ARTICULATORY, ACOUSTIC AND PERCEPTUAL ASPECTS OF FRICATIVE/STOP COARTICULATION

Noël Nguyen†, *Alan Wrench*‡, *Fiona Gibbon*‡, *and William Hardcastle*‡
† Laboratory for Psycholinguistics, FPSE, Univ. Geneva, Switzerland
‡ Depmt of Speech & Language Sciences, Queen Margaret College, Edinburgh, UK

ABSTRACT

Stops are not identified in the same way depending on preceding fricatives. According to Mann and Repp [1], such context-dependent variations in the perception of stops originate from the influence of fricatives on how stops are produced. This study aimed further to explore this hypothesis. A first experiment showed that the effect of fricatives on the identification of stops tends to be confined to the most ambiguous stimuli, when a large range of acoustic cues to place of stop articulation is provided to the listener. In a second experiment, articulatory and acoustic data were gathered in the production of fricative-stop sequences. Although on the whole consistent with previous findings, our results indicate that many articulatory dimensions are brought into play in fricative-stop coarticulatory patterns, thus making it difficult to establish a direct link between the articulatory and perceptual levels.

1 INTRODUCTION

This paper deals with the influence of a preceding fricative on the production and perception of a stop consonant. Mann and Repp [1, 2] have shown that, when preceded by a fricative, an ambiguous stop acoustically halfway between /d/ and /g/ is identified differently depending on the place of articulation of the fricative. Specifically, listeners tend to identify stops more frequently as velars following /s/ than following /ʃ/. This effect was found to occur regardless of the presence/absence of a syllable boundary between the two consonants. According to Mann and Repp, such fricative-dependent variations in the perceived place of articulation of a stop demonstrate that some tacit knowledge of articulatory dynamics is involved in the perception of speech sounds. Mann and Repp hypothesized that fricatives have an influence on how a following stop is produced and, more particularly, that the place of articulation of a velar stop is shifted forward in the context of /s/. As the perceptual effect observed goes, as it were, in the opposite direction (more velar responses following /s/ than following /ʃ/), this would indicate that compensatory adjustments are made in the identification of a stop depending on the adjacent fricative.

Two experiments will be reported. Experiment 1 aimed to replicate Mann and Repp's main finding, namely that stop consonants acoustically midway between /d/ and /g/ are perceived more frequently as velars following /s/ than following /ʃ/, and to determine whether this context effect extends to synthetic stimuli containing a greater variety of acoustic cues to place of stop articulation than those used in [1]. The goal of Experiment 2 was to assess the hypothesis that velar stops are produced differently depending on the preceding fricative, by examining coarticulatory patterns in the production of fricative-stop sequences both at the articulatory and acoustic levels.

2 PERCEPTION EXPERIMENT

In this experiment, we asked to what extent the influence of a fricative on the perception of the following stop depends on the range of acoustic cues to place of stop articulation provided to the listener. In [1], the influence of the fricative was found to be quite pervasive, extending beyond the most ambiguous stop stimuli up to the velar endpoint. However, the synthetic stimuli used in [1] were periodic sounds that differed in the onset frequencies of F_2 and F_3 only. One issue of interest to us was whether the perceived place of articulation of stops would vary to the same extent as in [1] depending on the preceding fricative, when the alveolar-velar continuum is synthesized by manipulating a larger set of perceptually-relevant acoustic parameters.

2.1 Method

Stimuli. Each stimulus was composed of two monosyllables. The first monosyllable was either the word "plus" or the word "dash", thus ending either in /s/ or /ʃ/. The second monosyllable was a CVC synthetic sequence with the initial consonant taking place on a /d/-/g/ continuum, and the vowel being either /ɑ/ or /i/.

These stimuli were constructed as follows. First, a male native speaker of Southern British English was asked to produce four target words, "dark", "guard", "deed", and "geese", along with the two context words, "plus" and "dash", a number of times. The utterances were recorded on a DAT tape, low-pass filtered, and digitized (sample rate: 10 kHz) on a Unix workstation.

