Consonant Context Effects on Vowel Sensorimotor Adaptation

Jeff Berry[1], *John Jaeger IV*[2], *Melissa Wiedenhoeft*[1], *Brittany Bernal*[1], *and Michael T. Johnson*[2]

[1] Marquette University Speech Pathology & Audiology, Milwaukee, WI, USA
[2] Marquette University Electrical and Computer Engineering, Milwaukee, WI, USA

`jeffrey.berry@marquette.edu`

Abstract

Speech sensorimotor adaptation is the short-term learning of modified articulator movements evoked through sensory-feedback perturbations. A common experimental method manipulates acoustic parameters, such as formant frequencies, using real time resynthesis of the participant's speech to perturb auditory feedback. While some studies have examined phrases comprised of vowels, diphthongs, and semivowels, the bulk of research on auditory feedback-driven sensorimotor adaptation has focused on vowels in neutral contexts (/hVd/). The current study investigates coarticulatory influences of adjacent consonants on sensorimotor adaptation. The purpose is to evaluate differences in the adaptation effects for vowels in consonant environments that vary by place and manner of articulation. In particular, we addressed the hypothesis that contexts with greater intra-articulator coarticulation and more static articulatory postures (alveolars and fricatives) offer greater resistance to vowel adaptation than contexts with primarily inter-articulator coarticulation and more dynamic articulatory patterns (bilabials and stops). Participants completed formant perturbation-driven vowel adaptation experiments for varying CVCs. Results from discrete formant measures at the vowel midpoint were generally consistent with the hypothesis. Analyses of more complete formant trajectories suggest that adaptation can also (or alternatively) influence formant onsets, offsets, and transitions, resulting in complex formant pattern changes that may reflect modifications to consonant articulation.

Index Terms: sensorimotor adaptation, coarticulation, auditory feedback.

1. Introduction

The speech motor control system affords great precision and coordination in the execution of running speech. This is accomplished through integration of internal models for speech production and peripheral, sensory feedback. Notably, when the sensory feedback information does not match the predicted speech target, the error is detected and minimized in subsequent productions by modifying the feedforward command for that target [1]. Over time, the compensatory behavior is replaced with a persistent, learned motor behavior. This phenomenon, referred to as *sensorimotor adaptation,* has been observed during studies of upper limb movement in which visual feedback is perturbed within "virtual reality" environments [2, 3]. Results suggest that repeated adaptation appears to change the brain's prediction of how limb movements will be carried out and effectively re-calibrates existing patterns for motor behavior. Similar findings have been documented in studies of gait [4-6] and hemineglect [7]. Such results have substantial implications for rehabilitation. Specifically, such work supports novel approaches to treatment aimed at improving motor control through adaptation-based therapies that exploit involuntary sensorimotor learning mechanisms [8].

A prominent question of interest in speech motor control is concerned with whether motor "re-learning" can be extended to articulatory movements. Experimental evidence suggests that speakers compensate for altered auditory feedback [9-15]. A common experimental method manipulates acoustic parameters, such as formant frequencies, using real time resynthesis of the participant's speech. This auditory perturbation induces a compensatory motor response, measured by an opposing shift in the formant frequency from baseline. When modified articulator movements persist under a condition of masking noise, the speaker demonstrates adaptation. Speech sensorimotor adaptation has been demonstrated in sustained vowels, "neutral" contexts (/hVd/), and time-varying acoustic patterns [15]. Additionally, persistent articulatory changes have also been documented in untrained (generalization) contexts.

One limitation of existing studies of speech sensorimotor adaptation is that little is known about articulatory compensation and adaptation at levels of linguistic complexity appropriate for clinical application. If sensorimotor adaptation is to be utilized for speech rehabilitation, we must have an understanding of adaptation in consonant-vowel sequences that occur in natural speech. The aim of the current research is to investigate the coarticulatory influence of consonants in the processes of compensation and adaptation for perturbed vowels, as well as generalization to untrained vowels.

