ACOUSTIC CORRELATES OF GLOTTAL ARTICULATIONS IN SOUTHERN BRITISH ENGLISH

Joanna Przedlacka & Michael Ashby

University College London (UCL), UK
j.przedlacka@ucl.ac.uk; m.ashby@ucl.ac.uk

ABSTRACT

This paper reports the results of a number of corroborative acoustic measures applied to an extensive corpus of recordings which had previously been analysed auditorily for sociophonetic research, gathered from teenage (14-16 yr-old) speakers of Southern British accents of English, with a focus on 'glottal stops' and 'glottally-reinforced' plosives. It is already known that glottalisation is typically realized as creaky or irregular voicing rather than complete closure resulting in a stop gap, and research with synthetic stimuli has pointed to changes in F0 and amplitude as perceptual cues. By contrast, the present results indicate that for natural speech, marked local minima in the autocorrelation function are reliably linked with auditory impressions of glottalisation, and that F0 and amplitude play only secondary roles. This points to local decline in regularity of vocal fold vibration as the chief cue to glottal articulation.

Keywords: glottal stop, glottalisation, voicing degree, autocorrelation, British English

1. INTRODUCTION

Both 'glottal stops' and 'pre-glottalisation' of oral stops are well attested in classic phonological descriptions of English, the glottal events being heard, described and symbolised as if complete stops, though acoustic studies have repeatedly found that the stops are commonly realized as various events falling short of closure, including creaky or irregular voicing. For both glottal and other places of articulation, even expert listeners may report the auditory impression of a stop in relation to events which turn out to deviate markedly from the canonical norms of plosive-like articulation [1, 2]. This raises uncertainty over the reliability of auditory judgements of glottalisation (as in the transcription of speech data in sociophonetic investigations) and doubt over what acoustic parameters might be appropriately used to assist or corroborate such judgments. Experiments using synthetic stimuli [3, 6] show that 'stop'-like percepts can be cued by various factors, including substantial changes in F0 and amplitude, with the former being more potent, though the factors and the ranges over which they are varied are modeled on relatively few natural tokens.

This study exploited an existing body of Southern British English speech data which had already been analysed auditorily for categories noted as glottal stop [ʔ], preglottalised alveolar [ʔt] or plain alveolar stop [t]. Only glottals associated with realizations of underlying /t/ were considered (i.e. excluding so-called 'hard onsets'). A range of acoustic parameters, including F0, energy, and degree-of-voicing measures [4] were retrospectively added to the files in the database. The research aims were: (1) to seek acoustic correlates of glottal events, (2) to compare glottal stops, glottally reinforced, and plain alveolar stops, and (3) to identify acoustic characteristics potentially useful in corroborating auditory classification. The general prediction was that auditory category judgements could be corroborated by appropriate acoustic analysis, and specifically that dips in the autocorrelation function (a measure of regularity in the speech wave) [4] might be expected to accompany the (incomplete) glottal constriction which is presumably utilized both in glottal stops and pre-glottalisation manoeuvres.

2. METHOD

2.1. Data

Speech data addressing a range of sociophonetic variables had previously been gathered using an elicitation methodology [7]. Subjects were 22 teenage speakers (10 male, 12 female) of broadly similar accents. The present subset of data consists of almost 600 tokens of non-word-initial /t/, where glottal or preglottalised variants are possible realizations. The original analog recordings, made using a Beyer CP340 microphone and Marantz recorder, were subsequently digitized at a sampling rate of 16kHz.

2.2. Analysis

Analysis was performed using the Speech Filing System, SFS [5]. Sound files were annotated manually, following segmentation protocols [8], and scripts in Speech Measurement Language (SML) were then used to aggregate the time-varying data: autocorrelation function (expressed as a fraction between 0 and 1), F0 (Hz), and energy (in dB relative to an arbitrary reference).

Annotations were required for (i) intervocalic examples of [ʔ], as in one possible realization of *butter*, (ii) for the 'reinforcing' [ʔ] which accompanies [t] in one possible realization of a word such as *meat*, (followed by pause), and (iii) for plain [t] which optionally occurs in either type of context.

As the glottal events are typically realized as somewhat diffuse intervals of disturbed voicing, neither the waveform nor the spectrogram lend themselves directly to segmentation and annotation. It was observed that the autocorrelation function commonly exhibits clear local minima in the region of the glottal events. This was used to place annotations at the beginning and end of the autocorrelation minima for events heard as intervocalic glottal stops. For the tokens heard as preglottalised, the energy track (in conjunction with the spectrogram) was used to place a single annotation where a major drop in energy indicated the estimated location of the onset of alveolar closure, and the same procedure was followed for tokens heard as non-preglottalised alveolar stops. A single annotation point does not permit time normalization, so the analysis script gathered data from the 100 ms interval leading up to that point.

