SPEAKER NORMALIZATION OF FRICATIVE NOISE: CONSIDERATIONS ON LANGUAGE-SPECIFIC CONTRAST

Martine Toda

ENST LTCI (CNRS UMR 5141) - 46, rue Barrault 75634 Paris
LPP, Université Paris III (CNRS UMR 7018) - 19, rue des Bernardins 75005 Paris
martinetoda@yahoo.co.jp

ABSTRACT

Both frication noise and vowel formants cue the place of articulation of sibilant fricatives (e.g., /s/ and /ʃ/ in English).However, only few studies have examined the effect of speaker-specific factors. This acoustic study of sibilant fricatives examines how speaker-specific formant information can improve the distinctness of two phonemic categories of sibilants: /s/ vs. /ʃ/ in French and /s/ vs. /sʲ/ in Japanese. The results show that the center of gravity of the frication noise, normalized with respect to the subject-specific coefficient of vowel onset or vowel center formants, provides an appreciable improvement in the sibilant distinctness in Japanese. While the distinctness score of the noise is generally higher in French than in Japanese, the F2 onset patterns (/s/</sʲ/) are throughout consistent only in Japanese, and strongly improved with a speaker normalization. These language-specific behaviors are discussed in relation with the respective phonological system.

Keywords: speaker normalization, sibilant fricatives, frication noise, formant transitions.

1. INTRODUCTION

It has been shown that the acoustic and perceptual cues for sibilant fricatives' place of articulation include frication noise but also vocalic context and formant transitions (e.g. [2]; [8]). Furthermore, the categorical boundary in the perception of a [s] – [ʃ] noise continuum shifts according to the sex of the speaker who produced the following vowel [2]. These results are consistent with [9], where the frequency of voiceless fricatives' noise permitted to predict the speaker's gender in 90 %. Likewise, children, whose vocal tracts are shorter than adults, produce frication noise that is higher in frequency [3]. Having in mind the relationship between vocalic formants and gender or age groups, and the underlying vocal tract lengths [6], these findings raise the question as to whether the normalization of frication noise with respect to vocalic information is a general perceptual mechanism based on constant acoustic relations between frication noise and vocalic formants.

In order to test this hypothesis, this study compares the acoustic distance of contrasting sibilants with or without a normalization involving individual formant information.

Two languages, French and Japanese, were investigated. These languages both possess a set of two voiceless sibilant sounds, but they differ in their phonological status. In French, /s/ and /ʃ/ are well described by the distinctive feature of place, [+/- anterior]. In Japanese, the phonemic status of the sibilant sounds 's' and 'sh' (Romanized according to the *Hepburn* system) is controversial. While some authors treat 'sh' as an allophone of 's' occurring when followed by /j/ or /i/ [5], others consider their contrast to involve a palatal alternation that affect the whole consonantal system [4]:5-6. For simplicity, the two sounds will be represented here /s/ and /sʲ/, respectively.

2. METHOD

Seven native speakers of French (6 males and 1 female) and nine native speakers of Japanese (5 males and 4 females) provided the acoustic data as a part of a larger study including articulatory measurements.

2.1. Material

The target fricative-vowel (FV) sequences, where F={/s/, /sʲ or ʃ/} and V={/a/, /i/, /u or ɯ/}, were contained in the following words or word sequences and read one time each in citation context. French: 'Assam' /a**sa**m/, 'ici' /**si**/, 'vous soulevez' /vu**su**ləve/, 'achat' /a**ʃa**/, 'Clichy' /kli**ʃi**/, 'doux choux' /du**ʃu**/ ; Japanese: 'あっさり' /a**s:a**ri/, 'うっすら' /ɯ**s:ɯ**ra/, '滑車' /ka**sʲ:a**/, 'びっしり' /bi**sʲ:i**ri/, 'プッシュ' /pɯ**sʲ:ɯ**/. (In Japanese native lexicon, /s/ never occurs before /i/.)

The acoustic recordings were made in a soundproof room, using a portable digital recorder (Marantz, type PMD671; 48 kHz sampling).

2.2. Description of noise and formants

2.2.1. Center of gravity (COG) of frication noise

Ten Fourier transform power spectra were calculated by using Hamming windows of 8 ms with 2 ms overlap, corresponding to the central 62 ms of each fricative noise. The center of gravity (Eq. 1) was calculated for the frequency interval of 1k to 24 kHz on the time-averaged spectrum. The lower frequency limit is intended to avoid artifacts due to possible residues of voicing.

$$f_{COG} = \sum P_n f_n \Big/ \sum P_n \qquad \text{(Hz)}, \qquad (1)$$

where P_n denotes the power of frequency content at f_n (Hz). The index n covers the frequency bound from 1 to 24 kHz.

