

Sam Tilsen*, Debarghya Das and Bruce McKee

Real-Time Articulatory Biofeedback with Electromagnetic Articulography

Abstract: This paper presents an articulatory biofeedback system and discusses new research methods made possible by this technology. The real-time electromagnetic articulography biofeedback system (RT-EMA) enables speakers to observe a visual representation of the movements of their speech articulators while they are speaking. Investigators can dynamically control the visual display of virtual targets or other objects in vocal tract space, track events involving interactions between virtual objects and articulators, and define custom actions in response to such events. Preliminary findings from experimental studies and games employing biofeedback are reported, with emphasis on the potential applications of articulatory biofeedback for investigating questions of linguistic interest.

Keywords: Phonetics, speech production, motor control, biofeedback, electromagnetic articulography

DOI 10.1515/lingvan-2014-1006

1 Introduction

Learned motor behaviors are fundamentally constrained by the nature of available sensory feedback. To learn and maintain normal speech production, auditory, tactile, and proprioceptive feedback are required, but not visual feedback. Because speech does not provide visual feedback regarding articulator positions, we do not know how such feedback could influence speech motor control and motor learning. The aim in developing a visual feedback system for speech is to open new areas of experimental investigation in articulatory phonetics and speech motor control, addressing questions regarding the cognitive representation of speech motor categories, movement parameters, and control over movement trajectories.

The technology presented here provides speakers real-time visual feedback of articulatory movements using electromagnetic articulography (EMA). The system can be considered a “biofeedback device” by virtue of providing sensory information that is not typically available in the context of an action. The system tracks the motion of sensors adhered to articulators such as the tongue tip, lower lip, upper lip, jaw, etc. and displays the positions of those sensors on a computer monitor. It allows for dynamic control over the visual display of targets or other objects in articulator space and tracks events involving interactions between articulators and those objects.

Electromagnetic articulography collects a time-series of 2D or 3D positions of sensors adhered to articulator fleshpoints (Berry 2011; Perkell et al. 1992; see the Appendix for detailed descriptions of EMA for linguists). In phonetic research EMA has been used to investigate relative timing of speech movements (Byrd 2000; Chitoran et al. 2002; Marin and Pouplier 2010; Shaw et al. 2011), coarticulation (Hoole and Nguyen 1999; Katz et al. 1991), effects of prosodic boundaries on articulation (Byrd 2000; Parrell et al. 2013), interspeaker articulatory entrainment (Tiede et al. 2012), rhythmic-articulatory coupling (Tilsen 2009), speech-halting behavior (Tilsen and Goldstein 2012), and oral-manual movement coupling (Parrell et al. 2013). EMA has also been used to infer synthesis parameters for audio or audio-visual models of speech (Beskow et al. 2003; Steiner et al. 2013) and to estimate control parameters in articulatory modeling and acoustic-articulatory inversion (Toutios et al. 2011). Clinical applications include diagnosis/therapy for

*Corresponding author: Sam Tilsen, Cornell University, Ithaca, NY, USA, E-mail: tilsen@cornell.edu,

Debarghya Das, E-mail: dd367@cornell.edu, Bruce McKee, E-mail: bwm55@cornell.edu Cornell University, Ithaca, NY, USA

various disorders such as aphasia and apraxia of speech (Katz et al. 1999; Van Lieshout et al. 2007), Parkinson's disease (Ackermann et al. 1993), dysarthria induced by traumatic brain injury (Goozee et al. 2000; Jaeger et al. 2000), influence of orthodontics on speech (Schwestka-Polly et al. 1995), swallowing disorders (Steele and Van Lieshout 2004), and cleft lip and palate (Hoenig and Schoener 1992; Van Lieshout et al. 2002).

Biofeedback is sensory information provided to a person regarding some aspect of their physiology. Typically the information is in a sensory modality that is not normally available as feedback and is enhanced in some way. For speech behaviors, biofeedback has been used in various ways. Chest- and abdominal-strap plethysmographs can provide feedback regarding speech breathing patterns (Murdoch et al. 1999). Visual feedback of acoustic spectra and spectrograms can allow speakers to compare a visual representation of their acoustic output to a model (Byun and Hitchcock 2012). Ultrasound images can allow speakers to observe the shape of their tongue and attempt to match a correctly articulated target (Preston et al. 2013; Shawker and Sonies 1985). EMG feedback has been used for correcting excessive laryngeal tension (Prosek et al. 1978), and nasopharyngoscopic feedback has been used for velopharyngeal dysfunction (Brunner et al. 2005; Witzel et al. 1988; Ysunza, Pamplona, Femat, Mayer, & García-Velasco et al. 1997).

Biofeedback with EMA has been deployed for clinical purposes and second language training. Katz and colleagues (2010, 2007, 1999, 2003) have used EMA biofeedback for treatment of articulatory deficits in people with apraxia of speech. In Katz et al. (2003), participants saw the position of their tongue tip on a monitor and were required to move their tongue tip to circular target regions located at the lower and upper lip. It was observed that an apraxic participant achieved the target with greater accuracy by the end of the training. Bock (2009) used EMA feedback to improve vowel production in congenitally deaf speakers. Participants aimed to produce vowels with their tongue located in target regions determined by the experimenter. EMA biofeedback has also been used to aid second language learning: Levitt and Katz (2008) observed that articulatory biofeedback facilitated the learning of a non-native post-alveolar flap articulation in mono-lingual English speakers, and Suemitsu et al. (2013) used EMA biofeedback to improve production of unfamiliar vowels for Japanese learners of English.

Although the aforementioned studies have used EMA biofeedback for clinical purposes and second language training, many possible areas of investigation remain open, particularly with regard to theoretical questions involving the cognitive representations of phonological categories. To that end, Section 2 below describes and analyzes data from two preliminary studies that investigate the representation of movement targets and control over movement trajectories. Section 3 describes two games that were developed to investigate articulatory control, and Section 4 summarizes the potential contributions of EMA biofeedback methodology.

2 Research applications of articulatory biofeedback with RT-EMA

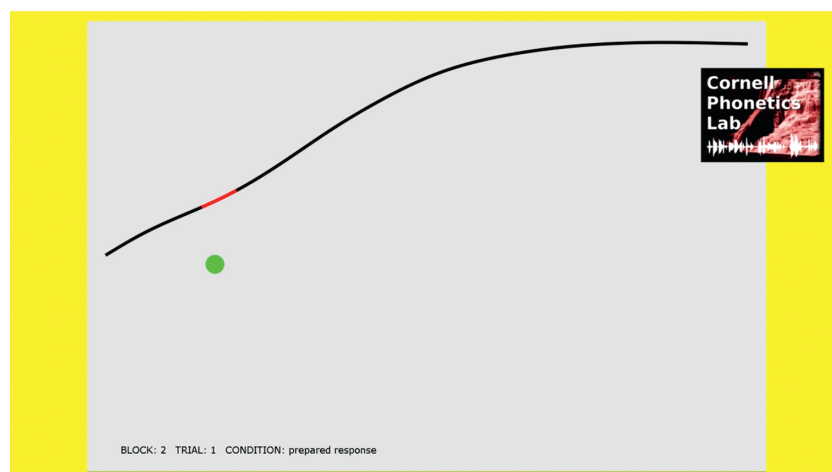
This section describes research applications of articulatory biofeedback with the RT-EMA system. Pilot experimental data were collected from the first author, a native speaker of Midwestern American English. Details regarding the design of the RT-EMA system and data collection procedures are described in the Appendix. The analyses presented below adopt the theoretical perspective of the task-dynamic model of articulatory phonology (Goldstein et al. 2006; Saltzman and Munhall 1989), in which articulatory gestures constitute the basic units of phonological representation. A gesture is conceptualized as a dynamical system with a one-dimensional point attractor, where the attractor corresponds to a target value of a variable describing vocal tract geometry. Thus the production of an alveolar stop such as /t/ is associated with a target tongue tip constriction degree (TTCD), which is typically zero or a negative value, and with a target tongue tip constriction location (TTCL), which can vary in the anterior–posterior dimension of the palate. Articulatory phonology holds that gestural targets of this sort are important parameters of the phonological representation of speech. Thus it is crucial to understand the factors that influence target accuracy and learning of new targets.

