Onset-rime coarticulation in the production of /dai/ and /stai/ by four-year-old Scottish English speaking children: preliminary results

JAMES M SCOBBIE

Department of Speech and Language Sciences, Queen Margaret University College, Edinburgh.
8 Clerwood Terrace, Edinburgh EH12 8TS, Scotland.

ABSTRACT

Patterns of coarticulation between onset and rime in the speech of six Scottish English speaking children aged 4 years is examined in the words dye /dai/ and sty /stai/. F2 is 272Hz lower at the burst following /st/ than following /d/. The lower F2 after /st/ in the onset-rime transition is accompanied by a higher F2 during /ai/'s diphthongal offglide. Moreover, the overall shapes of the F2 trajectories are similar, and the frequency difference in F2 at the stop burst correlates with difference in duration between complex and singleton onset. These facts suggest that in the /st/-initial word the diphthong is apparently initiated earlier with respect to the stop burst — and therefore that the intial portion of the vowel opening gesture is masked acoustically by the preceding stop to a greater extent after the cluster than after the singleton. This interpretation supports models of intergestural timing in which hierarchical prosodic nodes (in this case, the onset) enable non-local gestural organisations. The syllable onset provides a fixed anchoring point for the following vowel (a "C-centre") which results in greater acoustic masking when the onset is complex.

1. Introduction

1.1 Background

This study is part of a long term programme motivated by a desire to understand the acquisition of (a) phonetic systems, (b) phonological systems and in particular, (c) their inter-relationship. To put it simply, the central goal of the programme is to address how children acquire an adult-like phonetics/phonology interface. This simple statement is rather misleading, however. The first complicating factor is that we do not know what the *adult* interface is like in any detail. The second is that there is so little work of the appropriate kind on the linguistic speech motor development of children or on language-specific phonetics in child speech that it can be hard to disentangle aspects of phonetic development that are under linguistic control (i.e. language specific) from those that are universal. In this short paper I present preliminary results of work in progress, reporting only on child speech.

One aspect of phonetic/phonological development of interest is the relative weighting of different cues to phonological contrasts. For example, vowel duration is a cue in most English dialects to both the phonological voicing of a following consonant and to the moraic/bimoraic status of the vowel itself. Stoel-Gammon, Buder & Kehoe (1995), for example, report that at 2½ years old, American English children vary the duration of /i/ to signal the voicing of a following consonant but not to differentiate it from /ɪ/. Swedish children do basically the opposite. Though the American English vowels realising /i/ and /ɪ/ are approximately the same duration before voiceless stops, the contrast is conveyed using large spectral differences. The Swedish /i/ and /ɪ/ were produced with only a very small spectral difference.

Mastery of the inventory of cues and their relative weighting (the « cue-grammar ») appears to take many years, and relies crucially on more general developments in gestural co-ordination that extend from pre-speech behaviour right through into skilled adult speech. In canonical babbling (and in the early stages of phonological development) a fairly small set of relatively rigid articulatory routines are available to the child (Menn 1983; Vihman and Velleman 1989). The subsequent rapid expansion of phonological diversity towards an adult inventory reflects improvements in articulatory skill and the possibilities arising from the availability of more complex gestural organisations (e.g. Davis & MacNeilage 1990 ; Studdert-Kennedy and Goodell 1995 ; Nittrouer, Studdert-Kennedy and Neely 1996; Browman and Goldstein 1998). A typical view is that « the holistic, undifferentiated syllable appears to be the initial unit of speech production from which (we hypothesise) segments gradually emerge, first by differentiation of the syllable into its gestural components, then by integration of those gestures into the recurrent articulatory-acoustic patterns that we know as consonants and vowels » (Nittrouer et al. 1996).[1] Naturally, much of the research into these developments compares child and adult productions using instrumental methods, but indirect evidence from transcribed child speech is also highly relevant (e.g. Ferguson & Farwell 1975 ; Waterson 1971 ; Stoel-Gammon 1983; Macken 1979).

The emergence of intergestural co-ordination has usually been approached in instrumental studies through studies of vowel-vowel and consonant-vowel anticipatory coarticulation. Observed age-related differences in coarticulation (or « co-production ») are taken as evidence that all speech sounds develop towards stable adult-like articulatory constellations (Goodell and Studdert-Kennedy 1993 ; Nittrouer 1993 ; Nittrouer et al. 1996). The literature is rather inconclusive : see Goodell and Studdert-Kennedy (1993) and Kühnert and Nolan (1999) for reviews. This is partly because coarticulation is not a unitary phenomenon, and cannot be expected to develop in a unitary way (Repp 1986). It also appears that even when one type of coarticulation is examined, there is a great deal of inter-subject variability and lack of agreement between different studies (Kühnert and Nolan 1999). With respect to anticipatory coarticulation from vowel to onset, however, it seems likely that the large, well-controlled studies of Studdert-Kennedy and his colleagues can be relied on. They suggest that children from the ages of 2-3 have a great deal more intra-syllabic co-production of consonants and vowels than older children and adults.

