**Proceedings of the Institute of Acoustics**

# ANALYSIS AND MODELLING OF A PLOSIVE/AFFRICATE CONTRAST

S.J.Mair (1) and C.Scully (2)

(1) Department of Linguistics and Phonetics
(2) Department of Psychology
The University of Leeds, LEEDS, LS2 9JT.

## Abstract

Studies of [t] and [tʃ] in VCV sequences have suggested consistent differences between the two consonants, for example with regard to the rate of release of the constriction. An articulatory model is used to try to capture some of these distinctions. A successful perceptual contrast is effected by making three articulatory distinctions - a later glottal opening in relation to the release, a slower increase in constriction area during the release phase, and a slightly more posterior place of constriction, for the affricate compared to the plosive.

## 1. INTRODUCTION

A phonological contrast holds between /t/ and /tʃ/ in English. The processes of speech production generate complex acoustic patterns in plosives which provide multiple cues for perception (for example, Stevens and Blumstein [1]). It seems probable that a multiplicity of acoustic cues may be used in perceptual judgements of a plosive versus affricate contrast also, and that their origins may lie in the production processes.

Quantitative data in the literature regarding affricates are somewhat limited. Here, analysis of real speech data is combined with articulatory modelling in order to understand better which articulatory processes might be important in contributing to the phonemic and phonetic distinction.

Traditional phonetic descriptions classify the English phoneme /t/ as a voiceless alveolar plosive and the English phoneme /tʃ/ as a voiceless palato-alveolar affricate (for example, Jones [2]; Gimson [3]). Differences in place of articulation can arise from differences in the anatomical structure of the articulators, as well as from the speaker making a choice from various physiological options (Keating [4]). Perhaps a more important contrast between the two sounds lies in the type of articulatory release.

During the articulation of [t], a fast release of the oral closure results in an 'explosive' release of the pressure of the enclosed volume of air. For [tʃ], the articulators move away from each other more slowly, in such a way that homorganic frication is produced (O'Connor [5]).

## 2. QUANTITATIVE STUDIES

### 2.1 Production

Some data which compare plosives and affricates for three North-German speakers are provided by Fischer-Jørgensen [6]. Affricates were found to have a shorter closure interval than plosives for the same place of articulation. Peak oral air pressure values did not differ for the two types of sounds, though the affricates showed a significantly slower decay in the pressure trace following the release. In stressed position, oral airflow peaks following the release were found to be higher for plosives than for affricates, and airflow also rose more quickly in the case of plosives. The intensity of the 'explosion' was weaker for the affricates.

### 2.2 Acoustics

There is evidence that the rise time of the frication, measured from the onset of frication to the point of maximum amplitude in either a positive or negative direction (Howell and Rosen, [7]) is sufficient for listeners to distinguish between affricates and fricatives. The duration of the rise time was shorter in [tʃ]; the average mean rise time for the affricates measured from 'nonsense' syllables was 61ms. Measures of rise times for corresponding plosives were not made. The release of the affricate [tʃ] is described as consisting of a burst, followed by a silent interval of about 15ms, followed by frication (Howell and Rosen, [7]).

Fischer-Jørgensen [6] found longer voice onset times (VOT) for affricates than for plosives. VOT is defined as the time interval from the release to the onset of voicing. It is likely that the coordination between the vocal fold abduction-adduction gesture and the supraglottal articulatory release may be an important aspect which distinguishes between different VOTs, although Löfqvist [8] discusses other factors which may determine initiation of voicing following voiceless stops.

### 2.3 Articulatory modelling

Vowels and consonants may be modelled in terms of the vocal tract (VT) area function. Plosives have been successfully simulated using two articulatory parameters; the position of the constriction along the VT and the shape of the occlusion. Both parameters will manifest acoustic effects in the transition from the burst to the vowel (Maeda, [9]). In order to model affricates, in particular [tʃ], Stevens [10] suggests the need for two constriction areas which can be manipulated somewhat independently, one released more slowly than the other.

## 3. ANALYSIS AND MODELLING OF [t] AND [tʃ]

The aim of the present study is to gather data for the production of [t] and [tʃ] in English and to represent any differences found in the articulatory block of a composite model of speech production. The acoustic output of the model is compared to data described in section 2.

### 3.1. Experimental procedure

The data consisted of repeated sequences of [pV1CV2] spoken by 10 R.P. English speakers (5 female and 5 male). V1 and V2 are the same vowel and V2 is intended to be longer than V1. Subglottal pressure is kept as constant as possible throughout each utterance, by avoiding pitch and stress contrasts. Speech data recorded in a more natural setting may reveal different and equally interesting results, but here a highly-controlled experimental context is necessary as a first step to ensure consistency across repetitions and in order that more accurate estimates of parameters can be made for use in the modelling.