The target words were used as models to synthesize four CVC stimuli with the same phonetic structure, by means of the Klatt formant synthesizer [3]. The Klatt parameter values were estimated automatically from the acoustic signal in a first step, then adjusted interactively until the similarity between the synthesized signal and the natural one was judged to be satisfactory. The synthesizer output sampling rate was set to 10 kHz.

From each of the four initial CVC sequences, a 7-step continuum was generated, with the initial consonant ranging from /d/ (Stim. 1) to /g/ (Stim. 7). For each continuum, the parameters manipulated were the onset frequencies of F_2, F_3 (for both /ɑ/ and /i/) and F_4 (for /i/), the duration of the transition for F_1 (shorter at the alveolar endpoint), and the duration of the interval between the burst onset and the beginning of the vowel (shorter at the alveolar endpoint). The duration of the formant transition had a fixed value of 40 ms for F_2 and the upper formants.

Following Stevens and Blumstein [4], the burst's main spectral peaks were considered to be continuous with formants at the onset of voicing in the following vowel. For alveolars, the noise burst was produced by exciting a resonator continuous with F_4 at the onset of /ɑ/, and with F_4/F_5 at the onset of /i/. For velars, the burst main peak was continuous with either F_2 (/ɑ/) or F_3 (/i/). For intermediate stimuli on each /d/-/g/ continuum, the noise burst had two spectral peaks, whose relative amplitudes were interpolated linearly (on a dB scale) between the amplitude values associated with the two end-points. The duration of the burst was set to 10 ms.

Subjects. Eight native speakers of English took part in the experiment, which was carried out in the Department of Linguistics of the University of Cambridge, UK.

Procedure. Disyllabic sequences were formed by combining "plus" or "dash" with each of the synthetic stimuli. The duration of the interval between the fricative and the onset of the following burst was set to 70 ms.

Subjects were asked to identify the first sound of each CVC sequence as /d/ or /g/, by mouse-clicking on a computer screen. Each stimulus was presented 8 times. The test itself was preceded by a dry run in which the subject was twice presented with the /d/ and /g/ endpoints for each continuum, and was given the correct response after having heard the stimulus.

2.2 Results

Average percentages of "d" responses are displayed in Figure 1 depending on the position of the stimuli on the /d/-/g/ continuum, the preceding fricative, and the following vowel.

As Figure 1 shows, the endpoint stimuli were clearly perceived as belonging to different phonemic categories. Stimuli 1 and 2 were categorized as alveolars in most cases, regardless of the phonetic context, whereas Stimuli 5, 6 and 7 were consistently categorized as velars. The categorical boundary between /d/ and /g/ fell approximately in the middle of each continuum, as expected, although some differences were observed depending on the following vowel.

Figure 1 also reveals that stops were categorized differently depending on the preceding fricative, being more

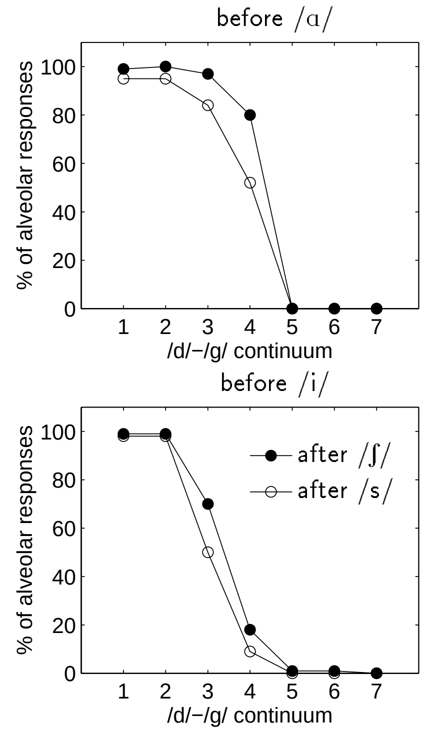

Figure 1: Percentages of "d" responses for each stimulus in each context.

frequently identified as alveolars following /ʃ/ than following /s/. The difference in the percentages of alveolar responses following /ʃ/ and /s/ was significant for Stimulus 4 in the context of /ɑ/ (diff.: 28.1%, $F(1.7) = 13.186$, $p < .01$), and for Stimulus 3 in the context of /i/ (diff.: 20.3%, $F(1,7) = 13.598$, $p < .01$).

