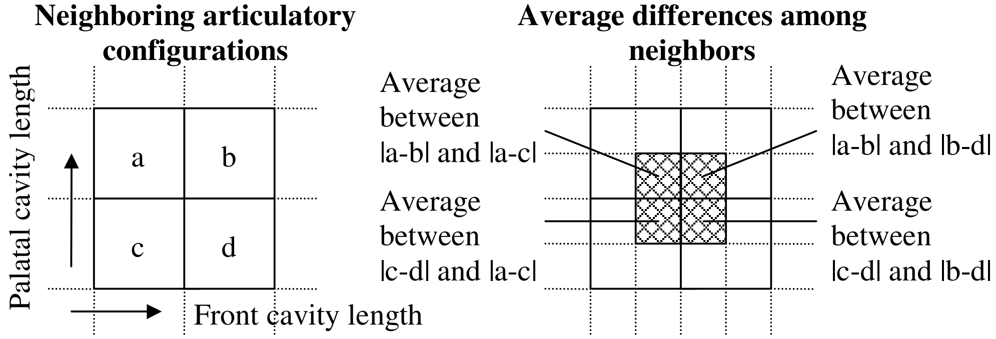
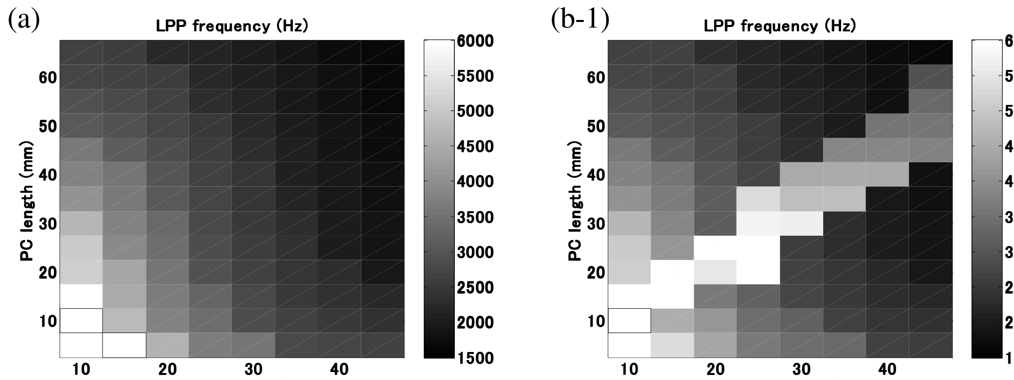


Figure 2. Average differences among neighbors calculated in order to quantify the degrees of stability in the articulatory to acoustic mapping. The symbols a, b, c and d represent the LPP frequency in Hz or in Bark converted using Traunmüller’s (1990) formula.

3. Results

Figure 3 illustrates the frequency distributions of the LPP for the three source conditions. The frequency distribution of the LPP is very similar between the dental pressure source, in Figure 3a, and the palatal flow source condition, in Figure 3b-2. In these two configurations, the length of FC and the length of PC behave in a complementary fashion. This is due to the reciprocal interaction between the first resonance of FC and that of PC when they come close together. At the top left corner of these figures, the lowest resonance is more likely to arise from PC, whereas at the right bottom, the first FC resonance is the lowest of both and is detected as the LPP.

For the same model configurations, the transfer functions calculated with a dental pressure source only differ from the palatal flow source conditions in that a free zero of FC appears at the left side of the LPP (see Figure 4). This explains why the LPP distributions in these source conditions are similar to each other. In addition, with a dental source, the spectra at the left bottom of Figure 3a, where PC and FC have very short lengths, are typically flat or rising, without a sharply-edged cut-off frequency (see Figure 5). For those vocal-tract configurations, the small size of the cavities gives rise to resonances and antiresonances of higher frequencies that are not captured within the frequency range of the simulations.