The current work compares vowel adaptation across consonant environments that vary in coarticulatory demands. *Coarticulation* refers to the fact that a phonological segment is not realized identically across phonetic environments but often becomes more like an adjacent or nearby segment during running speech [16]. Coarticulation occurs in both spatial and temporal domains of speech. Coarticulatory effects are especially large during intra-articulator coarticulation in which the tongue is subject to competing demands for production of the vowel as well as adjacent consonants. We hypothesize that these "coupled" contexts may interfere with vowel adaptation. During inter-articulatory coarticulation, competing demands upon the tongue are reduced, since consonant production is primarily controlled by other articulators, such as the lips. We hypothesize that these "de-coupled" contexts allow greater opportunity for adaptation to occur. Furthermore, we extend this hypothesis to address the temporal constraints during running speech. Specifically, we hypothesize that longer, more continuous, static consonants, such as fricatives, may be less facilitatory to vowel adaptation than shorter, non-continuous, dynamic consonants, such as stops.

Interactions between the spatial and temporal domains of coarticulation upon vowel adaptation are likely. Thus, we

Copyright © 2014 ISCA

hypothesize that intra-articulatory contexts with more continuous manners or articulation (i.e., lingual fricatives) will show the greatest interference with vowel adaptation, while inter-articulatory contexts with non-continuous manners of articulation (i.e., bilabial stops) will show the least interference, offering the most facilitatory context for vowel adaptation.

2. Methods

Each experimental run in this study required a participant to repeat words or near-words with varying consonant and vowel contexts. Each participant wore a headset microphone and audiology-grade insert earphones. As the participant spoke, auditory feedback was systematically manipulated using the Audapt LPC resynthesis software [1] to shift the 1st and 2nd formant center frequencies (F1 & F2) with the goal of eliciting involuntary changes in vowel articulation (see Figure 1). Participants in the study were 4 typically-functioning (neurologically healthy) American English talkers (1 male) with no history of speech, language, or hearing pathology. Participants ranged in age from 20-22 years and had upper-Midwestern dialects.

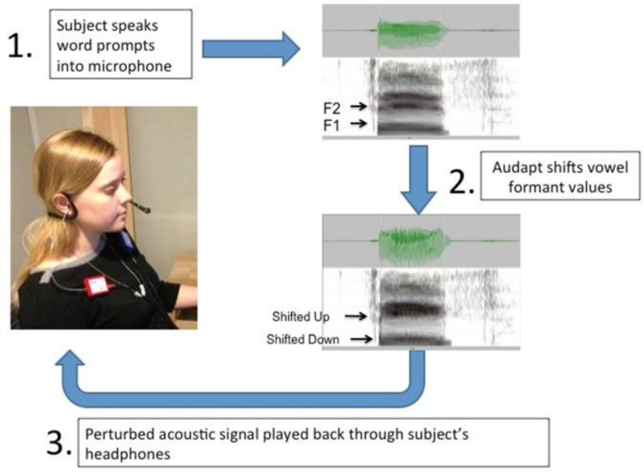

Figure 1: *Schematic of experimental setup.*

The vowel /e/ was used for "training" in the study. Thus, participants only heard formant perturbations to utterances containing /e/. The vowel /o/ was used to assess generalization. Thus, participants never heard formant perturbations during utterances containing the vowel /o/. Figure 2 schematizes the perturbations and expected consequences within a vowel quadrilateral. Axes indicate the orientation of (approximate) acoustic and articulatory relations. Specifically, increases in F2 are presumed to roughly correspond with increases in tongue advancement and decreases in F1 are presumed to roughly correspond to increases in tongue height. Utterances containing the training target /e/ were perturbed to create the perception of the vowel /i/ at full perturbation (F1 was decreased and F2 was increased). The magnitude of the perturbations varied by participant and was determined by the formant shift required to mostly closely transform a participant's habitual /e/ production into her habitual /i/, based on values obtained during a brief screening. Across participants, F1 shifts ranged from 175-200 Hz, and F2 shifts ranged from 200-400 Hz. In response to these perturbations, it was hypothesized that each participant's productions would demonstrate compensatory acoustic changes (increased F1 and decreased F2 relative to baseline) corresponding to a lower and more posterior tongue position for the target vowel (closer to the vowel /ae/). Utterances with the vowel /o/ were not acoustically perturbed and were used to determine if compensatory articulatory changes generalized to "untrained" vowel contexts. The words or near-words spoken in this study were constructed from symmetric CVC forms and included voiceless stops (/p/ or /t/) or fricatives (/f/ or /s/) flanking the training and generalization vowels. Voiceless consonants were chosen to avoid unintended triggering of acoustic changes in consonants during LPC resynthesis. The resulting stimuli set included 8 CVCs with contrasting consonant manner and place of articulation that allowed assessment of the spatial and temporal manifestation of coarticulation on vowel adaptation.

