AN INSTRUMENTAL STUDY OF ALVEOLAR TO VELAR ASSIMILATION IN CAREFUL AND FAST SPEECH

Lucy Ellis and William J. Hardcastle
Department of Speech and Language Sciences, Queen Margaret University College, Edinburgh, U.K.

ABSTRACT

The assimilation of a word-final alveolar to a following velar has been traditionally described as a discrete phonological process. That is, the place of articulation features for the alveolar have been completely swapped for those of the velar. More recently electropalatographic (EPG) studies have shown empirically that this process is sometimes gradual, providing evidence of intermediate 'residual' alveolar articulations. These conflicting perspectives raise the question: at what level in the generation and execution of an utterance *does* assimilation occur? A number of speakers' productions of /n#k/ were recorded using EPG. The major finding is that while some subjects produce gradient assimilations, others clearly demonstrate categorical assimilations. The lack of residual movement in the latter group was confirmed in a pilot study using EPG in combination with EMA (electromagnetic articulography), a technique which complements EPG contact-only data. On the basis of this, a speaker-specific model of assimilatory behaviour is proposed.

1. EPG STUDY

1.1 Introduction

Assimilation describes the variation in the phonetic description of a speech unit as it becomes more like an adjacent speech unit. For example *red coat* may be realised as [ɹɛg kəʊt]. This process is particularly prevalent in fast speech.

Standard phonological theory traditionally describes place assimilation as a discrete phonological process [1]. So, in the case of alveolar to velar assimilation either there is no assimilation and the alveolar target is preserved or an assimilation takes place where the alveolar target is completely replaced by a velar target. An assimilation is said to be caused by a cognitive substitution rule, utilising stored phonetic/phonological information, which results in categorical variation. More recently, however, the level of phonetics, previously thought of as mere implementation of phonological form determined at a higher cognitive level, has been regarded by some as having considerable explanatory power in accounting for assimilatory processes. Research into the articulatory mechanisms underlying assimilation have shown that this process is sometimes gradual [2,3]. In the case of alveolar to velar assimilation there may be intermediate assimilatory forms which indicate 'residual' alveolars where the tongue's target has been 'undershot'. Thus, phonological theory and instrumental research have tended to locate assimilation within the abstract planning stage and the concrete physical execution stage respectively.

Previous studies in this area were based on small subject numbers (typically one or two) and were concerned more with identifying types of assimilatory pattern than with how speakers might differ with regard to preferred assimilatory strategy. For these reasons the EPG study reported below is an attempt to systematically investigate the distribution of types of alveolar assimilation across a number of speakers. Another way in which work reported in this paper sought to overcome some limitations of previous EPG studies was through the use of Electromagnetic Articulography (EMA) in combination with EPG in a follow-up pilot study. EPG data in isolation gives information on tongue-palate contact only and not on underlying movements. Thus apparently complete alveolar assimilations, indistinguishable from lexical velar patterns, may involve some tongue tip raising short of contact with the alveolar ridge, indicating a reduced but preserved alveolar gesture. This pilot study will be presented in section 2.

In this paper we will show the relevance of speakers' asssimilatory strategy to internal modeling of speech planning.

1.2 Method

1.2.1 Stimuli Speech material captured a potential site of alveolar assimilation and a neutral velar control sequence. These experimental combinations were embedded in the sentences: *"It's hard to believe the ba<u>n c</u>uts no ice "* /n#k/ and *"I've heard the ba<u>ng c</u>omes as a big surprise"* /ŋ#k/. Another experimental sentence was devised to capture the alveolar to alveolar sequence /n#t/:*"I'm not surprised the ba<u>n t</u>ouched a raw nerve"* so that coarticulatory effects on /n/ before /k/ can be compared with /n/ in a non-coarticulatory context. The patterns for the velar to velar control sequences served as a comparison for apparent cases of complete alveolar assimilation and as a yardstick for the identification of residual alveolars. The vocalic environment for the sequences was kept as consistent as possible. Bordering vowels were /a/ & /ʌ/ and the /a/ vowel was preceded by a bilabial stop to eliminate the possibility of any lingual coarticulatory effects on the target consonants. A further 4 filler sentences of no experimental interest were added to the original 3. 10 repetitions of each sentence were required bringing the total stimuli to 70 sentences.

