ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

Characterizing Covert Articulation in Apraxic Speech Using Real-time MRI

Christina Hagedorn[1], *Michael Proctor*[1,2], *Louis Goldstein*[1],
Maria Luisa Gorno Tempini[3], *Shrikanth S. Narayanan*[1,2]

[1] Department of Linguistics, University of Southern California, USA
[2] Viterbi School of Engineering, University of Southern California, USA
[3] Department of Neurology, University of California-San Francisco, USA

chagedor@usc.edu http://sail.usc.edu/span

Abstract

We explore the use of real-time magnetic resonance imaging (rtMRI) as a tool to investigate apraxic speech, in particular, by examining articulatory behavior. Our pilot data reveal that covert (silent) gestural intrusion errors (employing an intrinsically simple 1:1 mode of coupling) are made more frequently by an apraxic subject than by fluent speakers. Covert intrusion errors are also found to be pervasive in non-repetitious apraxic speech. We demonstrate that acoustically silent periods observed before the initiation of apraxic speech oftentimes contain completely covert gestures that occur frequently with multigestural segments. Covert gestures corresponding to entire words are also observed. These data demonstrate that rtMRI can provide important new insights into apraxic speech that are not available using traditional methods of transcription based on acoustic data alone.

Index Terms: Apraxia, speech production, covert articulation, speech error, real-time MRI, disordered speech, gestures

10.21437/Interspeech.2012-309

1. Introduction

Using real time magnetic resonance imaging (rtMRI) to study speech production is particularly advantageous in that it allows for nearly all components of the vocal tract to be observed at once, over time. While other methods of articulometry (e.g. electro-magnetic articulography [1] or X-ray microbeam [2]) offer high temporal and spatial resolution, they provide information about specific flesh points only, and do not allow for a global view of the vocal tract in which the coordination patterns of all articulators can be observed.

Apraxia of speech (AoS) is a phonologic disorder affecting the selection, programming and execution of speech motor commands specified in a target sequence [3][4]. Apraxic speech is often claimed to be characterized by the presence of distorted perseverative and anticipatory substitutions, errors of stress assignment, multiple initiation gestures, and difficulty producing consonant clusters [5][6]. rtMRI is an ideal modality with which to identify and further characterize these and other attributes of apraxic speech, as it is minimally invasive to the subject and allows for unobstructed viewing of articulatory activity in all parts of the vocal tract. Because of this, rtMRI allows for the identification of unintended, covert speech gestures that cannot typically be detected in the acoustic speech signal that is traditionally used to transcribe disordered speech [7].

Using rtMRI and an analytical method of estimating constriction kinematics based on pixel intensity, we aim to see if rtMRI can shed light on some aspects of apraxic speech articulation that are not evident from acoustic studies alone.

2. Method

An adult male apraxic speaker of English was imaged, using a custom MRI protocol [8][9], while producing a diverse corpus of spontaneous speech, repeated lexical repetition tasks, and self-paced repetitions of word-pairs designed to elicit speech errors. The subject lay supine in the scanner bore throughout the experiment, and was able to interact with the experimenter through an intercom system.

Spontaneous speech was elicited by asking the subject to respond to questions about general topics of interest. The subject was then prompted to repeat a series of short phrases and single words (presented orally by the experimenter) ten times in random order. Stimuli were selected from a standard lexical set commonly used by clinicians in determining a patient's degree of motor speech impairment. Target words contained a variety of consonant sounds and consonant cluster sequences. Finally, the subject repeated the phrase 'cop-top' as many times as possible, over two separate trials.

2.1. Data Acquisition

Image data were acquired on a 1.5T GE Sigma scanner, using a 13-interleaf spiral gradient echo pulse sequence (TR = 6.5 msec, FOV = 200 × 200 mm, flip angle = 15◦) and a head and neck receiver coil. The scan plane (3 mm slice thickness) was located midsagittally; image resolution in the sagittal plane was 68 × 68 pixels (2.9 × 2.9 mm). New image data were acquired at a rate of 18.52 frames per second, and reconstructed as 22.41 frames/sec. video using a sliding window technique. Audio was recorded inside the scanner at 20 kHz simultaneously with the MRI acquisition, and subsequently noise-reduced [10]. The resulting companion video and audio recordings allow for dynamic visualization of the entire midsagittal plane of the subject's vocal tract during speech.

