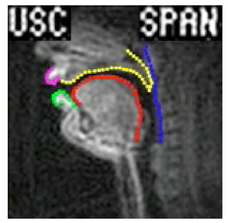


Beyond acoustic data: Characterizing disordered speech using direct articulatory evidence from real time imaging

Ed Holsinger[a], Vikram Ramanarayanan[b], Dani Byrd[a], Louis Goldstein[a], Maria Luisa Gorno Tempini[c], Shrikanth Narayanan[b]
[a]USC Linguistics Department, [b]USC Viterbi School of Engineering, [c]University of California, San Francisco

Motivation

A major avenue of research in phonetics investigates the causes and properties of atypical speech. Tokens of atypical speech, such as speech errors, may occur for a variety of reasons:

- Brain Lesions
- Speech Rate
- Processing Difficulty
- Semantic & Phonetic Activation

To date, a large majority of research into atypical speech, and particularly speech errors, has primarily focused upon the acoustic record. Most often acoustic recordings are used to produce text transcriptions which are then analyzed.

However, recent articulatory research of both normal and abnormal speech suggests that in many cases of atypical speech the acoustic record is at best incomplete. Given the large body of literature and theory founded upon transcription analyses, these new articulatory findings are problematic.

We examined the articulatory record of an individual with degenerative brain lesions, which lead to an apraxia of speech. We set out to investigate the degree to which the speech record produced by trained transcribers reflected the actual underlying speech behavior of the participant.

Data & Method

We collected real-time MRI with synchronous noise-canceled audio in a speak-after-me shadowing task. Our participant suffered from a degenerative brain disorder which resulted in apraxia of speech as well as other cognitive impairments. They produced a total of 18 words, with 5 repetitions of each for a total of 90 tokens.

Each token was transcribed by two separate trained transcribers. The transcribers disagreed on 4 of the 90 tokens. These were excluded from further examination.

The transcriptions were then examined and compared by hand to the articulations evident from in the real-time MRI videos. Where there was a discrepancy, the token was marked as inaccurate and noted for the source of the inaccuracy.

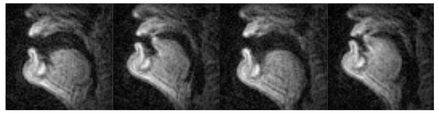
Where possible the acoustic record was also examined, with the idea that the participant's actual acoustic record might alleviate some of the inaccuracies in the transcription. Due to the noise created by the MRI, even with our noise-cancled audio signal, the results of this investigation are partially inconclusive.

In addition to our shadowing task our speaker also produced several spontaneous utterances. These were not used in our analyses, as the nature of spontaneous speech makes it difficult to separate articulatory problems from other higher-level issues with language production.

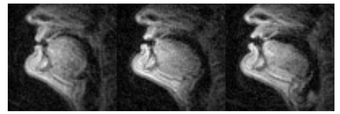
Examples

Overt Articulations

Overt articulations are those articulations that could potentially have some reflex in the acoustic signal, but did not appear in the acoustic or transcription records. This is problematic for analyses of abnormal speech behavior which rely upon such records.



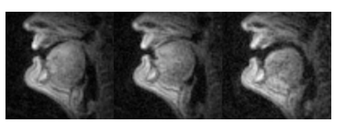
Production of initial /n/ in the word *know*. Speaker produces two nasals, only the second has sufficient airflow to be heard. Stutters of this sort were very common in this individual's speech, this case being particularly problematic, as airflow considerations made the first gesture invisible to the acoustic record.



Production of a medial /st/ cluster in the word *statis**t**ical* (middle) with a typical /s/ (left) and typical /t/ (right) for comparison. Despite blending of the two phonemes across the syllable boundary, the transcription appears normal. *Slurred speech* was a common note in the transcription, in this particular instance the articulatory data allows us to make claims regarding the *source* of the slurred speech.

Covert Articulations

Covert articulations are those articulations that are invisible to the acoustic record for reasons of aerodynamics. Our participant exhibited a large number of coronal intrusions during normal speech, sometimes even for words with no coronal gesture. Covert articulations occur when an articulatory gesture is made and released while the oral cavity is closed for independent reasons.



