Proceedings of the Institute of Acoustics

AERODYNAMIC PATTERNS AS INDICATORS OF ARTICULATION AND ACOUSTIC SOURCES FOR FRICATIVES PRODUCED BY DIFFERENT SPEAKERS

K. Stromberg (1), C. Scully (2), P.Badin (3), C. Shadle (4).

(1)(2) Department of Psychology, University of Leeds, England.
(3) Institut de la Communication Parlée, Grenoble, France.
(4) Department of Electronics and Computer Science, University of Southampton, England.

1. INTRODUCTION

Acoustic patterns for fricatives are difficult to characterise in spectral terms. They show consistency within one speaker but much cross-speaker variation (Hughes and Halle [1]). For our two main speakers, analysis of anechoic recordings has shown this within-speaker consistency while it suggests that the lowest front cavity resonance is associated with a different formant number in the two cases (Shadle et al. [2]). The location in time of perceptual cues may vary for different fricatives, with /f/ and /θ/ for example distinguished more by transitions at the edges of the frication noise segments than by the spectra within this segment (Harris [3]).

Aerodynamic processes are a crucial factor in the production of fricatives. Aerodynamic conditions and the configurations of the articulators at and near a vocal tract constriction combine to produce the acoustic output associated with that fricative. To gain more understanding of the mechanisms that are involved we need to have information about these aerodynamic processes and about the vocal tract constriction size, shape and location. At present it is extremely difficult to infer three-dimensional shapes. The articulatory data presented here derive from measures which are aerodynamic in nature and can therefore be said to be an approximation of the real constriction area. Aerodynamic conditions in the vocal tract reflect the totality of actions including those of the respiratory system and larynx.

2. AIM

By focusing on aerodynamic and aerodynamically-derived traces we hope to be able to characterise some aspects of the mapping from articulation to acoustics in three speakers. Each speaker is subject to the same physical laws, such as continuity of mass flow. The acoustic source-generating mechanisms for all three speakers presumably use aerodynamic forces in a similar way. However, each speaker operates within a different set of constraints - anatomical, physiological and phonological for example. Our aim is to characterise the degree of common properties across the speakers as possible indicators of common processes amongst speakers, but with individual values for the parameters. It should be emphasized here that this is not a cross-language study nor is it aiming to identify male/female differences.

3. METHOD

The two main subjects were CS, a female American English speaker and PB, a male French speaker. A third subject, CD, a male speaker of German, was also studied. All three speakers were recorded as part of the EC project "Mesure, caractérisation et modélisation des sons fricatifs" and were recorded in Leeds using the aerodynamic and acoustic set-up described below. This study focuses on fricatives produced during slow, phonetically-controlled sequences. Measurements were taken from Corpus 3 of the data set recorded for the above project. This corpus was designed to elicit fricatives within a strictly-controlled phonetic environment. Each speech item was of the form /p V1 F V2/ where V1 and V2 were chosen from the cardinal vowel set /i, a, u/ and F from the set /f, v, θ, ð, s, z, ʃ, ʒ, ç, j, x, ɣ/. For this study items containing the first eight fricatives in /iFi/ and /aFa/ vowel contexts were used for the two main speakers and /f, v, s, z/ for the German speaker. All the fricatives analysed here except for PB's [θ] and [ð] are phonemes of the speaker's language. Oral air pressure for most of the back fricatives was not high enough to produce a measurable trace, due to the fact that the constrictions for these fricatives are generally upstream of the pressure tube opening. Therefore, back fricatives have not been analysed here. The subject was required to produce 8 - 12 repetitions of each item on a single breath, aiming for equal stress on both syllables and monotone pitch. Auditory checks were made for the two main speakers. The only confusions noted were between /f/ and /θ/, and between /v/ and /ð/. Acoustic transmission across the mask may well account for these confusions. A reasonable conclusion is that all the speech signals produced by PB and CS were near-normal.

