ACOUSTIC CORRELATES OF VOICING-FRICATION INTERACTION IN FRICATIVES

Jonathan Pincas & Philip J. B. Jackson

Centre for Vision Speech and Signal Processing, University of Surrey, UK
eem1jp@surrey.ac.uk

1. ABSTRACT

This paper investigates the acoustic effects of source interaction in fricative speech sounds. A range of parameters has been employed, including a measure designed specifically to describe quantitatively the amplitude modulation of frication noise by voicing, a phenomenon which has mainly been qualitatively reported. The signal processing technique to extract this measure is presented. Results suggest that fricative duration is the main determinant of how much the sources overlap at the VF boundary of voiceless fricatives and that the amount of modulation occurring in voiced fricatives is chiefly dependent on voicing strength. Furthermore, it appears that individual speakers have differing tendencies for amount of source-source overlap and degree of modulation where overlap does occur.

2. INTRODUCTION

It has been generally apt to study speech sound sources independently from one another to gain a better understanding of the mechanisms at work. However, voicing and frication sources co-occur in voiced fricatives and in the vowel-fricative (VF) transition regions of voiceless fricatives (Heid and Hawkins1999). This raises the possibility of bi-directional source interaction effects. The most commonly cited effect is the mutual reduction in overall source amplitudes (compared to when only a single source is present) resulting from the superposition of constrictions in the vocal tract (VT) and the concomitant pressure drop and flow rate changes. These amplitude reductions are both theoretically predicted (Stevens et al. 1992) and practically attested. Less commonly discussed effects include the modification of voicing quality from the vowel due to formation of a constriction in the upper VT (Löfqvist et al. 1995), spectral modification of the frication source (Shadle 1995) and the amplitude modulation of aperiodic energy by voicing. The last represents the chief focus of this paper and can be described as a "periodic and synchronous" modulation of the aperiodic component. Jackson and Shadle (2000) used a specialized signal processing technique – pitch-scaled decomposition – to separate periodic and aperiodic sources in voiced fricatives. Their results revealed a significant pulsing of the aperiodic component which was pitch-synchronous, as well as a place dependent phase delay between the aperiodic modulation signal and glottal vibration. The mechanism behind this particular source interaction is assumed to stem from the sensitivity of unstable jets to an acoustic field, meaning that the low-frequency pressure wave induced by phonation 'forces' the jet formed during frication and imposes or increases regularity on the structure of the turbulence. Crow and Champagne (1971) give an in-depth treatment of this phenomenon, although application of their results to speech sounds is problematic. Fricatives are produced using various combinations of constrictions, walls and obstacles, resulting in differing jet structures. Given the complexity of turbulent jet structures and the mechanics of forcing, it is hard to predict how individual fricative jets will respond. The problem is further complicated by

difficulties in obtaining measurements from the VT. We hope, then, to suggest a direction for future theoretical work by establishing whether and how various types of interaction effect pattern with vowel context and place of articulation.

3. METHOD

Corpus. Results presented here are based on the analysis of recordings from 2 speakers in the 20-25 age group (JP, male, and AT, female); both approximating British RP accents. The set of recordings consisted of sentences of the format *"What does* /VFə/ *mean?"*, where vowel context V=/a, u, i/ and F was one of the 8 English fricatives. With 9 repetitions of each possible VF combination, 216 sentences were recorded for each speaker. Recordings were made in an acoustically sheltered cubicle using a Beyerdynamic M59 dynamic microphone linked directly to a PC with a Creative Labs Audigy sound card. Audio was captured in mono, at a sampling rate of 44.1kHz, and with 16-bit quantisation. Speakers were simultaneously presented with two prompts: a randomised list of the sentences to be read and an audio recording of the list played through single-ear headphones, consisting of the sentences being read out in time to a metronome beat with sufficient pause for the sentence to be repeated. The latter was designed to reduce variation in speaking rate and intonation.