2.2.2. Formant measurements

The frequencies of the first four formants were measured manually on a broadband spectrogram at the vowel onset (F_{onset}, or F_{on}) and at the vowel center (F_{vowel}, or F_{vo}).

2.3. Estimate of the acoustic distance (distinctness score) and normalization

The acoustic distance between contrasting sibilant pairs (for COG and F_{onset} for F1 through F4), called *distinctness score*, was estimated by using the Student's t equation (independent groups). Since the number of samples was equal for the two phonetic categories (/s/ and /ʃ/ in French, and /s/ and /sʲ/ in Japanese), the equation can be simplified as follows:

$$Ds = \left(M_a - M_b\right) \Big/ \sqrt{\left(S_a^2 + S_b^2\right)/n}\,, \qquad (2)$$

where $M_{a,b}$ denotes the inter-subject mean frequency (of either COG or F1-F4 onset) for the phonetic categories *a* and *b* (/s/ and /sj or ʃ/), *S* the standard deviation and *n* the number of data points in each category.

If we assume that noise and formants generally tend to vary inversely to the lengths of speaker's vocal tract structures as suggested by the normalization behaviors and gender- and age-correlated data, then we expect that a normalization involving a speaker-specific scaling coefficient (the vocal tract scale is assumed to be inversely proportional to the formant frequency) shall neutralize the variation due to subject, and indeed reinforce the phonemic information. If the absolute value of the distinctness score of the two categories of sibilants is higher *using* a normalization procedure, then it would suggest that the noise and formants entertain a relation that is relatively invariant among speakers, and thus could be used as a cue in identifying the fricative.

The normalization of frication noise was thus performed by Eq. 3:

$$f_{COG\_norm} = s \times f_{COG}\,, \qquad (3)$$

where the speaker dependent scaling coefficient *s* is defined in Eq.4. Actually s is calculated at the vowel onset and at the center.

$$s_i = 1/\left(F_i / F_{ave}\right), \qquad (4)$$

where F_i denotes the formant frequency (either at F_{onset} or F_{vowel} and for F1 through F4), *i* the subject's index and F_{ave} the inter-subject average of the formant.

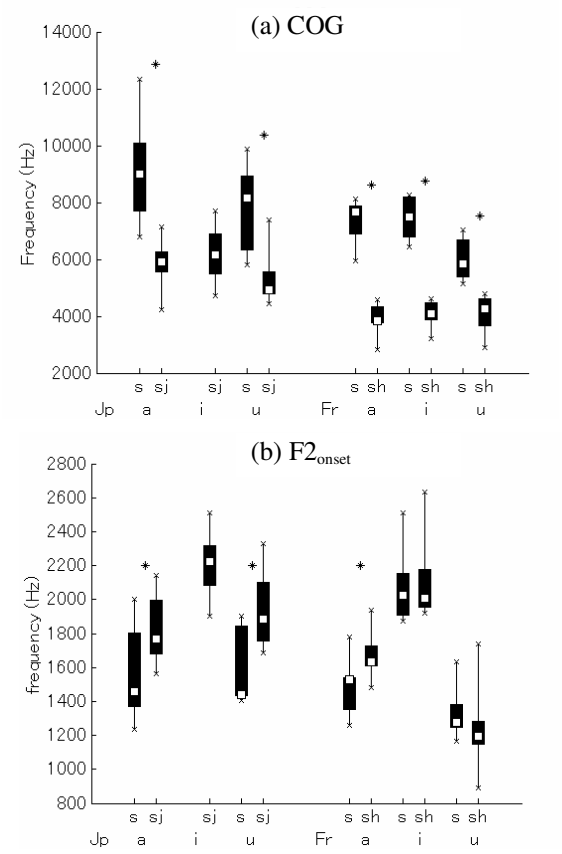
In addition, the normalization of F_{onset} was also performed by the same procedure for comparison, with the scaling coefficient derived from F_{vowel}. The F2 onset has been shown [1] to be significantly lower in American English /s z/ than /ʃ ʒ/, and also lower for males than females. Therefore, we expect that the normalization of F2 would lead to a better distinctness score.

3. RESULTS

As expected, the COG of the noise spectrum (see Figure 1a) exhibits a consistent pattern with a higher frequency for the noise measures for /s/ in comparison with /ʃ/ (French) or /sʲ/ (Japanese) within each subject. (Wilcoxon's paired rank test at α (bilateral) = 0.05). The frequency of the most prominent spectral peak was also examined but is not reported here, since the patterns of contrast as well as the distinctness scores were very similar to those of COG.