2.3. Limitations

The relatively noisy recording environment (schools), speaker behaviour (shuffling papers, speaking away from the microphone) and the use of natural prosody made the recordings potentially challenging for acoustic analysis. Nevertheless, the annotation protocols could be applied successfully in 555 out of 598 cases. In some of the rejected tokens the sought-for indicators are seemingly masked by the prosodic pattern of the utterance, such as fall into creak or whisper. In others a period of frication replaces the expected canonical alveolar stop gap.

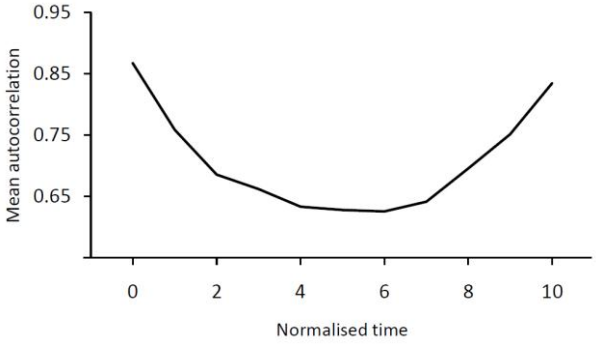
3. RESULTS AND DISCUSSION

3.1. Autocorrelation

3.1.1. Intervocalic

Fifty-three of the tokens (about 9.5% of the total) are intervocalic examples heard as [ʔ]. Durations (mean=131 ms; s.d.=43 ms) were time-normalised to facilitate comparison.

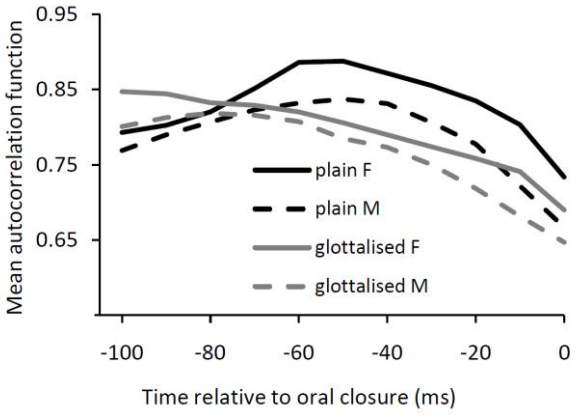
Figure 1: Mean autocorrelation function for intervocalic tokens (n=53).



The average autocorrelation function[1] (Figure 1) suggests that the glottal events are characterized as a class by a decline in regularity of vocal fold vibration, followed by recovery. This would be consistent with a production mechanism consisting of a glottal constriction gesture applied and then relaxed. The shape of this curve – from good voicing to less good and back – may help to explain why glottal stops are perceived as 'voiceless' even though they fail to exhibit the relatively long period without glottal pulses which is typical for voiceless segments.

3.1.2. Pre-glottalised and plain /t/

Figure 2: Mean autocorrelation function over 100ms for tokens judged to have oral closure.



The hypothesis is that tokens judged as preglottalised should as a class exhibit lower autocorrelation in the period leading up to oral closure than the plain alveolars. Figure 2 shows this is indeed the case. The Mann-Whitney U test was applied. At time -50, plain>preglottalised for females (*U*=3,777.5; *n1*=96; *n2*=119; p=0.000) and males (*U*=7,967.59.5; *n1*=164; *n2*=129; p=0.000).

3.1.3. Male-female differences

Though males and females have dynamic patterns which are closely congruent, the males have somewhat lower autocorrelation values than females overall, in this comparison and in all 3 types analysed, presumably indicating the use of creakier baseline voice quality (widely reported for males, though normative data for 14-16 yr-olds is lacking).

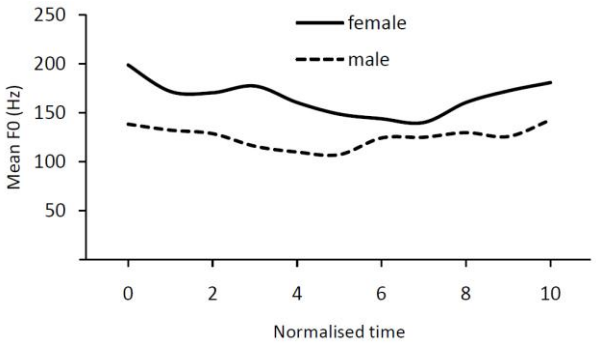
3.2. F0

It has been reported that with synthetic stimuli an F0 dip cues the percept of glottal closure [3]. Auditory judgment for the present natural data might therefore have been based partly on F0, and analyses similar to those used for the autocorrelation data were repeated for F0.

3.2.1. Intervocalic

The F0 data is somewhat noisy, partly as a result of unavoidable errors in F0 estimation applied to field recordings, and the expected intervocalic dip is only weakly indicated (Figure 3). Non-parametric analysis of variance (Kruskal-Wallis) indicates that F0 sampled at beginning, middle and end of the interval are different for males (p=0.02), though not females (p=0.231).