Duration measures. From each /VFV/ sequence, the following measurements were taken: frication onset time (t_f^+), voicing offset time (t_v^-), frication offset time (t_f^-), and voicing onset time (t_v^+). Measurements were taken by hand from waveforms visualized with *SFS*. Frication onset and offset were taken as the appearance/disappearance of *any* noticeable frication. Voicing onset and offset were taken as the disappearance/appearance of *any* periodic low-frequency oscillation around f_0. Irregular low-frequency spikes and dips as occasionally attested in voiceless fricatives were ignored. Temporal readings were used to calculate *total frication duration* ($t_f^- - t_f^+$, *TFD*) as well as *source overlap duration* ($t_v^- - t_f^+$, *SOD*), which, in a voiced fricative where voicing persists throughout, is simply the total frication duration.

Modulation amplitude, voicing strength and aperiodic energy. For each /VFə/ waveform, the following were obtained from frication onset to offset: a high-pass (3kHz, order 4) filtered version u_n, a low-pass (350Hz, order 4) filtered version, and an f_0 trace estimated using the *fxanal* algorithm supplied by SFS (based on a method using autocorrelation and tracking). Tests indicated the algorithm to be sufficiently accurate in medial fricative regions for our subjects that traces did not require manual correction.

An algorithm was designed to measure the *amplitude* of *any voicing-modulation of frication* (*AVM*) in areas of source overlap. Our measure characterizes the degree of fluctuation around the mean aperiodic power and is thus independent of overall frication power which allows for comparison across place of articulation. Figure 1 is a diagrammatic overview of the technique.

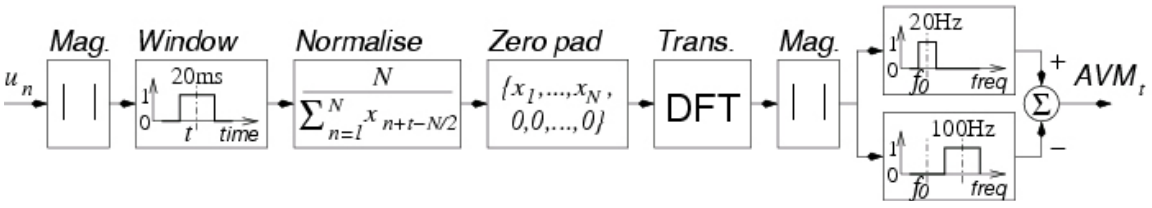

Figure 1. Algorithm for extraction of frication-voicing modulation measure, *AVM*.

An aperiodic energy envelope x_n is derived from the magnitude of the high-pass waveform (Fig. 2, left) whose mean forms the *aperiodic energy (AE)* for each 20ms frame (at 10ms intervals). Having normalized x_n against *AE*, the spectrum of amplitude modulation is computed by Discrete Fourier Transform of each zero-padded frame (Fig. 2, middle). Components within 10Hz of f_0 are aggregated relative to those in a 100Hz band centred 80Hz above f_0 to yield *AVM* (Fig. 2, right). An indicator of voicing strength (*VS)* was taken as the difference between minimum and maximum values in the low-pass filtered waveform, for each frame. Scores were normalized against the amplitude of the vowel 60ms before frication onset to compensate for variations in recording and speaking level.

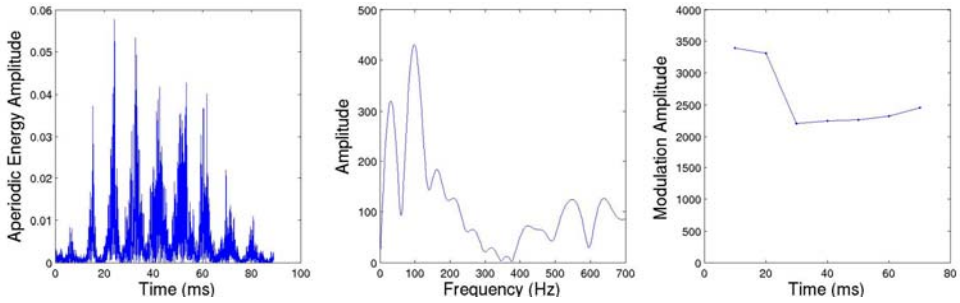

Figure 2. Illustration of modulation amplitude measurement algorithm applied to [z]. Left: Aperiodic energy envelope; Middle: Modulation envelope spectrum; Right: Modulation amplitude plot.