The $F2_{onset}$ (Figure 1b), on the other hand, is consistently higher for /sʲ/ compared to /s/ in Japanese, but not in French where $F2_{onset}$ for /ʃ/ is consistently higher than that of /s/ only in the /a/ context. In addition, some of the other F_{onset} values also exhibit a consistent (and significant) relation between the fricatives (cf. table 1b).

Figure 1: Boxplot of (a) COG and (b) $F2_{onset}$ values (minimum, first quartile, median, third quartile and maximum). The Japanese and French sibilants uttered in different vocalic contexts are plotted along the x-axis. Significant differences between the sibilants, obtained with a within-subject Wilcoxon's paired rank test (alpha=0.05), are indicated by an asterisk.



As shown in Table 1a, the distinctness score is rather high in the noise measurements, especially in French /a/ and /i/, where the original score is already higher than 3.50, which corresponds to less than 1 % overlap in Student's t distribution. In French, a few improvements are obtained from the normalization (highlighted in boldface), but they seem to depend both on vowels and formants. In Japanese, the normalization (either with s_{Fonset} or s_{Fvowel}) somehow improves the distinctness score, especially with s_{F2}, s_{F3} and s_{F4}. The highest values (2.10) corresponds to 3 % overlap in Student's t distribution.

The distinctness score for F_{onset} (Table 1b) is low in general but slightly better with an F_{vowel} normalization (bold entries) in both languages. In Japanese, $F2_{onset}$ is significantly higher in /s^j/ than /s/ context for both /a/ and /u/. The distinctness scores for $F2_{onset}$* s_{Fvowel} are -1.39 and -2.27, respectively, which correspond to 9 and 2 % overlap in Student's t distribution. In French, only /a/ exhibit a consistent relation between /s/ and /ʃ/ contexts as for $F2_{onset}$ and $F4_{onset}$ with or without normalization.

Table 1: (a) COG and (b) F_{onset} distinctness scores calculated on original (in italics) and normalized data with respect to (a) s_{Fonset} or s_{Fvowel}, and (b) s_{Fvowel}. The absolute normalized scores that are higher than the original scores are highlighted in boldface. The symbol '*' indicate that there is a significant difference between the two categories according to Wilcoxon's paired rank T test (bilateral, alpha = 0.05). Negative values indicate that /s/'s values are lower than /s^j/ or /ʃ/ .

(a)	V	COG		F1	F2	F3	F4
Jp	**a**	*1.66**	* s_{Fon}	1.25*	**1.99***	**2.02***	**2.10***
	u	*1.41**		**1.42***	**1.51***	**1.58***	**1.46***
		a	* s_{Fvo}	**1.74***	**2.13***	**1.77***	**1.88***
		u		1.38*	**1.62***	**1.56***	1.41*
Fr	**a**	*3.54**	* s_{Fon}	3.54*	**5.51***	3.67*	**3.93***
	i	*3.80**		3.03*	3.59*	**4.56***	**3.94***
	u	*1.91**		1.26*	1.61*	**2.11***	1.60*
		a	* s_{Fvo}	**3.70***	**3.93***	3.21*	**3.73***
		i		3.28*	3.47*	**4.52***	3.53*
		u		**2.14***	1.77*	1.90*	1.45*

(b)	V	F_{on} or F_{on}* s_{Fvo}	F1	F2	F3	F4
Jp	**a**	F_{on}	*0.27*	*-0.74**	*-0.11*	*0.2*
	u		*-0.07*	*-1.10**	*-0.16*	*0.71**
	a	F_{on}* s_{Fvo}	**0.33**	**-1.39***	**-0.3**	**0.34**
	u		**-0.1**	**-2.27***	**-0.55**	**1.46***
Fr	**a**	F_{on}	*0.37*	*-0.83**	*0.25*	*0.88**
	i		*0.05*	*-0.13*	*0.06*	*0.08*
	u		*0.19*	*0.28*	*0.47*	*0.32*
	a	F_{on}* s_{Fvo}	0.37	**-1.41***	0.2	**3.11***
	i		0.05	**-0.26**	**0.18**	**0.27**
	u		0.19	**0.38**	**0.62**	**0.54**

4. DISCUSSION

It appears from the results of COG distinctness that there is a strong tendency for speaker-normalization to enhance the acoustic distance between the two categories of sibilants in Japanese. It is therefore likely that frication noise and formants entertain a relatively constant relation within the sibilant categories in this language. It is difficult to assume such a formant-noise relation in

French, where only localized improvements are observed with a normalization, although the original distinctness scores are already rather high, and higher than Japanese in overall.