The recording procedure involved the use of a Rothenberg mask placed over the speaker's mouth and nose (Rothenberg, [11]), and an orally inserted pressure tube with its open end pointing across the airstream behind the oral constriction, to measure airflow through the constriction (UC) and the pressure drop across the constriction (PC) respectively. We assume that there is no nasal airflow during the sequences to be analysed. This was confirmed in preliminary recordings for each speaker. Hardcopy mingogram traces were made simultaneously during the recording session. Spectrograms (Voice Identification, 700 Series) were made at a later date.

### 3.2. Estimating a minimum constriction area for the vocal tract

The aerodynamic methods require us to represent the front of the VT as a single constriction. Electropalatography (EPG) data gathered for speaker SM (the first author) did not suggest the need for two constrictions to represent affricates. Therefore, throughout this study, a single VT constriction is used to represent both [t] and [tʃ]. Calculation of the area of this constriction is based on the orifice equation (Warren and Dubois, [12]), which was originally developed for estimation of the velopharyngeal orifice area in cleft palate patients, but has been adapted for other uses in speech research (Scully, [13], and references therein). The working equation used is:-

$$AC = \frac{k.UC}{\sqrt{PC}}$$

where **AC** is the constriction area and ***k*** is an empirical constant (0.00076), with area **AC** in $cm^2$, oral volume flowrate of air **UC** in $cm^3/s$, and oral air pressure **PC** relative to near-atmospheric pressure (air pressure inside the mask) as zero, in $cmH_2O$.

Estimates of the minimum cross-sectional area of the VT constriction during the release phases of [t] and [tʃ] were made for each speaker and for several vowel contexts; [i:], [ɜ:]. [ɑ:], [ɔ:] and [u:].

Five repetitions of each sequence were analysed for each speaker. An example of the contrast between [t] and [tʃ] repetitions for one (female) speaker for an [ɜ:] vowel context is shown in Figure 1.

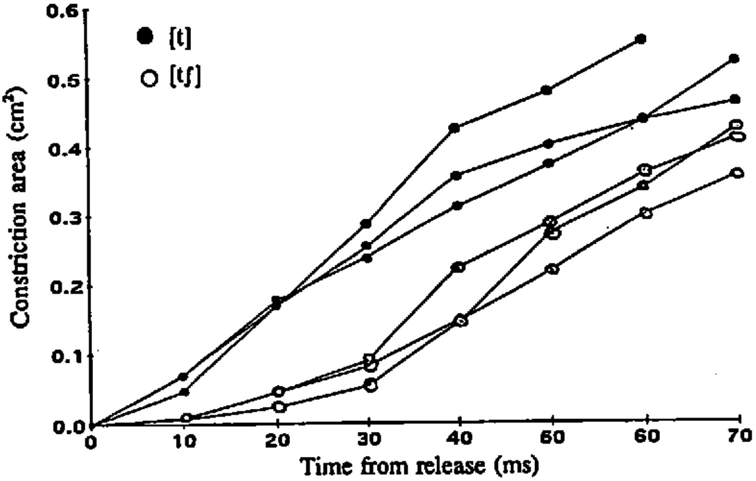

**Figure 1. Constriction area release paths for speaker GB, [ɜ:tɜ:] and [ɜ:tʃɜ:].**

The graph indicates that the increase in area is slower following the release of the affricate compared to that for the plosive. For all the speakers and all the sequences analysed, this consistent difference was found, up to at least 40-50ms following the release, although it was more clear-cut for some speakers than others. Data for a further speaker, SM, also indicated a similar contrast in the constriction area release paths for [t] versus [tʃ].

### 3.3 The articulatory model

Some data for speaker SM were used as a basis for articulatory modelling.

The computer-implemented articulatory synthesiser used here has been described in detail elsewhere (Scully, [13]). Articulation comprises all the relevant actions, including those of the respiratory and laryngeal systems.

Successive stages of the model are: articulation, aerodynamics, generation of acoustic sources, and filtering of the sources. Fundamental to our approach is that each stage is dependent upon conditions in the previous one. Thus, for example, VT constriction area and glottal area provide the orifices for the aerodynamic system. The same orifice equation is used here for synthesis as is used for inversion of real speech.