1.2.2 Data collection The technique of electropalatography (Reading EPG3 system) was used to record the timing and the location of tongue contact with the hard palate during continuous speech. 10 speakers with EPG palates were recorded.

The experiment fell into two parts. The aim of the first part was to elicit careful speech and all subjects were instructed to read each sentence *slowly and clearly.* The aim of the second part was to elicit fast and casual speech. The material for this second part was identical to the careful speech part but the 70 sentences were arranged in groups of 3 and filler sentences were purposely distributed to avoid 'clustering' of experimental sentences. 2 filler items were repeated an extra time each to make 24 groups of 3 sentences. A time limit of 5 seconds on the delivery of each group of sentences was imposed on the speakers. The time constraint automatically ruled out any attempt on the part of speakers to pause between each sentence or to impose too complex an intonational structure on the sentences. The time limit was successful in eliciting speech appreciably faster than the careful speech condition and in preventing over-awareness of test items they had already encountered in the first half of the data collection. The sentences in the first half of the experiment and the groups of sentences in the second were individually cued during recording with a pause between each.

1.2.3 Data analysis The criteria for labelling a contact pattern as an alveolar stop closure was the presence of mid-sagittal contact in the first three rows of the EPG palate. Thus according to this definition, Figure 1 (a) is an example of an EPG contact pattern annotated as an alveolar stop articulation, while in 1 (b) we see a contact pattern annotated as an assimilation because there is no mid-sagittal contact in the alveolar region. Individual palate diagrams in 1 (a) and (b) are 10ms apart and tongue-palate contact is indicated by filled circles. The top of each diagram is the alveolar region and the bottom is the velar region. 1 (b) is an example of a residual alveolar articulation whereby closure across the alveolar ridge is absent but the tongue has still made the supporting lateral gesture. A residual alveolar is considered to be an intermediate assimilatory stage where the target closure is undershot due to time constraints. This interpretation follows Lindblom's duration-dependent undershoot model [4]. On the basis of acoustic evidence of vowel reduction he proposed that articulatory and acoustic undershoot of vowels is a function of *reduction of movement* towards the vowel target due to physiological limitations.

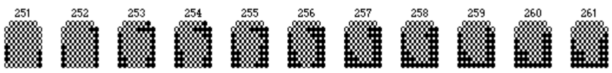

Figure 1 (a) EPG patterns for /n#k/ fast speech annotated as alveolar stop closure starting at frame 253 (Subject b)

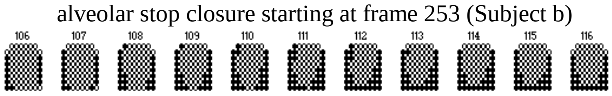

Figure 1 (b) EPG patterns for /n#k/ fast speech annotated as alveolar assimilation (Subject a)

1.3 Results and Discussion

In the fast speech condition a number of speaker specific assimilatory strategies were identified. 2 subjects produced only non-assimilations, 4 subjects always produced what appeared to be complete assimilations and the remaining 4 subjects each produced varying forms of /n#k/. Figure 2 shows the occurrence of assimilations for all speakers. All tokens here are categorized as either non-assimilations or assimilations with residual alveolars/partial assimilations subsumed into the assimilation category, as they were for the annotation procedure. Speakers a-j are ranked from left to right according to frequency of assimilations.

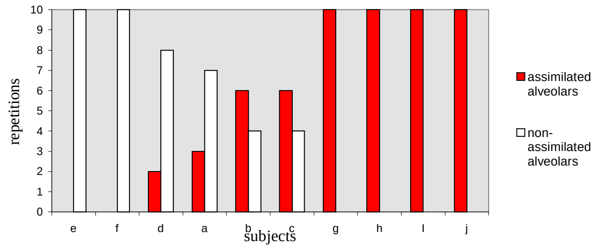

Figure 2 distribution of assimilations and non-assimilations for individual speakers in productions of /n#k/ fast speech

Fast speech /n#k/ EPG patterns for those speakers who produce habitual 'complete' assimilations (g, h, i and j) were indistinguishable from their fast speech lexical /ŋ#k/ productions. It is clear from the graph, however, that assimilation is not motivated by fast speech alone for all speakers (i.e. subjects e and f).

