

TALKING WHILE CHEWING: SPEAKER RESPONSE TO NATURAL PERTURBATION OF SPEECH

Connor Mayer¹, Bryan Gick^{1,2}, and Elizabeth Ferch¹

¹Dept. of Linguistics, University of British Columbia, 2613 West Mall, Vancouver, BC, Canada, V6T 1Z4

²Haskins Laboratories, 300 George Street, Suite 900, New Haven, Connecticut, USA, 06511

1. INTRODUCTION

Previous studies of speech motor control have employed various types of mechanical perturbations to investigate the goals of speech production. While the types of perturbations used have varied in terms of being static or dynamic, all have been under external control. Studies of static perturbations have used artificial palates (McFarland, Baum, & Chabot 1996; Aasland, Baum, & McFarland 2006), bite blocks (McFarland & Baum 1995; McFarland, Baum, & Chabot 1996), and dental prostheses (Hamlet, Cullison, & Stone 1979; Jones & Munhall 2002); studies of dynamic perturbations have employed loads on the lower lip or jaw (Abbs & Gracco 1984; Munhall, Löfqvist, & Kelso 1994), horizontal force applied to the jaw (Tremblay, Shiller, & Ostry 2003), and changes in the thickness of an artificial palate (Honda, Fujino & Kaburagi 2002).

These studies have collectively found articulatory, acoustic, and perceptual effects on speech production that underscore both the spatial-motor and acoustic-auditory goals of speech. Tremblay et al. (2003) find that compensation to displacement of the jaw occurs despite the absence of acoustic or perceptual effects, suggesting that “a somatosensory goal is pursued independent of the acoustics” (868). Others, such as Honda et al. (2002) and Jones and Munhall (2002), find evidence that auditory feedback is important, suggesting an acoustic goal; this is consistent with clinical findings and studies of properties such as pitch and vowel quality (Jones & Munhall 2002).

The present study examines the acoustic and articulatory effects of chewing during speech. Chewing, viewed as articulatory perturbation, is a variety of dynamic perturbation that is different from previous approaches in several important ways: It is naturalistic, experienced every day by speakers; it is under speakers’ control; it changes over time, requiring constant readjustment; and it interferes with both the movement of the articulators and the shape of resonating cavities in the mouth.

We propose that speakers in control of this type of highly complex articulatory perturbation during speech will show evidence of optimizing to maintain acoustic-auditory speech goals. Specifically, we expect acoustic distinctions between sibilants to be maintained even when perturbation forces significant articulatory differences in tongue shape.

2. METHODS

Subjects were seated in a modified dentist's chair. An Aloka Prosound SSD-5000 ultrasound machine with a 180 degree probe was used to record midsagittal images of the tongue. Profile video was taken using a Sony Mini-DV Handicam. Subjects wore a pair of sunglasses; two sticks covered in blue construction paper, each with two pink dots affixed, were attached to the sunglasses and to the probe. These dots showed the position of the head relative to the probe, allowing for correction of head movement once tongue shapes had been traced (Mielke, Baker, Archangeli & Racy 2005). The chromakey feature of a Videonics MXPro DV video mixer was used to superimpose the subject's face and the dot positions over the ultrasound video. Audio was recorded using a Sennheiser MK66 short shotgun microphone. The video and audio signals were fed through a Canopus ADVC-110 advanced digital video recorder into a MacPro computer, where the combined video and audio were captured using iMovie.

The stimuli consisted of the carrier phrase "I'm a ___", followed by one of three words containing the phonemes of interest: 'saw', 'shaw', or 'raw'. For the present paper, only the sibilants /s/ and /ʃ/ are considered for comparison. There were two conditions: an experimental condition, in which speakers produced the stimuli while chewing a large bolus consisting of four pieces of Wrigley's™ spearmint gum, and a no-gum control condition. Stimuli were presented in four blocks for each condition, with each block containing four repetitions of each word, resulting in 32 tokens for each word (16 in each condition). These stimuli were produced by 7 native speakers of English. The boundaries and midpoints of /s/ and /ʃ/ were marked using Praat. Acoustic centre of gravity (COG) measurements were made with a 30ms window around the midpoints. Still frames were extracted from the midpoints using ELAN. Palatoglossatron was used to trace tongue shapes and align the tracings.

3. RESULTS

S2, S3, and S4 displayed no significant differences in COG between the two conditions, while S1, S5, and S6 did display significant differences. The remaining subject, S7, displayed significant differences in /s/ but not /ʃ/. These results are summarized in Table 1. Of the three subjects who had significant differences between the two conditions, S6

had COG frequencies that were significantly higher in the with-gum condition, while those of S1 and S5 were significantly lower. Although absolute acoustic targets were compromised by the presence of gum for these three subjects, the relative distances between /s/ and /□/ were maintained. S7 displayed lower COG frequencies for both conditions, but although a distinction was still preserved between /s/ and /□/, relative acoustic distances were not as closely maintained. The relative distances between /s/ and /□/ for subjects with significantly different COG frequencies are shown in table 2. A 2-way factorial ANOVA found no significant interaction effects in any subject.