The voice waveshape varies with glottal area (articulatory component), a vocal fold stiffness/mass factor and air pressure drop across the glottis (aerodynamic component). The turbulence noise sources - aspiration noise at the glottis and frication noise at the VT constriction - use Stevens' [14] formula. The amplitude envelope of frication depends on the relation $PC^{1.5}.AC^{0.5}$. A similar equation applies at the glottis for aspiration noise. Frication noise is inserted just in front of the constriction at a position which corresponds approximately to that of the teeth. The teeth play an important part in the strength of frication in real speech in "obstacle" cases (Shadle, [15]). For calculation of transient sources, we use the time derivative of the calculated oral air pressure trace. Information regarding the modelling of transients is lacking in the literature, though it is noted that the results of many perception experiments showed successful identification of plosives with different places of articulation even in the absence of the transient burst (for example, Blumstein and Stevens, [16]). The acoustic filtering of the sources relies on the K-L method for calculating the reflection and transmission of pressure waves (Kelly and Lochbaum, [17]). Time-varying losses through the glottis related to glottal area, both articulatory and acoustic components, are included, but acoustic outputs through the VT walls are not. After filtering through the VT, the signal is down-sampled and low-pass filtered, and can then be played back as audio output, with a bandwidth of 5.9kHz.

### 3.4. Modelling [i:ti:] and [i:tʃi:]

Articulatory files were created for [V1CV2] sequences where V1 = V2 = [i:], and C = [t] or [tʃ]. Three main distinctions were made between the two files. Firstly, the place of constriction was set slightly further forward in the mouth for [t] than [tʃ], as indicated by traditional phonetic descriptions of articulation, and by preliminary analysis of contact regions from EPG data. Secondly, constriction area release paths for representative examples from speaker SM's data were copied by linear interpolation between values calculated at 10ms intervals. Finally, published data on vocal fold abduction-adduction gestures were used as a basis for the shape of the glottal area transition, which was the same for both consonants (for example, Löfqvist, [8]). The maximum glottal area was the same for both [t] and [tʃ], but was made to occur at 10ms following the release of [t] and 40ms following the release of [tʃ]. A later glottal opening for [tʃ] compared to [t] is suggested in the literature (Kagaya, [18]) and additional evidence was provided by SM's traces of airflow for the particular examples used in the modelling.

A plausible VT area function for the [i:] vowel was obtained using an inversion program. The input to the program consisted of the frequencies of the first 4 formants for Speaker SM's [i:] vowel, together with a specified VT length of 15cm.

Several runs of the articulatory and aerodynamic stages of the model were performed. Airflow and air pressure traces generated by the model were compared to those for SM.

### 3.5. Results of the modelling

Traces of area of constriction (AC), pressure drop across the constriction (PC) and flow through the constriction (UC) are shown for the model output and for the real speech (Figures 2a and 2b). A good qualitative match is achieved with regard to the smaller and later airflow peak and a slower decline in pressure following the release for [tʃ].

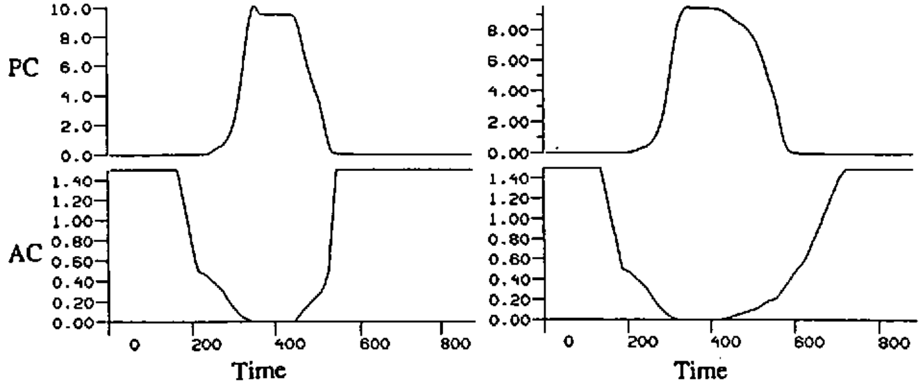

Figure 2a. Airflow, air pressure and constriction area traces for the model output.