Oral airflow (Ut) measurements were obtained via a single mesh Rothenberg mask and a differential pressure transducer; the pressure drop across the mesh of the mask being proportional to the volume flowrate through it. Oral air pressure (Po) measurements were obtained via a polyethylene tube inserted through the mask and into the subject's mouth, behind the teeth. A reference tube was placed inside the mask. Both tubes were connected to a differential pressure transducer, which measured the pressure drop across the vocal tract constriction, teeth and lips. These signals, together with an acoustic signal from a B&K microphone outside the mask and a signal from a portable laryngograph, were recorded onto FM tape and betamax cassette. The acoustic signal was also used to obtain a fundamental frequency and two intensity level signals. Preliminary checks showed that there was no significant nasal airflow in the vicinity of the fricatives in non-nasal contexts, allowing Ut to be used as an accurate measure of oral airflow. Calibration procedures were performed for the airflow, air pressure and area traces, all of which have linear calibrations. A Siemens' mingograph produced hard copy traces (mingograms) for all these signals, including one from the Aerodynamic Speech Analyser (Electronic Instrument Design, Leeds). This produces a signal for cross-section area by hardware processing of the flow and pressure signals, based on the orifice equation :

$$A = (0.00076)Ut/Po^{0.5} \tag{1}$$

where A is the constriction area (in cm^2), Ut is the volume flowrate of air through the mouth and nose combined (in cm^3/s) and Po is the oral air pressure drop across the tongue, teeth and lips (in cm H_2O). This method of calculating area function cannot be considered to be an absolute measure of area function, only an aerodynamic approximation to it (Scully, [4]).

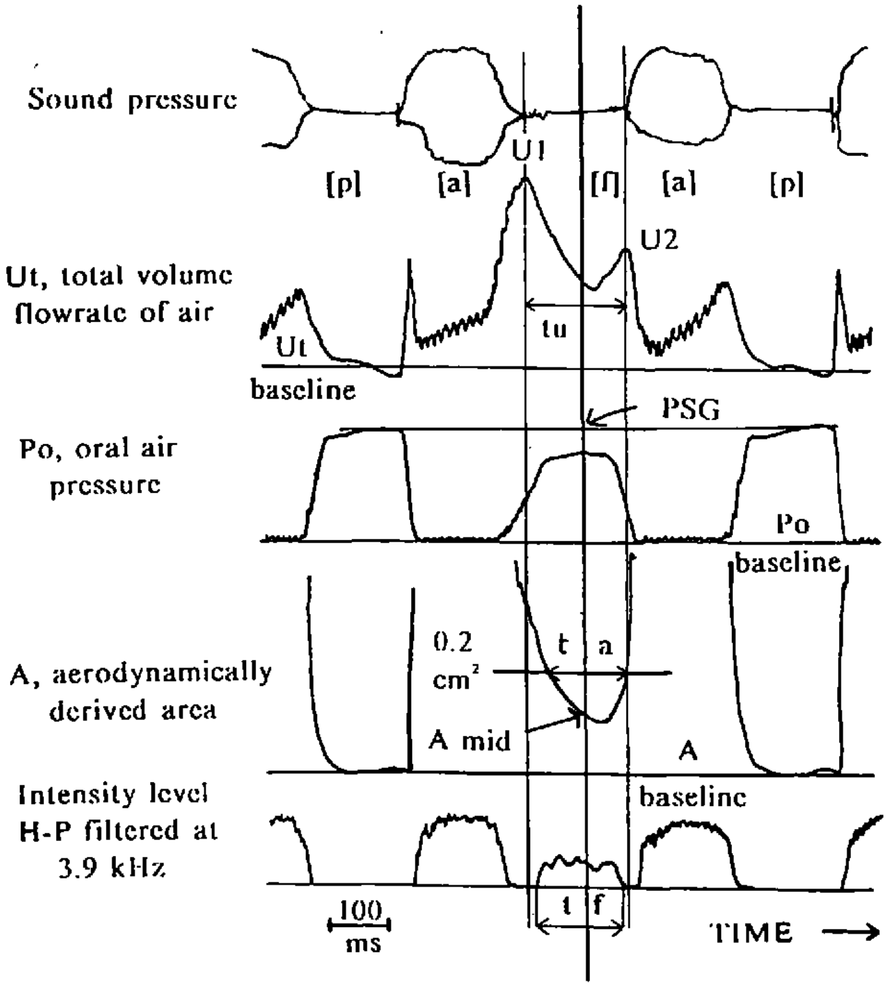

Figure 1. One repetition of [pafa], speaker PB, showing the parameters referred to in the text.