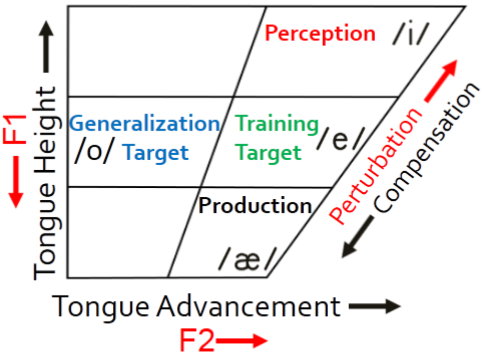

Figure 2: *Schematic of experimental manipulations.*

Each experimental run included multiple, continuous phases (see Figure 3). First, a baseline phase established habitual performance for both the training and generalization vowels. During this phase, participants repeated 80 CVCs (40 per vowel) and heard resynthesized productions without formant frequency changes. Data from the baseline phase served as the reference point for assessing compensation and adaptation. The training process was comprised of a ramp phase, in which formants were gradually shifted away from baseline values, and a full perturbation phase, during which formants were maximally shifted. Comparisons of performance between baseline and full perturbation were used to assess sensorimotor compensation. Following full perturbation, during the masking phase, speech-shaped white noise was used to eliminate auditory feedback. Participants produced CVCs with both training and generalization vowels during masking. Comparisons of performance between baseline and masking were used to assess sensorimotor adaptation and generalization. Finally, during the return phase, auditory feedback was returned to the baseline condition to facilitate de-adaptation.

To analyze the patterns of articulatory change throughout the different experimental phases, the TF32 software [17] was used to mark the boundaries of the vowels on the spectrogram based on the occurrence of the first and last glottal pulses associated with the vowel. TF32 was also used to generate pitch-synchronous LPC tracings (26 coefficients) of the time-varying F1 & F2. Apparent LPC tracking errors were manually corrected to define complete, time-varying patterns of change for F1 & F2. Formant center frequency values were extracted at the temporal midpoint of the vowel for all participants. For a subset of data, complete time-varying formant patterns were extracted for analysis.

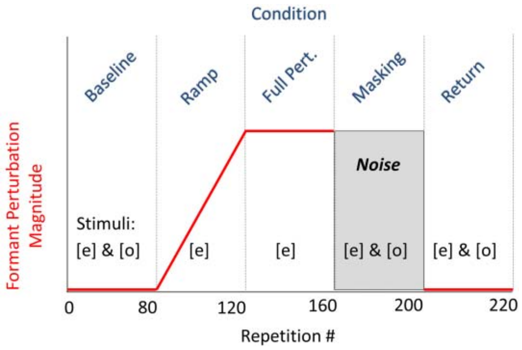

Figure 3: *Experimental phases.*

3. Results

Initial analysis focused on the discrete F1 & F2 values at the temporal midpoint of the vowel. Figure 4 shows example data from one talker and context. Formant axes are oriented to roughly approximate the articulatory working space of talker in right-facing profile. Thus, increases in F2, along the x-axis correspond roughly with forward movement of the tongue and decreases in F1, along the (inverted) y-axis correspond roughly with increases in tongue height. 20 replicates of the training vowel /e/ are shown from each condition (except "return" during which only 10 replicates occurred). Baseline /e/ productions are indicated by black circles. Compensation is reflected in the shift from baseline to full perturbation, indicated by red triangles. Adaptation is reflected in shift from baseline to masking, indicated by green squares. De-adaptation is reflected by the return condition, indicated by yellow triangles.

Following Figure 4, this talker demonstrates the general compensation and adaptation effect hypothesized since acoustic perturbations that increased perceived F2 and decreased perceived F1 elicited changes in articulation that reduced F2 and increased F1. Thus, compensatory (down and back) changes in tongue position are indicated by the change from baseline to full perturbation. Slightly lesser magnitude adaptation of tongue position (also down and back) are indicated by the change from baseline to masking.