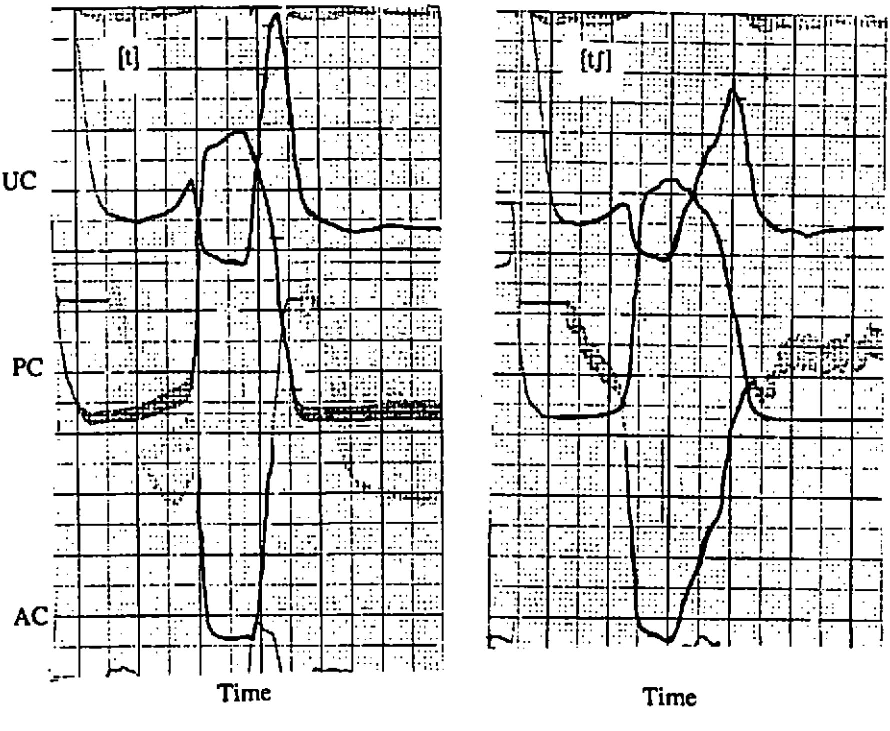

**Figure 2b. Airflow, air pressure and constriction area traces for the real speech data.**

Figure 3 shows the corresponding sources created in the model (voice, aspiration, frication and transient). The results are consistent with the published findings described in section 2. There is a longer rise time, a greater duration of frication noise and a longer VOT for [tʃ] as compared to [t]. Frication noise duration for [tʃ] following the release is of the order of double that for [t]; this ratio is in agreement with spectrograms of the real speech data for speaker SM.

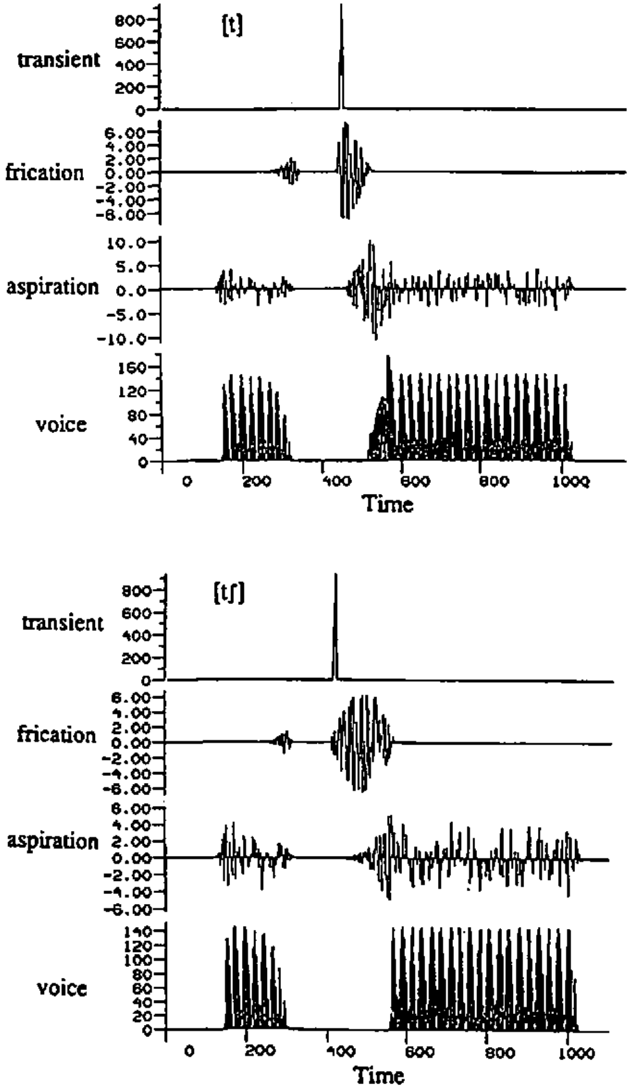

Figure 3. Voice, aspiration noise, frication noise and transient sources (unfiltered) for [i:ti:] and [i:tʃi:] (voice in $cm^3$/s, others are in arbitrary units).

