The Timing of Voicing In British English Consonant Clusters as a Function of Medial Boundary Status

Gerard J. Docherty

Linguistics Department and
Centre for Speech Technology Research,
Edinburgh University, Scotland

An experiment was performed to examine the timing of voicing during CC sequences, in which C_1 is a stop, and C_2 is either a liquid or a glide. These sequences were studied with different types of medial boundary ranging from a segmental boundary, to a clause boundary. The frequency of ocurrence and the degree of progressive assimilation of voicelessness were measured as a function of boundary type. Progressive assimilation of voicelessness from C_1 to C_2 was found to occur consistently when the cluster was in syllable-internal position, but less frequently when there was a deeper medial boundary. The type of boundary had no significant effect on the degree of assimilation.

1. Introduction

The work reported below is part of a study of the timing of voicing in southern British English (SBE) obstruents. Despite the work which has been carried out on voice onset time, and on physiological aspects of voicing, there are still a number of gaps in our knowledge regarding the way in which voicing is realised, and its relation to underlying phonological categories. This study has three main strands. Firstly it provides descriptive data on the way in which voicing is timed in relation to the supralaryngeal events marking the onset and offset of obstruents in SBE. This will allow experimental evaluation of claims which have been made via auditory analysis with regard to the timing of voicing in SBE. Secondly, the investigation will contribute to the debate surrounding the status of fine phonetic detail within models of vocal performance. An increasing number of reports show that many aspects of the fine detail of articulatory coordination are neither universal or automatic. Therefore, it is difficult to attribute them to mechano-inertial performance constraints. Attempts to provide alternative explanations are hampered by the fact that many models of speech production do not have the means of incorporating such low level controllable behaviour [2,9]. Experimental evidence on the timing of voicing in obstruents will be looked at in this light. Thirdly, the increasing need for high quality speech synthesis in advanced speech output applications has lead to the requirement for very detailed phonetic knowledge bases to be available, incorporating comprehensive detail of systematic segmental and subsegmental variation in a language or accent [6]. The data gathered in this experiment will be exploited in this way in order to improve the allophonic rule base for a SBE text-to-speech system.

An earlier experiment [4,5] has shown that in SBE, voice onset time and the occurrence of voicing during obstruents are affected by the phonological category of the segment concerned, and by the nature of the phonetic environment within which it is embedded. The experiment reported below was designed to investigate whether the timing of voicing in relation to a stop closure was affected by the status of the boundary between the stop and a following sonorant, and to test the strong claims that are made in the auditory phonetic literature about SBE, namely that sonorants which follow voiceless stops will only be devoiced if the sequence occurs word internally, or across a sequences of words which 'forms a close-knit entity – a phrasal word or a rhythmic group' [8].

2. Experimental Procedure

The sequences investigated in this experiment were as follows (sequences commencing with a phonemically VOICELESS segment, will be referred to as VOICELESS sequences, those beginning with a phonemically VOICED segment are referred to as VOICED sequences);

VOICELESS Sequences	VOICED Sequences
/pr/ /pl/	/br/ /bl/
/tr/ /tw/	/dr/ /dw/
/kr/ /kl/ /kw/	/gr/ /gl/

These sequences were examined under six different conditions.

1. CCV word initial position (i.e. syllable-internally).
2. C$C across a syllable boundary which is also a morpheme boundary
3. C$C across a syllable boundary which is also a morpheme and compound boundary (this may or may not involve a word-boundary due to the fact that compound nouns are not consistent in their structure).
4. C#C across a word boundary phrase-internally (between an adjective and noun)
5. C#C across a word boundary phrase-internally (between the subject of a verb and the verb)
6. C#C Across a word boundary which is also a clause boundary.

In conditions 1, 3, 4, 5, and 6, all the initial clusters occur before a stressed syllable. In condition 2, it was not possible to maintain this control throughout the data due to the large number of unstressed affixes which were used in order to produce the correct sequences.

In addition, tokens of [sm] and [sn] occurring under all the six conditions were obtained from all the subjects. The purpose of including these sequences was to investigate whether the delay in onset of voicing which occurs between the fricative and nasal in these sequences was affected by the different boundary conditions in the same way as the delay in voicing onset in the [stop]-[sonorant] sequences.