4. RESULTS

Duration measures. Results from the measurement of *TFD* confirmed the traditional voiced/voiceless distinction: mean frication durations were longer for voiceless tokens than for voiced (122ms vs. 90ms for JP; 135ms vs. 72ms for AT). Comparing trends across place of articulation and vowel context with the trends for *SOD* reveals a high degree of correlation for both speakers for voiceless fricatives. AT's data is typical and is illustrated in Figure 3. For both *TFD* and *SOD* there is a clear vowel context effect: times for /a/ are significantly shorter than for /i/ and /u/. Furthermore, a divide is suggested between labiodental and dental fricatives on the one hand, and alveolar and postalveolar fricatives on the other, the latter displaying significantly longer mean durations overall.

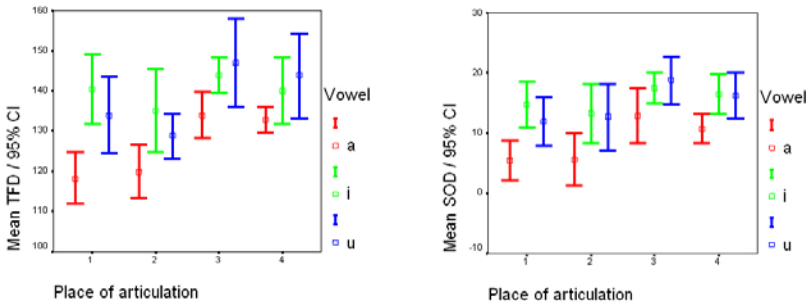

Figure 3. Mean *total frication durations* (left) and *source overlap durations* (right) with 95% Confidence Intervals. Speaker: AT. Voiceless fricatives across vowel contexts /a, i, u/. Place: 1 labiodental; 2 dental; 3 alveolar; 4 postalveolar.

Given that correlations with place and vowel context are similar for both measures, we would assume that the main determinant of *SOD* in voiceless fricatives is the *TFD*. Although this holds within speaker, it does not hold across speaker. It is interesting to note that AT's mean *TFD* for voiceless fricatives (135ms) is longer than JP's (122ms), whereas her mean *SOD* (13ms) is significantly lower than JP's (25ms). Figure 4 shows the *TFD/SOT* correlation for both speakers.

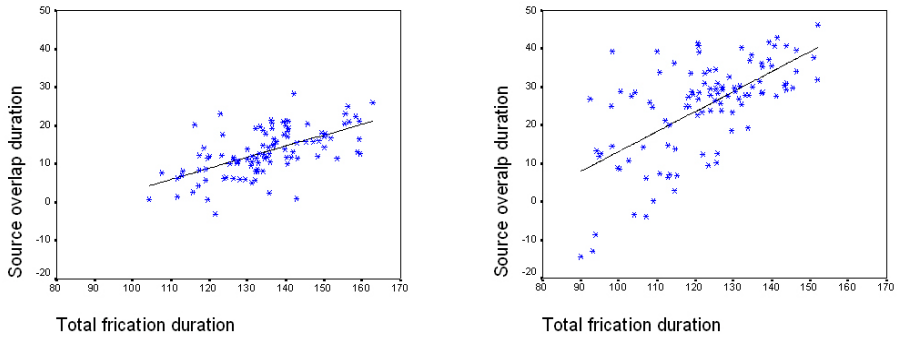

Figure 4. Relationship between SOT and TFD for AT (left) and JP (right). Points represent individual fricatives. Lines are linear best-fit estimations.

As suggested, the positive *TFD*/*SOT* correlation holds for both speakers. Equations characterizing the linear models are: $d_{SOD} = 0.29d_{TFD} - 25.81$ for AT (R = 0.60) and $d_{SOD} = 0.52d_{TFD} - 39.06$ (R = 0.63) for JP. For typical 140ms fricatives, the model predicts 15ms of overlap for AT and 34ms for JP.

Modulation amplitude, voicing strength and aperiodic energy. Figure 5 shows place of articulation ensemble-averaged contours for A*VM, AE* and *VS*. These were constructed by resampling the 27 contours from each place of articulation. The ensemble averaged contours represent good summaries from fricatives of slightly varying lengths.