2.2. Articulatory Analysis

For each task in the experimental corpus, audio and MRI video recordings, and MR image frame sequences of the subject's speech were examined. Disfluencies in production and prosodic abnormalities were noted, along with any speech errors detectable in the audio signal and/or articulatory data. For each speech error observed, articulatory coordination of the lips, tongue tip and tongue body was examined in the temporal vicinity of the target lexical item. Acoustic and articulatory data

from word repetition trials was compared to data reported in past studies [11] from fluent speakers performing the same tasks. Incidence of speech errors made by the apraxic speaker was estimated by calculating the ratio of onsets containing a deletion or intrusion error to the total number of onsets, using the time series described below [12].

For selected tasks, time series showing the articulatory activity in regions of interest (lips, tongue tip, tongue body; Figure 1) were automatically generated by calculating the mean intensity of highly correlated pixels in each region – a metric which has been found to provide a robust estimate of constriction degree in noisy data [12].

Figure 1: *Vocal tract regions (l-to-r: labial, alveolar, velar) within which articulatory activity is estimated from correlated pixel intensity* (details in [12]).

3. Results and Analysis

3.1. Gestural Coordination in Speech Errors

By examining the frequency and nature of gestural intrusion and deletion errors made by speakers, much can be inferred about their ability to coordinate vocal tract gestures appropriately such that linguistically meaningful segments are formed. When producing repeated sequences such as 'top-cop', in which /t/ and /k/ gestures are in 1:2 frequency locking with the /p/ gesture, normal speakers produce intrusion errors, in which gestures for /t/ and /k/ are coproduced [11]. The number of intrusions was found to increase with speech rate: with acceleration, it becomes increasingly difficult to maintain the 1:2 mode of coupling, and speakers shift into the simpler mode of 1:1 coordination, in which both onsets /t/ and /k/ are produced simultaneously before each coda /p/ iteration [11]. Past studies using EPG [12] provide indirect evidence that such gestural intrusion error or "misdirected articulatory gesture" frequency is higher for apraxic speakers than for normal speakers.

3.1.1 Apraxic Speech vs. Normal Speech

Audition of the acoustic recordings of the word pair repetition task by apraxic and normal controls reveals no audible difference in error frequency. However, comparison of articulatory time functions in labial, coronal, and dorsal regions reveal striking differences between the fluent and apraxic speakers with respect to intrusion error frequency. Compared to published speech error data from normal speakers performing the same linguistic task at a comparable speech rate [11], a much higher frequency of covert intrusion errors is evident in the apraxic speech (~40% error rate; c.f. ~15-20% for fluent speakers). Error rates were higher in the apraxic speech, even in trials produced with fewest errors by the apraxic subject.

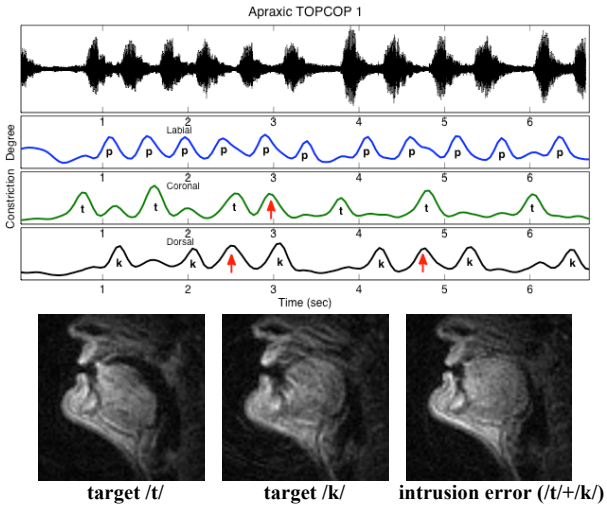

Figure 2, Top: *Acoustic waveform and time-aligned estimated constriction functions (labial, coronal, dorsal) in* /kɒp-tɒp/ *repetition task: apraxic speech.* Bottom: *MRI frames showing articulatory postures for target* /t/*, target* /k/*, and first intrusion error (coproduced* /t-k/*).*

Articulatory analysis of apraxic speech in these trials (Fig. 2) reveals that these are not simple *substitution* errors, (e.g. tongue tip gesture for /t/ replaced by a tongue dorsum gesture for /k/). Rather, they are true *intrusion* errors (red arrows), whereby the tongue tip and tongue dorsum gestures are coarticulated, with neither being reduced in amplitude. Instead of producing the target alternation elicited in this task – /k/ and /t/ each occurring once for every two /p/s, exhibiting 1:2 frequency locking – the apraxic speaker shifts momentarily to the intrinsically simpler mode of 1:1 frequency locking, producing the gestures for /p/, /t/, and /k/ simultaneously, so that the lingual consonants no longer occur before every *second* repetition of /p/ (1:2), but before *each* repetition of the labial coda consonant (1:1).