The fifteen sequences were read in random order twice each in the six different experimental conditions by seven subjects. The subjects were all male speakers of SBE, aged between 18 and 25. The speech waveform was recorded on channel one of a REVOX A77 tape recorder and the output of a throat microphone (FJ Electronics) attached to the neck at the level of the thyroid cartilage was recorded on channel 2. The signals were digitised and stored on VAX 11/750 computer. The two signals were then aligned (using the ILS signal processing package), and manually controlled cursors were used for segmentation and durational measurements.

The criteria used for extracting the measurements from each token were as follows. The amount of devoicing of the sonorant was measured from the point at which the stop closure was released (visible as a sudden burst of noise in the time-waveform) to the first peak indicating periodicity in the throat microphone signal. It is assumed that at the moment at which the stop is released, the vocal tract would have adopted the configuration required for the following sonorant, therefore the delay in voice onset corresponds to an interval of devoiced sonorant. In the cases in which devoicing was not observed to occur, speakers used a range of different strategies. These were identified as follows. The replacement of the stop by a glottal stop, and glottalisation of the stop were detected auditorily. A pause was taken to have occurred if there was an interval of silence between the release of the stop closure and the onset of voicing for the sonorant. An incomplete stop closure was identified by a continuous noisy signal throughout the period in which the stop was being produced (i.e. there was no silence to mark a complete closure). [1]

Data analysis consisted of two main tests. A CHI^2 test was used to examine whether devoicing occurs with significantly greater frequency in any of the boundary conditions. Analysis of variance was used to see if the degree of devoicing was significantly affected by the boundary condition.

[1] The problems encountered in segmentation, measurement, and instrumentation are discussed in detail in [5]

3. Results

The measurements presented describe the FREQUENCY with which devoicing or partial devoicing of the sonorant took places, and the DEGREE to which it was devoiced.

3.1. Frequency of Occurrence of Progressive Devoicing

The results show that progressive devoicing of sonorants in a stop-sonorant environment occurs with varying frequency according to the status of the boundary which is present between the two components of the CC sequence. Figure 1 shows the number of cases of VOICED and VOICELESS sequences in which devoicing occurred in the six conditions investigated (all of the results presented are from data which has been pooled across all the subjects).

In VOICELESS sequences occurring syllable-initially (i.e. with only an intervening segmental boundary), there is always progressive devoicing of the following VOICED sound. In clusters with a deeper medial boundary (conditions 2 through 6), devoicing of the following VOICED sound does occur, but less frequently. Statistical analysis using the CHI^2 method has revealed no significant effect of boundary condition in the pooled data, or in the individual subject data. There are however, two tendencies which are consistently present across all the subjects and which are noteworthy. Firstly, there is a consistently greater frequency of devoicing when clusters occur in syllable-initial position compared to when they occur in non-syllable-initial position. This is due to the fact that in the former boundary condition, progressive devoicing always occurs, whereas in the other conditions, a range of different realisation strategies are observed, only one of which is progressive devoicing of the following sonorant. The second tendency emerging from the results is that the clause boundary condition produces by far the fewest cases of progressive devoicing.

In the cases in which progressive devoicing did not occur, the most frequent strategies were insertion of a pause between the release of the stop and the onset of voicing for the following sonorant (especially in the clause boundary condition), replacement of the stop by a glottal stop or glottalisation of the stop, and an incomplete stop closure. The relative frequency of occurrence of the various alternative realisation strategies observed is shown in Figure 2.

In syllable-initial VOICED clusters, devoicing of the sonorant occurs almost as reliably as it does in VOICELESS clusters under the same boundary conditions (this is a reflection of the fact that phonemically VOICED stops in SBE are frequently not accompanied by vocal fold vibration throughout their duration [5]). However, it is not the case that devoicing *always* occurs in VOICED clusters in syllable-initial position. When deeper boundaries intervene between the components of the cluster progressive devoicing does occur, but less frequently than in the segment-boundary condition. The main effect to emerge from the CHI^2 analysis is that the frequency of devoicing is significantly greater in VOICED clusters occurring in syllable-initial position than in VOICED clusters occurring in the other five boundary conditions ($p < 0.001$).