But the most notable result comes from the assimilatory data for those speakers who, in Figure 2, appear to vary between non-assimilation and assimilation. A more detailed examination of the type of patterns produced such as in Figure 1, reveals a fundamental contrast in assimilatory strategy between two groups of subjects, namely Subjects a and b and Subjects c and d.

The contact patterns for subjects c and d suggest the adoption of a binary segmental strategy. That is, either full alveolar contact is achieved for target /n#k/ or an assimilation takes place whereby the place features for /n/ have been completely swapped for those of /ŋ/. The most plausible explanation is that this is achieved for these speakers by accessing a phonological rule. By contrast the EPG data for subjects a and b show a non-binary continuum of assimilatory patterns. Patterns for these speakers can be ranged from full alveolar stop closure through intermediate residual patterns to velar patterns indistinguishable from lexical controls. Figure 3 shows all 10 repetitions of /n#k/ fast speech produced by subject b. Each line captures the tongue-palate contact for a single realization of the target sequence. All 10 realisations are ordered to show an articulatory continuum from full alveolar stop closure at the top to apparently complete assimilation at the bottom. Repetition 3 on line 3 of Figure 1 shows a slightly shorter alveolar articulation with less contact in the alveolar region (rows 1-3) than repetitions 1 and 2. By the time we get to repetition 4, alveolar closure is only partial. Repetition 5 is a residual alveolar articulation whereby the tongue is stretching as far forward as row 3, frames 316-320. The next 5 repetitions follow a progression of lessening side contact until repetition 10, which is the result of either a spirantised /k/ or a closure made too far back for the EPG palate to sample.

In fact, the only 3 residual 'undershoot' articulations yielded from the entire experiment were produced by these two subjects, a and b. Thus while some subjects produce gradient assimilations, others demonstrate categorical assimilations. This finding suggests that variation on basic speech units arising from assimilation is for some speakers determined at a higher level in the form of stored phonemic and thus categorical alternatives, while for others variation is 'computed on line'.

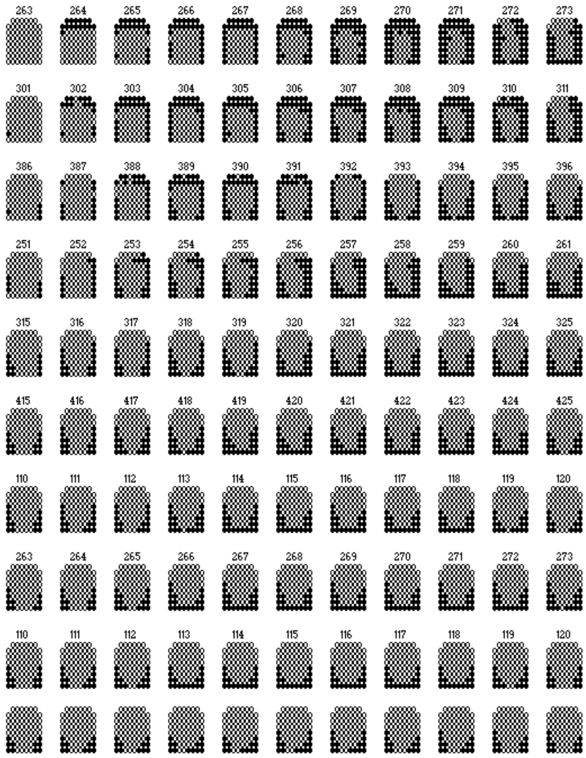

Figure 3 All 10 repetitions of /n#k/ fast speech subject b. Each line captures tongue-palate contact for a single realisation of /n#k/

2. PILOT EPG/EMA STUDY

2.1 Introduction

EPG data is limited because it gives information on tongue-palate contact only and not underlying lingual movement which does not necessarily result in contact with the hard palate. This information is important in the light of speculation that apparently completely assimilated /n#k/ sequences may be accompanied by some residual movement in the form of vertical tongue-tip raising [3]. Kühnert [5] found that identical EPG patterns arising from assimilated alveolar to velar sequences and velar control sequences *may* show different tongue trajectories in the EMA display. This finding undermines the view that differences in lexical form will always result in distinct phonetic output.