The filtered voice, aspiration, frication and transient sounds for the [t] and [tʃ] files were combined together with equal weightings in both cases. Both VCV sequences were recorded onto cassette together with outputs from two other modelling experiments based on [ɜ:tɜ:] and [ɜ:tʃɜ:] for another female speaker. Five repetitions of each of the 4 sequences were recorded in random order, with "dummy" VCV sequences placed at the beginning and end. Forced choice listening tests via headphones were conducted; the response had to be either "[t]" or "[tʃ]". Five listeners, all native speakers of English, were able to distinguish clearly between the sequences of [i:ti:] and [i:tʃi:] presented, and only one "incorrect" response was made. This suggested that the model output was an acceptable representation of the sequences modelled.

## 4. CONCLUSIONS

Even the highly simplified representations of speech production processes used by us are capable of generating some of the complexities of natural speech. Perturbation of the two basic articulatory plans obtained here should give insight into the rules governing the multiplicity of acoustic differences observed both within and across the speakers analysed.

The results of the listening tests have suggested that a perceptual distinction between [t] and [tʃ] can be achieved by varying the rate of increase in a single VT constriction area following the release and the glottal/supraglottal constriction coordination. Further modelling may reveal whether inclusion of two constrictions for an affricate of the kind proposed by Stevens [10] could enhance the contrast.

## 5. ACKNOWLEDGMENTS

This research was supported in part by a Science and Engineering Research Council studentship.

## 6. REFERENCES

[1] STEVENS, K.N. & BLUMSTEIN, S.E., 'Invariant cues for place of articulation in stop consonants', JASA 64, pp.1358-1368 (1978).
[2] JONES, D., 'An Outline of English Phonetics', 9th edition, W.Heffer and Sons Ltd., Cambridge (1967).
[3] GIMSON, A.C., 'An introduction to the pronunciation of English', Edward Arnold Publishers Ltd., (1962).
[4] KEATING, P.A., 'Coronal places of articulation', UCLA WPP 74, pp.35-60, (1990).
[5] O'CONNOR, J.D., 'Phonetics', Harmondsworth, Penguin, (1973).
[6] FISCHER-JØRGENSEN, E., 'Some data on North German stops and affricates', Annual Report of the Institute of Phonetics, University of Copenhagen, 10, pp.149-200 (1976).

[7] HOWELL, P. & ROSEN, S., 'Production and perception of rise time in the voiceless affricate/fricative distinction', JASA 73, pp.976-984 (1983).

[8] LÖFQVIST, A., 'Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants, Language and Speech 35, pp.15-28 (1992).

[9] MAEDA, S., 'Articulatory-acoustic relationships in unvoiced stops - a simulation study', Proc.11th Intl.Cong. on Phon.Sci., pp.11-14 (1987).

[10] STEVENS, K.N., 'Modelling affricate consonants', Speech Communication 13, North-Holland, pp.33-43 (1993).

[11] ROTHENBERG, M., 'A new inverse-filtering technique for deriving the glottal air flow waveform during voicing', JASA 53, pp.1632-1645 (1973).

[12] WARREN, D.W. & DUBOIS, A.B., 'A pressure-flow technique for measuring velopharyngeal orifice area during continuous speech', Cleft Palate Journal 1, pp.52-71 (1964).

[13] SCULLY, C., 'Speech production simulated with a functional model of the larynx and the vocal tract', Journal of Phonetics 14, pp.407-414 (1986).

[14] STEVENS, K.N., 'Airflow and turbulence noise for fricative and stop consonants: static considerations', JASA 50, pp.1180-1192 (1971).

[15] SHADLE, C.H., 'Articulatory-acoustic relationships in fricative consonants', in W.J.Hardcastle and A. Marchal (Eds.) Speech Production and Speech Modelling, Kluwer Academic Publishers, pp.187-209 (1990).

[16] BLUMSTEIN, S.E. & STEVENS, K.N., 'Perceptual invariance and onset spectra for stop consonants in different vowel environments', JASA 67, pp.1001-1017 (1980)

[17] KELLY, J.L. & LOCHBAUM, C.C., 'Speech synthesis', Proc.4th Intl.Cong. on Acoustics, Copenhagen (1962).

[18] KAGAYA, R., 'A fiberscopic and acoustic study of the Korean stops, affricate and fricatives', Journal of Phonetics 2, pp.161-180 (1974).