Coproduction of a coronal stop during a velar closure. The coronal stop was not transcribed as it was released before the velar closure, and was thus invisible to the acoustic record.

Coproduction of a coronal stop during a labial closure. Again, the coronal stop was not transcribed and invisible to the acoustic record.

Deviant Articulations

Deviant articulations are those that are very dificult to account for given the transcription or acoustics. For our speaker, examination of the transcriptions showed a pattern in which the speaker habitually replaced initial /s/ with /v/. The articulatory data suggest, however, that this error is not simply replacement of one phoneme with another, and allows us to suggest a hypothesis about why this error occurs.

Production of what was reliably transcribed as a /v/ (middle) with a typical /v/ (left) and typical /s/ (right). Looking at the articulatory record, we can see that his production of /v/ in these cases is very different from his production of a normal /v/ or of a normal /s/. Indeed we hypothesize from the video, that he is attempting to make an /s/ (or a /z/, as the speaker has trouble with voice onset independently of this issue), but is only moving the jaw, not the tongue tip, during the gesture.

Results

Comparing the transcriptions to the articulatory images revealed the transcription to be an incomplete record of articulatory motion in 34% of tokens (31 tokens). Additionally, for the majority of these tokens the inaccuracy cannot be ascribed to the transcriber.

14% of tokens included clear gestures which were missing from the transcription for various reasons. Many of these gestures were silent coronal intrusions, or silent stutters.

9% of the tokens included a gestural blend that did not appear in the transcription. Blending occurs when multiple gestures are produced simultaneously or overlapping to a larger degree than normal. In most cases the blend was transcribed as one phoneme.

The remaining 11% contained direct errors in the transcription. In many of these cases the phoneme in question was transcribed as a /v/ or /f/ while the participant's articulatory records show a very different gesture.

Several example cases are presented to the left.

Conclusions

Based upon this case study it seems clear that an accurate characterization of atypical speech requires articulatory information. In this regard real-time MRI provides some great advantages:

- Transcription and audio records can still be created and examined, the MRI video only provides an additional source of information
- It does not require researchers to make decisions about which articulators to examine
- It provides detailed information regarding articulator motion and shape, not just point information
- Speaking while prone is somewhat unnatural, however there are no objects inserted into the vocal tract that may introduce additional issues, particularly with speakers with articulatory disorders

Like any method, however, real-time MRI does have its drawbacks:

- While information regarding the upper vocal tract is very rich, the lower vocal tract, particularly glottal activity, is more difficult to examine
- The noise generated by the MRI degrades the acoustic signal, and this noise is difficult to remove

Despite these drawbacks, it is evident that care must be taken when inferring speech behavior from transcriptions alone. This is particularly true when analyzing atypical speech.

Selected References

Byrd & Harris. (2007). Identifying and evaluating apraxic speech deficits using magnetometry. Proceedings of the XVIth International Congress of Phonetic Sciences (Saarbrucken, Germany).

Fowler. (1996). Listeners do hear sounds, not tongues. Journal of the Acoustical Society of America, 99, 1730-41.

Browman & Goldstein. (1995). Dynamics and articulatory phonology. In van Gelder & Port (Eds.), Mind as Motion (pp. 175-193). Cambridge, MA: MIT Press.

Goldstein, & Fowler. (2003). Articulatory phonology: a phonology for public language use. In Meyer & Schiller, Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities (pp. 159-207). New York: Mouton.

Goldstein, Pouplier, Chen, Saltzman & Byrd. (2007). Dynamic action units slip in speech production errors. Cognition, 103, 386-412.

Hartsuiker, Pickering & de Jong. (2005). Semantic and phonological context effects in speech error repair. Journal of Experimental Psychology. Learning, Memory and Cognition, 5, 921-32.

Ladefoged. (1967). Three areas of experimental phonetics. London: OUP.

McGurk & MacDonald. (1976). Hearing lips and seeing voices. Nature 264, 746-8.

Pouplier. (2007). Tongue kinematics during utterances elicited with the SLIP technique. Language and Speech, 50.