2.1 Experiment 1: effects of target location and response preparation on articulatory movements

The aim of this experiment is to investigate how movement accuracy is influenced by (1) the location of an externally provided constriction location target and (2) the time available for preparing an articulatory movement. The speaker is tasked with hitting a target line along the palate with their tongue tip in the course of producing [ta]. In addition, the speaker must precede the target movement with a prolonged [a], i.e. a low/central vowel. This pre-response articulation is initiated upon a ready-signal that is cued by the margins of the screen turning yellow. After a 1.5–2.5 s delay, the margins of the screen turn green, cueing the speaker to initiate the response as soon as possible. The palate is displayed on the screen and the targets are highlighted sections of the palate (cf. Movie 1).

Each experimental session is organized into blocks of 18 trials. In the first block, no feedback or palate is displayed. The location of the sensor during the [t] closure (specifically, the position of the tongue tip sensor when the velocity of its first principle component reaches a minimum) is calculated for each trial. The average horizontal position along the palate of the closure location in the first block serves as an intrinsic target from which alternative targets are determined for use in subsequent blocks. The alternative targets are nine points located along the palate, equally spaced horizontally in a range of –5 to 5 mm from the intrinsic target. From these nine points, nine target lines are defined along the palate, each having a horizontal length of 2 mm. Note that the fifth target is centered horizontally on the intrinsic target, and there are four targets anterior and four targets posterior to the intrinsic one.

In each block all targets are presented twice in random order, and blocks alternate between prepared and unprepared responses. Preparation is manipulated by controlling the timing of the appearance of the target. In the prepared response blocks, the target becomes visible when the ready-signal is given, and hence the speaker has 1.5–2.5 s to pre-plan the articulatory movement before they must initiate it. In the unprepared response blocks, the target becomes visible when the go-signal is given, and hence the speaker has minimal time to pre-plan the movement. Movie 1 shows the stimulus screen during a prepared and unprepared block.



Movie 1 [thumbnail hyperlinked to movie]* Demonstration of Experiment 1. Shows feedback monitor from a block of prepared response trials and a block of unprepared response trials

The experiment aims to test two hypotheses. First, response preparation is hypothesized to facilitate movement accuracy, since more time is available for planning. This predicts that speakers will be more accurate in hitting the target in prepared responses compared to unprepared responses. Second, speakers

*Movie available online: <http://dx.doi.org/10.1515/lingvan-2014-1006>.

are hypothesized to adjust pre-response vocalic posture to facilitate movement accuracy, a phenomenon likely related to VC coarticulation. This predicts that prior to the prepared responses speakers will subtly adjust the posture of their tongue during the ready phase.

Analyses of data support the above hypotheses and suggest several further avenues of research. A number of methods for defining accuracy were examined in analyzing these data, but for brevity just one is reported here, in which accuracy was defined as the horizontal displacement of the closure from the center of the target. Overall accuracy was highest for the intrinsic target and the two nearest anterior targets, as shown in Figure 1 (top). Furthermore, the speaker was more accurate in the prepared responses than the unprepared responses, particularly for the two most posterior targets. Repeated measure ANOVAs corroborate these observations, as there was a significant effect of target location ($F(8,303) = 3.77, p < 0.001$) and a significant effect of response condition ($F(1,303) = 8.5, p = 0.004$). Although there were relatively large fluctuations in accuracy between some blocks, Figure 1 (bottom) shows that there was no overall trend in accuracy across the session, and a linear regression of accuracy by block results in block number accounting for only a small proportion of variance in accuracy ($R^2 = 0.008, F(1,316) = 3.54, p = 0.06$).

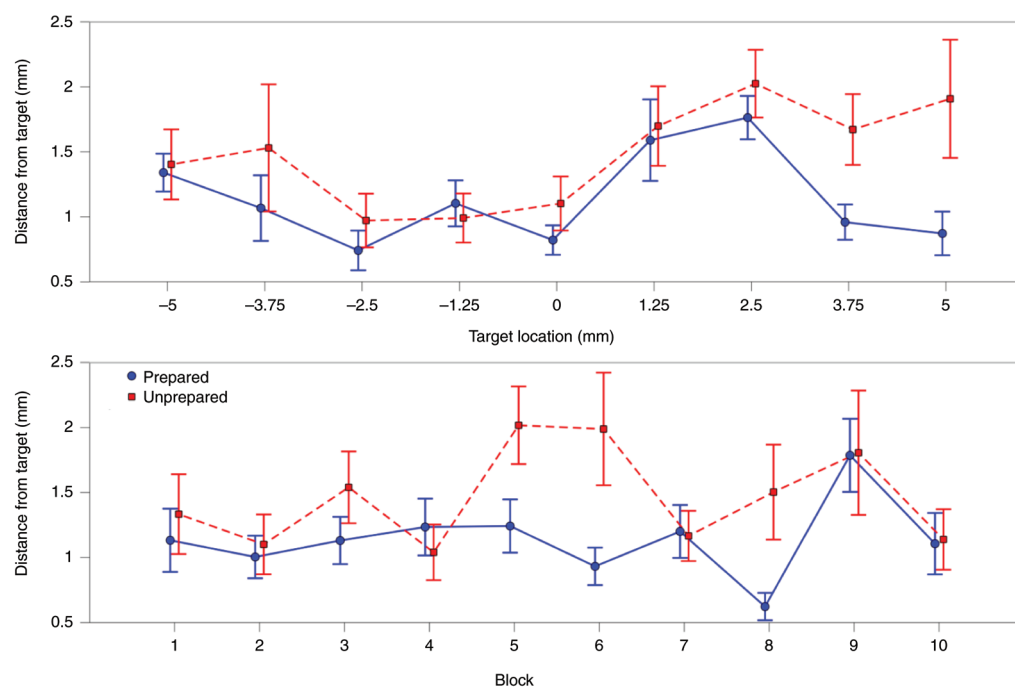


Figure 1 (Top) Accuracy as a function of target location. The speaker was less accurate in the unprepared trials, particularly for targets located posterior to the reference. (Bottom) Accuracy as a function of block. No clear trend was observed for prepared or unprepared blocks over the session. Circles/solid line: prepared trials; squares/dashed line: unprepared trials. Error bars indicate ± 2.0 standard error

Figure 2 shows the movement trajectories produced for each target location, which reveals a number of interesting patterns. Trajectories for the anterior targets consistently exhibit broad loops in a posterior-to-anterior curve. As the target becomes more posterior the loops exhibit a higher degree of curvature, and for the more posterior targets, some loops exhibit anterior-to-posterior curves or acute angles in the transition from closure to release. In addition the distributions of closure and pre-response positions for a given target were not the same between prepared and unprepared responses. The density distributions of closure locations (Figure 3) show that despite an approximately uniform distribution of target locations within each condition, the speaker did not produce closures uniformly distributed along the palate. Instead, there appear to be three modes of closure location: one that is 3–4 mm anterior to the intrinsic target, a second near the intrinsic target, and a third about 5 mm posterior to the intrinsic target.

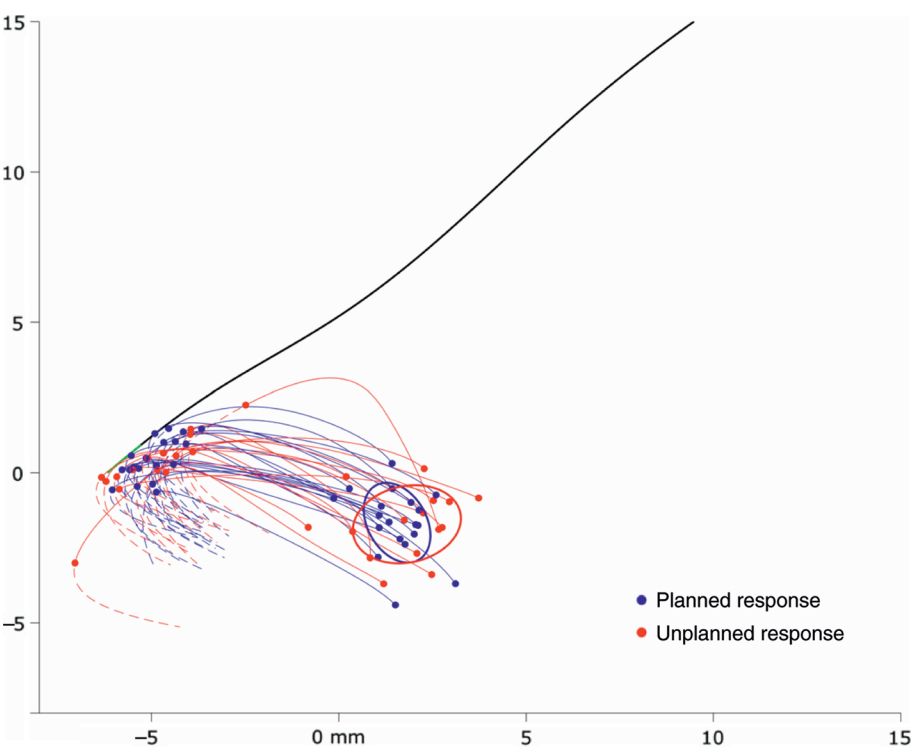


Figure 2 [this is an animated .gif] Movement trajectories for each target. Trajectories from prepared/unprepared responses are shown in blue/red. Onset and closure positions are indicated by small dots, and 95% confidence regions for the onset positions are indicated with ellipses