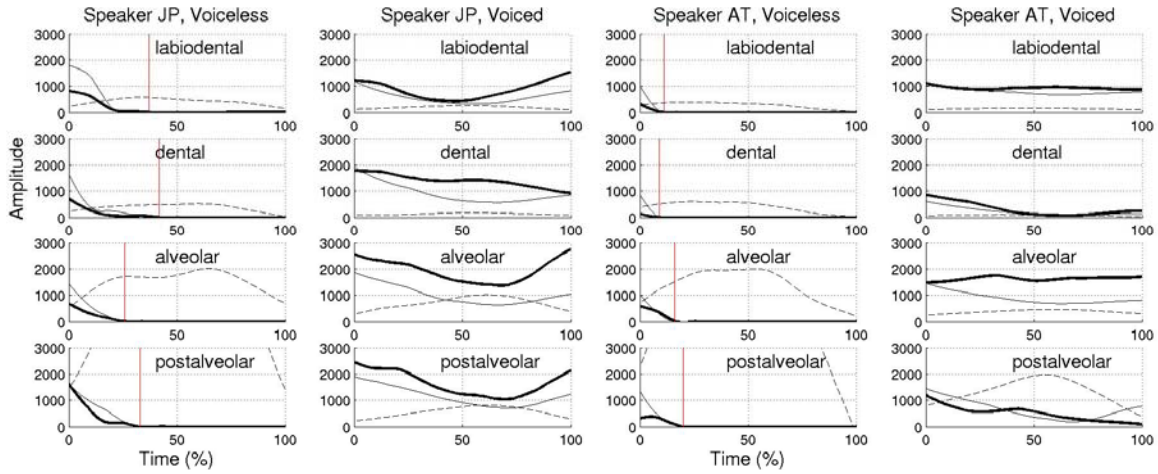

Figure 5. AVM (thick), VS (thin) and AE (dashed) contours for JP (left) and AT (right). VS and AE are scaled by 2000 and 20000 respectively. Vertical red lines show mean voicing offset on a time axis scaled to mean fricative duration.

In most cases, modulation follows normalized voicing strength closely; notable exceptions are voiced dentals for JP and voiced alveolars for AT. In voiced fricatives for both speakers an overall pattern is discernable: modulation and voicing strength fall during the first half of the fricative and then begin to rise again until the end of the fricative. AE contours follow the reverse pattern; rising into the middle of the fricative and falling towards the end. Means computed from the ensemble averaged contours for voiced fricatives appear in Table 1.

Place of	*Speaker: JP*			*Speaker: AT*		
articulation	AVM	VS	AE ($x10^{-3}$)	AVM	VS	AE ($x10^{-3}$)
Labiodental	840	0.28	9.9	930	0.39	7.1
Dental	1400	0.45	6.5	310	0.11	3.8
Alveolar	1920	0.51	36.3	1630	0.44	19.4
Postalveolar	1620	0.57	28.6	490	0.32	67.9

Table 1. Mean *AVM*, *VS* and *AE* values for both speakers for voiced fricatives.

At each place of articulation, mean VS scores generally correlate with mean AVM scores. For both speakers, alveolar fricatives produce strong normalized voicing and the strongest modulation. However, at other places there appear to be discrepancies between speakers. JP, for example, produced strongly modulated postalveolar voiced fricatives whereas AT's were weakly modulated in comparison to the other places. The reverse is true of labiodental fricatives, which, in AT's case, were relatively strongly modulated in comparison to JP's.