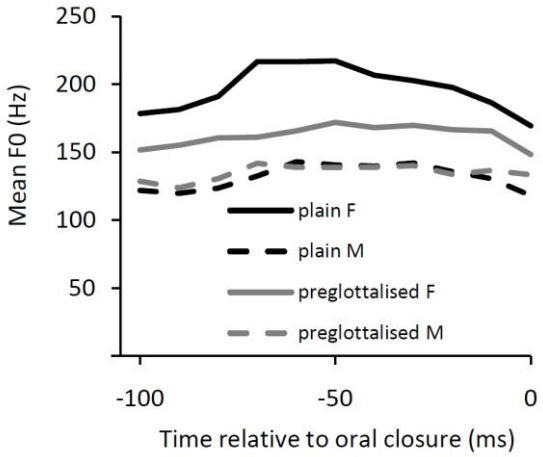
Figure 3: Mean F0 in intervocalic tokens.



3.2.2. Pre-glottalised and plain /t/

The preglottalised and plain tokens (Figure 4) differ in F0 only for female speakers (*U*=42,000.5; *n1*=96; *n2*=119; p=0.001) For male speakers there is no difference in F0 between the two token types (*U*=10,648; *n1*=164; *n2*=129; p=0.923). Hence auditory categorization can hardly have been based heavily on F0 judgements. The cause of this male-female difference is unclear. It might be the case that a glottal manoeuvre in the direction of increased creakiness is automatically accompanied by F0 lowering for females, or the result might partly be an artefact resulting from the way the F0 estimation algorithm operates.

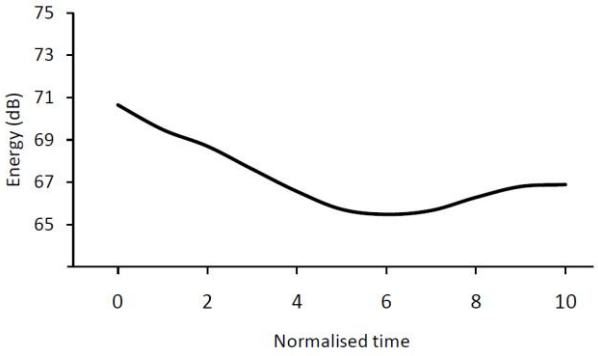
Figure 4: Mean F0 over 100ms interval preceding oral closure in plain and preglottalised [t] tokens.



3.3. Energy

The mean energy over time in intervocalic tokens (Figure 5) reveals a minimum which is small compared with the ranges employed in [3]. Furthermore, the means at times 5 and 10 are not significantly different (*U*=1,629; *n1*=53; *n2*=53; p=0.156), so the natural examples effectively show a shallow unidirectional drop, not the dip in amplitude found effective in synthesis. It is safe to conclude that this cue can have played no more than a minor role in determining auditory classification.

Figure 5: Mean energy in intervocalic tokens.



4. CONCLUSIONS

Of the parameters examined, the autocorrelation function is apparently the most effective acoustic property for corroborating auditory impressions of glottal closure. This is true for both intervocalic examples of single glottal stop, and for glottal closures which accompany (reinforce) oral articulatory closures. Dips in F0 and energy, found to be effective cues in synthesis, are evidenced weakly, and to some extent inconsistently, in the natural speech data.

5. REFERENCES

[1] Ashby, M., Przedlacka, J. 2011. The stops that aren't. *Journal of the English Phonetic Society of Japan* 14-15.

[2] Docherty, G.J., Foulkes, P. 1999. Derby and Newcastle: Instrumental phonetics and variationist studies. In Foulkes, P., Docherty, G.J. (eds.), *Urban Voices*. London: Arnold.

[3] Hillenbrand, J.M., Houde, R. 1996. Role of f0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech and Hearing Research* 39, 1182-1190.

[4] Holmes, J.N. 1998. Robust measurement of fundamental frequency and degree of voicing. In Mannell, R.H., Robert-Ribes, J. (eds.), *Proc. ICSLP (Interspeech)*, 1007-1010.

[5] Huckvale, M. 2010. Speech filing system [computer software]. *http://www.phon.ucl.ac.uk/resource/sfs/*

[6] Pierrehumbert, J., Frisch, S. 1996. Synthesizing allophonic glottalisation. In van Santon, J., Sproat, R., Olive, J., Hirschberg, J. (eds.), *Progress in Speech Synthesis.* New York: Springer-Verlag, 9-26.

[7] Przedlacka, J. 2002. *Estuary English? A Sociophonetic Study.* Frankfurt am Main: Peter Lang.

[8] Przedlacka, J., Ashby, M. 2010. *An Acoustic and Auditory Analysis of Glottals in Southern British English.* Poster presented at BAAP London.

[1] By "autocorrelation function" is meant the size of the peak in the autocorrelation function computed at 5 ms intervals, using the SFS program 'vdegree'. Tests with sweep tones indicate that the autocorrelation function is independent of F0 within ±0.7% over the range 50 Hz-300 Hz.