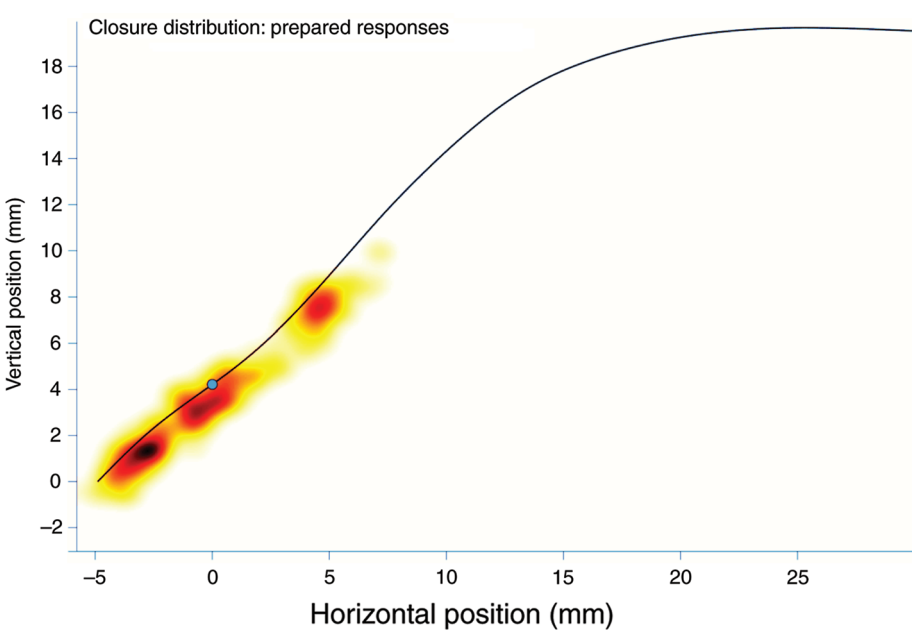


Figure 3 [this is an animated .gif] 2D-density distributions of closure locations produced in each condition, and the difference between density distributions. Closure locations tend to be shifted anteriorly in the prepared responses compared to the unplanned responses. Gray dot shows intrinsic target location

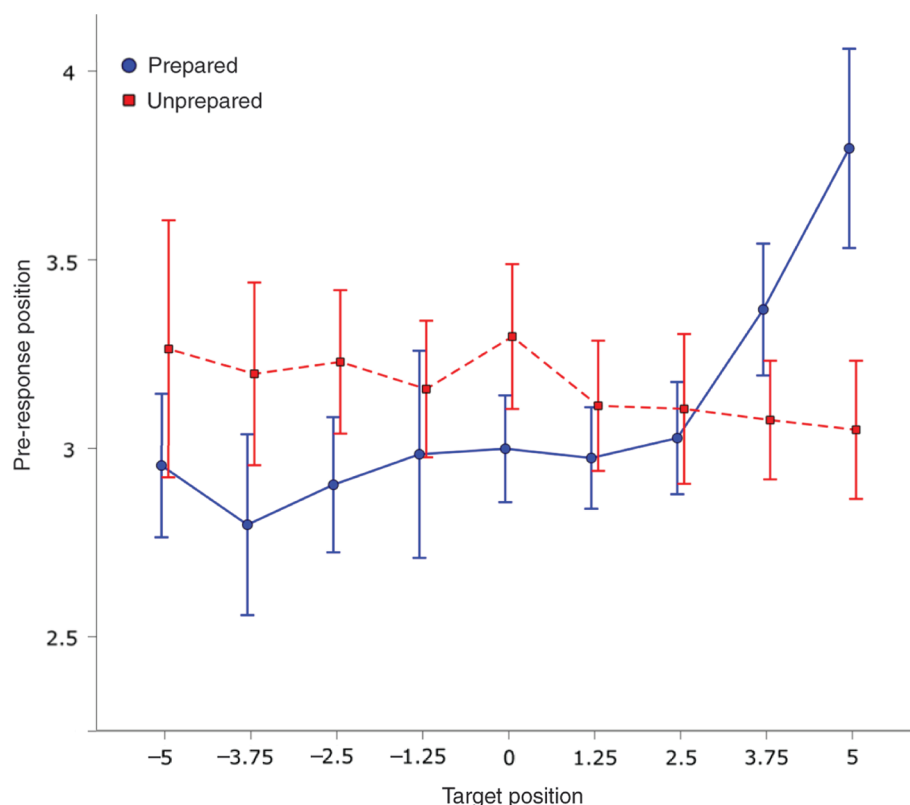


Figure 4 Pre-response horizontal position as a function of target horizontal position. Target position has no effect on pre-response position for unprepared responses; in prepared responses, the tongue tip is slightly more anterior for anterior targets, and is substantially retracted for posterior targets. Circles/solid line: prepared trials; squares/dashed line: unprepared trials. Error bars indicate ± 2.0 standard error

The closure location density distributions are somewhat more concentrated in the prepared responses, and the difference in density distributions shows that closures at each of the three modes were more anterior in the prepared responses than in the unprepared responses. This might be attributable to differences in pre-response posturing between conditions. Figure 4 shows that in prepared responses, the pre-response horizontal position of the tongue tip was more anterior for more anterior targets, and was substantially retracted for the posterior targets. This effect may represent an effort by the speaker to facilitate accuracy by decreasing the required movement range, or it may represent an interaction between trajectory planning and the articulatory posture of the pre-response vowel.

The above observations warrant further investigation because the mechanisms behind them are of linguistic relevance. Specifically, the trimodal distribution of closure locations observed in both conditions is of particular interest for understanding alveolar stop categories and their associated gestural targets. A post-hoc interpretation of the results is that the more anterior mode may be associated with a dental [t], the mode closest to the intrinsic target with an alveolar [t], and the posterior one with a retroflex [ɬ]. Although the speaker's native phone inventory (i.e. American English) does not include a retroflex alveolar, the speaker is a trained phonetician with some prior experience in producing this sound. A pertinent question in this regard is whether the stronger biases toward target modes in the prepared condition is consistent with a role of previously learned gestural targets. On the one hand, precision control over movement targets might be expected to improve with more time for preparation, resulting in a stronger bias toward constriction location modes in unprepared responses. On the other hand, with more time for preparation, target categories might exert a relatively stronger bias on the response.

Also of linguistic interest is the effect of preparation on pre-response posture, since this can be directly related to anticipatory VC coarticulation. The prepared responses may more closely mimic the situation of conversational speech, where the speaker has knowledge of upcoming gestural targets during the production of preceding gestures. In prepared responses the vocalic posture was altered to locate the tongue tip more anteriorly or posteriorly in accord with the target location. Yet the degree of this anticipatory posturing was not linearly related to the target location; this suggests the presence of non-linear constraints on coarticulation that deserve further examination.

2.2 Experiment 2: articulatory target learning

The aim of this experiment is to investigate the use of biofeedback to facilitate learning novel targets. The task and pre-response constraints are identical to Experiment 1, except that there are only two targets, one shifted 3 mm anterior to the intrinsic target, and one shifted 3 mm posterior. After the intrinsic target is estimated in the first block, subsequent blocks are organized into sets with paired training and memory blocks. In training blocks the speaker receives feedback and sees a visual representation of the target; in the memory blocks the speaker receives no feedback and sees no target or palate, in which case they are tasked with achieving the target in the immediately preceding training block, relying on their memory. The anterior and posterior targets alternated between sets.

The experiment is designed to test the hypothesis that biofeedback facilitates the learning of new articulatory targets. This hypothesis predicts that the location of tongue tip closure on memory blocks will deviate from the intrinsic target toward the trained target, and that over the course of the session, accuracy in closure location will increase and variance in closure location will decrease. Figure 5 shows the trajectories produced in each block of the experiment.