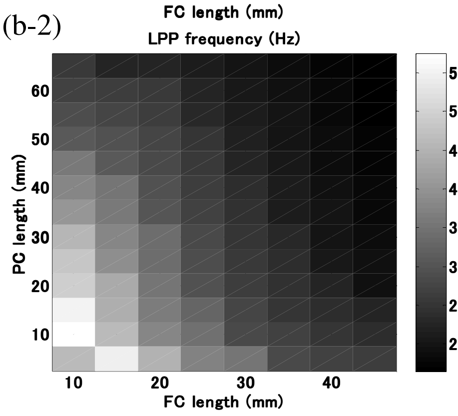


Figure 3. Distribution of the LPP frequency (dark: low; bright: high) in the articulatory space defined by the length of front cavity (horizontal dimension) and that of the palatal channel (vertical dimension). (a) dental pressure source; (b) palatal source; (b-1) pressure source; (b-2) flow source. White rectangles with borders indicate that no LPP was detected.

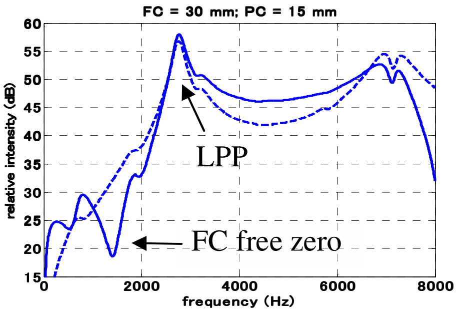


Figure 4. Example of transfer functions obtained for the same vocal-tract model configurations. Dental pressure source (plain line) and palatal flow source (dashed line) conditions.

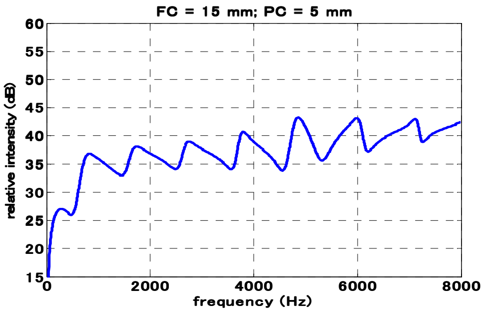


Figure 5. Example of a flat rising transfer function (dental pressure source), where both PC and FC are short and no LPP is detected.

The LPP frequency distribution with the palatal pressure source exhibits a quite original pattern (see Figure 3 b-1). A jump into higher frequencies is observed in a white/light gray diagonal band as the length of FC and PC are proportionally increased. Within this band, the first FC resonance coincides with a free zero arising from PC, and thus is weakened or canceled, so that LPP is then affiliated to PC. Figure 6 shows a spectrographic nomogram representing the variation of peaks and dips of the transfer functions as the length of PC is varied, while FC is fixed at 30 mm. A big jump of the LPP into higher frequencies can be seen at the PC lengths of 30 to 40 mm.

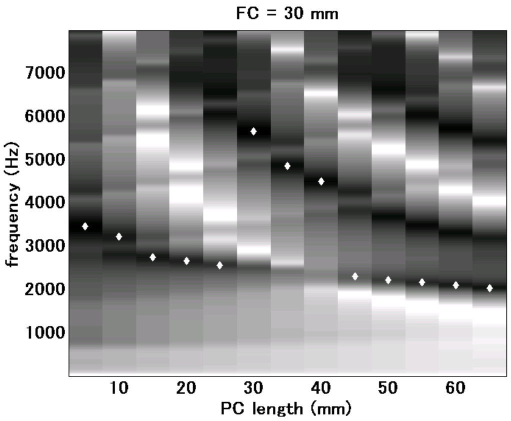


Figure 6. Spectrographic nomogram (dark: peaks; bright: dips) with FC length = 30 mm; palatal pressure source. The palatal channel varies in length from 5 to 65 mm. The white dots indicate the frequency of the LPP detected for each of the transfer functions.

For the dental pressure source condition, Figure 7 represents the pattern of discontinuities estimated by the average difference of LPP among neighbors. In this figure, several regions of stability (small LPP frequency differences) as well as several regions of instability are clearly visible.