Table 1. Average COG for bolus (B) and no bolus (NB) conditions with p values. Significant results are bolded.

	/s/ NB	/s/ B	/s/ p	/□/ NB	/□/ B	/□/ p
S1	8863 Hz	8138 Hz	< 0.01	6387 Hz	5717 Hz	0.04
S2	7088 Hz	6763 Hz	0.42	4889 Hz	4920 Hz	0.9
S3	8422 Hz	8160 Hz	0.32	4758 Hz	4628 Hz	0.49
S4	7825 Hz	7546 Hz	0.14	6787 Hz	6643 Hz	0.62
S5	8242 Hz	7643 Hz	0.03	5895 Hz	5454 Hz	0.01
S6	6984 Hz	7731 Hz	< 0.01	5352 Hz	5687 Hz	0.047
S7	7106 Hz	6308 Hz	0.02	4581 Hz	4404 Hz	0.43

SS ANOVA tests were performed on the ultrasound tongue tracings (Davidson 2006). For /s/ and /□/ for all subjects, interaction plots show significant interaction effects with BCI for the no bolus condition. Only S3 /□/ and S4 /□/ had significant interaction effects with BCI for the bolus condition. This is due to increased variability of tongue shape in the bolus condition across subjects. The location of the differences varies somewhat across speakers, indicating different strategies for handling the gum during speech, but tends to be primarily in the blade and root of the tongue.

Table 2. Average COG frequency ratios of /s/ to /sh/.

	NB Condition	B Condition
S1	1.39	1.42
S5	1.40	1.40
S6	1.31	1.36
S7	1.43	1.55

4. DISCUSSION

For all subjects in our study, the presence of a large gum bolus in the mouth interfered with the shapes and movements of the articulators during speech. Despite this perturbation, all subjects maintained the relative acoustic distance between /s/ and /□/ in terms of centre of gravity. In the cases where a speaker's articulatory strategies were not sufficient to maintain normal acoustic targets, relative acoustic targets were still met. S7 provides an interesting case, as her strategy for producing /□/ was sufficient to maintain acoustic targets while her strategy for /s/ was not.

This is not altogether surprising, as the more anterior articulation of /s/ makes it more susceptible to interference from the gum, and adaptation more difficult. As well, /s/ may require more precise tongue control in general (Stone, Faber, Raphael, & Shawker 1992). That all speakers displayed significant differences in tongue shape between conditions strongly suggests that speakers were employing articulatory-acoustic tradeoffs (Guenther, Espy-Wilson, Boyce, Matthies, Zandipour, & Perkell 1999), adjusting their chewing strategies to optimize acoustic-auditory targets. These results show that speakers in control of their own articulatory perturbations – even ones that are highly complex and subject to constant change – adjust their strategies to maintain the acoustic goals of speech.

REFERENCES

- Aasland, W., Baum, S. & McFarland, D. (2006). Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation. *J. Acoust. Soc. Am.*, 119, 2372-2381.
- Abbs, J.H., & Gracco, V.L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. *J. Neurophys.*, 51(4), 705-723.
- Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am.*, 120(1), 407-415.
- Guenther, F. H., Espy-Wilson, C.Y., Boyce, S.E., Matthies, M.L., Zandipour, M. & Perkell, J.S.. (1999). Articulatory tradeoffs reduce acoustic variability during American English /t/ production. *J. Acoust. Soc. Am.*, 105, 2854-65.
- Hamlet, S.L., Cullison, B.L. and Stone, M.L. (1979). Physiological control of sibilant duration: Insights afforded by speech compensation to dental prostheses. *J. Acoust. Soc. Am.*, 65, 1276-1285.
- Honda, M., Fujino, A. and Kaburagi, T. (2002). Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics*, 30, 281-302.
- Jones, J. and Munhall, K. (2002). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *J. Acoust. Soc. Am.*, 113, 532-543.
- McFarland, D. and Baum, S. (1995). Incomplete compensation to articulatory perturbation. *J. Acoust. Soc. Am.*, 97, 1865-1873.
- McFarland, D., Baum, S. and Chabot, C. (1996). Speech compensation to structural modifications of the oral cavity. *J. Acoust. Soc. Am.*, 100, 1093-1104.
- Mielke, J., Baker, A., Archangeli, D., & Racy, S. (2005). Palatron: a technique for aligning ultrasound images of the tongue and palate. *Coyote Papers*, 14, 147-161.
- Munhall, K.G., Löfqvist, A. and Kelso, J.A.S. (1994). Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip. *J. Acoust. Soc. Am.*, 95, 3605-3616.
- Stone, M., Faber, A., Raphael, L.J. & Shawker, T.H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [□], and [l]. *Journal of Phonetics*, 20, 253-270.
- Tremblay, S., Shiller, D. M. & Ostry, D. (2003). Somatosensory basis of speech production. *Nature*, 423, 866-869.

ACKNOWLEDGEMENTS

Thanks to Donald Derrick, Amanda Miller, Cathy Lin, Denise Tom, and the members of the UBC ISRL.