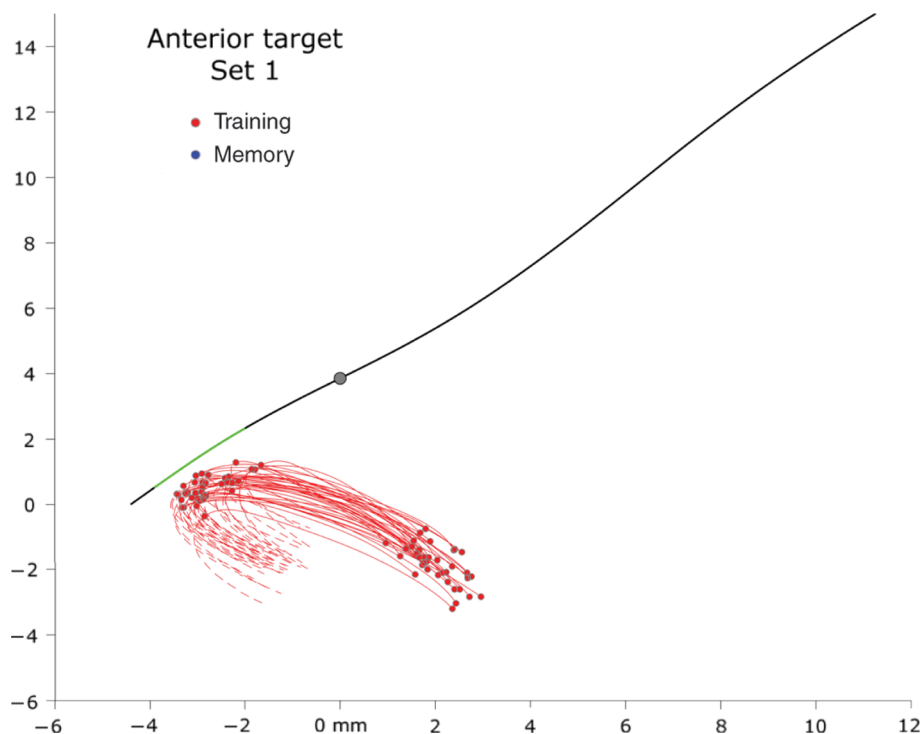


Figure 5 *[this is an animated .gif]* Tongue tip trajectories from training and memory blocks across Experiment 2. Gray circle shows location of intrinsic target

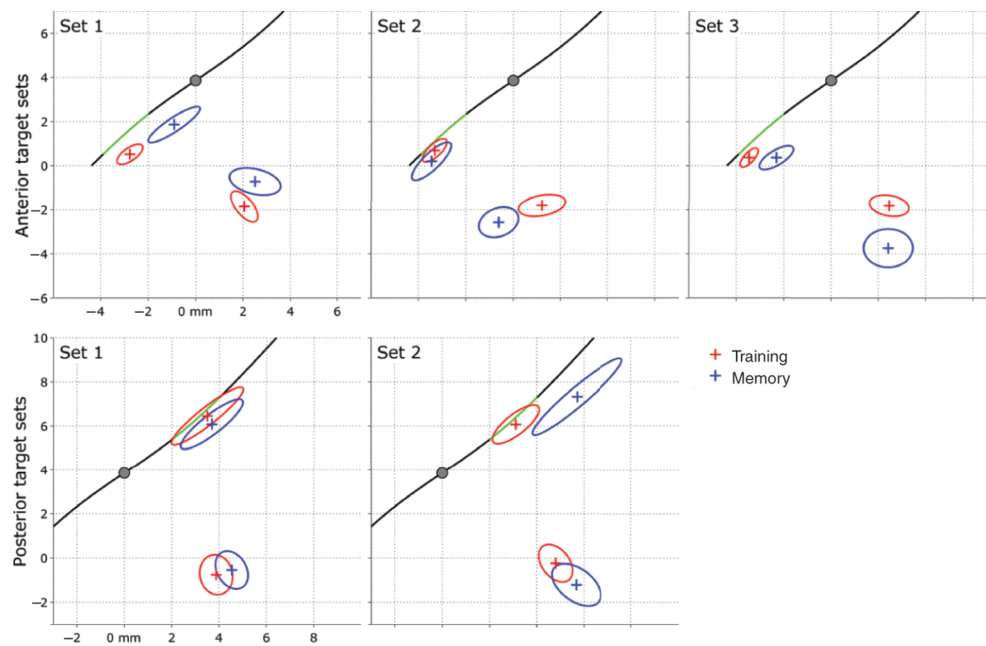


Figure 6 95% confidence regions for starting position and closure location for training and memory blocks in each set. Gray circle shows the location of the intrinsic target

The results support the hypothesis that biofeedback facilitates learning new tongue tip closure targets, but also indicate a difference in learning dynamics between the anterior and posterior target. Figure 6 shows 95% confidence regions for the pre-response and closure locations of the tongue tip on training and memory blocks in each set. It is clear that the speaker did not fully return to the intrinsic target during the memory blocks in any set. In the first anterior target set, closure locations in the memory blocks returned partway toward the intrinsic target. Yet in the second and third anterior target sets, the closure location remained fairly close to the trained target. The change in performance after set 1 was accompanied by changes in the starting position of the movement, such that it was initiated from a more anterior position on memory blocks in set 2 and a more inferior position on memory blocks in set 3. These patterns indicate that learning across the session extended to the vocalic posture as well as the consonantal target, but it is not clear why changes in starting position were less substantial between training blocks.

To further assess learning, accuracies and dispersions of closure location across training and memory blocks were compared separately for each target. For the anterior targets, a repeated measures ANOVA of accuracy showed significant effects of condition ($F(1,214) = 71.1, p < 0.001$), set ($F(2,214) = 12.1, p < 0.001$), and a set-condition interaction ($F(2,214) = 13.0, p < 0.001$). Figure 7 (top) shows that accuracy was higher in training blocks and that the set effect was driven mostly by an increase in accuracy in memory blocks between the first and second set. There were also significant effects of condition ($F(1,214) = 31.8, p < 0.001$) and set ($F(2,214) = 5.8, p = 0.004$) on dispersion of closure location (Figure 7, bottom). This measure represents the Euclidean distance of closure location from the within-block centroid and hence serves as an index of variability in closure location. The effects indicate that closure location was less variable in the training blocks and became less variable across sets. Hence both accuracy and variability patterns suggest that the speaker acquired a more robust representation of the anterior target over the session.

For the posterior target, however, the second set memory block showed an unexpected increase in variability and decrease in accuracy. A repeated measures ANOVA of accuracy showed significant effects of condition ($F(1,144) = 49.5, p < 0.001$), set ($F(1,144) = 13.6, p < 0.001$), and an interaction ($F(1,144) = 50.7, p < 0.001$). Figure 7 (top) reveals that the closure location became less accurate and more variable; this reflects an exaggeratedly posterior location of the closure in the second memory set (cf. Figures 5 and 6). One possible account of the difference between the anterior and posterior targets is that the posterior target