The work outlined above is still at a relatively preliminary stage, and only examines a limited range of coarticulatory phenomena in segmentally simple structures. This reflects a lack of knowledge about gestural coordination in more complex structures in adult language. Within the theory of Articulatory Phonology (which is compatible in many respects with the developmental work of Studdert-Kennedy and his colleagues) a number of articulatory studies indicate that there are different methods of phasing gestures with respect to each other in adult speech. Browman and Goldstein (1988), Honorof and Browman (1995) and Byrd (1995) found that the most stable timing relationship between an onset and a following vowel was defined on the articulatory midpoint of the onset (the « c-centre »), because the c-centre generalises over singleton and cluster onsets. This « global » relationship appears to differ from the « local » intergestural timing applicable to coda-vowel or vowel-onset sequences which cross a word boundary, in which the most stable timing relationship is defined between the vowel and the nearest *edge* of the consonant sequence. From articulatory (x-ray microbeam) data, it appears as if the c-centre is half-way through the onset. Byrd's (1995) study suggests that tautosyllabic vowel and coda may also be globally timed, at least for some speakers, but Honorof and Browman (1995) find no evidence for this. Recent work by Browman and Goldstein (1998) attempts to reconcile these differences using a « bonding » analysis in which the c-centre effect is derived as the optimal configuration in a model which minimises the violation of several local CV relationships resulting from articulating consonant clusters in the correct sequence. Whatever the exact explanation underlying the c-centre, in adult speech, the important point here is that it has been demonstrated in adult speech that the onset retains a stable articulatory relationship with the following vowel irrespective of the number of consonants in it.

We might hypothesise that complex gestural phasing of this kind might be harder to learn and more liable to misacquisition than simpler types of intergestural timing, such as local timings. Local timings are presumably sufficient for words containing no clusters. Under this hypothesis, the later age of acquisition of consonant clusters could therefore be due in large part to their complex articulation. Unfortunately there are very few instrumental studies of cluster acquisition and very few adult studies of gestural coproduction in clusters, so such a hypothesis must remain tentative, despite its superficial attractiveness. Of course, standard phonological accounts which rely at heart on the intuitive appeal of the greater structural complexity of branching vis-a-vis non-branching structures are equally tentative.

1.2 Cluster acquisition in normally-developing and phonologically disordered children

The work reported here forms part of an ongoing research programme looking at the normal and disordered acquisition of word-initial consonant clusters by Bill Hardcastle, Fiona Gibbon and Jim Scobbie at Queen Margaret University College, and Paul Fletcher at the University of Hong Kong. This paper examines second formant (F2) transtions from the stop burst into the following vowel in words with initial /d/ (a voiceless unaspirated coronal stop) and /st/ (a voiceless coronal fricative followed by a voiceless unaspirated coronal stop). At this preliminary stage there is no attempt to compare child and adult productions. Rather, this study addresses normally developing children with a view to gaining a greater understanding of developmental phonological disorder (p.d.). I follow up a result reported by Baker (1998), who investigated onset/rime formant transitions in a four year old child (DB) with phonological disorder (previously described in Scobbie, Gibbon, Hardcastle & Fletcher 1998, in press). DB produced /st/ as an unaspirated stop [t] which was homophonous with /d/ as part of a pattern of cluster reduction. Our working hypothesis was that, rather than phonological neutralisation of the /st/-/d/ contrast, DB might have had a disorder of phonetic implementation rendering /st/ and /d/ nearly identical in production such that both would be *transcribed* [t]. Nevertheless, imperceptible differences might have existed in duration, constriction location, voicing characteristics or degree of stricture — a « covert contrast. »[2] To test this hypothesis using spectral information, Baker sampled F2 at the burst, at burst+16ms and burst+30ms, but in fact no covert contrast was discovered for this particular parameter.

However, since the p.d. subject was followed longitudinally, Baker was also able to examine the transitions associated with /d/ and /st/ in those sessions in which DB had acquired the « correct » production of /st/ as [st]. Interestingly, F2 was lower after /st/ than after /d/ in the two minimal pairs studied in detail: *dye*/*sty* and *door*/*store*. A measurement of F2 at a point 100ms after voicing onset did not reveal any difference, suggesting that the same vowel target was attained, albeit via different transitions.

There are a number of possible explanations for this difference. Most obviously, the stops in /st/ and /d/ could have been articulated at different place of articulation. This could be due to coarticulation of the cluster stop to the child's production of /s/, or to a greater perseverative coarticulation on the stop exponent of /d/ from the preceding vowel /i/ (in the carrier phrase *give me___please*). Another explanation might be that there was a greater amount of tongue blade lowering posterior to the constriction in /st/ at the time of the burst. This could be due to greater anticipatory coarticulation with the vowel following /st/ following an earlier onset to the vowel opening gesture.