If tongue tip raising could be found for the complete assimilations of subjects c or d, then their assimilatory strategy would be gradient and not categorical as previously assumed. This would also mean that gradient movement, on an assimilatory continuum, is observable in both EPG and EMA data since we have already seen gradience with EPG for subjects a and b. Intermediate undershoot forms involving lateral extension of the tongue body visible on EPG patterns as in Figure 1 (b) are surely quite different articulatory events to those forms involving vertical tongue tip raising.

Two subjects who produced categorical assimilations, d and h, were selected to be recorded using combined EPG/EMA to look again for any gradience. Velar to velar sequences /ŋ#k/ were used as controls assuming that these involve minimal or no tongue tip raising.

2.2 Method

Subjects d and h were re-recorded using EPG in combination with EMA (Carstens AG100 Electromagnetic Articulograph). EMA is a transduction device which tracks x-y movement of coils attached to the mid-line of the tongue, typically on the tip/blade, tongue body and tongue dorsum.

Stimuli for this experiment were identical to that of the EPG-only study.

2.3 Results and Discussion

On analysis of the EPG data, it was found that subjects d and h replicated the assimilation strategy they each used in the EPG-only study. On analysis of the EMA data, there was no evidence of gradience in the form of tongue tip displacement for assimilated /n#k/ sequences of a greater magnitude than that for a neutral velar control articulation.

The articulatory positions of tongue tip and tongue dorsum coils for fast speech /n#k/ and /ŋ#k/ tokens at the beginning of the consonant cluster for subject d and h are shown in Figures 4 and 5 below. Beginning of the cluster was defined as the moment of maximum tongue tip displacement (regardless of whether a stop closure was achieved or not) and this was taken from minimum tangential velocity of the coil as it reaches its maximum height and changes direction. x-axis and y-axis position is shown in millimetres. The left hand cluster on each graph shows tongue tip positions and the right hand cluster shows tongue dorsum positions. Subject d's articulatory positions are plotted for non-assimilated alveolar sequences (numbering 5), apparently assimilated alveolar sequences (numbering 5) and all velar control sequences (numbering 10). For subject h, articulatory positions are plotted for apparently assimilated alveolar sequences (10) and velar control sequences (10) only. Subject h produced only 'complete' assimilations.

In Figure 4 maximum vertical displacement for the tongue tip is defined by full alveolar closure for the 5 non-assimilations produced by Subject d. For this speaker, however, it is clear that the tongue tip cluster for the 'complete' /n/ assimilations overlaps with the cluster for the underlying velars. This means that there is no vertical displacement for any assimilated /n#k/ sequence beyond that which accompanies a neutral velar control sequence and thus no intermediate partial assimilation stage in between full alveolar stop closure and complete assimilation. In fact the tongue tip cluster for assimilated /n/ is more constrained spatially than the cluster for underlying /ŋ/ in a way more characteristic of a target articulation. The 'target' in this case could be the result of a high-level instruction to delete the coronal gesture.

For Subject h, a similar picture emerges with overlap of tongue tip positions for assimilated /n/ and lexical /ŋ/. Since this subject produced no full alveolar stop closures for /n#k/ no maximum tongue tip raising is defined, but it seems that the raising movement for underlying velars is surprisingly advanced for some tokens. Once again, the height range for underlying forms is greater than that for assimilated /n/ forms suggesting a constraint on movement variability for the latter. These results confirm that Subject d and h are genuinely operating categorical assimilatory strategies.

