ABSTRACT

Acoustic analysis of CC sequences in colloquial Syrian Arabic shows that voicing assimilation is a gradient phonetic phenomenon correlated with speaking rate. Clusters in which the voicing values for the two consonants are opposed undergo regressive assimilation, to a greater extent at faster speaking rates. In [+voi][-voi] sequences devoicing extends progressively ‘leftwards’, i.e. as speaking rate increases, voicing in C₁ ceases at progressively earlier points in time; while in [-voi][+voi] sequences C₁ undergoes voicing which extends ‘rightwards’, i.e. as rate increases, carryover voicing from a preceding vowel increases in duration during C₁, leaving a progressively shorter voiceless portion of C₁ immediately prior to C₂. The complexity of the temporal patterns observed point towards a revision of the Window Model of coarticulation involving explicit representation of the time domain in the specification of window width. Both syllabic position and speaking rate may influence the window size for a given speech sound.

1. OBSTRUENT DEVOICING AND VOICING ASSIMILATION

The aim in this paper is to examine time-domain patterns in Syrian Arabic obstruent sequences which are said to undergo assimilation of voicing, and to consider the implications of these and other data for models of speech production. While the literature concerning voicing assimilation is relatively slender, a number of studies in the last fifteen years of the related phenomenon of final devoicing of obstruents have shown that some instances of so-called ‘neutralisation’ of the voicing contrast in final position display phonetic characteristics which indicate that an underlying contrast may still be discerned in close scrutiny of one or more temporal parameters associated with the voicing contrast. It has been argued [1, 2] that parameters such as consonant duration, duration of glottal pulsing and duration of preceding vowel pertaining to a final ‘devoiced’ (underlyingly voiced) consonant may show values distinct from those observed in final consonants which are underlyingly voiceless. The issue is controversial, and a number of sources (cf. [3]) suggest that reported instances of so-called incomplete neutralisation may in fact be accounted for by a range of factors pertaining to the experimental method such as the elicitation procedure and the choice of lexical and textual materials, as well as by variability within and among speakers.

Assimilation of voicing occurs in a cluster of two consonants which differ underlyingly in terms of voicing, when one of the two consonants adopts the value for voicing for the other consonant. The process may be either anticipatory or perseverative in character, and typically involves a sequence of two obstruents. In some cases, however, a consonant which in other respects patterns as a sonorant may be susceptible to assimilation to the voicing value of an adjacent obstruent. This is the case for Russian /v/, which may undergo assimilation before a voiceless obstruent, but does not itself trigger voicing of a preceding voiceless obstruent [4]. Empirical investigation of the phenomenon in Russian [3] confirmed the standard account of the process as anticipatory (or regressive) in character, and identified some time- and frequency-domain characteristics of assimilated tokens as different from those observed in underlyingly homogeneous sequences, thus giving support to the claim that in processes such as this the voicing contrast is neutralised only incompletely. One of the parameters measured was the duration of voicing into the C₁ stop closure or fricative phase in CC clusters showing voicing assimilation, in which small (non-significant) differences were observed for underlyingly voiced and voiceless C₁.

Obstruent voicing assimilation has also been studied in other languages, e.g. Dutch [5] and Catalan [6]. In the latter study the assimilation process is held to be complete rather than incomplete.

2. VOICING ASSIMILATION IN SYRIAN ARABIC

A recent study [7] investigated the phonetic facts of voicing assimilation in obstruent clusters in Syrian Arabic. In informal accounts two parallel phenomena are said to occur, namely regressive voicing and regressive devoicing by assimilation across word boundaries. An example of voicing is given in (1) and an example of devoicing is given in (2) below:

(1) /ʃtaret dar/ → [ʃtaret dar] “I bought a house”

(2) /lwalad tami/ → [lwalat tami] “another child”
The study formed part of a larger investigation into the description and modelling of connected speech processes in Syrian Arabic, encompassing aspects of overlap and assimilation in supraglottal articulatory activity as well as the laryngeal features discussed further in the present paper. The larger study also aimed to explore aspects of assimilation sensitive to speaking rate and speaking style, since assimilatory phenomena of the type under discussion are typically evidenced in informal and rapid speech rather than in the more formal and monitored style of slow careful speech.