If the lower F2 following the cluster were due only to small differences in the place of articulation of the stop, the F2 movement schematised in the left panel of Figure 1 would be more likely : F2 following /st/ would begin at a lower point but approach the vowel target at approximately the same moment relative to the burst as F2 following /d/. The right panel of Figure 1 would be expected if the attainment of the rime target were not timed locally to the stop, but globally, to the midpoint of the onset, say. Given the small sample of points, it was not clear whether the equivalent low vowel targets following /d/ and /st/ were reached at the same time relative to the burst, indicating different starting points and rates of transition (which would have looked like the left panel of Figure 1), or whether the rates of transition were the same for each onset, leading to an earlier target following /st/ (right panel).

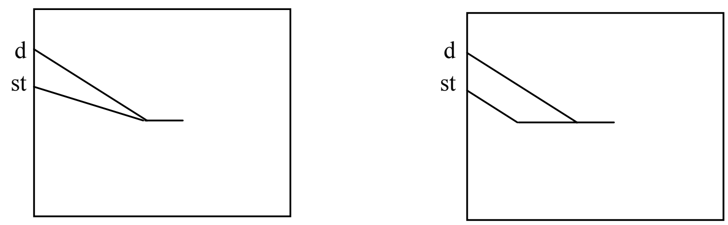

Figure 1. Two models of F2 transition from different y-intercepts. Vertical axis indicates frequency (Hz), horizontal axis indicates time from burst (ms). Left panel : both words time the vowel target to the burst of the onset stop. Right panel : both words have the same rate of transition.

As part of a more extensive study of onset-rime coarticulation prompted by Baker's finding, I decided to examine the speech of normally developing children. In this paper I will look at the movement of F2 in a sample of four year olds' productions of /dai/ and /stai/, sampling F2 at the burst and throughout the *entire* word. This should reveal whether the sample of normally developing children behave like DB in having a lower frequency of F2 after /st/ than /d/ and the extent of normal intra-subject variability. Furthermore, if the control subjects *do* display F2 lowering after the cluster, we may be able to uncover the underlying mechanism. This point is hard to determine, because the CV transitions are not linear as in Figure 1 but gradually reduce their rate of movement as the vowel target is attained.

As noted above, the literature on coarticulation between a complex onset and a following vowel is somewhat limited. Browman and Goldstein's (1988) model of gestural coordination posits an anchor to which the vowel opening gesture is attached, called the c-centre, a concept we will find useful below. The particular gestural sequences studied in the c-centre literatue have tended to be hetero-organic, enabling detection by the instrumentation of the end-points of the gesture even under conditions of overlap. Since a homorganic cluster is addressed here (/st/), the mutual influence of /s/ on /t/ and vice versa may involve a rather different underlying articulatory mechanism. Furthermore, only the acoustic record is available for study. Nevertheless, with care, the results can be interpreted in the light of what is already known about intergestural coordination.

2. Method

2.1 Data collection

The data discussed here form part of a larger study designed to investigate the developing phonetic/phonological systems of children with phonological disorder. Other results, relating to subjects with phonological disorder, can be found in Scobbie, Hardcastle, Gibbon & Fletcher (1995) and Scobbie, Gibbon, Hardcastle & Fletcher (1997, 1998, in press). In this paper, as part of a larger project on the acquisition of Scottish Standard English, I consider data from six children, aged approximately four years old, living in Edinburgh, with no known speech or hearing difficulties.[3,4]

In the dataset the /d/-/st/ contrast appears in three rime contexts : before /ai/, /or/ and /ir/. It was felt that looking at a complex rime would provide more information than is usually available from a monophthongal vowel. The diphthong /ai/ has two targets, and two transitional phases: the CV transition between onset and the first target, and an internal VV transition between the two diphthongal targets. Since /ai/ is realised as a diphthong by all the children in what appears to be a relatively adult-like manner, a meaningful measurement of the entire F2 contour of /ai/ in *dye* and *sty* was feasible. The other available minimal pairs in the dataset were *door*/*store* and *deer*/*steer*. The adult form of these in Scottish Standard English is a rhotic monosyllable, but children of this age frequently produce non-rhotic monosyllables and disyllables, so the decision was made to exclude these data from general analysis at this stage and focus on /ai/.

The children played games involving picture naming, and had to speak the word associated with a picture in the phrase « Give me ____ please. » Full details of data collection and annotation are given in Scobbie et al (1995), but some points are of note here. Six repetitions of each word were digitised using KAY CSL at 40,960Hz. Annotation of various acoustic landmarks was made, the ones relevant here being the onset of stop closure (after a preceding vowel or fricative), the release of closure (nearly always with a noticeable burst), the onset of voicing to the vowel, the end of the vowel (corresponding to the onset of the labial stop in *please*) and additionally, in *sty*, the onset of friction. Durational analysis of these acoustic segments was performed, but will only be referred to here when relevant. All annotation was performed by the author.

