QUANTITATIVE ANALYSIS OF SUBPHONEMIC FLAP/TAP VARIATION IN NAE

Donald Derrick[1] and Bryan Gick[1, 2]
[1]Dept. of Linguistics, University of British Columbia, Vancouver, 2613 West Mall, British Columbia, Canada V6T1Z4
[2]Haskins Laboratories, 300 George St., 9th Fl., New Haven, CT 16511, USA

1. INTRODUCTION

Extreme 'covert' categorical subphonemic variation has been thought to occur only in rare cases such as American English 'r' [1]. The present study demonstrates that English flaps/taps are produced using up to four distinct kinematic variations: up-flaps, down-flaps, alveolar taps and postalveolar taps. Surface distinctions between up-flaps and down-flaps, and between alveolar taps and post-alveolar taps, have not been previously described for any language. Our research expands on preliminary research by Gick [2,3] to include B/M mode ultrasound measures that capture details of flap kinematics with higher temporal resolution.

Based on our pilot work, we expect that in words with one flap, speakers will produce categorically distinct kinematic alternations primarily based on resolution of articulatory conflict. Articulatory conflict in flaps/taps arises based on the tongue positions required for surrounding vocalic sounds. Vowels are produced with the tongue tip below the alveolar ridge, while vocalic 'r's are produced with the tongue tip above the alveolar ridge.

Therefore, when a flap is preceded by a vocalic 'r' and followed by a vowel, as in the word 'Berta', we expect the flap to be produced by the tip of the tongue coming from above the alveolar ridge, hitting the ridge and continuing down. That is, we expect a down-flap.

Similarly, when a flap is preceded by a vowel and followed by a vocalic `r', as in the word 'otter', we expect the flap to be produced by the tip of the tongue coming from below the alveolar ridge, hitting the ridge and continuing up. That is, we expect an up-flap.

When a flap is preceded and followed by a vowel, as in the word `autumn', we expect the tap to be produced by the tip of the tongue coming from below the alveolar ridge, hitting low on the ridge and returning. That is, we expect an alveolar tap, like in Spanish [4].

When a flap is preceded and followed by a vocalic `r', as in the word 'murder', we expect the tap to be produced by the tip of the tongue coming from above the alveolar ridge, hitting above the ridge and return. That is, we expect a post-alveolar tap.

If speakers do not produce taps, we expect speakers to produce flaps favoring a suitable tongue position for the end of the word, so we expect down-flaps for `autumn' and up-flaps for 'murder'.

2. METHOD

2.1 Participants

Twenty-four participants were recorded, 12 female and 12 male. The present paper reports results from the first 4 of these.

2.2 Stimuli

38 unique stimuli were recorded in twelve randomized blocks. These were presented in carrier phrases designed to contain labial and glottal consonants only (except at phrase end) and induce stress on the first syllable of the stimuli phrases. Of these 38, 10 phrases contained a single flap. The stimuli with vowel-flap-vowel context included `autumn', 'edit the', 'audit the', 'edify', 'audify', 'vomit a' and 'acerbity'. The stimulus with vocalic `r'-flap-vocalic `r` context included 'murder'. The stimulus with a vocalic 'r'-flap-vowel sequence was 'Berta'. The stimulus with a vowel- flap-vocalic 'r' sequence was 'otter'.

2.3 Experiment procedure

Participants were seated comfortably in an American Optical Co. (1953) opthalmic chair with a twin-cup headrest. A 180^{o} EV tranducer, attached to an Aloka ProSound SSD-5000 ultrasound machine, was placed under the chin of the participant. A Sennheiser MKH-416 short shotgun microphone attached to an M-Audio DMP3 via XLR cable was used to record audio. The ultrasound and audio were channeled into an ADVC110 Canopus A/D video converter and recorded on a MacPro using iMovie HD (2006).

Ultrasound data were recorded in B/M mode. B-mode was set to record the tongue along the mid-sagittal plane. Three parallel M-mode lines were set following the line of the anterior palate, so as to maximally intersect movement of the tongue tip through the alveolar region (see Figure 1).

Stimuli was presented on an LCD monitor using PXLab [5] in two groups of six blocks of 38 stimuli each, with a break between the two groups.

2.4 Analysis procedure

The ultrasound audio and video were separated to allow for accurate acoustic boundary identification. The audio recordings were labeled and segmented using PRAAT [6]. Audio segmenting was imported in ELAN [7] and the type of tap/flap was labeled and transcribed using both the audio and ultrasound video recordings. Results were exported back to PRAAT textgrids and extracted using PERL scripts into statistics tables used in JMP 5.1 [8].

3. RESULTS

3.1 Flap-tap Identification

Using the M-mode information allowed for a relatively straightforward identification and distinction of the four kinematic variations of taps and flaps. The flaps are the easiest to identify. In down-flaps, a fuzzy white line, indicating the tongue surface trajectory, in the M-mode ultrasound image moves downward, as seen in Fig. 1a and highlighted by the red, green and blue arrows.