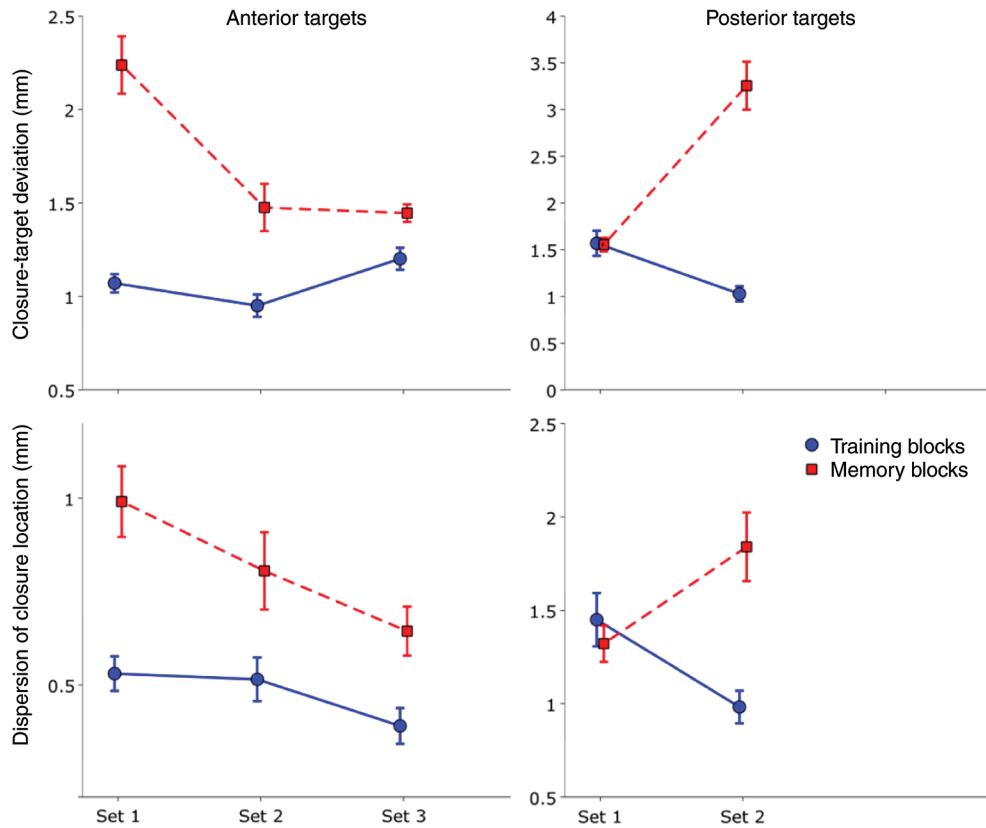


Figure 7 (Top) accuracy by block as measured by deviation of closure location from the target. (Bottom) dispersion of closure location by block. Circles: training blocks; squares: memory blocks. Error bars indicate ± 2.0 standard error

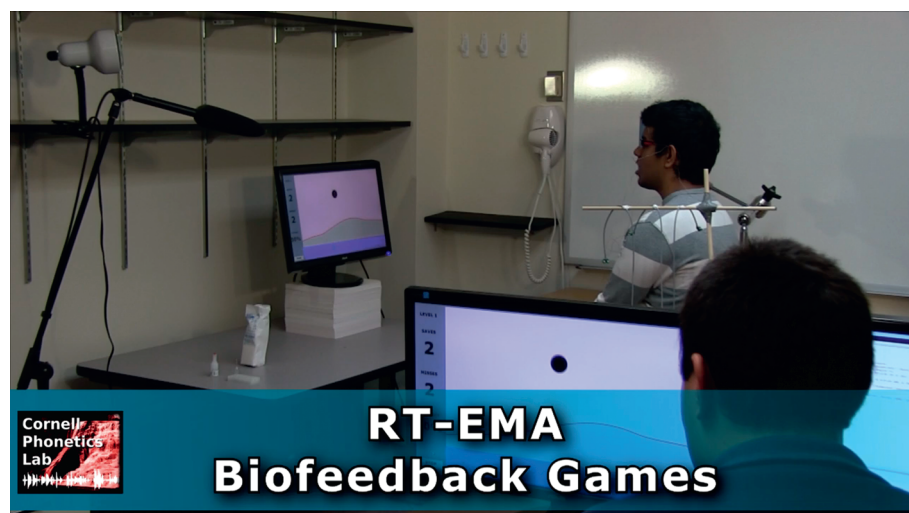
may have been attracted to a previously learned category target, which would be akin to the retroflex-like posterior mode observed in Experiment 1. Indeed, the posterior mode observed in Experiment 1 was located 5 mm posterior to the intrinsic target, whereas the posterior target in Experiment 2 was located just 3 mm posterior to the intrinsic target.

In sum, the results of Experiment 2 indicate that visual biofeedback induces motor learning, but detailed analyses reveal a dissociation in learning behavior between anteriorly and posteriorly shifted targets. The sources of this difference warrant further investigation, and could involve previously acquired motor categories, biomechanical asymmetries, and/or other factors. It is also possible that performance in Experiment 2 may have been influenced by participation in Experiment 1; this possibility can be avoided in future studies by recruiting speakers separately for different experiments. Future designs should also assess learning of targets with greater and lesser degrees of displacement, and should include a training condition without feedback, where target shifts are guided by verbal instructions alone. It is anticipated that training without feedback will be substantially less effective, due to the inherent difficulty of communicating anatomical parameters to naive speakers.

3 RT-EMA biofeedback games: QuickLips and QuickTongue

Many experimental paradigms in phonetic research are not very engaging for participants, and hence the participants tend to become inattentive over the course of a session. By designing an experiment in the form of a game, this loss of attention can be mitigated. RT-EMA biofeedback is particularly suitable for game-like experimental designs, due to the novelty of visual feedback regarding articulation. To this end, two RT-EMA

games were developed, which are designed to assess limits on articulatory control parameters and to induce noncanonical patterns of relative timing between closure and release gestures. Gameplay is illustrated with demonstration videos in which the second author was a participant; the first author also participated in preliminary tests of the games. The quantity of gameplay data currently available is not sufficient for quantitative analyses, so the focus here is on qualitative characterizations of behavior during gameplay and on how future quantitative analyses can be used to address questions of linguistic relevance. An overview of the games can be seen in Movie 2.



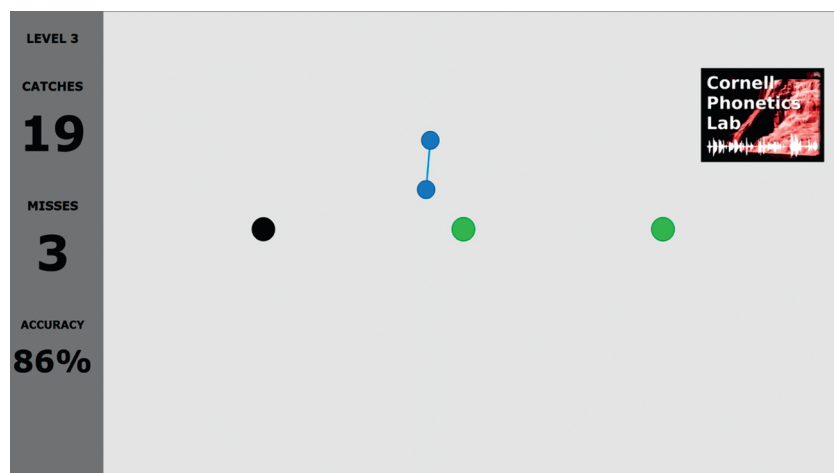
Movie 2 [thumbnail hyperlinked to movie]* Overview of RT-EMA biofeedback gameplay

QuickLips is an RT-EMA biofeedback game designed to elicit bilabial closure and release gestures under varying time constraints. In *QuickLips* a stream of balls moves horizontally toward the lips of the speaker. Sensors attached to the upper and lower lip are represented on the screen by circles, and a line connecting them represents the aperture between the lips. The speaker tries to catch each ball between their lips by saying “pa”. When a ball hits the lip aperture line it is caught, but if the ball hits one of the lips or goes past the aperture line it is missed. A further constraint is that once the lips are opened, the aperture line discharges after approximately half a second and will no longer catch a ball. In order to recharge the line, the speaker must close their lips. This constraint ensures that speakers employ a bilabial release gesture to catch the balls. Movie 3 shows how the game is played.

QuickLips is of interest for linguistic research because it elicits movement behaviors in which speakers must exert independent control over bilabial closure and release gestures, which tend to be precisely coordinated in normal conversational speech. Notably, previous literature presents some disagreement regarding whether the closure and release gestures in a CV response are both associated with the consonant, or whether the release is associated with a following vocalic segment (cf. Tilsen and Goldstein 2012). Hence inducing a dissociation between them can inform our understanding of how these gestures are organized phonologically.