3. METHODS

Test materials were prepared containing instances of five voiceless stops, five voiced stops, five voiceless fricatives, each in a potential ‘assimilation site’, that is, followed (across a word boundary) by an obstruent the more formal and monitored style of slow careful speech. Typically evidenced in informal and rapid speech rather than in assimilatory phenomena of the type under discussion are

<table>
<thead>
<tr>
<th>Sp. rate</th>
<th>Voicing</th>
<th>Devicing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v by ass</td>
<td>v by ass</td>
</tr>
<tr>
<td></td>
<td>(N = 5)</td>
<td>(N = 5)</td>
</tr>
<tr>
<td>SS</td>
<td>5.7 (0.8)</td>
<td>80.4 (7.2)</td>
</tr>
<tr>
<td>NS</td>
<td>10.8 (2.8)</td>
<td>92.7 (4.5)</td>
</tr>
<tr>
<td>FS</td>
<td>12.6 (2.4)</td>
<td>95.4 (10.9)</td>
</tr>
<tr>
<td>X</td>
<td>9.7 (3.6)</td>
<td>89.5 (8.0)</td>
</tr>
</tbody>
</table>

Table 2: Voice assimilation in Syrian Arabic: Mean percentages and standard deviations (within brackets) pooled across all speakers, in both voicing and devoicing, at all three speaking rates. Source: [7]

<table>
<thead>
<tr>
<th>Sp. rate</th>
<th>Voicing</th>
<th>Devicing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v by ass</td>
<td>v by ass</td>
</tr>
<tr>
<td></td>
<td>(N = 5)</td>
<td>(N = 5)</td>
</tr>
<tr>
<td>SS</td>
<td>7.5 (1.9)</td>
<td>78.6 (3.7)</td>
</tr>
<tr>
<td>NS</td>
<td>12.0 (3.8)</td>
<td>95.9 (9.2)</td>
</tr>
<tr>
<td>FS</td>
<td>12.7 (2.0)</td>
<td>95.9 (3.9)</td>
</tr>
<tr>
<td>X</td>
<td>10.7 (2.8)</td>
<td>90.1 (10.0)</td>
</tr>
</tbody>
</table>

Table 1: Voice assimilation in Syrian Arabic: Mean percentages and standard deviations (within brackets) of voicing into closure pooled across all speakers, in both voicing and devoicing, at all three speaking rates. Source: [7]

The results here show that for all speaking rates the mean percentage duration of voicing in C₁ in an underlyingly heterogeneous CC cluster (the assimilation site) is intermediate between the expected value for C₁ in either a voiced-voiced or a voiceless-voiceless cluster. The difference between C₁ voicing in the assimilation site and in the voiced-voiced cluster or voiceless-voiceless cluster ( whichever represents the cluster that would be created by complete assimilation taking effect at the
assimilation site) was significant at the slow speaking rate alone for the stops and non-significant at the normal and fast rates; and significant at the slow and normal rates for the fricatives, and non-significant for fast speech. The results thus closely mirror the findings for Russian [3], suggesting at least some evidence that assimilation of voicing operates in a manner yielding only an incomplete neutralisation of the voicing contrast at slower speaking rates.

5. THE NON-CONTIGUOUS NATURE OF VOICING ASSIMILATION

Of significance to the discussion here is that the data showed evidence of a spread of values for voicing into the C₁ stop closure or fricative phase. If voiced-voiceless clusters are ordered according to the extent of voicing in C₂, a continuum of forms emerges in which voicing ceases progressively later; that is, devoicing may be argued to be spreading leftwards from the assimilation site (underlyingly voiceless) C₂. For voiceless-voiced clusters, however, although a similar pattern emerges (shown schematically in figure 1), the interpretation is less straightforward, since in this continuum what we observe is the progressive spread rightwards (i.e. ostensibly from the preceding vowel) of voicing into C₁. Thus the process of voicing by assimilation appears to involve the spread of a phonetic property that might be termed non-contiguous, in that the process by which C₁ acquires voicing is triggered by a voiced C₂, but a voiceless portion of C₁ (of variable duration, and shorter in more completely assimilated tokens) always intervenes between the two voiced portions.

