ACOUSTIC CHARACTERISTICS OF THE FRONT FRICATIVES \([f, v, \theta, \delta]\)

Christine H. Shadle\(^1,2\)        Sheila J. Mair\(^1\)        John N. Carter\(^1\)

\(^1\)Dept. of Electronics and Computer Science, University of Southampton, SO17 1BJ, U.K.
\(^2\)ATR HIP Research Laboratories, 2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto 619-02, Japan

ABSTRACT

Une analyse spectrale détaillée d’un corpus important de fricatives pour deux sujets a été réalisée dans le but de déterminer plus précisément les différences acoustiques entre les fricatives frontales, et ainsi comprendre mieux les mécanismes de production impliqués. Des différences apparaissent entre les moyennes ensemblistes des spectres des fricatives centrales, ces différences variant en fonction du contexte voyelle. Ces différences sont réduites, et les spectres sont plus plats, dans le contexte [a-a]. Les résultats des modèles mécaniques indiquent que cet effet pourrait être dû à l’effet de la forme des lèvres sur une très petite cavité frontale.

A detailed spectral analysis of an extensive fricative corpus for two subjects was performed to attempt to specify further the acoustic differences between the front fricatives, and thereby illuminate the production mechanisms involved. Differences were found between the ensemble-averaged spectra mid-fricatives; these vary with vowel context. Differences are least apparent, and spectra are most flat, in the [a-a] context; mechanical model results indicate this may be due to effect of lip shape on a very short front cavity.

1. INTRODUCTION

There have been several studies of the front fricatives of English, \([f, v, \theta, \delta]\), seeking differentiating acoustic characteristics. Nearly every parameter investigated differentiates these as a group from the group consisting of \([s, z, j, s]\): the front fricatives are more confusable (Miller and Nicely, 1955), shorter in duration, lower in amplitude, and have a more variable spectral shape (Behrens and Blum-stein, 1988). Harris (1958) noted that the frication portions extracted from \([f, \theta]\) could not be used to discriminate them from each other, unlike \([s, j]\), but the transitions to and from adjacent vowels could be used.

Wilde (1995) compared amplitudes at the edge and in the center of the fricatives \([f, \theta, s, j]\). She found a significant difference between \([f, \theta]\) and \([s, j]\) – essentially the latter are of higher amplitude – and also found that the amplitude in the region 1-4 kHz is higher at the edges than mid-fricative. She did not find significant differences between \([f]\) and \([\theta]\). Jongman and Sereno (1995) used spectral moments and locus equations to search for differences between \([f, v, \theta, \delta]\). They found that \([f, v]\) spectra had smaller kurtosis, indicating greater spectral dispersion over the range of 0-11 kHz, and \([\theta, \delta]\) had smaller skewness, indicating greater high-frequency noise. However, these differences were small; prediction of the fricative based on these measures was poor.

In this paper detailed spectral analysis of an extensive fricative corpus for two subjects was used to attempt to specify further the acoustic differences between the front fricatives, and thereby illuminate the production mechanisms involved.

2. METHOD

The speech corpora consisted of the fricatives \([f, v, \theta, \delta]\) in two environments: (1) preceded by the vowel [a] and sustained for 3 s, and (2) inserted into the nonsense words \([pV_1FV_2]\) and repeated 10 times on a single breath, for \(V_1, V_2\) chosen from \([a, i, u]\). Two subjects were recorded: a man speaker of French (PB; recorded twice, three years apart) and a woman speaker of American English (CHS; recorded once). PB’s recordings were used in
this study, even though he does not speak [ŋ, ñ] natively, in part because of the more extensive acoustic, articulatory and airflow data available for him.

The acoustic recordings in all cases were made under the 'High-Fidelity conditions' reported previously (Shadle and Scully, 1995). Averaged power spectra from 0 to 17 kHz were computed at beginning, middle and end of the steady-state portion of the fricatives in vowel context, using ensemble averaging (Shadle et al., 1992a,b), and in the center of the sustained fricatives, using time averaging. In both cases eight Discrete Fourier Transforms, each computed from a 20-ms Hanning windowed portion of the speech signal, are averaged to form the averaged power spectrum. For ensemble averages, each window is located at the same 'event' (mid-fricative, or beginning of steady state) in eight successive tokens of the same [pVFV] item; for time averages, the windows are placed adjacent to each other (i.e. with no overlap) in the center of the fricative.

3. RESULTS AND DISCUSSION

Spectra of all six tokens of each sustained fricative were graphed together in order to check for variability. These graphs (omitted here due to space constraints) reveal consistent spectral shape cross-token for each fricative-subject combination, particularly up to 6 kHz; peaks and troughs are of nearly identical frequency and amplitude. The overall levels are low, as expected. The spectra are not flat; the unvoiced fricatives show a difference of the order of 20 dB across the frequency range shown, and the voiced fricatives show a bigger range, largely because of the significant energy appearing in the fundamental and the first few formants, which are presumably primarily voice-excited. Differences are apparent in the spectral shapes between [ŋ] and [ñ], particularly for PB, but the differences are certainly small compared to those between [s] and [ʃ], for instance.