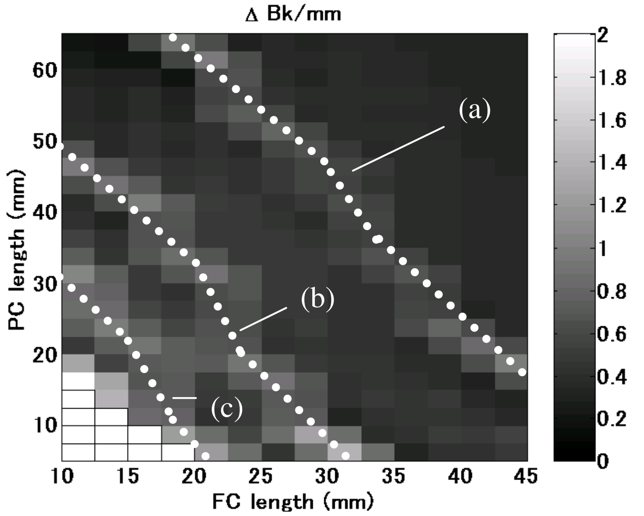


Figure 7. Average difference between neighboring vocal-tract configurations (in Δ Bark), for dental pressure source. Dark patches indicate stable regions (differences are small) and bright ones unstable regions, highlighted by white dotted lines labeled (a) to (c).

We assume the discontinuities that are shown in this representation would have some impact on the perception since those neighboring regions are apart from one another by 1 to 2 Barks (300 to 1500 Hz). The examination of nomograms such as that shown in Figure 6 indicates that three regions of instability, highlighted under the labels (a) to (c) in Figure 7, are due to the crossing of a BC pole-zero pair over the pole responsible for the LPP, affiliated to either FC or PC. When the LPP resonance crosses these regions, it changes in formant affiliation. For the palatal pressure source, in addition to this pattern, the boundaries surrounding the diagonal band of higher frequency appear as very marked unstable regions.

4. Discussion

We didn't examine, for individual subjects, in which region of their vocal-tract the monopole flow and dipole pressure noise sources should be located, nor in what proportion they should be mixed in order to explain the spectral distribution of the radiated sound.

Nevertheless, from our experiment, the estimation of discontinuities in the articulatory space defined by the front cavity and the palatal channel lengths indicates the presence of several stable or 'quantal' regions delimited by unstable regions where LPP crosses the back cavity resonances and a change in LPP formant affiliation occurs. These regions are assumed to provide robust phonetic prototypes. With a dental pressure source, the quantal regions for coronal fricatives' places of articulation would be at least four (delimited by the lines (a) to (c) in Figure 7). Some more regions might exist in the high frequency areas ([s]-like configurations) that are not captured by our procedure. In addition, it is likely that the spectacular discontinuities resulting from the LPP jump in the palatal flow source setting would add two more unstable regions further dividing this tableau.

The absolute location of the unstable regions, however, is most likely variant. The unstable 'boundaries' will be slightly shifted when the back cavity is longer or shorter, depending on the individual subject's vocal-tract size and morphology. A secondary constriction within the pharynx when the main constriction is back and apical (retroflex), would also lead to some changes. In addition, when FC comprises a sublingual cavity, a part of the FC shall be best modeled as a side branch. This would not change significantly the LPP frequency, but BC would need to be modeled longer than in the current model and thus the unstable regions would be shifted upwards and to the right in Figure 7. All these sources of complexity are assumed to be responsible for the inter-individual variation.

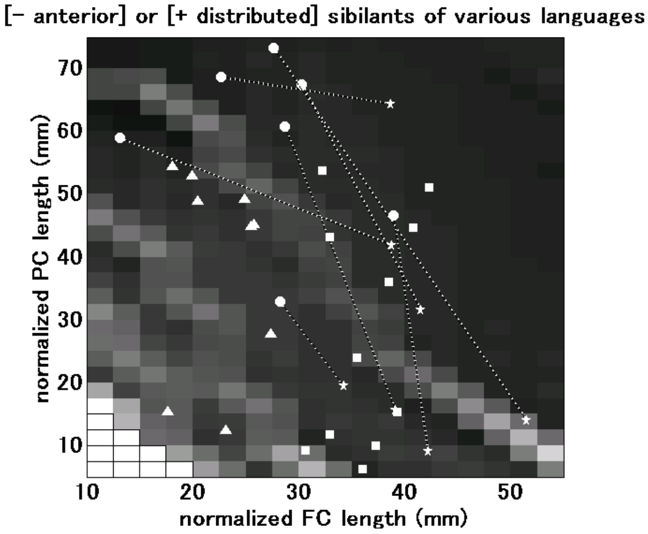


Figure 8. Normalized FC and PC data (measured on MRI mid-sagittal profiles). Triangles: Japanese /s^j/; squares: French and English /ʃ/; circles: Chinese, Polish and Swedish /ɕ/; stars: Chinese, Polish and Swedish /ṣ or ʂ/. For this last group of languages, the data from the same subjects are connected with dotted lines.