3.1.2 Variability in Apraxic Speech

Speech errors due to Apraxia of speech are often noted to be made inconsistently, with a speaker making an error on one instantiation of the utterance, then producing it according to target on the next [3].

Articulatory data in Figure 2 (top) reveal that in the first *top-cop* repetition trial, alternating /t/ and /k/ onsets before coda /p/ produced by the apraxic speaker are accompanied by coronal and dorsal intrusion errors (red arrows). In the second elicitation of this sequence (Figure 3), the speaker does not at any point make regular alternations between /t/ and /k/ onsets. Instead, tongue tip and lip gestures are made synchronously throughout, exhibiting an intrinsically stable 1:1 coordination pattern, while tongue dorsum gestures rarely appear at the expected times.

3.1.3 Covert intrusion errors in repeated speech

Covert gestural intrusion errors are not limited to the context of repetitive speech tasks, but also surface regularly in speech produced during a speak-after-me shadowing task. Figure 4 illustrates the presence of a tongue tip intrusion error during the labial closure for the initial /b/ of "bow" in the phrase *"I can type* 'bow know' *five times"*.

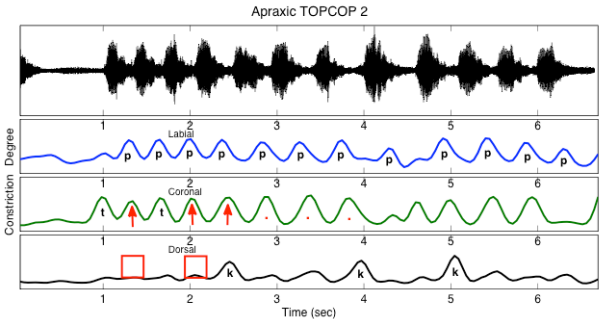

Figure 3: *Acoustic waveform* (top) *and time-aligned estimated constriction functions (labial, coronal, dorsal) in second* /kɒp-tɒp/ *repetition trial: apraxic speaker. Labial and tongue tip gestures coordinated in-phase (synchronously)* (arrows)*. Dorsal gestures are missing at expected times* (boxes)*.*

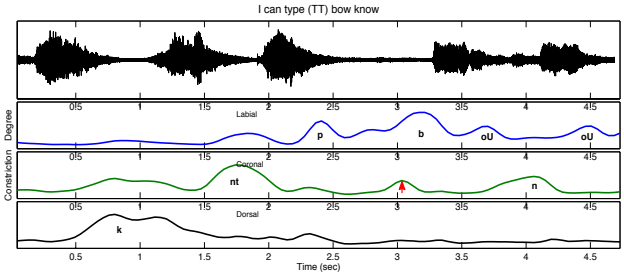

Figure 4: *Audio signal* (top) *and time-aligned labial, coronal & dorsal constriction functions reveal covert tongue tip intrusion during labial closure for* /b/ *in apraxic utterance* "I can type 'bow know' five times."

Some of the individual words produced by the subject also reveal evidence of gesture intrusions. Acoustic analysis of one of the subject's responses to the stimulus item "federation" suggests a form which might be represented in close transcription as [ɹɛdəɹeɪʃən]. Articulatory analysis of the same utterance using rtMRI provided additional insights: the segment transcribed as [ɹ] was found not to arise from simple "anticipatory substitution" (segment [ɹ] replaces segment [f]); rather, the initial labial gesture of target /f/ was observed to be synchronously produced with an anticipatory lingual intrusion gesture for [ɹ]. Target and erroneous productions of the initial portion of the word "federation" are compared in Figure 5.

3.1 Multiple Initiation Gestures in Spontaneous Speech

Consistent with earlier findings [5], multiple initiation gestures by this subject were found to be more frequent in spontaneous speech than in imitated speech. Segments used word-initially by the apraxic subject in spontaneous speech included /t/, /d/, /g/, /k/, /b/, /dʒ/, /w/, /f/, /s/, /n/, /m/, /l/, /ð/, and /h/. Tokens exhibiting multiple initiation gestures are defined as those in which a visible articulatory gesture occurs at least once before complete production of the word.