2.2 Spectral analysis

For spectral analysis, the digitised tokens were low-pass filtered and downsampled at 10,240Hz. A section of waveform corresponding to the vowel was displayed, from 1ms before the burst annotation point up to the end-of-vowel annotation point. A spectrogram was drawn of the displayed portion of the waveform, supplemented by automatic formant analysis using the FMT command. The spectrogram showed frequencies up to 5kHz, with the expectation that three formants would be displayed. The parameters of the LPC-based formant tracker were as follows. The autocorrelation method was used, with 20ms frames at 25ms steps. Pre-emphasis of 0.8 was used, and 8 poles used in the calculation of the formants. Visually, the formants were indicated superimposed on the spectrogram as points, and if it was felt that they successfully tracked the formant that had been identified impressionistically as F2, then the values of F2 determined by the CSL formant tracker were copied into a spreadsheet unaltered, though every value had to be checked. This was the appropriate action for more than half the tokens (evenly spread among the two targets and six subjects). On many occasions, however (particularly during transitions and in tokens with a high fundamental frequency) the formant tracker did not appear to give good results, so the token was measured by hand from the spectrogram.[5] In total, 801 measurements were made, approximately 11 measurements per token (corresponding to 250ms). The average duration of the rimes was in fact 241ms for *sty* and 256ms for *dye*.

This is not to imply that 11 measurements were made from every token. On the contrary, there was a great deal of token-to-token and subject-to-subject variation in rime duration, and, consequently, in the number of points measured. To illustrate the former case, consider that some tokens were as long as 425ms, some as short as 100ms. As an example of the latter, MM's rimes were short, at about 170ms long, while EC's were long, at about 310ms. Measurements were made at 25ms intervals for as long as possible. It might appear, therefore, that a better measurement technique would be to sample F2 at, say, 10% intervals throughout the rime, which *would* give the same number of analysis points for every vowel. While this normalisation would undoubtedly give more homogenous results for the measurement points late in the rime, it is a more time-consuming procedure and offers no normalisation benefits during the initial CV transition immediately after the burst which was expected to provide the most important data.

Each measurement frame was 20ms long. The so-called « burst » measurement actually began 1ms before the burst annotation point, so was centered 9ms after the burst. At this point in the rime the formant was excited by a mixture of aperiodic and periodic energy in most cases. This position was chosen as the centre of the first frame rather than the burst annotation point itself after some experimentation, because it provided a more reliable measure of F2. LPC measurements made right on the burst were too variable to be relied upon, probably because they reflecting burst spectra rather than the cavity-related spectra which were being tracked. This first F2 measurement was therefore more strictly speaking based on F2 in the period just after the burst, while the articulators are separating. Subsequent measurement frames follow at 25ms intervals, so the second was centred 34ms after the burst, and so on. In presenting the results, I will stick with the simpler naming scheme of 0ms, 25ms, 50ms etc unless greater accuracy is required.

3. Results and discussion

The mean value of F2 at the burst measurement point for /dai/ (2660Hz) was 272Hz higher than for /stai/ (2388Hz), reflected as a strong main effect in a 2-way ANOVA, $F(1,60)=54.54$, $p<0.001$. Each of the six individual subjects displays a similar pattern, although there is some variation : the smallest difference in F2 is 101Hz (subject 4EC) and the greatest 434Hz (subject 4GB), and this variation is reflected in the statistical analysis in a significant interaction between subject and onset factors $F(5,60)=3.28$, $p<0.05$, as well as a strong main effect for subjects, $F(5,60)=8.21$, $p<0.001$.

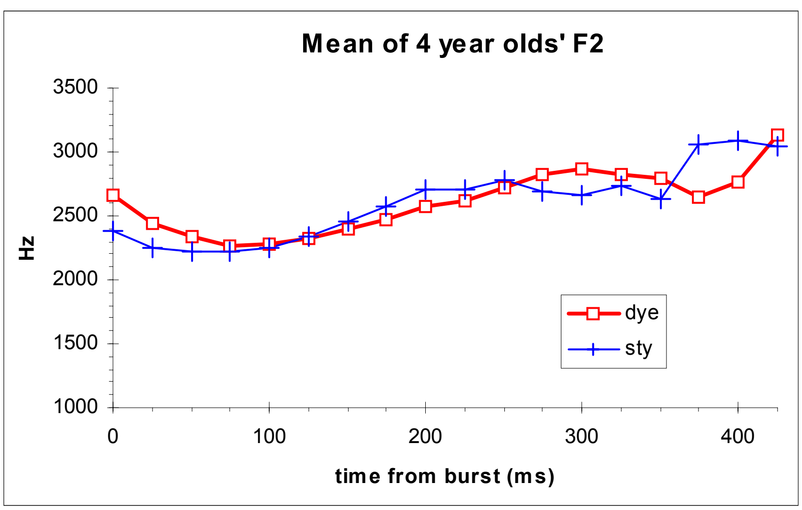

Figure 2. Mean F2 trajectories for two different onsets preceding /ai/.