In the instances in which progressive devoicing did not occur, the most frequent strategy which was observed in its place was continuation of voicing unbroken right through the closure and release phases of the stop (see Figure 2). This occurred in almost one third of the non-syllable-initial VOICED clusters (it was also subject to large between-speaker differences -- some subjects used this strategy frequently, others hardly at all).

3.2. Extent of devoicing

The amount of devoicing in VOICELESS clusters is affected by the place of articulation of the stop and the identity of the sonorant, thus confirming the results of [5]. In every case, with the exception of /pl/ there is a greater delay in voice onset when the sequence occurs in syllable-initial position compared to when the same sequence occurs across deeper boundaries. There is no consistent trend observable across the remaining boundary conditions. In some cases, the delay in voice onset becomes gradually less as the boundary is deepened, whilst in others the values vary seemingly at random. There is no trend for the shortest values always to occur under the deepest boundary (this is only found in /tw/, /kl/, and /kw/.

The amount of devoicing in VOICED clusters is affected by the place of articulation of the stop and the identity of the sonorant, as established in [5] It is particularly noteworthy that the voice onset times for /dr/ sequences are almost double those for other voiced stop clusters. With regard to the effect of the boundary condition, no clear tendencies emerge from the data. Unlike the findings for VOICELESS clusters, it is not the case that the highest delays in voice onset are observed when the sequence occurs in syllable initial position.

4. Discussion

The results show that fine details of voicing timing in these clusters in SBE are affected by medial boundary status. Specifically, the nature of the boundary between a stop and a following sonorant has an effect on the probability of devoicing of the sonorant taking place. The crucial factor determining this aspect of voicing timing in stop-obstruent sequences seems to be that it should occur syllable-initially; when a syllable boundary intervenes devoicing occurs less frequently. It is instructive to compare these results with data, given in Figure 3, concerning the frequency of occurrence of progressive devoicing in /s/-nasal sequences under the same boundary conditions. The frequency of progressive devoicing is approximately the same under all of the different boundary conditions. This suggests that the devoicing observed in these clusters is a different sort of context-sensitivity to that observed in the stop-sonorant sequences (possibly a lower-level form of coarticulatory activity since it is not inhibited by the different grammatical boundaries).

The fact that devoicing of the sonorants has been shown to occur under all of the boundary conditions investigated provides counter-evidence to claims made in descriptions of the pronunciation of SBE. It was not found that devoicing occurred only within words or between words that formed 'closely-knit units' This backs up the findings of [3] and [1] who both provided a small amount of data on the devoicing of /l/ following stops in English.

A more accurate description of what is taking place in SBE is as follows In VOICELESS sequences, the default CV implementation strategy involves considerable progressive devoicing of the sonorant. In VOICED sequences, the default CV implementation strategy is to commence voicing just after the release of the stop (on the assumption that the stop closure is not accompanied by voicing throughout its duration). Across all the sequences, in the $CV condition (boundary condition 1), the default case is the one which occurs most frequently. In the other conditions (boundary conditions 2-6) a good deal of free variation takes place, and the default case occurs less frequently, the deeper the boundary condition. In both VOICED and VOICELESS sequences a range of different realisation strategies is used, apparently in free variation. In both types of sequences, but especially noticeable in the VOICED sequences, it appears that there are considerable between-speaker differences regarding the likelihood of one particular strategy occurring rather than another one.

The fact that the default case occurs less frequently under deeper boundary conditions may reflect the fact that under those conditions, a greater number of alternative strategies are available to the speaker (e.g. the pause option would presumably only normally be available to boundary conditions 4-6) Alternatively it may reflect a reduction in the perceptual weight which is being carried by the devoicing in the deeper conditions. It will be possible to evaluate whether any perceptual weight is carried by this aspect of timing, by observing subjects' response to synthetic stimuli which are constructed in such a way as to countervene the findings of this experiment.