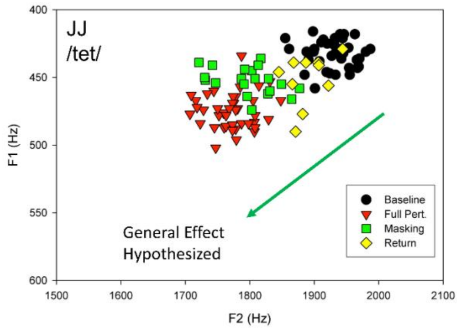

Figure 4: *Example data from one participant & context.*

Since adaptation changes reflect short-term sensorimotor learning, and have the most relevance to speech rehabilitation, the remainder of this work will focus on articulatory changes between baseline and masking phases. Figure 5 summarizes the adaptation shifts exhibited by all participants. Significant effects of manner of articulation were obtained for all participants ($p<0.005$). Significant effects for place of articulation ("coupling") were obtained for 2 participants ($p<0.005$).

While significant effects of place and manner of articulation were observed for a majority of the participants, the magnitudes of adaptation effects did not strictly follow the hypothesized directions and were not necessarily consistent across F1 and F2. For example, the "stop-decoupled" condition (/pep/) was hypothesized to have the largest adaptation magnitudes. However, this was never strictly true. Specifically, results for subject 2 were closest to this expectation, though F2 shifts for the stop-decoupled condition were indistinguishable from those for the fricative-decoupled condition (/fef/). Also, results for subject 1 showed the predicted effect (stop-decoupled maximum shift across contexts) for F2, but had maximum F1 shifts for the fricative-decoupled context. Subjects 3 and 4 deviated the most from expectations, with "coupled" (lingual consonant) contexts tending to show larger adaptation effects than decoupled contexts.

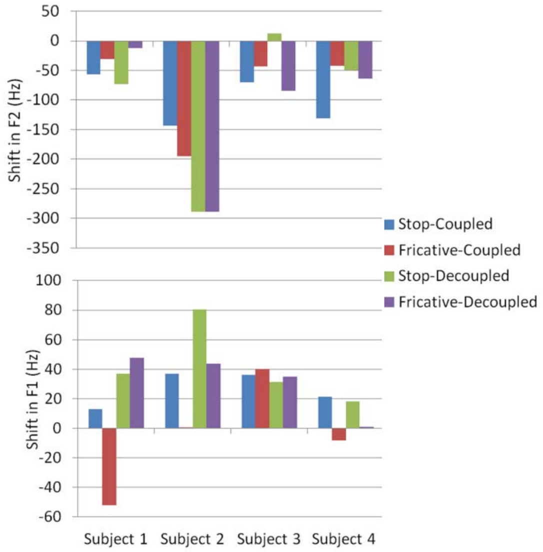

Figure 5: *Summary of adaptation formant shifts.*

Vowel midpoint measures of acoustic change provide general support for the hypothesis that consonant context effects influence vowel sensorimotor adaptation, but only limited support for the specific hypotheses that effect magnitudes should follow a systematic pattern of change based on place and manner of articulation. Qualitative appraisal of complete formant tracks suggest potentially complex changes in articulation that may not be characterized by discrete, temporal midpoint measures. Figure 6 shows an example of time-varying formant patterns obtained for one talker for baseline (black) and masking (green) phases for the near-words /fef/ (adaptation) and /fof/ (generalization).

The upper panel of Figure 6 shows F1 and F2 patterns for the training CVC /fef/ and the lower panel shows patterns for the generalization CVC /fof/. Baseline productions are in black and masking productions are in green. Since formant tracking was pitch synchronous, the x-axis of these plots is glottal cycle count. A noteworthy observation from these plots is that formant shifts may occur at the vowel temporal midpoint, as can be seen for /fef/. However, other effects, (on transitions for /fef/) and possible changes in the overall pattern of formant movement not affecting the temporal midpoint