Figure 2 clearly demonstrates that sampling F2 at the absolute interval of 25ms and then taking an average at each interval gives rise to noisy results towards the end of the vowel. The number of tokens being averaged remains at the maximum 36 till 100ms for *sty* and 150ms for *dye*, after which the number falls away to two tokens of *sty* and a single token of *dye* at 425ms. The alternative approach, sampling each token a fixed number of times in intervals defined relative to the duration of that token, would have given a more meaningful F2 contour for the second half of the vowel. Figure 2 contains clear artifactual sudden jumps in F2. There is also a general lowering of the F2 peak, due to brief but large lowering offglides in F2 being included in the mean F2 track. However, we are more concerned with the earlier parts of the diphthong, and in particular the first meaurement point, which is unaffected by the sampling method. The main effect of onset is the same in both cases, and since both words pattern in a similar way, the absolute 25ms interval method still allows useful comparison to be made. As for intersubject differences, these affect the *size* of the F2 lowering effect, not its occurrence. Note that the difference in F2 is not due to differences in VOT. Mean VOT for *sty* is 21ms, and for *dye* it is 20ms.

The general shape of the F2 contours in Figure 2 in the first 200ms is particularly interesting. The difference in the CV transition can be seen quite clearly, and, as predicted, is difficult to model linearly in the manner of Figure 1. The F2 contour in /stai/ begins lower and then reaches the low target of the diphthong earlier. Inspection of individual subjects confirms this, except in the case of 4EC, where there appears to be no onset-determined timing difference in the attainment of the low F2 target. Examining data from all tokens, it appears that both words have similar low F2 targets for the diphthong : the mean /a/ F2 target for /st/ is 2086Hz (n=36) and for /d/ is 2168Hz (n=36) (a difference which is just insignificant in a t-test).

So, the diphthong following the cluster has a lower F2 at the burst than the diphthong following the singleton, and the first F2 target following the cluster seems to be achieved earlier than the F2 target for the diphthong following the singleton, but is at the same frequency.

Consider now the VV transition within the diphthong (from /a/ to /i/). The diphthongal transitions begin some 75ms-100ms after the burst, and move in the opposite direction to the CV transitions. The transition to the second vowel target begins earlier in /st/-initial words. Moreover, it shows the opposite effect to the CV transition: because it rises sooner, F2 is higher in the /st/-initial words. Note that this pattern is less uniform, and 4DB and 4EC appear to have no difference in the VV transition,[6] but in general F2 is higher in the cluster-initial word from 125ms-200ms after the burst (after which point the contours become too irregular and unreliable to use). From this I conclude that diphthongal opening gesture *and* the diphthongal closing gesture are both begun earlier, relative to the stop burst, in the cluster initial word. This is a very important point, since it implies a difference between *dye* and *sty* which is not merely located in the articulation of the stop, for example, without reference to the following rime.

My interpretation of the results is that the diphthongal rime as a whole is timed to begin with respect to the syllable *onset* and not the preceding segment, the stop. This hypothesis holds that the gestural timing in each word is, at a prosodic level, identical. A fixed coordination of the rime to some point in the onset (perhaps an immature correlate of the c-centre) will derive apparent differences in the timing of F2 relative to the burst, because complex onsets are longer than simplex onsets in duration. For example, the mean difference in duration between the obstruent intervals /st/ and /d/ is 96ms, so if the acoustic midpoint of the onset were the appropriate locus for timing, we might expect the diphthong gestures in *sty* and *dye* to be out of sync by about 50ms. In fact, this figure is a little high, on the basis of the results, but is in the right approximate area. The children’s timing anchor may be nearer the end of the /st/ onset. When analysis of all 12 subjects is complete, a more accurate figure may emerge, but since it will, by necessity, be calculated acoustically, it's not clear how comparable it will be to the results of the articulatory studies in any case.

Meantime, consider Figure 3, which shows that the mean durational difference in the onset (/st/-/d/) correlates with the mean difference (/d/-/st/) at the burst in F2. The c-centre model proposed above predicts that a subject with a larger mean durational difference between /st/ and /d/ would therefore time the onset of the vowel gesture earlier than a subject with a smaller durational difference, giving rise in turn to a larger mean difference in F2 at the burst. (The five-year-olds’ results also seem to fit this pattern and it appears that there is a strong statistical correlation when all 12 subjects are examined, although with so few speakers, this correlation must be treated cautiously.)

As a final point, note that a c-centre type model might predict that the overall duration of the vowel /ai/ is less after /st/ than after /d/. This prediction is not upheld, however. As was mentioned above, /ai/ is is 256ms in /dai/ and 241ms in /stai/, and indeed all six four-year-olds follow the pattern of /ai/ being longer after /d/, but the five-year-olds appear to pattern the opposite way, giving rise to overall vowel duration means that are very similar. Perhaps this is due to the overall variability in the data and the small size of the predicted difference.