It is clear that in SBE, this particular feature of interarticulator coordination is not wholly determined by peripheral, mechano-inertial effects. Phonetic implementation is subject to a constraint which is introduced at a more central level of utterance planning. These findings can be added to the ever-increasing body of data (see [7] and [9] for further examples) which suggests that there exists a wide range of complex, fine-grained, language-specific, systematic realisational detail which cannot be attributed to universal effects, and yet which so far have failed to be adequately accounted for in models of speech production.

References

[1] Bladon, R.A.W., & Al-Bamerni, A. (1976) Coarticulation Resistance in English /l/. *Journal of Phonetics,* 4:137-150

[2] Browman, C., & Goldstein, L. (1986) Towards an Articulatory Phonology. *Phonology Yearbook,* 3:219-252

[3] Dent, H. (1984) Coarticulated Devoicing in English Laterals. *Work In Progress, Phonetics Laboratory, Reading University,* 4:111-134

[4] Docherty, G.J. (1985) The Timing of Voicing in Stops and Fricatives in British English. *Edinburgh University Linguistics Department, Work in Progress,* 18:27-41

[5] Docherty, G.J. (1987) An Experimental Study of Voicing in Obstruents in British English Unpublished ms.

[6] Docherty, G.J., & Shockey, L.S. (1987) Speech Synthesis: Weighing Factors of Task Complexity and Linguistic Sophistication. Paper submitted to *Proceedings of the Institution of Electrical Engineers.*

[7] Fourakis, M., and Port, R. (1986) Stop Epenthesis in English. *Research in Phonetics and Computational Linguistics, Indiana University,* 5:38-71

[8] Gimson, A.C. (1980) *An Introduction to the Pronunciation of English* Longman.

[9] Port, R. (1986) Translating Linguistic Symbols into Time. *Research in Phonetics and Computational Linguistics, Indiana University,* 5:153-173

Figure 1
The frequency of progressive devoicing in voiced and voiceless stop - sonorant sequences in the six boundary conditions.

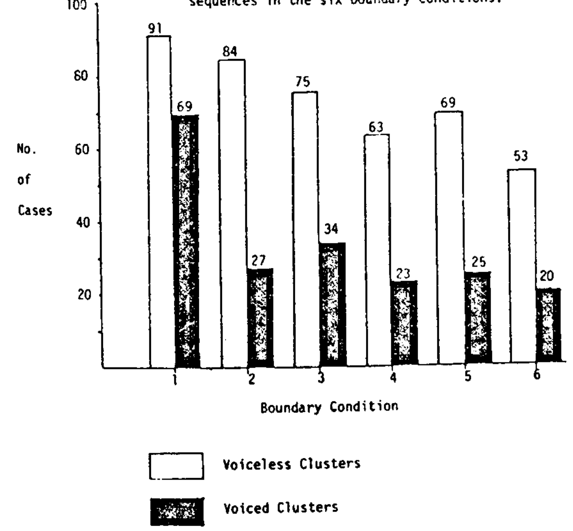

Figure 2
The frequency of occurrence of alternative implementation strategies in the six boundary conditions.

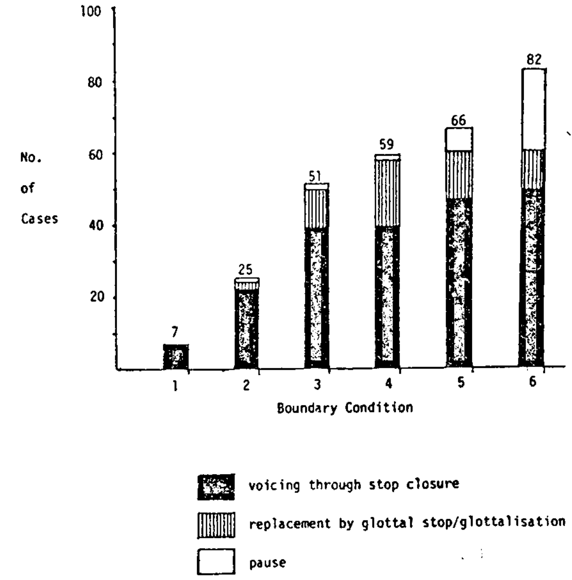

Figure 3
The frequency of occurrence of progressive devoicing in /s/ - nasal sequences in the six boundary conditions.