Similarly, in up-flaps the tongue-surface trajectory moves upward, as seen in Fig. 1b.
As positions on the hard palate are rarely identifiable in ultrasound images, the two taps must be identified relative to each other. In alveolar taps the tongue-surface trajectory is flat and lower than the top of a down-flap, as seen in Fig. 1c. In post-alveolar taps the tongue-surface trajectory is flat and higher than the trajectory of the alveolar tap, as seen in Fig. 1d.

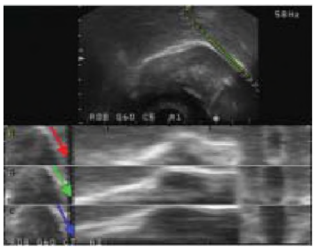

Fig. 1a. 'Berta' – down-flap

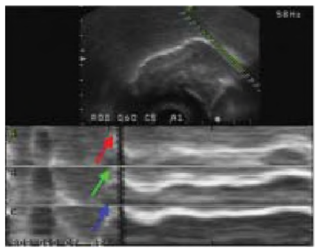

Fig. 1b. 'otter' – up-flap

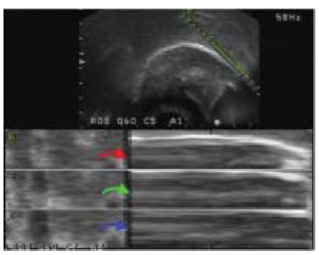

Fig. 1c. 'autumn' – alveolar tap

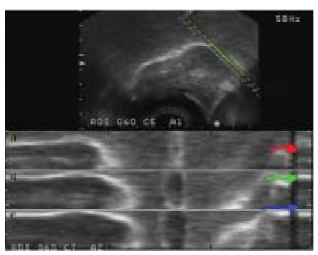

Fig. 1d. 'murder' – PA tap

3.2 Kinematic variation frequency by token

The results show that there is a highly significant relationship between the categorical kinematic variation of the tap or flap produced and the vocalic context for all four subjects individually and as a group. The within-subject chi-square test likelihood ratio was a χ^2 of 732.4, $P < 0.001$, with an $R^2 = 81.9\%$.

Context	Dn-flap	Hi-tap	Low-tap	Up-flap	Exclude
RR	0	46	2	0	0
RV	47	0	0	0	1
VR	0	0	0	48	0
VV	1	0	246	18	23

Table 1. Kinematic variation frequency by token
(Excluded tokens = speech errors and dropped frames)

The most variation was found in the vowel-flap-vowel context, and occurred only with the tokens containing the words 'edit-the' and 'audit-the'. In all of `edit-the' and 'audit-the' tokens and for all of the participants, the body of the tongue raised higher in the mouth than in other vowel-vowel context tokens. For 18 of the tokens, all from the same subject, there was an up-flap instead of the more common low-tap.

4. DISCUSSION

The results of the experiment strongly support the hypothesis that there are four categorical kinematic variations of flaps and taps in English. All four of the participants demonstrated all four variations during the experiment. The results also support the hypothesis that in words with one flap, articulatory conflict resolution, constrained by vowel and vocalic 'r' context, has the largest effect on the kinematic variation. We expect there to be more variation as we measure more of our participants. In particular, based on previous work, we expect speech errors to influence variation. We also have participants whose data has not been measured yet who do not produce all four kinematic variants either at all or with as much frequency.

This research also contains multiple measures of words and phrases with two consecutive flaps, such as 'Saturday'. In these cases, we believe and have already observed that participants use preferred tap sequences that will interact with articulatory conflict, such that in words like 'auditor' and 'editor', the sequence is 'up-flap, post-alveolar tap' rather than 'alveolar-tap, up-flap', to avoid ending a sequence on an up-flap.

REFERENCES

[1] Westbury, J. R., Hashi, M., and Lindstrom, M. J. (1999). Differences among speakers in lingual articulation for American English `r'. *Speech Communication*, 26:203–226.
[2] Gick, B. (2002). Exercises in 'absurdity': Interactions between flapping and tongue position in American English with implications for speech motor planning. *Journal of the Acoustical Society of America*, 112(5),pt.2:2416.
[3] Gick, B. (2006). Errors and strategy shifts in speech production indicate multiple levels of representation. *Journal of the Acoustical Society of America*, 119(5),pt.2:3302.
[4] Ladefoged, P. and Maddieson, I. (1996). The Sounds of the World's Languages. Blackwell, Oxford
[5] Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online]. Version 2.1.11. Mannheim (Germany): University of Mannheim. [cited 2007 June 19]. Available HTTP: http://www.pxlab.de
[6] P. Boersma and D. Weenik, (2008). "Praat: Doing phonetics by computer (version 5.0.27)" [Computer program], [cited 2008 Jul 05], Available HTTP: http://www.praat.org/
[7] Max-Planck-Institute. (2008). ELAN Linguistic Annotator [online]. Version 3.5.0. Nijmegen (The Netherlands): Max-Plank-Institute [cited 2008 Jul 05]. Available HTTP: http://www.lat-mpi.eu/tools/elan/
[8] JMP, Version 5.1. (2003). SAS Institute Inc., Cary, NC, 1989-2003.

ACKNOWLEDGMENTS

Special thanks to Aislin Stott for labeling and transcribing audio recordings and for help in conducting experiments. This research was funded by NSERC.