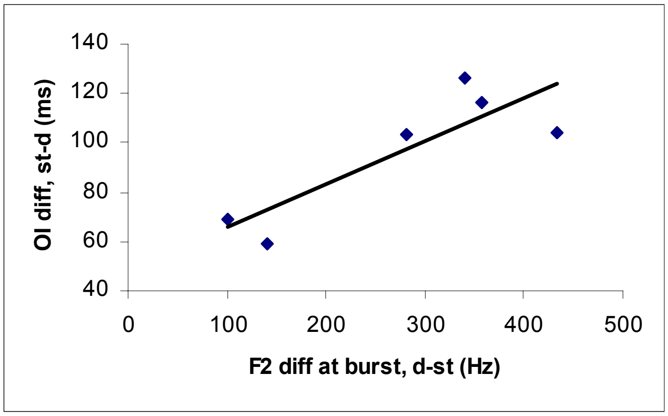

Figure 3. Scatterplot of the mean difference in duration between /st/ and /d/ (y-axis) against the mean difference in F2 at the burst (x-axis) for six four-year-old subjects.

Two studies of fricative-stop coarticulation by Noel Nguyen and colleagues (Nguyen, Gibbon & Hardcastle, 1996; Nguyen, Wrench, Gibbon & Hardcastle, 1998) provide some normative data with which to compare the results presented above. Noting the lack of published coarticulatory data on clusters, Nguyen et al undertake perception and production studies to investigate further the results of Repp & Mann (1982) (which need not concern us here).

Nguyen et al. (1996) investigated a variety of CV and FCV sequences using EPG in the speech of an adult male phonetically trained speaker of Southern British English. Of interest here is the difference between /da/ and /sta/. They found that the fricative [s] exerts a coarticulatory influence on the following stop [t] as follows : in the first frame of complete closure, [t]'s pattern of tongue-palate contact has a less anterior score for centre of gravity (COG) when it is a singleton (4.18) compared to when it is part of a cluster (5.32). Qualitatively, the difference is that [ta] has more lateral contact in the three backmost rows than [sta]. In the final frame of complete closure, there is almost no difference between the in-cluster and singleton stops.

Nguyen et al. (1998) is a perception and production study. It looks at two adult female phonetically trained speakers (Southern British English and Australian English). The speech analysis component makes use of EMA, EPG and acoustic analysis. The /d/ vs. /st/ materials (*pup dump* vs. *muss dump*) are not comparable with those in this study because the /s#d/ cluster is heterosyllabic and indeed heterolexical. Nevertheless, the results are worth mentioning. From the EMA analysis it appears that at the release of the stop consonant and at the onset of the following vowel, the midsaggital profile of the tongue dorsum was slightly *higher* in the cluster environment *muss dump* than in the singleton environment *pup dump*. Acoustic analysis of F2 at the start of the vowel showed a very slight lowering after /s/ (1714Hz vs. 1766Hz for speaker 1, 1738Hz vs. 1747Hz for speaker 2). Nguyen et al. conclude that their results are compatible with findings that there is hollowing of the tongue immediately behind an [s] constriction. In other words, the differences in word-initial /d/ caused by a preceding word-final fricative are not limited to primary place of articulation, but also involve factors influencing the shape of the tongue posterior to the constriction.

One interpretation of these results, mentioned in Section 1.2, is that the stop in /st/ is simply articulated in a different way to the stop in /d/ without the temporal dimension introduced above. The influence of /s/ may have had a direct effect in that coarticulation of the stop to /s/, resulting in a different articulatory pattern of closure for the stop, could account for the lower F2 at burst. An analysis of the difference in F2 which is based only on constriction patterns without any aspect of intergestural timing will not account well either for the earlier attainment of the low F2 target nor the earlier timing of the VV transition following /st/.

The presence of /s/ may also have had a more indirect effect, lowering F2 in this particular experiment by *blocking* coarticulation. The segment in the carrier phrase immediately preceding the target words was /i/ : the stop is not adjacent to /i/ in *sty* and so again a lower F2 is not unreasonable just due to the greater temporal distance from that vowel in the cluster case. Clearly, data with a different preceding vowel context would ideal to test this hypothesis, but that is not possible with this dataset. I do not think this hypothesis is partiularly compelling however. It assumes that the children exhibit consistent intermorphemic, intersyllabic coarticulation. As discussed in Section 1.1, this is not likely. I think an explanation which rests on strong, syllabically timed intrasyllabic coarticulation sits better with the known findings about children's intergestural timing and adult coarticulation within the syllable.

Nguyen et al.'s EPG results are, moreover, entirely compatible with the timing model outlined above, because they suggest that the tongue dorsum is actively lowering during the stop closure in both the clustered and singleton words, but that the lowering and hollowing out begins earlier in [sta]. In the case of [ta], there is no lowering in the first closure frame, but in [sta], on the other hand, the first frame of closure follows the fricative phase, and the tongue dorsum has had time to begin lowering already by the time the first frame of closure is reached. The lowering means that there is a reduced amount lateral contact in the posterior rows of the palate. In the final frame of closure, EPG records the same amount of tongue-palate contact in both cases, though on the basis of the F2 results reported above, I presume that the cavity would be larger in [sta].

