AN ULTRASOUND STUDY OF CORONAL STOP DELETION IN PERSIAN

Reza Falahati and Jeff Mielke
Department of Linguistics, University of Ottawa, 70 Laurier Avenue East, Ontario, Canada, K1N 6N5, rfala042@uottawa.ca

1. INTRODUCTION

This study examines the production of word-final consonant clusters in Persian in order to provide an account of their optional acoustic simplification. Persian optionally allows the simplification of word-final coronal stops when preceded by obstruents or the homorganic nasal /n/. The words /ræft/ 'went', /duχt/ 'sew' and /qæsd/ 'intention', for example, are optionally simplified to [ræf], [duχ], and [qæs] in fast or casual speech. What is unclear is whether there is any coronal gesture left after simplification occurs. This type of issue was first discussed in the Articulatory Phonology literature. For example, Browman and Goldstein (1990) hypothesized that the alveolar closure gesture for the apparently deleted /t/ in *must be* [mʌsbi] in fluent/casual speech is still present but completely overlapped or hidden by the labial closure gesture. This means that the gesture for the alveolar /t/ (raising the tongue tip towards the alveolar ridge) is masked during the time that lips are closed or narrowed for labial /b/ and that is why the closure for /t/ is not audible. The current study aims to investigate this issue in Persian.

2. METHODS

Audio and real time ultrasound video recordings were made while subjects had a guided conversation with a native speaker of Persian. The subjects of the study were 10 Persian-speaking graduate students from Carleton University and the University of Ottawa, five male and five female. Three of these were excluded from the analysis due to poor imaging and some other technical issues. Subjects were seated in an optometry chair located in a sound booth in the Sound Patterns Laboratory. After subjects held water in their mouth (for palate imaging) and produced three clicks (for audio-video synchronization), they engaged in a casual conversation with the first author. Each conversation lasted 30-35 minutes. At the end of the conversation, the subjects were asked to read a short passage in a casual style.

Real time images of the tongue were collected using a Terason T3000 ultrasound system. Mid-sagittal ultrasound images of the tongue were recorded at a scan rate of 60 frames per second and a focal depth of 10 cm. The transducer probe was fixed under each participant's chin using a headset made for this purpose by Articulate Instruments. Audio was recorded in Audacity, using a Shure condenser microphone mounted close to the subject's mouth and a USB preamplifier.

All coronal stops and their preceding segments were marked in TextGrids, resulting in a total of 662 word-final coronal clusters for the conversational speech of the seven subjects.

Tokens of singleton /t/ in coda position (referred to here as real V[t]) were selected as a reference point for unreduced /t/, and control tokens with singleton /f q χ/ in coda position were selected as a reference point for the complete absence of /t/, because the complete deletion of /t/ in word-final consonant clusters involving non-coronals would result in singleton [f q χ] as the only remaining segment in the clusters. The underlying assumption here is that if gestural simplification is complete, then the gestures of the final segment of the resulting word should be similar to those of singletons. The non-simplified tokens are expected to have gestures similar to real V[t]. Finally, several palate images showing the alveolar ridge clearly were selected for each subject.

A python script was used to extract the ultrasound images corresponding to the speech intervals marked in the TextGrids. Due to the challenge of separating the coronal gestures belonging to two adjacent coronals from each other, only final coronal stops in a non-coronal environment were selected for further analysis. This resulted in the exclusion of about 20% of the captured cluster tokens. For each token, the frame which showed the highest position of the tongue tip was analyzed. Palatoglossatron (Baker, 2005) was used to trace the tongue tip and blade in each tongue image and alveolar ridge in each palate image.

3. RESULTS

Table 1 shows the rate of acoustically simplified C[t] for each subject (i.e., whether /t/ was inaudible in each recorded token). The seven subjects in the study showed different rates of [t]-simplification. On the two extremes, S1 simplified almost 87% of the C[t]-clusters whereas S7 simplified only 2%.

Table 1. Number of C[t] tokens and C[t] simplification rate, for each subject.