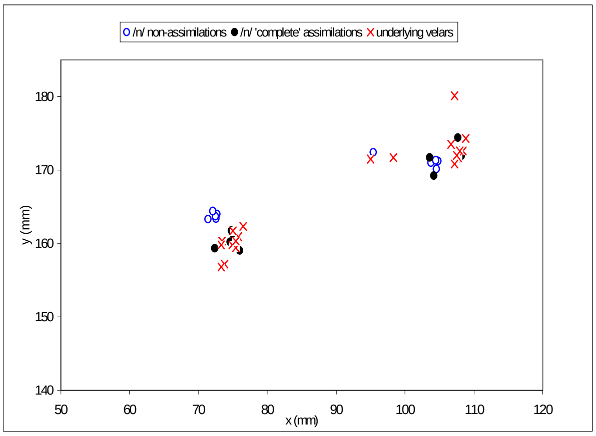

Figure 4 Subject d: articulatory positions (mm) for tongue tip (left cluster) and tongue dorsum (right cluster) at the moment of minimum tangential velocity for all non-assimilated alveolar tokens, assimilated alveolar tokens and all underlying /ŋ/ tokens

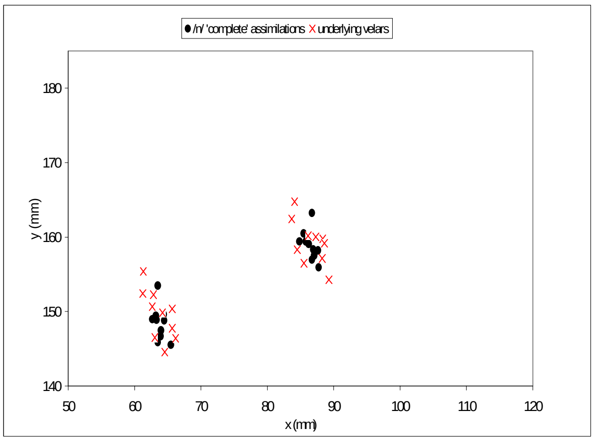

Figure 5 Subject h articulatory positions (mm) for tongue tip and tongue dorsum at the moment of minimum tangential velocity for all assimilated alveolar tokens and all underlying /ŋ/ tokens

3. CONCLUSIONS

There are two principal conclusions to be made from the results of these studies. The first concerns a possible speaker specific basis for assimilatory behaviour and the second addresses the success of the methodologies employed.

The specific contribution of this work to knowledge about assimilatory behaviour is not that speakers have a preferred assimilatory strategy but that these preferred strategies can be fundamentally different. For some speakers assimilation appears to have a mechanical basis, motivated by phonetic factors such as speech rate. For others optional assimilation of this kind seems to be governed by the application of a cognitive rule. We can surmise from this that speaker specific and language specific aspects of speech can override language universal, biomechanical and other constraints on assimilatory processes.

It is clear that EMA used in combination with EPG is a powerful and promising tool in the pursuit of answers to questions about assimilatory strategy. It provides complementary information on the midsagittal and the lateral plane and combines more fine-grained resolution in the anterior region of the vocal tract from EPG with more extensive information on velar movement from EMA. The addition of EMA data to EPG data collected in the pilot study reported above, however, did not result in a radical redefinition of Subject d and h's assimilatory strategy. For this reason the limitations of EPG data may not be as serious as previously assumed.

REFERENCES

[1] Spencer, A. 1996 *Phonology: Theory and Description* Oxford: Blackwell

[2] Hardcastle, W.J. 1994 Assimilation of alveolar stops and nasals in connected speech. In *Studies in General and English Phonetics in honour of Professor J.D. O'Connor* J. Windsor Lewis (editor). Routledge, 49-67

[3] Wright, S. and Kerswill, P. 1989 Electropalatography in the study of connected speech processes. *Clinical Linguistics and Phonetics* 3, 49-57

[4] Lindblom, B. 1963 Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-81

[5] Kühnert, B. 1993 Some kinematic aspects of alveolar to velar assimilations. *Forschungsberichte des Institut für Phonetic und Sprachliche Kommunikation der Univerität München* 31, 263-72