The vocal tract closure for [p] gave the baseline for the area A trace. Where the airflow trace dips below the baseline, as for both the [p] examples in Figure 1, the A trace is reflected slightly upwards, which helps to clarify the location of the zero area baseline. At the start of each run the speaker was asked to hold his/her breath and from this baselines for airflow (Ut) and oral air pressure (Po) were obtained. The airflow trace often dips below the baseline during the [p] closures, as seen in Figure 1. This negative airflow may be associated with jaw lowering which precedes the release of the plosive. Apart from artifacts associated with the use of a mask, reductions in oral airflow, even negative values, are associated with real effects such as sudden enlargement of the volume of the vocal tract. At the time point defining Amid these effects are expected to be small, with the jaw and other vocal tract articulators moving very slowly, if at all. The maximum value of Po for [p], when the whole respiratory system was closed, gave an estimate of subglottal air pressure, PSG. The estimated value near the fricative was obtained by linear interpolation between two successive [p] peaks (Scully [4]). The fricative could be identified by rising-falling air pressure and a falling-rising area trace. Figure 1 is an example of the traces for speaker PB. In some cases Po rises above the estimated value of PSG during the production of the fricative. This is seen for speaker CD and to a lesser extent for speaker CS, but not for speaker PB. There may be several reasons for this; for example, for some fricatives, where the whole respiratory tract is severely constricted, pressure throughout the whole tract may rise. Another possibility is that the speaker may allow subglottal pressure to rise, so that linear interpolation is not valid.

Two parameters were used to characterise the A trace : ta, the durational width of the trace at an arbitrary threshold value of A = $0.2cm^2$; and Amid, the value of A at a time point midway through ta. The threshold value chosen was based on previous analyses and modelling, as being measurable in most cases. The Amid value thus defined was very close to the minimum of the trace, Amin, in nearly every case. Amid was preferred as a parameter to be used as a basis for approximations to the natural A trajectories in subsequent modelling of the processes. In some cases ta could not be defined and Amin was measured instead. A characteristic of most of the voiceless fricative sequences is a double peak of airflow (U1 and U2), more prominent for [a-a] than for [i-i] contexts: tu was defined as the time separation of these two peaks. Intensity level, high-pass filtered at 3.9 kHz, from the microphone signal, was taken to indicate the domain of frication noise for the fricative, with duration tf.

4. RESULTS AND DISCUSSION

Figure 2 shows mean values and 95% confidence intervals for the Amid measurements for voiced and voiceless fricatives respectively in two vowel contexts for each speaker.