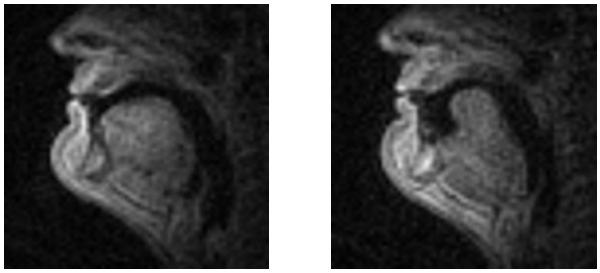

Figure 5: *Two productions by apraxic subject of initial fricative in lexical item "federation."* Left: *target production;* Right: *labial constriction co-produced with intrusive lingual gesture corresponding to tongue posture observed during* [ɹ] *production later in same word. Labialization not discernible in acoustic signal of second token.*

Words exhibiting multiple initiations included those starting with /w/ (100% of /w/ tokens), /dʒ/ (50% of tokens), /s/ (16%), /m/ (20%), /l/ (20%), /t/ (50%), and /d/ (50%). It is noteworthy that the majority of these segments that were problematic for the subject require more than one vocal tract gesture. This suggests that the added complexity of gestural coordination required for the production of multi-gestural segments might present additional challenges in planning, with the result that they exhibit false starts, or perhaps explicit articulation as part of the planning process. In many cases, the repeated initiation gestures are covert, being articulated without any phonation, and could not, therefore, be captured by acoustic analysis or adequately represented in standard phonetic transcription.

In Figure 6, a silent tongue tip gesture (arrow) can be seen to precede full (and audible) production of the coronal-initial word *no* [noʊw] in the utterance *"I can type 'bone no' five times"*. Under acoustic analysis, the duration of the covert gesture might be interpreted as a pause, however using rtMRI, we find that a complete tongue tip gesture is present.

Evidence for covert gestural rehearsal is also shown in Figure 7, where three silent tongue tip gestures can be observed in the interval before the vocalized production of the coronal stop which initiates successful production of the complete lexical item *temperatures*.

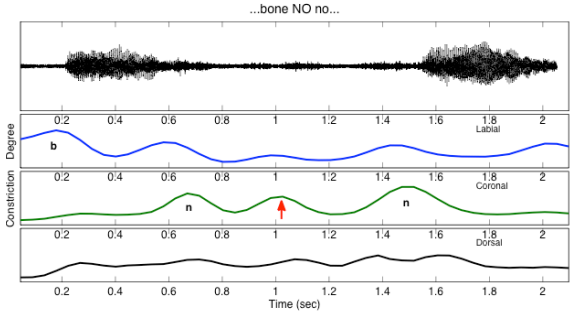

Figure 6: *Covert tongue tip gesture during first (silent) attempt at producing coronal-initial word* 'no' *in the utterance "I can type 'bone no' five times."*

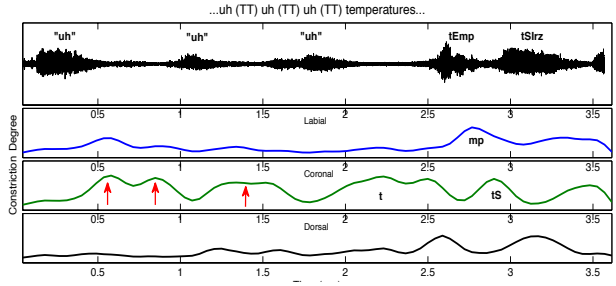

Figure 7: *Three covert tongue tip gestures proceeding successful (vocalized) production of coronal-initial word* 'temperatures.'

3.2 Covert Articulation in Imitated Speech

Perhaps the most striking feature observed in the speech of this subject is the covert articulation of entire lexical items. In multiple instances, the apraxic speaker fully produces all supralaryngeal gestures for entire words, although these articulations were not accompanied by phonation. Articulatory organization during the production of the sentence *'I can TYPE KNOW know how (stumbles) bow know five times'* is illustrated in Figure 8, where the constituent tongue tip and labial gestures corresponding to the words "type" and "know" can be clearly identified, though, as evidenced by the acoustic waveform, are produced silently.

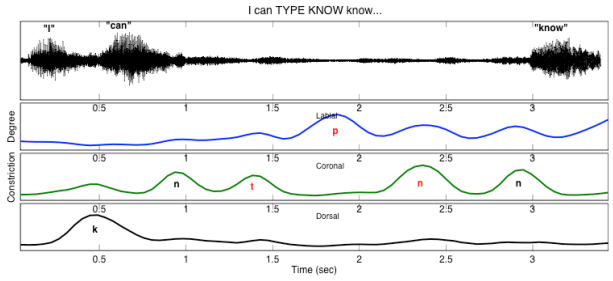

Figure 8: *Covert production (see attenuated acoustic signal,* top*) of entire lexical sequence* 'type know' *in utterance "I can type know how…"*

4. Discussion

A major contribution of this work is to demonstrate that there are many components of apraxic speech that are not readily detected by traditional behavioral diagnostic methods relying on acoustic data and impressionistic phonetic transcription.