6. CONSTRAINTS ON VARIABILITY IN THE COARTICULATORY WINDOW FOR VOICING

This non-contiguous character of the spread of voicing in a CC cluster presents a problem for any theory of speech production or of sound patterning which seeks either to provide an account rich in low-level detail of the phonetic character of the spreading, or to provide an explanation for the phenomenon grounded in what has been termed ‘the inertial properties of the speech apparatus’ [8: 297]. One such model is the Window Model of coarticulation [9], which proposes that articulator paths (or equivalently, though less concretely, acoustic paths such as formant trajectories) are specified not by monotonic interpolation between fixed target points specified for each ‘segment’ in a sequence of sounds, but by specifying a ‘window’, that is a range governing all the observed variation for a particular articulatory parameter for a given speech sound. The articulator path is then generated not by interpolation but by the tracing of a path within limits specified by the window for successive segments. The mechanisms governing how the precise articulator path might be determined are unclear, but the implication in the model is that the speaker has a certain amount of freedom within a given window, mirroring the fact that the window is itself determined empirically from observations of variability in speaker behaviour. The author of the model leaves open the question whether values within a given window have a flat distribution, or whether the window is subject to some function indicating the probability of the articulator path taking any particular value.

For the voicing assimilation phenomena discussed in this paper a version of the model is proposed which is illustrated in figure 2. On the vertical axis voicing is specified in a notional space with two states (vibrating and non-vibrating vocal folds), following Keating’s suggestion that the model may specify speaker behaviour at a more abstract level than the movements of specific articulators. The horizontal axis incorporates the time domain, again in a manner consistent with the original model, in which it is left open whether the horizontal axis reflects real time (in some sense) or merely sequential ordering among segments. In the present proposal the observed details of timing on which the discussion hinges require the assumption that the horizontal axis incorporates real time, at least in that windows of different horizontal extent must be taken to have different temporal extent; and in that windows situated closer together horizontally in the representation must relate to events that are closer together in time in the utterance being modelled.

Figure 2 also illustrates the range of tokens observed for the voiceless-voiced sequences. ‘Voicelessness’ in underlying C₁ is specified by a window of narrow ‘width’ to use Keating’s term for vertical extent within the representation, stipulating the non-vibrating glottal state. As tokens become successively more subject to assimilatory voicing, the duration, that is, horizontal extent within the representation progressively decreases,
leaving successively shorter voiceless portions within C₁, but
(as the diagram indicates) temporally adjacent to C₂. The
limiting case is the reduction in the duration of the window for
voicelessness to zero duration (indicated by the Ø symbol in
figures 1 and 2), whereupon the ‘path’ for voicing follows only
the windows for the surrounding vowels and the cluster C₂, all of
which (in this simplified representation) show vocal-fold
vibration.

Thus crucially this account proposes two modifications to
the Window Model as originally set out [9]. Firstly, it is
proposed to build time explicitly into the representation by
allowing windows to differ one from another by virtue of their
temporal extent in the representation. This allows us to model
the progressive encroachment of voicing (and also of the
simpler case of devoicing) in C₁ as tokens show more evidence
of assimilation. Secondly, we propose a significant departure
from the nature of the ‘window’ as originally defined in the
model in Keating’s formulation, the range of values specified
by a window of given width serves to encompass the full range
of variability encountered in the behaviour of a given articulator
path (or other parameter trajectory), regardless of the source of
this variability. It is then left to mechanisms underdetermined
by the model to generate a particular path in a given utterance.
This is evidently potentially a weakness in the model, since
phonetic variability, which the model is devised precisely to
account for, becomes entirely extrinsic to the representation.
Our proposal is instead to disaggregate the body of factors
governing variability in articulatory trajectories, by expressly
identifying particular contextual factors as having a predictable
effect on window widths. Two such factors proposed in this study
are speaking rate and syllable-position. Both a faster speaking
rate and the coda syllable position are phonetic environments
associated with weakening of articulatory gestures, and it seems
to us appropriate to adapt the window model so as to show
global reduction effects on window sizes for segments in these
environments. In the data under consideration in the present
study, this allows us to predict a reduction in magnitude of the
temporal extent of the window specifying the non-vibrating
voicing windows for vls-vd clusters showing (a) no
assimilation together with the surrounding vowels,
(b) incompletely assimilated and (c) completely
assimilated (with reduction of window width to
zero).

Figure 2. Sequence of schematic amplitude of
voicing windows for vls-vd clusters showing (a) no
assimilation together with the surrounding vowels,
(b) incompletely assimilated and (c) completely
assimilated (with reduction of window width to
zero).

REFERENCES
in German. Journal of Phonetics, 13, 455–471.
UK: Cambridge University Press.
word boundaries and sex of speaker and listener. Journal of Phonetics,
14, 311–326.
neutralization. Phononia, 50, 28–43.
of Manchester, UK.
UK: Cambridge University Press.
and Beckman, M. (eds.) Papers in Laboratory Phonology I: Between the
Grammar and Physics of Speech. Cambridge, UK: Cambridge University
Press.