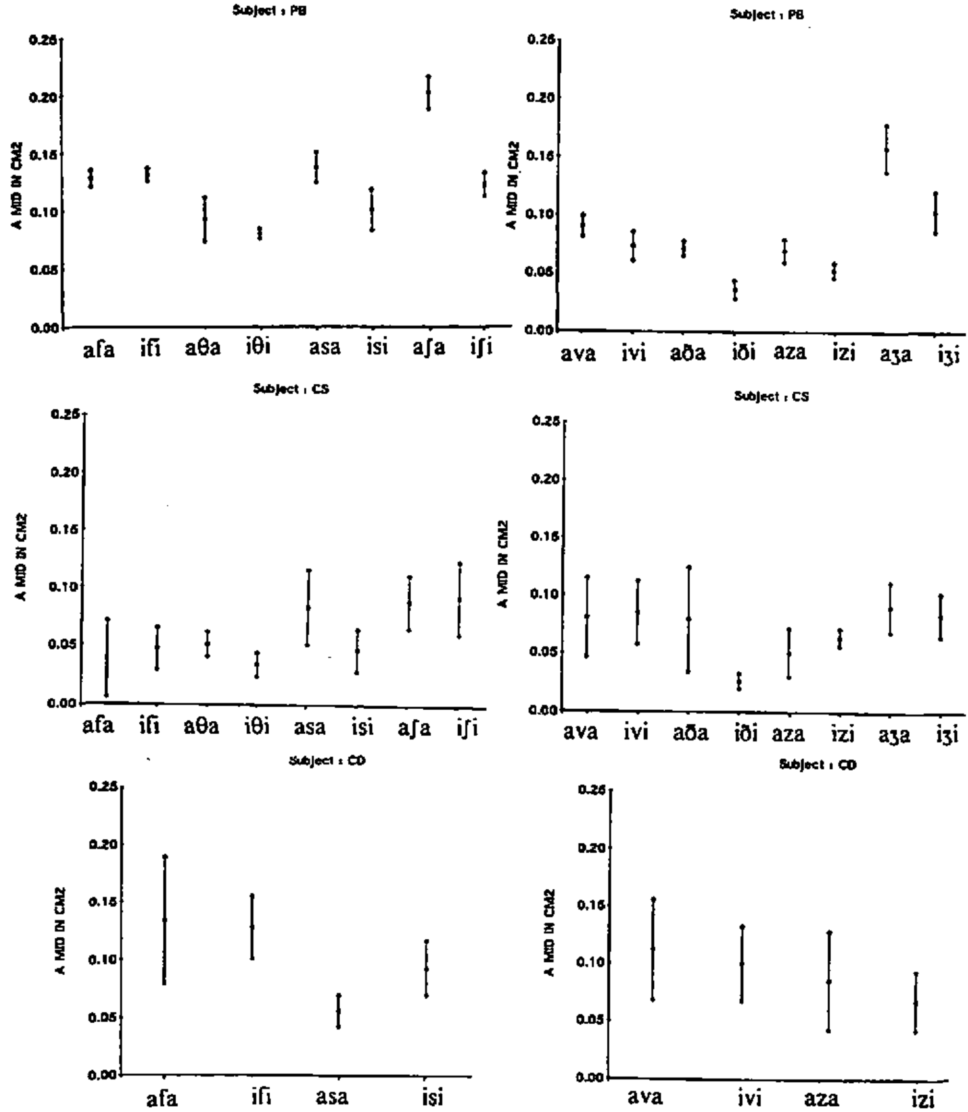

Figure 2. Means and 95% confidence intervals for Amid measurements for speaker PB (top, n = 7), speaker CS, (middle, n = 7) and speaker CD (bottom, n = 5). Left hand figures represent voiceless fricatives, right hand figures, voiced fricatives.

4.1 Area values

Considering the variability shown in Figure 2, the implication seems to be that for each speaker a wide range of vocal tract constriction areas is suitable for the production of these fricatives. Overall, the range is from about 0.02 cm^2 to about 0.2 cm^2, although CS has some values below this for [afa], for one repetition of [iθi], and for one repetition of [aza]. PB has some values above this range for [ʃ]. Similar ranges of values were found for minimum area from aerodynamic measures for four speakers of British English saying "hiss", "his", "peace", "peas", "pass", and "parse" in the frame sentence "A - it said", as follows (in cm^2): (1) 0.01 to 0.12; (2) 0.01 to 0.08; (3) 0.01 to 0.16; (4) 0.02 to 0.10 (unpublished data). The 95% confidence intervals are smaller for PB than for CS and CD, which suggests that some patterning may be obscured by the high variability. The only clear and consistent vowel context effect suggested by the data is for PB [ʃ] and [ʒ]: the Amid values are lower in [i-i] context than in [a-a] context. A similar effect is seen for PB's [ð] but not for his [θ]. Comparing voiced and voiceless fricatives in homorganic pairs, there seems to be no evidence for different Amid values for speakers CS and CD. For the majority of PB's fricatives the voiceless one has a larger A value than the homorganic voiced one in the same vowel context. Highly significant differences were found for [f] versus [v] and [s] versus [z] in both vowel contexts and for [θ] versus [ð] in [i-i] context (2-tailed t-test, $n = 7$, $p \le 0.001$). Significant differences were found for [ʃ] versus [ʒ] in [a-a] context ($p \le 0.01$). The remaining homorganic pairs were not found to be significantly different.