4. Conclusion

The specific impetus for this study was Baker's (1998) study of a child with developmental phonological disorder (DB). I wanted to see whether his « phonological » disorder was responsible for his having a lowered F2 at the stop burst of /st/ (compared to the related singleton /d/). It appears that it was not, because the six normally developing children beteween 4 and 4½ years of age reported here also have lowered F2 in the cluster context.[7] Both the initial and the final vowel targets appear to be unaffected by whether the onset is /st/ or /d/. Given the overall pattern of F2 movement, both in the initial CV transition and in the VV transition inherent to the diphthong, as revealed by measuring F2 in 25ms increments, it is likely that the articulatory vowel gestures underlying /ai/ are temporally advanced (relative to the burst) after /st/ compared to /d/. Further analysis (of six more subjects, aged five years old) is on-going, and the results so far are comparable to those mentioned here.

These results support models of gestural timing in which some consistent point defined by reference to a prosodic entity such as the syllable onset can serve as a temporal anchor. The greater duration of /st/ suggests a model in which aspects of the opening gesture of the diphthong (such as the increase in cavity size posterior to the closure) have progressed further in a cluster in the time available before stop closure is released. Hence when F2 is measured, it is lower during the first 75ms or so following a cluster. The second transition, between the two diphthongal targets, provides evidence that there is an apparently earlier initiation of the diphthong during the cluster-intitial word (though with more individual variation). The exact point during the obstruent interval at which the vowel is initiated remains to be discovered, but even without any articulatory data, the correlation between the durational difference between /st/ and /d/ and their F2 difference at the burst supports the contention that there is a consistent location during that interval which is an anchor for intergestural timing.

EndNotes

1. Katz, Kripke & Tallal (1991), on the other hand, holds that children's speech is more segmental than adult speech.
2. See Scobbie (in press) for a review of the covert contrast literature.
3. Three of the children (4DB, 4RM, 4EC) were within a month of their fourth birthday and three (4CK, 4MM, 4GB) were within a month of 4;6 at the time of recording. The age difference is collapsed here, and the subjects referred to as being « four years old, » although more accurately the mean age of the subjects is 4;3. Six other control subjects (in groups approx five years old and five and half years old) have been analysed, and the results are extremely similar to those presented here. Further analysis is ongoing. Of particular interest are the word pair *store* and *door*, which, on the basis of measurements by Lesley Baker, exhibit a similar pattern of F2 lowering.
4. In broad phonological terms, there are no differences in the consonant systems of Scottish Standard English (SSE), GenAm and RP that is relevant here, nor are there any relevant onset/rime phonotactics that require mention. I will therefore refrain from presenting a description of SSE. For details, see the papers on SSE in Foulkes & Docherty (1999). Note, however, that the patterns described in this paper might not occur in other dialects of English, considering the typologically unusual phonological and phonetic characteristics of the diphthong /ai/ (Scobbie, Turk & Hewlett, 1999).
5. All the spectral measurements reported here were performed by the author. Similar measurements were performed by Lesley Baker, and though they are not reported here, are in broad agreement with my results.
6. This is not the p.d. subject DB but a normally developing child with the same initials.
7. Further analysis is required to determine whether the *degree* of DB's lowering was atypical and whether children with phonological disorder (at least, those who apparently have normal exponents of /st/ and /d/) typically master this aspect of the phonetic grammar. This is a topic for future research.

Acknowledgements

The data was collected as part of the MRC project « An instrumental investigation of consonant cluster acquisition in phonologically disordered and normally developing children » (G9117453), and analysed as part of the ESRC project « The Scottish Vowel Length Rule: the phonetics, phonology and acquisition of a marginal contrast » (R000237135). Thanks are due to Fiona Gibbon, Bill Hardcastle and Paul Fletcher for their original insights and to Fiona Gibbon in particular for continuing collaboration on the QMUC Cluster Acquisition Database, to Lesley Baker for her many formant measures of these and other subjects, to Alan Wrench, Cassie Mayo and Alice Turk for very useful discussions on issues related to these proposals, and to Ailbhe Ni Chasaide and Noël Nguyen for comments on an earlier draft. None of them should be held responsible for errors or omissions.

References

Baker, L. (1998). *Lingual onset-rime coarticulation in the speech of phonologically disordered child: a longitudinal study*. Unpublished undergraduate thesis, Dept of Linguistics, University of Edinburgh.

Byrd, D. (1995). C-centers revisited. *Phonetica*, *52*, 285-306.

Browman, C.P. & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140-155.

Browman, C.P. & Goldstein, L. (1998). Competing constraints on intergestural coordination and self-organisation of phonological structures. Paper presented at *CTIP-II, Second Conference on Current Trends in Phonology*. Royaumont Abbey, France, June 22 - 24 1998.

Davis, B.L. & MacNeilage, P.F. (1990). Acquisition of correct vowel production : a quantitative study. *Journal of Speech and Hearing Research*, *33*, 16-27.

Ferguson, C.A. & Farwell, C.B. (1975). Words and sounds in early language acquisition. *Language*, *51*, 419-439.