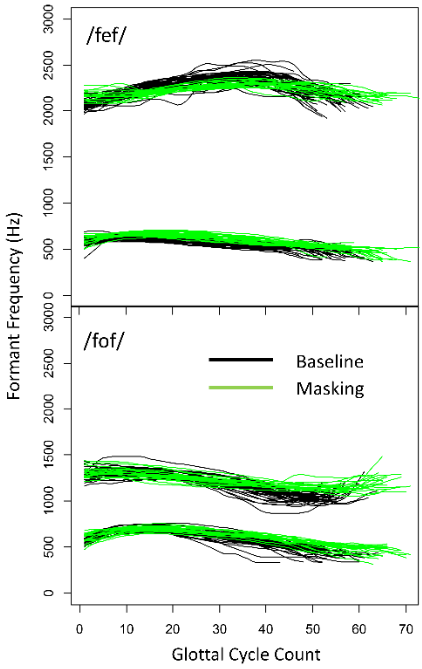

Figure 6: *F1 & F2 patterns for training and generalization.*

(note the F1 and F2 difference across conditions for the generalization context /fof/) would not be adequately characterized by a singular, discrete time measure.

To assess the possibility of more complex, formant pattern changes associated with adaptation effects in different consonant contexts, we time normalized formant patterns, calculated a mean pattern per formant and experimental phase, then calculated the zero-lag cross correlation (based on 5 proportionally-distributed points along each formant pattern). By comparing formant patterns of individual replicates to the mean pattern for each condition, we obtain a description of the likelihood of changes in formant pattern shape as an effect of sensorimotor adaptation.

Figure 7 shows a series of box and whisker plots characterizing zero-lag cross correlations for formant patterns from individual replicates compared to mean patterns for training CVCs. These plots essentially describe the variability in formant pattern within and across conditions. F1 and F2 pattern correlations were calculated separately. Comparisons are color coded as follows: 1) black boxes characterize pattern variation within the baseline condition; 2) red boxes characterize pattern variation within the full perturbation condition; 3) green boxes characterize pattern variation within the masking condition; 4) brown boxes characterize pattern variation within full perturbation compared to the baseline mean pattern (shape-change associated with compensation); and 5) blue boxes characterize pattern variation within masking compared to the baseline mean pattern (shape-change associated with adaptation). For each plot, the upper and lower bounds of the "box" are the 1st and 3rd quartiles and the "center bar" is the median. The "whiskers" indicate the highest and lowest values within 1.5 of the inter-quartile range. Formant pattern shape changes are indicated both by changes in the location of the box boundaries and center bar (implying "systematic" changes in the patterns) as well as the length of the whiskers (indicating increased pattern variability).

Clear context differences appear to be present in the extent to which formant pattern shape is affected in these data. Fricative contexts appear to elicit substantially larger changes in shape than stop contexts, as well as substantially greater changes in F2 shape compared to F1. In contrast, stop contexts show substantially smaller differences between formants, as well as examples (such as "compensation" for /pep/) of F1 patterns changing more than F2. While this analysis approach suggests that the shape of formant patterns may be affected in both compensation and adaptation, it is important to note that similar (though somewhat less) variability in pattern shape is often evident in "within condition" analyses. For example, the blue box and whisker plot for /fef/ suggests that adaptation elicits substantial change (and variability) in F2 pattern shape. Yet, the noteworthy, though somewhat smaller, green box plot suggests that F2 shape change and variability increase simply as a function of masking. Thus, some apparent "adaptation" effects on formant pattern change may reflect performance variation resulting from masking auditory feedback and are not completely related to sensorimotor learning.

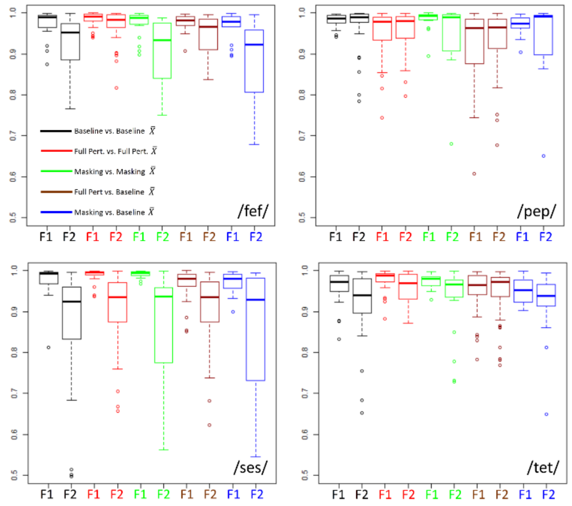

Figure 7: *Cross-correlation analysis of formant shape.*

4. Conclusions

Auditory feedback elicited speech sensorimotor adaptation is influenced by coarticulatory context. Context effects are complex and reflect the influences of adjacent consonant place and manner. The effects of vowel adaptation on formants cannot be adequately characterized by singular, discrete-time measures as changes in formant pattern shape are evident. These findings have important implications for the design of adaptation studies and methods of analysis. Moreover, the fact that formant transitions appear to be affected by vowel adaptations suggests that vowel-based adaptation paradigms may be used to affect consonant articulation. This conclusion has substantial implications for potential speech neurorehabilitation applications of sensorimotor adaptation, since individual vowel targets are rarely a primary or a sufficient focus in the treatment of motor speech disorders.