Comparing normal speech to apraxic speech by examining both the acoustic record and rtMRI data, we find that while the apraxic and normal speech in repeated speech sequences is very similar acoustically, there are considerable differences in the articulation patterns used in each. In particular, the apraxic speaker exhibits a substantial number of covert (silent) gestural intrusion errors, evidence of shifting into a more stable 1:1 mode of gestural coordination, while the normal speaker does not. Further, it is found that the apraxic speaker produces repetitions of the same token quite variably, though always exhibits a preference for the intrinsically stable mode of 1:1 coordination of gestures. Importantly, we find that these intrusion errors occur frequently in non-repetitive apraxic speech, as well.

Multiple covert initiation gestures are visible using rtMRI during periods of acoustic silence. Given that there is no phonation during the initiation gestures, and that these covert gestures occur for entire words, it is possible that rather than corresponding to true initiation attempts, these gestures might correspond to articulatory rehearsals of the lexical item at hand. Multiple initiation gestures are often found for segments requiring coordination of more than one vocal tract gesture. Perhaps due to the complex nature of this coordination, the motor programs or executions for these segments take longer to plan, and explicit articulatory rehearsal is beneficial. Expanding this study to include samples of apraxic speech from multiple speakers will provide more insights into whether particular segments (i.e. those requiring multiple gestures) are systematically problematic for speakers with Apraxia, and will in turn aid in tailoring therapy programs accordingly. Since the AoS of the patient at hand seems to be characterized by discoordination of articulatory gestures rather than inability to reach a gestural target, a tactile-kinesthetic treatment such as PROMPT would not likely be most effective. Treatments focused on articulatory tasks that the data suggest are most problematic for the speaker; namely, the production of segments or segment sequences requiring multiple gestures to be coordinated anti-phase or eccentrically, would likely be far more effective.

This study demonstrates that using rtMRI, covert speech gestures in apraxic speech can be observed and can be quantified. These data suggest that acoustic analysis of apraxic speech alone provides insufficient information into the characteristics of this type of speech, and that important new insights into the nature of this disorder can be obtained from the use of multi-modal phonetic sensing methods, including rtMRI.

5. Acknowledgements

Research supported by NIH Grants R01 DC007124-01 and DC008780-05

6. References

[1] J. Perkell, M. Cohen, M. Svirsky, M. Matthies, I. Garabieta, and M. Jackson, "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *JASA*, vol. 92, no. 6, pp. 3078-3096, Dec 1992.

[2] J. Westbury, G. Turner, and J. Dembowski, "X-Ray microbeam speech production database user's handbook," Univ. Wisconsin, Tech. Rep., 1994.

[3] R. Wertz, L. La Pointe and J. Rosenbeck, *Apraxia of Speech in Adults, The Disorder and Its Management*. Orlando: Grune & Stratton, 1984.

[4] W. Ziegler, "Differential diagnosis of AoS," In *Speech Motor Control in Normal and Disordered Speech*, B. Maassen et al., Eds., OUP, 2004.

[5] J. Duffy, *Motor Speech Disorders*. St. Louis: Mosby, 1995.

[6] J. Ogar, H. Slama, N. Dronkers, S. Amici and M. Gorno-Tempini, "Apraxia of speech: An overview," *Neurocase*, 2005, 11(6), 427-431.

[7] V. Fromkin, "The non-anomalous nature of anomalous utterances," *Language*, 47, 27-52, 1973.

[8] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *J. Acoust. Soc. Am.*, vol. 115, no. 4, pp. 1771-1776, 2004.

[9] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time MRI," *IEEESPM*, vol. 25, no. 3, pp. 123–132, 2008.

[10] E. Bresch, J. Nielsen, K. Nayak, and S. Narayanan, "Synchronized and noise-robust audio recordings during realtime MRI scans," JASA, vol. 120, no. 4, pp. 1791-1794, 2006.

[11] L. Goldstein, M. Pouplier, L. Chen, E. Saltzman and D. Byrd, "Dynamic action units slip in speech production errors," *Cognition* (103/3), 386-412, 2007.

[12] A. Lammert, M. Proctor, and S. Narayanan, "Data-driven analysis of realtime vocal tract MRI using correlated image regions," in *Proc. InterSpeech*, Makuhari, Japan, 2010.

[13] M. Pouplier and W. Hardcastle, "A re-evaluation of the nature of speech errors in normal and disordered speakers," *Phonetica* (62), 227-243, 2005.