For the fricatives in vowel context it was possible to make more comparisons; we looked in particular at differences due to vowel context, the general spectral shape of each fricative and of each fricative in different vowel contexts, and the change of spectral shape during the fricative.

Vowel context most noticeably affects the frequencies of the peaks in the fricative spectra, with the differences following those observed between the vowels involved. A spectral peak corresponding in frequency range to the second formant, which we shall call F2, is highest in frequency for the [i-i] context, and lowest for the [u-u] context. The higher peak in the [i-i] context is accompanied by the lowest amplitude low-frequency trough at approximately 1 kHz; one explanation of this is that the trough corresponds to a zero with a frequency related to source and constriction location, which do not change; the zero becomes more visible as the back-cavity pole (F2) moves away from it. Such vowel effects are most noticeable in the item [pui], where F2 clearly increases from 1.5 to 2 kHz between the beginning and middle of the fricative 'steady'-state (see Figure 1). [pu6i] does not show such a big difference, presumably because only rounding is changing in this case: the tongue position is constrained in [ŋ], and is not in [ʃ].

As documented in earlier studies (Shadle et al. 1992a,b), there were slight differences in the lower formant frequencies at the beginnings and ends of the fricatives for [pafa] vs. [pafa]. These differences consisted of [pa6a] having higher frequency formants near the transition regions, though not mid-fricative, than did [pafa]. This is consistent with the
tongue forming a necessary part of the constriction in that case and not in the [pafa] case. This was shown to hold for both subjects for the earlier recordings. It did not hold for PB's later recordings of [pafa-paea], but did for the later [pava-paeöa]. In addition to the time lag between the two recordings, the later recordings analyzed here were recorded with the EPG palate in place; perhaps these factors affected his production of the fricatives. Since the difference in frequency, when it exists, is small, it seems likely that it is both not always produced and, when produced, not always observed.

If differences in the transition regions are not always present, are there some observable differences in spectral shape mid-fricative? Differences in spectral shape between [f] and [θ] were observed, although these were subtle and varied considerably with vowel context. In general, [θ] has the lowest amplitude at low frequencies, rising fairly steadily above 3 kHz (see Fig. 2). However, near the edges of the steady-state region the low-frequency amplitude is often significantly higher; a time-averaged spectrum across the entire steady-state region would thus obscure the mid-fricative spectral shape. [f] tends to have a lumpier spectrum, with noticeably higher amplitude at low frequencies than [θ], even mid-fricative, and in some contexts a broad peak in the region 6-12 kHz.

Both fricatives show the least amplitude variation during the fricative steady-state, and the least dynamic range across the frequency range, in the [a-a] context (see Fig. 3). Spectra of [s] and [θ] are contrasted for two vowel contexts in Fig. 3. Although [s] exhibits some vowel context effects such as differences in peak frequencies and overall amplitude, it does not show any such effect on overall dynamic range. A possible explanation of this difference is that the noise generated and radiated in [f] and [θ] may be more influenced by small differences in lip shape because the length of the front cavity, that part anterior to the constriction, is so short. From mechanical model data it is apparent that the position of the jet relative to the wall can make a significant difference in the amount of sound radiated for a front cavity as short as 1 cm, and that the difference is much smaller for a cavity of length 4 or 6 cm (Shadle, 1985). It thus seems possible that changes in lip shape due to vowel context could affect the radiated spectrum of the fricatives, with greater jet-wall contact leading to higher amplitude and greater dynamic range, and that similar changes in lip shape would have less effect on the spectrum for fricatives with a longer front cavity.

Vowel context effects apart, the front fricatives show a smaller amplitude change during the fricative steady-state (i.e. comparing the beginning-middle-end spectra) than do the fricatives [z, j, s], even though the airflow data indicate that constriction areas and intraoral pressures are similar (Stromberg et al., 1994). These observations are borne out by comparison of the sustained fricative spectra at different effort levels for both sets of fricatives. It appears that the same variation in flow velocity results in a smaller variation in noise amplitude in the front fricatives. This should be reflected in the flow exponents for the two sets of fricatives. The exponents reported in Badin et al. (1994, 1995) do not bear this out — the p exponents for [f] are higher than those for [s], and in the sustained case (Badin et al., 1994) the p exponent for [f] is higher than that for [j] as well — but those exponents were computed based on the spectra up to only 6 kHz, which may well explain the difference.
4. CONCLUSION

In summary, differences between [f,v] and [θ,ð] are observable with detailed spectral analysis; they are intricately interwoven with both vowel context and timing of the analysis windows relative to VFV transitions. The differences are least apparent, and the spectra are most flat, in the [a-a] context, an effect which is consistent with mechanical model results. Variation of the spectral shape and amplitude with flowrate appears inconsistent with results for [s,s] (Badin et al., 1994), but this may be explained by the limited frequency range used for the earlier study, and will be investigated further.

[Work supported in part by a collaborative EC SCIENCE award, CEC-SCI*0147C (EDB), and by a European CEC-ESPRIT project Speech MAPS.]

REFERENCES