Manipulation of the timing and speed of the balls induces a variety of circumstances which perturb the closure and release gestures in different ways (see Movie 3). For example, in some cases balls approach the lips close together in time and the speaker must produce a very rapid release-closure-release sequence. In these cases gestural magnitudes, durations, and velocities can differ markedly from those occurring with less time pressure. Quantitative analyses of gestural kinematics may be able to determine what movement control parameters the speaker is manipulating in order to perform the task. Another noteworthy occurrence observed

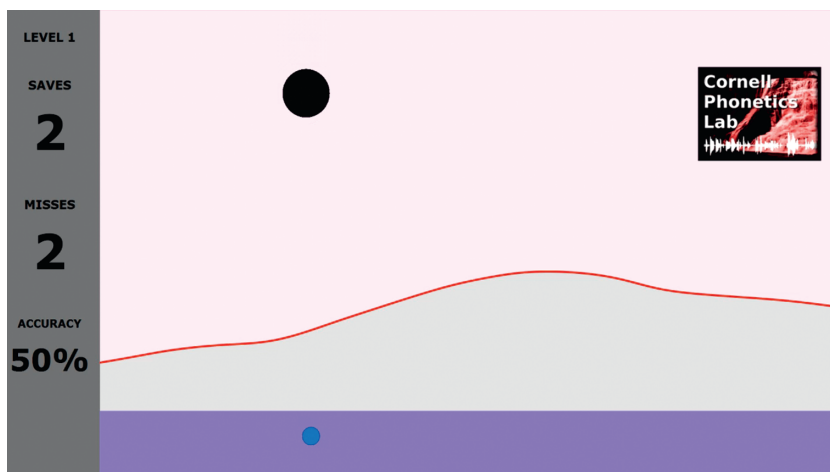
*Movie available online: <http://dx.doi.org/10.1515/lingvan-2014-1006>.



Movie 3 [thumbnail hyperlinked to movie]* Demonstration of QuickLips gameplay. The speaker attempts to catch each ball with the line between their open lips. The line discharges after half a second when the lips are open

during gameplay involves halted releases, where the speaker halts mid-release because the release was begun too early. Such events are of interest because they show that articulatory gestures are not beyond control once they have been initiated, a topic that was originally investigated in Ladefoged (1973) and has recently received renewed attention (Tilsen 2011; Tilsen and Goldstein 2012). Further manipulations of the experimental design—e.g. inducing more unpredictability in the stream of balls, instructing the subject not to speak out loud so that auditory feedback is unavailable, or requiring different gestures in response to different stimuli—can be used to investigate control over the selection of gestures and their execution.

The second game we have developed is *QuickTongue*, which is designed to investigate precision control over the timing and targets of tongue tip closure gestures. The speaker uses their tongue tip to hit balls that are falling from the top of the screen toward the palate; they must hit a ball before it falls below the palate, otherwise it is considered a miss. A short distance below the palate there is a recharge zone: the tongue tip discharges shortly after it leaves the recharge zone or when it makes contact with a ball. In order to recharge the tongue tip the speaker must bring it back to the recharge zone. The speaker is further required to say [ata] in the course of hitting a ball (see Movie 4).



Movie 4 [thumbnail hyperlinked to movie]* Demonstration of QuickTongue gameplay. The speaker attempts to hit balls with their tongue tip before the balls fall below the palate. The tongue tip discharges after a half a second and the speaker must lower it to recharge

*Movie available online: <http://dx.doi.org/10.1515/lingvan-2014-1006>.

Qualitative assessments of QuickTongue gameplay revealed several interesting phenomena. First, speakers have a strong tendency to dissociate the closure and release components of the [t]. Second, speakers appear to be less accurate with targets located very posteriorly along the palate or slightly posterior to the alveolar ridge. This observation may be related to the presence of distinct modes of closure observed in Experiment 1 and may speak to the influence of a set of previously learned motor targets. Third, speakers exhibit a strong propensity to prematurely initiate the movement toward the closure. One possible explanation for this is that speakers underestimate the velocity of their speech movements and hence tend to initiate them too early when unconstrained. Another possibility is that speakers simply have little experience in linking the control of lingual timing to visual stimuli; thus more practice might establish more accurate control over such timing.

4 Conclusion

This paper presented a system for providing speakers real-time visual feedback of speech articulation, thereby opening doors for new lines of research. The results of Experiment 1 suggested that tongue tip gestures to an arbitrary target closure location are more accurate when the speaker has more time to prepare the movement, and that speakers manipulate vocalic posture prior to the movement to facilitate accuracy. Further investigation of these phenomena is likely to shed light on our understanding of speech motor planning and vowel–consonant coarticulation. It was also observed that closures were attracted to three distinct target locations, which may speak to the influence of previously learned motor categories. Conducting this experiment on speakers of languages with different coronal stop inventories may inform our understanding of how phonological categories interact with phonetic implementation. The results of experiment 2 suggested that biofeedback facilitates learning and refining new closure targets, although we found that learning dynamics differed markedly depending on target location. The sources of these differences warrant further investigation, since it is unclear whether they are attributable to the presence of previously learned categories, biomechanical constraints on tongue movement, differential usage of vowel–consonant coarticulation, or some other factors.

Articulatory biofeedback games, such as QuickLips and QuickTongue, also provide rich environments for investigating speech motor control. Both games utilize biofeedback to enhance control over movement and our qualitative assessments show that players exhibit improvement in performance over time. Because of this the games provide a rich environment for investigating the influence of previously acquired motor control parameters on learning new parameters. For example, the general prediction can be made that speakers of languages with more coronal stop categories will perform differently from speakers with fewer categories. A larger motor repertoire of closure targets can be predicted to facilitate or hinder achieving an arbitrary target in the biofeedback paradigm, depending on assumptions about how previously learned motor categories influence the acquisition of new ones. Biofeedback games are also of potential value in clinical contexts. The software system allows for arbitrary selection of feedback information and definition of event detection/action functions, which can be readily customized to suit a specific purpose. Hence the system can be used not only to assess disordered function (as can arise in stroke-induced dysarthria, or in apraxia of speech), but also as a therapeutic tool, since it is fairly straightforward to introduce gameplay components that create additional motivation for improving performance.

In conclusion, RT-EMA biofeedback can create circumstances in which normal control is perturbed in various ways and thereby elicits noncanonical movement behaviors that are of linguistic interest. The perturbative nature of biofeedback and atypicality of the behaviors it induces are indeed its primary merits: by perturbing a system to extreme regimes of operation, one can better understand the dynamics of the system.

Acknowledgments: We would like to thank Mark Tiede for providing Matlab code for head movement correction and reference basis determination.

Appendix: RT-EMA biofeedback system design and setup

This section describes the system's hardware setup, software operation, and setup procedures. The basic physical principle behind EMA involves electromagnetic induction: transmitter coils generate alternating magnetic fields that induce electrical currents in small receiver coils; the amplitude of those currents varies as a function of receiver distance and orientation, and an optimization algorithm uses this variation to calculate receiver positions relative to the transmitter (cf. Berry 2011; Perkell et al. 1992). The EMA sensors consist of receiver coils housed in small ($\approx 3 \text{ mm}^3$) boxes and these are adhered to active articulators and fixed reference points. A wire from each sensor conducts the induced current to a control unit that calculates their position and orientation. The adhesion is accomplished by using a non-toxic cyanoacrylate adhesive commonly used in dental surgery.

The real-time EMA biofeedback system presented in this paper uses the NDI Wave Speech Research System (see Berry 2011). Figure 8 illustrates the hardware components. The speaker is seated approximately 1.5 m from a stimulus monitor. To the left of the stimulus monitor is a shotgun microphone pointed toward the speaker. The EMA transmitter box is mounted to the right of the speaker. Sensor wires adhered to the speaker's articulators and reference locations connect in pairs to sensor cables which run to interface units on a table to the speaker's left. The sensor cables are taped to a cantilevered bar anchored to a table. The sensor interface units are connected to the system control unit (SCU) on the table, and the SCU is connected through a serial-to-USB adaptor to a control PC. The control PC has two monitors: a control monitor for the experimenter and a stimulus-copy monitor that allows the experimenter to view what the speaker is seeing.

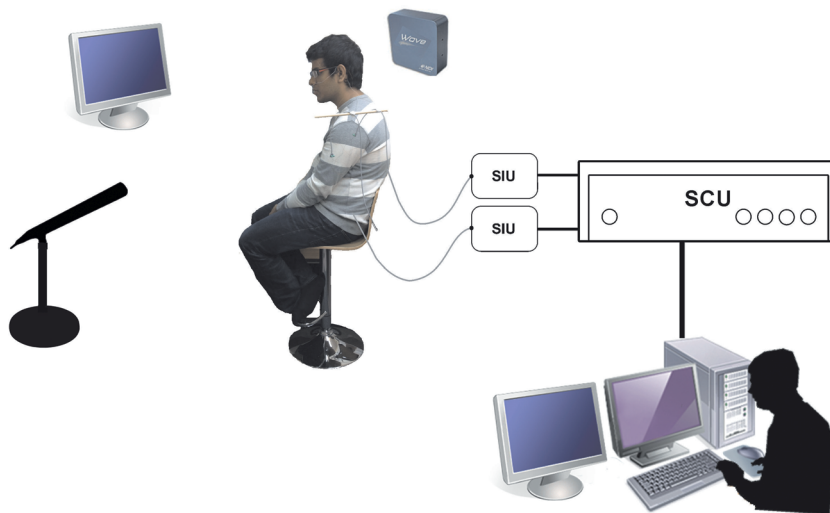


Figure 8 Hardware components in the RT-EMA biofeedback system

Common sensor locations in experimental studies include the upper lip (UL), lower lip (LL), mandible (JAW), lower incisors (LI), tongue tip (TT), tongue blade (TB), and tongue dorsum (TD). Sensors are also adhered to reference points which can include the upper incisors (UI), left and right mastoid processes (MPL, MPR), and nasion (NAS). The reference sensors are used to track head motion and correct for the effects of head motion on the positions of active sensors. In Experiments 1 and 2 and the QuickTongue game, only a TT sensor and reference sensors were used. The TT sensor was located approximately 1.5 cm from the

apex of the tongue. In the QuickLips game, UL, LL, and reference sensors were used. Movie 5 provides an overview of the hardware and speaker setup procedures.