Interestingly, the speaker $F2_{vowel}$ normalization of $F2_{onset}$ in Japanese leads to a rather good distinctness (high absolute value) of /s/ and /s^{j}/ contexts for both vowels, as high as those for COG.

From these results, it could be summarized that the acoustic cues involved in /s-s^{j}/ or /s-ʃ/ contrasts are not contained in the same proportion in formants and noise depending on the language. Japanese would have an F2+noise contrast, whereas French would have a noise contrast.

This tendency can be explained in relation with the phonological system. In Japanese, the 'plain'-'palatal' contrastive alternation is a productive process over the whole consonantal system, including plosives and nasals. For the 'plain'-'palatal' contrast to be perceivable for any consonant, a significant amount of the acoustic cues should be carried by vowel formants. Therefore, it is not surprising that the acoustic contrast of /s – s^{j}/ also rely on this cue.

In French, /s/ and /ʃ/ differ in place of articulation. Contrary to Japanese, while /s/ can be grouped together with the other alveolar sounds, there are no plosive series sharing /ʃ/'s place of articulation. Therefore, as long as a sufficient contrast between /s/ and /ʃ/ is achieved, there should be no particular constraint of symmetry in /ʃ/'s place of articulation. Since the French vocalic system is complex (16 vowels) when compared to Japanese (5 vowel qualities), it can be speculated that the interpretation of formant transitions is accordingly complex in French, so that the most robust acoustic cues rely on the frication noise.

Finally, it is worth noting that in French, the $F2_{onset}$ /s/ - /ʃ/ relations are consistent across the speakers in /a/ only (F2 and F4, Table 1b). In /i/ and /u/ contexts, /ʃ/ F2 is rising for some speakers and falling for the others. A possible explanation for that is a between-subject allophony of /ʃ/ in French, as suggested in [7]. It is known that there are no one-to-one acoustic to articulatory relations, and the front part of the vocal tract where the sibilant fricatives are articulated offer numerous ways to adjust the COG frequency of noise spectra. A palato-alveolar fricative with a sublingual cavity and protruded lips, or a palatalized palato-alveolar with a long palatal channel are allophones that are proper to produce a typical noise spectrum for /ʃ/, but they would result in different formant transitions.

5. CONCLUSION

This study examined whether beyond speaker variation there were a constant relation between frication noise and vocalic formants in FV sequences, so that a speaker normalization would permit to improve the distinctness of place-contrasting sibilants. The results, where a quantitative measure of distinctness was used, confirmed this hypothesis in Japanese. Moreover, the respective involvement of noise and vowel onset formants in the sibilant's contrast differed depending on the language. These trends are interpreted as being a response to the language-specific consonantal system's requirements.

6. ACKNOWLEDGEMENT

This study was supported, in part, by the ASPI project, EC 6th Framework Program n. 021324.

7. REFERENCES

[1] Jongman A., Wayland R., Wong S., 2000. Acoustic characteristics of English fricatives, *J Acoust Soc Am.* 108(3), 1252-63.

[2] Mann, V. A., Repp, B. H. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception and Psychophysics* 28(3), 213-228.

[3] McGowan, R. S., Nittrouer, S. 1988. Differences in fricative production between children and adults: evidence from an acoustic analysis of /ʃ/ and /s/. *J. Acoust. Soc. Am.* 83(1), 229-236.

[4] Mester, A., Ito, J. 2006. Systemic markedness and faithfulness. *CLS* 39 (Http://people.uscd.edu/~ito/pages/pubs.html).

[5] Okada, H. 1999. Japanese. In IPA (ed.) *Handbook of the International Phonetic Association: A guide to the usage of .the International Phonetic Alphabet.* Cambridge: Cambridge University Press, 117-119.

[6] Peterson, G.E., Barney, H.L., 1952, Control methods used in a study of the vowels. *J. Acoust. Soc .Am.* 24, 175-184.

[7] Toda, M. 2006. Deux stratégies articulatoires pour la réalisation du contraste acoustique des sibilantes /s/ et /ʃ/ en français. *Actes des XXVIèmes Journées d'études sur la Parole*, Dinard, 83-86.

[8] Whalen, D. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *J. Acoust. Soc. Am.* 69(1), 275-282.

[9] Wu, K., Childers, D.G. 1991. Gender recognition from speech. Part I: coarse analysis. *J. Acoust. Soc. Am.* 90, 1828-1840.