5. DISCUSSION

Some of the correlations from Section 4 have been well documented in the literature. Stevens et al. (1992) cover the relationship between frication duration and the voicing distinction and Jongman et al. (2000) cover fricative amplitudes across place of articulation. Our results are generally in agreement with these studies. We conclude from our investigation of *TFD* and *SOD,* that for a given speaker, the duration of voicing-frication overlap in a voiceless fricative is dependent on the length of the segment. Moreover, it seems that different speakers have different 'preferences' for overlap – compared with AT, JP produced, on average, twice the overlap duration for any individual fricative. Analysis of more results will reveal whether this is dependent on sex or is speaker-specific. In any case, it appears that equations derived from linear regression provide a convenient way of characterising speaker 'preferences' for overlap and could be easily employed in speech synthesis to model better /VF/ transitions. Regarding *AVM*, we observe, in the first place, that the measure is successful in characterizing that element of modulation considered most important; namely, increased fluctuation about a mean frication level at the frequency of f_0 . This is immediately visible from Figure 5 where a noticeably high overall frication amplitude does not guarantee a high level of modulation. Within fricatives, we suggest that the modulation contour is determined by the voicing strength contour, Given the theoretical background proposed in Section 1, this is unsurprising – a stronger acoustic field imposed on a turbulent jet will naturally induce greater fluctuation. The overall pattern of falling/rising *AVM* and *VS* is also expected and explicable in terms of aerodynamic theory. Analysis across place is slightly more complex: it appears that the chief determinant of modulation *across place* is also mean normalized voicing strength. However, casual inspection of the ratio *AVM: VS* reveals that for alveolars, the ratio is approximately 30% higher than the average ratio for other places. So, although voicing is generally strong in alveolars and we thus

expect a higher level of modulation, results may suggest that modulation is 30% greater than it should be assuming a simple relationship between *AVM* and *VS* without any peculiarities for place. Given the proposed relationship between *AVM* and *VS* it seems apt to investigate *VS* patterns across place. As pointed out, alveolars show the strongest normalized voicing for both speakers. This may be attributed to aerodynamic conditions: /z/ is produced by channeling a jet against a hard obstacle (the teeth), which appears to be the most efficient sound generation mechanism of all fricative possibilities. Given a more efficient sound production mechanism, it may be the case that speakers are able to maintain a greater pressure drop across the glottis and thus stronger voicing during the fricative. Excepting alveolars, the speakers appear to differ widely in their mean *VS* values across place, which might be attributed to different articulatory strategies. Finally, it seems relevant to compare mean modulation scores for the two speakers. JP (mean: 1443) produced significantly more modulation than AT (mean: 838) in voiced fricatives. Recall that JP also tended to produce longer periods of source overlap in voiceless fricatives. Only analysis of further results will reveal whether there is a relationship between these two parameters and, again, if it is sex or speaker dependent.

6. SUMMARY

Voiced and voiceless fricatives from a specially recorded corpus were analysed for acoustic correlates of source interaction. As well as traditional measures of duration and amplitude, a measure was devised to quantify amplitude modulation of frication noise by voicing. Results suggest that durations of source overlap can be modelled from total frication duration and a speaker-characteristic coefficient, and that amplitude modulation is dependent on voicing strength, speaker-specific tendencies and possibly frication type. Further analysis of the 6 remaining speakers in the corpus should provide grounds for more definite conclusions regarding place of articulation, vowel context, speaker-specific tendencies and gender effects.

REFERENCES

Crow, S. C. and Champagne, F. H. (1970) Orderly structure in jet turbulence, *Journal of Fluid Mechanics, 48 (3),* 547-591.

Heid, S. and Hawkins, S. (1999) Synthesizing systematic variation at the boundaries between vowels and obstruents, *Proc of the Int. Congress of the Phonetic Sciences*, 511-514.

Jackson, P. J. B. and Shadle C. H. (2000) Frication noise modulated by voicing as revealed by pitch-scaled decomposition, *Journal of the Acoustical Society of America, 108 (4),* 1421-1434.

Löfqvist, A., Koenig, L. L. and McGowan, R. S. (1995) Vocal tract aerodynamics in /aCa/ utterances: Measurements, *Speech Communication, 16*, 49-66.

Jongman, A., Wayland, R. and Wong, S. (2000) Acoustics characteristics of English fricatives, *Journal of the Acoustical Society of America, 108 (3),* 1252-1263.

Shadle, C. H. (1995) Modelling the noise source in voiced fricatives, *Proceedings of 15th International Congress on Acoustics, 3,* 145.

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M. and Kurowski, K. (1992) Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters, *Journal of the Acoustical Society of America, 91 (5),* 2979-3000.