Foulkes, P. & Docherty, G.D. (Eds.) (1999). *Urban Voices: Variation and Change in British Accents*. London, U.K. : Arnold.

Goodell, E.W. & Studdert-Kennedy (1993). Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: a longitudinal study. *Journal of Speech and Hearing Research*, *36*, 707-727.

Honorof, D.N. & Browman, C.P. (1995). The center or edge: how are consonant clusters organised with respect to the vowel? In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (vol.3, pp. 552-555). Stockholm, Sweden : Congress Organisers at KTH and Stockholm University.

Katz, W., Kripke, C. & Tallal, P. (1991) Anticipatory coarticulation in the speech of adults and young children : acoustic, perceptual and video data. *Journal of Speech and Hearing Research 34*, 1222-1232.

Kühnert, B. & Nolan, F. (1999). The origin of coarticulation. In W.J. Hardcastle & N. Hewlett (Eds.) *Coarticulation : Theory, Data and Techniques*, 7-30. Cambridge, U.K. : Cambridge University Press.

Macken, M.A. (1979). Developmental reorganisation of phonology: a hierarchy of basic units of acquisition. *Lingua*, *49*, 11-49.

Menn, L.(1983). Development of articulatory, phonetic and phonological capabilities. In M. Butterworth (Ed.) *Language Production*, 3-50. London, UK : Academic Press.

Nittouer, S. (1993). The emergence of mature gestural patterns is not uniform: evidence from an acoustic study. *Journal of Speech and Hearing Research*, *36*, 959-972.

Nittrouer, S., Studdert-Kennedy, M. & Neely, S.T. (1996). How children learn to organise their speech gestures: further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research*, *39*, 379-389.

Nguyen, N., Gibbon, F. & Hardcastle, W.J. (1996). Articulatory and perceptual aspects of fricative-stop coarticulation: a pilot study. In J.M. Scobbie (Ed.), *QMC Working Papers in Speech and Language Sciences*, *1*, 13-25. (Available electronically at « http://sls.qmced.ac.uk ».)

Repp, B.H. & Mann, V.A. (1982). Fricative-stop coarticulation: acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *71*, 1562-1567.

Repp, B.H. (1986). Some observations in the development of anticipatory coarticulation. *Journal of the Acoustical Society of America*, *79*, 1616-1619.

Scobbie, J.M. (in press). Interactions between the acquisition of phonetics and phonology. In Catherine Gruber, Derrick Higgins, Kenneth Olson and Tamra Wysocki (eds.) *Papers from the 34th Annual Regional Meeting of the Chicago Linguistic Society, Volume II: The Panels*. Chicago: Chicago Linguistics Society.

Scobbie, J.M., Gibbon, F., Hardcastle, W.J., & Fletcher, P. (1997). Longitudinal phonological and phonetic analyses of two cases of disordered /s/+stop cluster acquisition. In A. Sorace, C. Heycock & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition* (pp. 278-283). Edinburgh, U.K.: University of Edinburgh.

Scobbie, J.M., Gibbon, F., Hardcastle, W.J., & Fletcher, P. (1998). Covert contrast and the acquisition of phonetics and phonology. In W. Ziegler & K. Deger (Eds.), *Clinical Phonetics and Linguistics* (pp. 147-156). London, U.K.: Whurr Publishers.

Scobbie, J.M., Gibbon, F., Hardcastle, W.J., & Fletcher, P. (in press). Covert contrast as a stage in the acquisition of phonetics and phonology. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon*, Cambridge, U.K.: Cambridge University Press.

Scobbie, J.M., Hardcastle, W.J., Fletcher, P. & Gibbon, F. (1995). Consonant clusters in disordered L1 acquisition: a longitudinal acoustic study. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (vol.2, pp. 706-709). Stockholm, Sweden: Congress Organisers at KTH and Stockholm University.

Scobbie, J.M., Turk, A.E. & Hewlett, N. (1999). Morphemes, Phonetics and Lexical Items: The Case of the Scottish Vowel Length Rule. In J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A.C. Bailey (Eds.) *Proceedings of the XIVth International Congress of Phonetic Sciences* (vol.2, pp. 1617-1620).

Stoel-Gammon, C., Buder, E.H. & Kehoe, M.M. (1995). Acquisition of vowel duration: a comparison of Swedish and English. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (vol.4, pp. 30-36). Stockholm, Sweden: Congress Organisers at KTH and Stockholm University.

Sereno, J.A., Baum, A.R., Marean, G.C. & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *Journal of the Acoustical Society of America, 81,* 512-519.

Studdert-Kennedy, M. & Goodell, E.W. (1995). Gestures, features and segments in early child speech. In B. de Gelder & J. Morais, *Speech and reading: a comparative approach.* London, UK: Erlbaum, Taylor & Francis.

Vihman, M.M. & Velleman, S. (1989). Phonological reorganisation: a case study. *Language and Speech*, *32*, 149-170.

Waterson, N. (1971). Child phonology: a prosodic view. *Journal of Linguistics*, *7*, 179-211.