Subject	S1	S2	S3	S4	S5	S6	S7
Number of tokens	46	56	52	45	53	44	55
Simplified C[t] rate	.87	.70	.79	.38	.70	.86	.02

Table 2 shows the mean distance between the tongue and alveolar ridge for the two reference points (i.e., control and real V[t]). The control condition shows a range of 4.20-9.30 mm distance between the tongue and alveolar ridge whereas this comes to 0.40-1.80 for real V[t] condition. As the data related to real V[t] shows, small distances like 2mm in this table still show closure. The simplified C[t] and non-simplified C[t] can be compared with these two reference points, respectively. There is a clear difference for the tongue-alveolar ridge distance between non-simplified C[t] and simplified C[t] conditions. All the subjects show much less distance in the non-simplified C[t] condition than in the simplified C[t] condition. This could be expected if we assume that the coronal gesture is partially or fully deleted in the simplified C[t] condition. Except for S6, all the subjects show a positive correlation for the mean of tongue-alveolar ridge distance in control and simplified C[t] conditions. For example, the subjects who show the largest tongue-alveolar ridge distance in the control condition (i.e., S1, S4, and S7) also show the largest distance in the simplified C[t] condition. A similar correlation is apparent between non-simplified C[t] and real V[t] conditions. The distribution of different types of [t] shows that the simplified C[t] has the widest range of articulation.

Table 2. Mean and SD for tongue-alveolar ridge distance (mm) across different conditions.

	Subjects						
Conditions	S1	S2	S3	S4	S5	S6	S7
Control							
Mean	8.4	4.2	4.5	9.3	6.6	4.8	7.3
SD	3.4	2.6	1.7	3.8	3.7	1.8	2.7
Simplified C[t]							
Mean	6.1	5.4	5.2	10.8	5.4	7.9	7.7
SD	4.1	2.3	3.2	4.7	3.8	3.0	7.7
Non-simplified C[t]							
Mean	0.7	0.8	0.7	1.2	2.4	1.5	0.3
SD	0.2	0.4	0.4	0.9	1.3	1.3	0.2
Real V[t]							
Mean	0.4	0.6	0.9	1.6	1.8	1.1	0.7
SD	0.1	0.3	0.8	1.7	0.6	0.2	0.5

4. DISCUSSION AND CONCLUSIONS

In addition to variation for C[t]-simplification rate across subjects, the distance between the tongue and alveolar ridge is shown to have a wide range within and across subjects for this condition. This distance for most of the subjects started from 2 mm and extended to 17 mm as the maximum point for tongue tip raise in simplified [t] condition. This means that the tokens which are perceived as simplified clusters can be realized either with full coronal gesture or without it. This is in line with the predictions made by both Articulatory Phonology and perceptual accounts; however, perceptual accounts make further prediction regarding the simplified clusters. It is predicted that among C[t]-simplified clusters with 2-17 mm tongue tip-alveolar ridge distance, there are some tokens which are perceived as more [t]-like or less [t]-like than the others. This needs to be explored.

Despite the fact that there is substantial variation in the tongue-alveolar ridge distance, the general patterns of the simplified [t] look to be very similar across subjects. This suggests that variation and optional processes are not necessarily related to performance and that they behave in a systematic and predicable way (For more discussion see Kochetov and Pouplier 2008).

Additionally, the tongue-alveolar ridge distance shared between the simplified and non-simplified /t/ suggests that the cut-off point between categorical (i.e., non-simplified [t]) and gradient (i.e., simplified [t]) aspects of phonology is obscure. The fact that the border line between simplified and non-simplified [t] is very narrow could trigger the likelihood of their misperception. This also applies to the tokens within [t]-simplification category standing on either ends of a wide spectrum of tongue-alveolar ridge distance.

Zsiga (2000) stated that with the shortening of the gesture the chance of having that sound overlapped also gets increased. Perceived deletion of this type can lead to actual deletion. Ohala (1981) has emphasized that analyzing phonological processes in terms of articulatory gestures does not rule out a perceptual/acoustic component to these processes. Any change in gestures or their magnitude or timing can result in perceptual/acoustic changes.

REFERENCES

Baker, A. (2005). *Palatoglossatron 1.0*. University of Arizona, Tucson, Arizona. http://dingo.sbs.arizona.edu/~apilab/pdfs/pgman.pdf.

Browman, C. P., & Goldstein, L. (1990b). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physic* Press.

Kochetov, A. & Pouplier, M. (2008). Phonetic availability and grammatical knowledge: an articulatory study of Korean place assimilation. *Phonology,* 25, 399-431.

Ohala, J. (1981). The listener as a source of sound change. In Masek, C.S., R.A. Hendrik, M.F. Miller, eds., *CLS 17-2,* 178-203. Chicago: CLS.

Zsiga, E. C. (2000). Phonetic alignment constraints: Conso-nant overlap and palatalization in English and Russian. *Journal of Phonetics* 28 (1): 69–102.

ACKNOWLEDGEMENTS

The authors wish to express their thanks to Marc Brunelle and Marie Hélène Côté for their valuable comments on the earlier versions of this paper. We also wish to thank the participants of BWTL 14 for their comments and feedback. We also thank all Persian speaking students at Carleton University and the University of Ottawa who participated in this study.