Thus, the context effect found by Mann and Repp was basically replicated. However, this effect proved to be relatively small and confined to the most ambiguous stimuli. This may be partly due to our using a /d/-/g/ continuum with only seven steps, which means that the distance between two adjacent stimuli in the acoustic space was rather large. As a result, the subjects might have responded in a strongly categorical manner, by tending systematically to associate the same stimulus with the same phonetic label. The presence of a stop burst could also have contributed to further reducing that influence, by making the stop still less ambiguous. Note, however, that fricative-dependent variations in the identification of stops were quite systematic, occurring for all the subjects tested. The magnitude of the context effect did not show any significant variations depending on the following vowel ($F(1,7) = .612$, NS).

3 PRODUCTION EXPERIMENT

As indicated above, fricative-dependent variations in the identification of stops have been interpreted as a "compensation-for-coarticulation" effect. On this account, such variations would stem from the listener's attempting to compensate for shifts in place of stop articulation induced by a preceding fricative. It has been hypothesized

[1] that the place of articulation of a velar stop is further front following /s/ than in a neutral context (stop in utterance-initial position). At the acoustic level, this forward shift was expected to be reflected in a larger difference between F_2 and F_3 onset frequencies in the context of /s/ relative to a neutral context.

In an acoustic study involving eight speakers of American English, Repp and Mann [2] measured the formant onset frequencies following the stop release in utterances such as /stɑ/, /skɑ/, /ʃtɑ/, /ʃkɑ/. F_3 onset frequency was found to be higher following /s/ than following /ʃ/, as predicted. However, the fricative had no clear effect on the onset frequency of F_2, and the pattern observed for F_4 was at variance with what was expected. In the present study, this issue was addressed in a more direct manner, by looking into the articulatory movements involved in the production of fricative-stop sequences.

3.1 Method

Material. Our material was made up of six two-word sequences, "pup dump", "muss dump", "mush dump", "pup gum", "muss gum", and "mush gum", combining /p/, /s/, and /ʃ/ with /d/ or /g/ across a word boundary. The prevocalic consonant in the first word and the post-vocalic consonants in the second word were all bilabials, expected as such to have little effect on the position of the tongue in the production of the fricative+stop sequence. The vowel was identical in both words (/ʌ/) so as to put the F+S sequence in as symmetrical an environment as possible. The word "pup" was taken as a baseline context, the influence of fricatives on word-initial stops being related to this baseline. Each two-word sequence was read ten times within the carrier phrase "No, it's ... again", in response to a question displayed on a computer screen. (This set-up was intended to induce differences in stress pattern that will not be dealt with here.)

Speakers. One female speaker of British English and one female speaker of Australian English participated in the experiment. Both were trained phoneticians with an extensive experience of articulatory recordings.

Recordings. Articulatory data were gathered using electromagnetic articulography (EMA) and electropalatography (EPG). EMA coils were affixed to the lower and upper lips, the lower central incisors, and the tip, blade and dorsum of the tongue. Two further coils were attached to the upper central incisors and the bridge of the nose for a maxillary frame of reference. The coordinates of each coil relative to two orthogonal axes (x and y) were digitized at a sampling rate of 500 Hz, and were corrected for rotational misalignment as well as head movements. EPG data were recorded using the Reading EPG2 system, in which the artificial plate worn by the subject contains 62 electrodes arranged in eight horizontal rows. The acoustic signal was low-pass filtered and digitized at a sampling rate of 16 kHz.