Dots in Figure 8 (the background is similar to that in Figure 7, with additional FC and PC lengths) represent pairs of front and palatal cavity lengths measured on sibilant fricatives' MRI data reported in Toda and Honda (2003). These data comprise Japanese /s^j/ (9 subjects), French (7) and English (5) /ʃ/, and Chinese (4), Polish (2) and Swedish (1) /ɕ/ and /ṣ or ʂ/. The measured cavity lengths have been normalized to the ratio between the measured F4 value of an [a]-like open vowel (a coarse estimate of subject's vocal tract length) and that generated by the model having the 17 cm tract length and the area function appropriate for a theoretical posterior [ɑ] (Chinese, Polish and Swedish) or anterior [a] (Japanese, French and English). Although the methodology could be improved, these preliminary results tend to support the predictions of the simulation experiment. Globally, the data points are scattered within dark areas corresponding to stable regions. However, a phoneme in a given language does not necessarily occupy a single quantal region. Furthermore, the contrast between /ç/ and /ṣ or ʂ/ (where applicable) seems not to be achieved in the same way across different subjects. The investigation of individual subject's acoustic targets (including dynamic and stationary properties) may shed further light on inter-individual variation and diachronic sound change.

5. Conclusion

This study examined, through a systematic acoustic simulation experiment whether there exist unstable regions bounding the articulatory space of sibilant fricatives into stable, quantal regions. The results predict there are as many as four to tens of quantal regions. These regions, however, are subject to shift according to subject's vocal-tract size, a secondary constriction or a front cavity side branch. These factors of complexity are assumed to be partly responsible of the individual subject's articulatory strategy.

Acknowledgement

This work is in part supported by the ASPI project, Contract Number 021324, in the EC Sixth Framework Program (FP6).

References

Ladefoged, P. and Bhaskararao, P. Non-quantal aspects of retroflex consonants. In *Journal of Phonetics* 11: 291-302, 1983.

Ladefoged, P. and Wu, Z. Places of articulation: an investigation of Pekingese fricatives and affricates. In *Journal of Phonetics* 12: 267-278, 1984.

Maeda, S. A digital simulation method of the vocal-tract system. In *Speech Communication*, vol 1, pages 199-229, 1982.

Narayanan, S. and Alwan, A. Noise source models for fricative consonants. In *IEEE transactions on speech and audio processing*, vol 8, no. 2: 328-344, 2000.

Perkell, J. S., Boyce, S. E., and Stevens, K. N. Articulatory and acoustic correlates of the [s – š] distinction. In Wolf J. J. and Klatt D. H. (eds.) *Speech Communication Papers presented at the 97th meeting of the Acoustical Society of America.* American Institute of Physics, New York: 109-113, 1979.

Stevens, K. N. The quantal nature of speech: evidence from articulatory-acoustic data. In David, E. E. and Denes, P. B., *Human communication: a unified view*: 51-66, 1972.

Toda, M. Deux stratégies articulatoires mises en œuvre dans le contraste /s/ - /ʃ/ en français. Poster presented at the *IIVIèmes Journées d'Étude sur la Parole*, Dinard, June 12-16, 2006. See also http://www.cavi.univ-paris3.fr/ilpga/ed/student/stmt/.

Toda, M. and Honda, K. An MRI-based cross-linguistic study of sibilant fricatives. In *proc. 6th International Seminar on Speech Production*, Sydney, Australia, Dec. 8-10, 2003.

Traunmüller H. Analytical expressions for the tonotopic sensory scale. In *Journal of the Acoustical Society of America* 88: 97-100, 1990.