5. Acknowledgements

This paper is based upon work supported by American Speech-Language-Hearing Foundation New Century Scholar Grant and the National Science Foundation under Grant No. IIS-1142826 and IIS-1320892.

6. References

[1] Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. “Adaptive auditory feedback control of the production of formant trajectories in the mandarin triphthong /iau/ and its pattern of generalization.” *Journal of the Acoustical Society of America*, 128(4), 2033-2048, 2010.

[2] Suramanian, S.K., Massie, C.L., Malcolm, M.P., and Levin, M.F. “Does provision of extrinsic feedback result in improved motor learning in the upper limb poststroke? A systematic review of the evidence.” *Neurorehabilitation and Neural Repair*, 24(2), 113-124, 2010.

[3] Lehrer, N., Attygalle, S., Wolf, S.L., and Rikakis, T. “Exploring the bases for a mixed reality stroke rehabilitation system, Part 1: A unified approach for representing action, quantitative evaluation, and interactive feedback.” *Journal of NeuroEngineering and Rehabilitation*, 8, 51, 2011.

[4] Reisman, D. S., Wityk, R., Silver, K., & Amy, J. B. “Locomotor adaptation on a split-belt treadmill can improve walking symmetry post-stroke.” *Brain,* 130(7), 1861, 2007.

[5] Reisman, D. S., Bastian, A. J., & Morton, S. M. “Neurophysiologic and rehabilitation insights from the split-belt and other locomotor adaptation paradigms.” *Physical Therapy,* 90(2), 187-195, 2010.

[6] Yen, S. C., Schmit, B. D., Landry, J. M., Roth, H., & Wu, M. “Locomotor adaptation to resistance during treadmill training transfers to overground walking in human SCI.” *Experimental brain research*, 216(3), 473-482, 2012.

[7] Rossetti, Y., Rode G, Pisella L, et al. “Prism adaptation to a rightward optical deviation rehabilitates left hemispatial neglect.” *Nature*, 395, 166–169, 1998.

[8] Bastian, A. J. “Understanding sensorimotor adaptation and learning for rehabilitation.” *Current Opinion in Neurology*, 21(6), 628-633, 2008.

[9] Houde, J. F., & Jordan, M. I. “Sensorimotor adaptation in speech production.” *Science*, 279(5354), 1213-1216, 1998.

[10] Houde, J. F., & Jordan, M. I. “Sensorimotor adaptation of speech I: Compensation and adaptation.” *Journal of Speech, Language, and Hearing Research*, 45(2), 295-310, 2002.

[11] Purcell, D. W., & Munhall, K. G. “Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation.” *Journal of the Acoustical Society of America,* 120(2), 966-977, 2006.

[12] Tourville, J.A., Reilly, K.J., and Guenther, F.H. “Neural mechanisms underlying auditory feedback control of speech.” *NeuroImage*, 39(3), 1429-1443, 2008.

[13] Munhall, K.G., MacDonald, E., Byrne, S. & Johnsrude, I. “Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate.” *Journal of the Acoustical Society of America*, 125, 384-390, 2009.

[14] MacDonald E.N., Goldberg R., Munhall K.G. “Compensation in response to real-time formant perturbations of different magnitudes.” *Journal of the Acoustical Society of America*, 127(2), 1059-1068, 2010.

[15] Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. “Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing.” *The Journal of Neuroscience*, 31(45), 16483-16490, 2011.

[16] Hardcastle, W., & Hewlett, N .*Coarticulation.* Cambridge: Cambridge University Press, 1999.

[17] Milenkovic, Paul. "TF32 [Computer software]." Madison, WI: University of Wisconsin–Madison, 2004.