Data extraction. For each utterance, the release of the word-initial stop consonant (SREL) and the onset of the following vowel (VON) were located in the speech signal from the waveform and a wide-band spectrogram. The EMA and EPG frames at each of these two time locations were then extracted for further processing.

The EMA data were analyzed as follows. First, the principal components of the movements of the tongue were computed for each coil separately, over the entire set of data for each speaker. The orientation of the first principal component was taken to be parallel to the dorsal-ventral movements of the tongue at the coil's location along the vocal-tract midline. The second principal component was considered to be parallel to the anterior-posterior displacements of the tongue at the same location. Second, the average position of the coils in the mid-sagittal plane was computed for each target word at SREL and VON in the baseline context. Finally, we computed the distance of each coil to its corresponding baseline position along the dorsal-ventral and anterior-posterior dimensions at SREL and VON for each utterance. This distance was used as a measure of the influence of the preceding fricative on the position of the tongue at the beginning of the target word.

3.2 Results

Articulatory data. Figure 2 shows the mean position of each tongue coil relative to its baseline position at SREL and VON, following /s/ (open arrows) and /ʃ/ (filled arrows), for Speaker 1. The data for /dʌ–/ are shown in the upper half, and those for /gʌ–/ in the bottom half. Anterior is to the left. An outline is shown of the hard palate. Each arrow reflects the average size and direction of the coil's displacements attributable to the preceding fricative at the specified time location. Arrows are displayed only for the coils whose position following the fricative was found to be statistically different from their baseline position in a oneway ANOVA (with $p < .05$), along the dorsal-ventral or the anterior-posterior dimension. The average EPG frames (% of contact for each electrode) following each fricative at SREL and VON are also presented (posterior edge of the palate at the top).

Variations in the position of the tongue relative to the baseline were observed for both the alveolar and the velar stop, in /s#–/ sequences as well as /ʃ#–/ sequences. For /dʌ–/ first, the tongue midline was on the whole higher in the context of a fricative than in that of /p/, this trend being stronger following /ʃ/ compared with /s/. Fricative-dependent differences in the location of the coils were also found, with the mid and back coils being further back both at the release of the stop and the onset of the vowel in the context of /ʃ/, the baseline being taken as reference. For /gʌ–/, the front coil was higher in fricative-stop sequences as opposed to /p/-stop sequences. In addition, both the front and mid coils were further back at both time locations following /ʃ/ than in the context of /p/. Differences in the configuration of the tongue-palate contacts depending on the preceding fricative are in good agreement with these general trends. The data for Speaker 2 were in general consistent with those for Speaker 1.

Acoustic data. F_2 onset frequency was measured using DFT/LPC spectra computed over a 20-ms interval starting

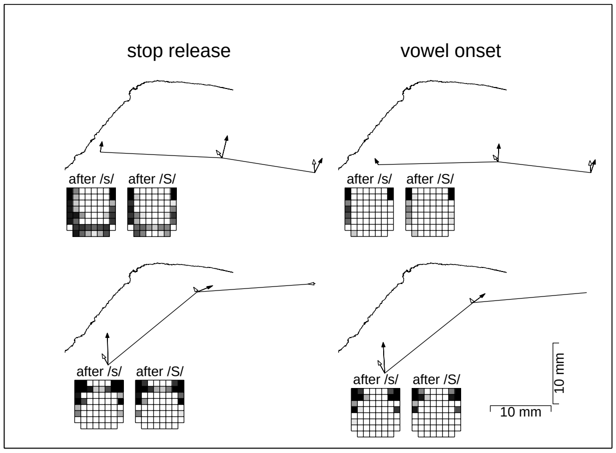

Figure 2: Influence of the preceding fricative on the position of the tongue at the release of the stop and the onset of the following vowel for Speaker 1.

at the onset of the vowel in each target word, together with a wide-band spectrogram. Average F_2 frequency values are given in Table 1, for each word type in each context, and for each speaker.