Preliminary analysis of some sustained fricatives (Corpus 2) produced by CS support the view that frication noise can be generated across a fairly wide range of aerodynamically-derived constriction areas, in agreement with the Quantal theory of Stevens [5]. For example, a constant level of frication noise for [s] was obtained for A values ranging from 0.04 to 0.07 cm^2. These and other values for the static fricatives of CS, both voiced and voiceless, lay within the 95% confidence intervals shown in Figure 2.

4.2 Comparison with other methods

We recognise that our A trace reflects patterns of air pressure for the whole of the portion of the vocal tract in front of the open end of the pressure tube and that caution is needed in interpreting airflow measured at a mask and not at the vocal tract constriction itself. We have used other approaches to the problem of inversion to estimate vocal tract areas, applied to some sustained, voiceless fricatives produced by the two main speakers, reported in Badin et al. ([6] pp. 35-41).The minimum area values are shown in Table 1. These data are based on midsagittal vocal tract contours drawn by hand from radiographs. Other articulatory data included photographs of the lips, dental impressions, direct palatography and electropalatography (EPG) and, for speaker CS, transverse slices in the region of the constriction from tomograms. The area functions were derived from the midsagittal contours in two ways. In Method 1 this was done by hand using all available information about the shape of the front of the vocal tract, including moulds made from dental impressions. Method 2 was an automatic procedure for the whole of the vocal tract. A video camera scan with software to define vocal tract sections and a midsagittal line were used to convert the contour into a midsagittal distance function. This was converted into an area function using a mapping developed by Perrier and Boë [7]. Comparing these geometrically-based values for Amin to the aerodynamically-derived Amid, which is close to Amin, the agreement between the two completely independent approaches seems good, especially for [θ] by Method 2 and for all the values obtained by Method 1. We may therefore have some degree of confidence in the aerodynamically-derived measures and their potential to provide anchor points for inversion from acoustic signals.

Table I. Minimum value for vocal tract area Amin in cm^2.

Method	Speaker	Fricative	Amin
Method 1	PB	[s]	0.12
		[ʃ]	0.20
	CS	[s]	0.05
		[ʃ]	0.09
Method 2	PB	[f]	0.175
		[θ]	0.063
		[s]	0.037
		[ʃ]	0.057

4.3 Voiceless fricatives: airflow patterns

Double peaks of airflow are consistently found for all the voiceless fricatives in [a-a] context. Although less prominent, double peaks are visible in [i-i] for [f], [θ] and [s] produced by CS and for [f] and [s] produced by CD, while the pattern is generally more variable for PB here. The double peaks are found in modelling (Scully et al. [8]; Castelli and Scully [9]). They arise where the respiratory tract contains two major aerodynamic resistances. In the cross-over from a high resistance, constricted glottis for a vowel to a high resistance, constricted vocal tract for a voiceless fricative, the total flow resistance reaches a minimum, and this is where the flow reaches a maximum. In [a-a] context the vocal tract trajectory moves rapidly into and away from a configuration with a significant flow resistance. Thus, for the same glottal area articulatory trajectory in each case, the total flow resistance reaches a lower minimum for [aFa] than for [iFi]. This is the explanation we suggest for the higher airflow peaks found in the [a-a] context.

4.4 Voiceless fricatives: duration measures

Figure 1 is representative of most of the voiceless fricatives as regards the relationships between the segments defined by tf, ta and tu. The values of tf and ta are close in most cases (with [aʃa] for PB being an exception). The segments defined are approximately symmetrical about Amid and the edges of the frication noise segment coincide approximately with the peaks of airflow. These lie near to the edges of the segment ta, bounded by A values of 0.2 cm^2.

In most [i-i] examples for PB the A trace never reaches a value higher than 0.2 cm^2, even for the vowels. The vocal tract palatal constriction for his [i] vowel offers a significant flow resistance which needs to be considered in series with the flow resistance for the fricative consonant constriction when the natural speech is represented in simplified models, as discussed elsewhere (Scully [10]).

4.5 Voiced fricatives

A dip in the fundamental frequency trace is seen for all the voiced fricatives. Fricative noise is clear and measurable in some, but not all, cases. It is to be expected that the requirement for voice and frication noise sources together should result in weak and variable noise, or weak and variable voice, or both.