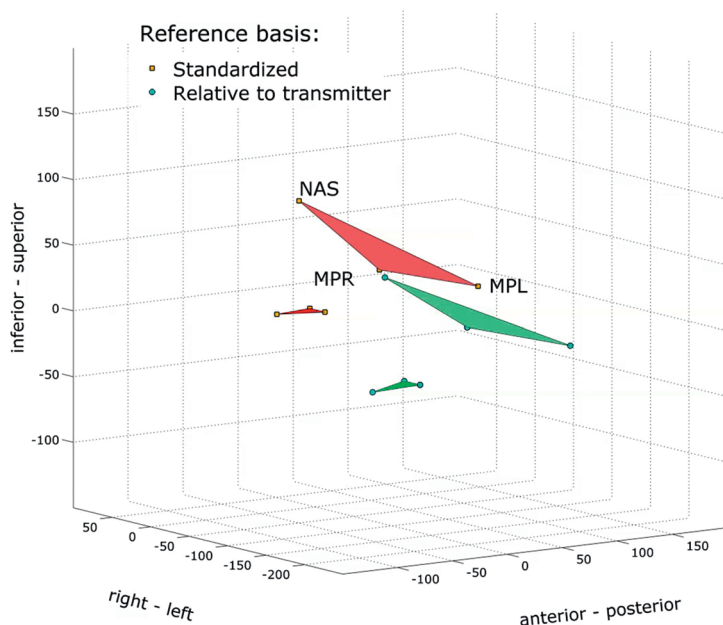
Movie 5 [thumbnail hyperlinked to movie]* Overview of hardware and speaker setup procedures for the RT-EMA biofeedback system

The software system is implemented in Matlab and utilizes an NDI process that enables data streaming from the articulograph's SCU to a PC through a USB TCP/IP port. Streamed data frames contain a timestamp and the 3D positions of each sensor in the transmitter coordinate space. The SCU calculates sensor positions every 10 ms, but the data received in each port read can represent a partial data frame or multiple frames, depending on the timing of communication with the SCU. To address this, the software repeatedly reads data from the port, stores the incoming data in a buffer, and reconstructs a full data frame when possible. When a data frame is reconstructed, customizable event detection and event action functions are called, and graphics objects are updated. Speakers receiving feedback observe no noticeable lag between movement and its visual display. Event detection functions assess whether a sensor is interacting with some other object or sensor. For example, an event might be defined to occur when the distance between the tongue sensor and a target point falls below a threshold value. Events can also involve interactions between objects, such as when a ball falling from the top of the screen goes below the palate. These events can be used to trigger customizable actions; these typically involve logging the event and changing the properties of graphics objects.

In order to correct for head movement while a participant is speaking, a reference basis is calculated at the beginning of each session. The reference basis is comprised of three points in Cartesian space, one for each of the reference sensors, i.e. the nasion (NAS) and left and right mastoid processes (MPL, MPR). Every data frame collected during an experiment is transformed such that the reference sensors are located at the reference basis. In order to standardize the reference basis across speakers the orientation of the occlusal plane is estimated using a bite plate in conjunction with calculating the reference basis. The occlusal plane is an imaginary surface that touches the edges of the anterior and posterior incisors. Its orientation is estimated by having the speaker gently bite down on a thin plastic plate on which 3 sensors are attached. The reference basis coordinates are rotated and translated such that the bite plate is perpendicular to the transmitter's vertical axis and the most anterior bite plate sensor is the origin of coordinate space.

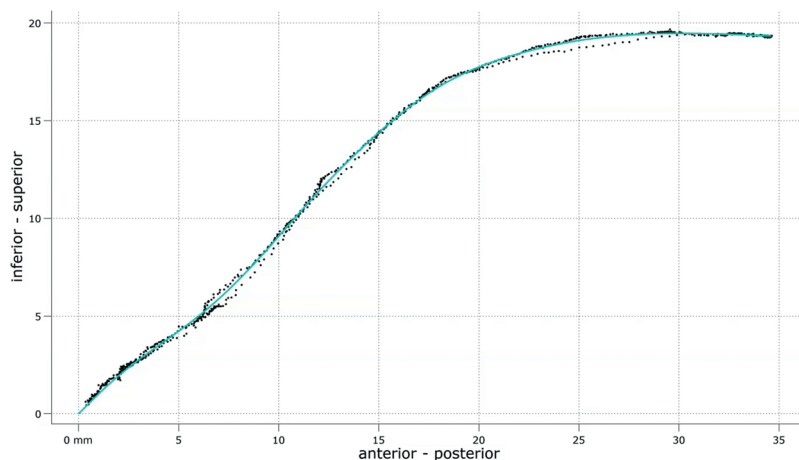
Movie 6 illustrates the reference sensors and bite plate before and after the transformation to a common reference frame. To correct for head movement during data acquisition, a rotation/translation matrix is

*Movie available online: <http://dx.doi.org/10.1515/lingvan-2014-1006>.



Movie 6 [thumbnail hyperlinked to movie]* Reference sensors and bite plate sensors before and after standardization. In the standardized reference basis the most anterior bite plate sensor is the origin and the horizontal coordinate axis is parallel to the occlusal plane. Note that horizontal coordinate increases from anterior to posterior.

calculated which locates the reference sensors at the reference basis, and the rotation/translation is applied to all sensor coordinates. Prior to beginning an experiment, a palate trace is collected to estimate the location of palate. The palate trace is acquired as follows (cf. Movie 7): the speaker places the tongue tip sensor against the back of their upper incisors and drags it first backwards and then forwards, maintaining gentle pressure against the roof of their mouth.



Movie 7 [thumbnail hyperlinked to movie]* Illustration of a palate trace. The speaker drags a tongue tip sensor from the back of their upper incisors along their palate. Data points from the palate trace are subsequently fit with a smoothing spline (red line).

*Movie available online: <http://dx.doi.org/10.1515/lingvan-2014-1006>.