	/dʌ–/			/gʌ–/		
speaker	/s/	/ʃ/	/p/	/s/	/ʃ/	/p/
1	1714	1810	1766	1960	1824	1865
2	1738	1767	1747	1748	1720	1686

Table 1: mean F_2 onset frequency (Hz) for /dʌ–/ and /gʌ–/ in each context for each speaker.

Table 1 basically shows that F_2 onset frequency was on average higher than the baseline following /ʃ/ for /dʌ–/, and following /s/ for /gʌ–/. This latter trend was significant for both speakers ($F(1, 18) = 5.070, p < .05$ for Speaker 1, $F(1, 18) = 9.953, p < .01$ for Speaker 2).

4 DISCUSSION

Contrary to previous findings, fricative-dependent variations in the perception of place of stop articulation were in Experiment 1 restricted to the central, most ambiguous part of the /d/-/g/ continuum. Although further investigations are clearly needed on that issue, we have suggested that the size of this effect may depend upon how many cues to place of stop articulation are made available to the listener. Importantly, our synthetic CV sequences contained a stop burst, which probably made the stimuli at both ends of the /d/-/g/ continuum more easily identifiable than they would be if place of articulation was cued by the formant transitions only. Indeed, the spectral characteristics of the stop burst are known to play quite an major role in the identification of velar stops.

Complex fricative-stop coarticulatory interactions were observed in the articulatory study. On the whole, variations in how stops were produced were in agreement with the predictions outlined above, insofar as the place of stop articulation appeared to be further back following /ʃ/ than following /s/, for both /d/ and /g/. However, these coarticulatory patterns appear to bring many other articulatory dimensions into play. For example, there seems to be a difference in the overall shape of the tongue in the midsagittal plane at the release of the alveolar stop, depending on the preceding fricative, with the mid coil being higher in the context of /ʃ/ as opposed to /s/. Such differences might stem from the fact that the part of the tongue immediately behind the point of constriction is domed for /ʃ/ and hollowed for /s/ in English [5]. In addition, fricative-dependent differences in the arrangement of the linguo-palatal contacts are visible on the EPG patterns. Thus, there was on average more medial contact in the anterior part of the palate following /s/ than following /ʃ/, particularly at the release of the stop (Fig. 2). This may derive from the anterior part of the tongue forming a narrow mid-sagittal groove in the production of /s/ [5].

Systematic fricative-dependent variations in the formant pattern were found for /gʌ–/. The observed trend (higher F_2 onset frequency following /s/ as opposed to ʃ) is at odds with previous findings [2], but it is in our view consistent with the differences in place of articulation for /g/ mentioned above. These differences yet seem rather subtle, particularly in the vicinity of the point of constriction for /g/. Small front-back displacements of the point of constriction in that region may however have quite a significant influence on the spectral structure of the speech signal [6].

In conclusion, our articulatory data provide good evidence that word-initial stops are produced differently depending on the preceding fricative. These differences are not limited to place of articulation, however, and include more general articulatory characteristics relating to the sagittal shape of the tongue and to the configuration of the linguo-palatal contacts. The potential impact of these articulatory variations on the acoustic output still needs to be more accurately determined.

REFERENCES

[1] V.A. Mann and B.H. Repp. Influence of preceding fricative on stop consonant perception. *JASA*, 69:548–558, 1981.

[2] B.H. Repp and V.A. Mann. Fricative-stop coarticulation: Acoustic and perceptual evidence. *JASA*, 71:1562–1567, 1982.

[3] D.H. Klatt. Software for a cascade/parallel formant synthesizer. *JASA*, 67:971–995, 1980.

[4] K.N. Stevens and S.E. Blumstein. Invariant cues for place of articulation in stop consonants. *JASA*, 64:1358–1368, 1978.

[5] P. Ladefoged and I. Maddieson. *The Sounds of the World's Languages.* Blackwell, Oxford, UK, 1996.

[6] K.N. Stevens. On the quantal nature of speech. *J. Phon.*, 17:3–45, 1989.