The phonologically voiced fricatives have different phonetic characteristics for the three different speakers. PB's voiced fricatives are always fully voiced, with little or no reduction in the amplitude of the laryngograph signal, suggesting that the vocal folds remain fully adducted. The difference between estimated subglottal pressure and oral air pressure remains above 1.7 cm H_2O and is much higher in most cases. Speakers CS and CD exhibit much more variability. In many cases their laryngograph signals become very low in amplitude suggesting incomplete closure of the vocal folds, even if voicing is maintained. For CD none of the fricatives is actually devoiced; where the laryngograph signal has a long reduced segment Po tends to rise to near, at, or above the estimated subglottal pressure. Similar effects are seen for CS, but in her case some of the fricatives are devoiced; these cases seem to be associated with long breaks in the full amplitude laryngograph signal and oral air pressure rising to within about 1 cm H_2O of estimated subglottal pressure. It seems highly likely that the different languages impose different constraints on the speaker in this respect. Much more information than we have here is needed to gain an understanding of how a speaker such as PB prevents oral air pressure from rising too high.

5. CONCLUSIONS

Our data suggest approximate values for parameters associated with a wide range of fricative types, from very fully voiced, to devoiced, to voiceless. The aerodynamic indicator of vocal tract articulatory trajectory seems to match the actual constriction size quite well, in some cases at least. The edges of the frication noise segment are associated with a localised falling-rising F0 contour or lie near the offset and onset of voicing, where peaks of airflow with still fairly high pressure differences across the glottis should combine to give aspiration noise. Rapidly changing formants associated with the steep slope of the A trace here are to be expected, providing information about both the vowel and the consonant. Modelling which copies the vocal tract constriction A path, and which is made to match the real speech pressure and airflow patterns also, may be able to support other evidence for larynx articulation, mode of vibration of the vocal folds, and interactions between voice and turbulence noise sources. Thus our approach associates acoustic sources with the aerodynamic conditions which help to determine their interactions with each other, and associates them also with the articulatory paths which help to determine formant frequencies and bandwidths, important properties of the filter. This approach has the merit of focusing on the covariations found in speech signals.

6. ACKNOWLEDGEMENTS

We thank Marion Shirt for the auditory analyses. The data were gathered with partial funding from EC SCIENCE contract SCI*0147-C(EDB). The present study is partially funded by ESPRIT Basic Research, project number 6975, SPEECH MAPS.

7. REFERENCES

[1] Hughes, G.W. and Halle, M. (1956) "Spectral properties of fricative consonants" Journal of the Acoustical Society of America, 28, 303-310.

[2] Shadle, C., Badin, P. and Moulinier, A. (1991) "Towards the spectral characteristics of fricative consonants" Proc. XIIth Intl Congr. of Phonetic Sciences, Aix-en-Provence, Aug. 1991, Vol 3, 42-45.

[3] Harris, K. (1958) "Cues for the discrimination of American English fricatives in spoken syllables" Language and Speech, 1, 1-7.

[4] Scully, C. (1986) "Speech production simulated with a functional model of the larynx and the vocal tract" Journal of Phonetics, 14, 407-413.

[5] Stevens, K. (1989) "On the quantal nature of speech" Journal of Phonetics, 17, 3-45.

[6] Badin, P., Castelli, E., Grabe-Georges, E., Guérin, B., Scully, C., Shadle C. and Stromberg, K. (1992) "Mesure, caractérisation et modélisation des sons fricatifs", Final report for EC Research Contract SCI*0147-C (EDB).

[7] Perrier, P. and Boë, L.J. (1989) "Passage de la coupe sagittale à la fonction d'aire" J. Acoustique, 2, 56-67.

[8] Scully, C., Castelli, E., Brearley, E. and Shirt, M. (1992) "Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives" Journal of Phonetics, 20, 39-51.

[9] Castelli, E. and Scully, C. (1992) "Mécanismes de production des consonnes fricatives: coordination entre glotte et constriction du conduit vocal", Journal de Physique, 4, Colloquium C1 Suppl. III, 2, 303-306.

[10] Scully, C. (1992) "L'importance des processus aerodynamiques dans la production de la parole" 19èmes Journées d'Etude de la Parole, Bruxelles, 7-12.