References

- Ackermann, H., B. F. Gröne, G. Hoch & P. W. Schönle. 1993. Speech Freezing in Parkinson's Disease: A Kinematic Analysis of Orofacial Movements by Means of Electromagnetic Articulography. *Folia Phoniatrica et Logopaedica* 45(2). 84–89.
- Berry, J. J. 2011. Accuracy of the NDI Wave Speech Research System. *Journal of Speech Language and Hearing Research* 54(5). 1295–1301.
- Beskow, J., O. Engwall & B. Granström. 2003. Resynthesis of Facial and Intraoral Articulation from Simultaneous Measurements. In *Proceedings of ICPhS*, Vol. 3, 57–60.
- Bock, R. 2009. Effectiveness of Visual Feedback Provided by an Electromagnetic Articulograph (EMA) System: Training Vowel Production in Individuals with Hearing Impairment.
- Brunner, M., A. Stellzig-Eisenhauer, U. Pröschel, R. Verres & G. Komposch. 2005. The Effect of Nasopharyngoscopic Biofeedback in Patients with Cleft Palate and Velopharyngeal Dysfunction. *The Cleft Palate-Craniofacial Journal* 42(6). 649–657.
- Byrd, D. 2000. Articulatory Vowel Lengthening and Coordination at Phrasal Junctures. *Phonetica* 57(1). 3–16.
- Byun, T. M. & E. R. Hitchcock. 2012. Investigating the Use of Traditional and Spectral Biofeedback Approaches to Intervention for /r/ Misarticulation. *American Journal of Speech-Language Pathology* 21(3). 207–221.
- Chitoran, I., L. Goldstein, and D. Byrd. 2002. Gestural Overlap and Recoverability: Articulatory Evidence from Georgian. *Laboratory Phonology* 7. 419–447.
- Goldstein, L., D. Byrd & E. Saltzman. 2006. The Role of Vocal Tract Gestural Action Units in Understanding the Evolution of Phonology. In Arbib, M. (Ed.), *Action to Language via the Mirror Neuron System*. 215–249. Cambridge: Cambridge University Press.
- Goozee, J. V., B. E. Murdoch, D. G. Theodoros, and P. D. Stokes. 2000. Kinematic Analysis of Tongue Movements in Dysarthria Following Traumatic Brain Injury Using Electromagnetic Articulography. *Brain Injury* 14(2). 153–174.
- Hoenig, J. F. & W. F. Schoener. 1992. Radiological Survey of the Cervical Spine in Cleft Lip and Palate. *Dentomaxillofacial Radiology* 21(1). 36–39.
- Hoole, P. & N. Nguyen. 1999. Electromagnetic Articulography. Coarticulation—Theory, Data and Techniques, Cambridge Studies in Speech Science and Communication, 260–269.
- Jaeger, M., I. Hertrich, U. Stattrop, P.-W. Schönle & H. Ackermann. 2000. Speech Disorders Following Severe Traumatic Brain Injury: Kinematic Analysis of Syllable Repetitions Using Electromagnetic Articulography. *Folia Phoniatrica et Logopaedica* 52(4). 187–196.
- Katz, W. F., S. V. Bharadwaj & B. Carstens. 1999. Electromagnetic Articulography Treatment for an Adult with Broca's Aphasia and Apraxia of Speech. *Journal of Speech Language and Hearing Research* 42(6). 1355–1366.
- Katz, W. F., D. M. Garst, G. S. Carter, M. R. McNeil, T. R. Fossett, P. J. Doyle & N. J. Szuminsky. 2007. Treatment of an Individual with Aphasia and Apraxia of Speech Using EMA Visually-Augmented Feedback. *Brain and Language* 103(1). 213–214.
- Katz, W. F., C. Kripke & P. Tallal. 1991. Anticipatory Coarticulation in the Speech of Adults and Young Children: Acoustic, Perceptual, and Video Data. *Journal of Speech Language and Hearing Research* 34(6). 1222.
- Katz, W. F., J. S. Levitt & G. C. Carter. 2003. Biofeedback Treatment of Buccofacial Apraxia Using EMA. *Brain and Language* 87(1). 175–176.
- Katz, W. F. & M. R. McNeil. 2010. Studies of Articulatory Feedback Treatment for Apraxia of Speech Based on Electromagnetic Articulography. *SIG 2 Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders* 20(3). 73–79.
- Ladefoged, P., R. Silverstein & G. Papçun. 1973. Interruptibility of speech. *The Journal of the Acoustical Society of America* 54(4). 1105–1108.
- Levitt, J. S. & W. F. Katz. 2008. Augmented Visual Feedback in Second Language Learning: Training Japanese Post-Alveolar Flaps to American English Speakers. In *Proceedings of Meetings on Acoustics*, Vol. 2, p. 060002. Acoustical Society of America.
- Marin, S. & M. Pouplier. 2010. Temporal Organization of Complex Onsets and Coda in American English: Testing the Predictions of a Gestural Coupling Model. *Motor Control* 14(3). 380–407.
- Murdoch, B., G. Pitt, D. Theodoros & E. C. Ward. 1999. Real-Time Continuous Visual Biofeedback in the Treatment of Speech Breathing Disorders Following Childhood Traumatic Brain Injury: Report of One Case. *Developmental Neurorehabilitation* 3(1). 5–20.
- Parrell, B., S. Lee & D. Byrd. 2013. Evaluation of Prosodic Juncture Strength Using Functional Data Analysis. *Journal of Phonetics* 41(6). 442–452.
- Perkell, J. S., M. H. Cohen, M. A. Svirsky, M. L. Matthies, I. Garabieta & M. T. T. Jackson. 1992. Electromagnetic Midsagittal Articulometer Systems for Transducing Speech Articulatory Movements. *The Journal of the Acoustical Society of America* 92(6). 3078–3096.
- Preston, J. L., N. Brick & N. Landi. 2013. Ultrasound Biofeedback Treatment for Persisting Childhood Apraxia of Speech. *American Journal of Speech-Language Pathology* 22(4). 627–643.
- Prosek, R. A., A. A. Montgomery, B. E. Walden & D. M. Schwartz. 1978. EMG Biofeedback in the Treatment of Hyperfunctional Voice Disorders. *Journal of Speech and Hearing Disorders* 43(3). 282–294.

- Saltzman, E. & K. Munhall. 1989. A Dynamical Approach to Gestural Patterning in Speech Production. *Ecological Psychology* 1(4). 333–382.
- Schwestka-Polly, R., W. Engelke & G. Hoch. 1995. Electromagnetic articulography as a method for detecting the influence of spikes on tongue movement. *The European Journal of Orthodontics* 17(5). 411–417.
- Shaw, J., A. I. Gafos, P. Hoole & C. Zerual. 2011. Dynamic Invariance in the Phonetic Expression of Syllable Structure: A Case Study of Moroccan Arabic Consonant Clusters. *Phonology* 28(3). 455–490.
- Shawker, T. H. & B. C. Sonies. 1985. Ultrasound Biofeedback for Speech Training: Instrumentation and Preliminary Results. *Investigative Radiology* 20(1). 90–93.
- Steele, C. M. & P. H. Van Lieshout. 2004. Use of Electromagnetic Midsagittal Articulography in the Study of Swallowing. *Journal of Speech Language and Hearing Research* 47(2). 342.
- Steiner, I., K. Richmond & S. Ouni. 2013. Speech Animation Using Electromagnetic Articulography as Motion Capture Data. arXiv Preprint arXiv:1310.8585.
- Suemitsu, A., T. Ito, and M. Tiede. 2013. An Electromagnetic Articulography-Based Articulatory Feedback Approach to Facilitate Second Language Speech Production Learning. In *Proceedings of Meetings on Acoustics*, Vol. 19, p. 060063. Acoustical Society of America.
- Tiede, M., R. Bundgaard-Nielsen, C. Kroos, G. Gibert, V. Attina, B. Kasisopa & C. Best. 2012. Speech Articulator Movements Recorded from Facing Talkers Using Two Electromagnetic Articulometer Systems Simultaneously. In *Proceedings of Meetings on Acoustics*, Vol. 11, p. 060007. Acoustical Society of America.
- Tilsen, S. 2009. Multitimescale Dynamical Interactions between Speech Rhythm and Gesture. *Cognitive Science* 33(5). 839–879.
- Tilsen, S. 2011. Effects of syllable stress on articulatory planning observed in a stop-signal experiment. *Journal of Phonetics* 39(4). 642–659.
- Tilsen, S. & L. Goldstein. 2012. Articulatory Gestures Are Individually Selected in Production. *Journal of Phonetics* 40(6). 764–779.
- Toutios, A., S. Ouni & Y. Laprie. 2011. Estimating the Control Parameters of an Articulatory Model from Electromagnetic Articulograph Data. *The Journal of the Acoustical Society of America* 129(5). 3245–3257.
- Van Lieshout, P. H., A. Bose, P. A. Square & C. M. Steele. 2007. Speech Motor Control in Fluent and Dysfluent Speech Production of an Individual with Apraxia of Speech and Broca's Aphasia. *Clinical Linguistics & Phonetics* 21(3). 159–188.
- Van Lieshout, P. H., C. A. W. Rutjens & P. H. M. Spauwen. 2002. The Dynamics of Interlip Coupling in Speakers with a Repaired Unilateral Cleft-Lip History. *Journal of Speech Language and Hearing Research* 45(1). 5.
- Witzel, M. A., J. Tobe & K. Salyer. 1988. The Use of Nasopharyngoscopy Biofeedback Therapy in the Correction of Inconsistent Velopharyngeal Closure. *International Journal of Pediatric Otorhinolaryngology* 15(2). 137–142.
- Ysunza, A., M. Pamplona, T. Femat, I. Mayer & M. García-Velasco. 1997. Videonasopharyngoscopy as an Instrument for Visual Biofeedback during Speech in Cleft Palate Patients. *International Journal of Pediatric Otorhinolaryngology* 41(3). 291–298.