QUANTIFYING ULTRASOUND DATA FROM A TONGUE TWISTER EXPERIMENT USING CURVE-TO-CURVE DISTANCE

Karen Reddick & Stefan A. Frisch

University of South Florida
karenr@mail.usf.edu; sfrisch@usf.edu

ABSTRACT

This study examined the production of four-word tongue twisters (e.g. *top cap cop tab*) by six English participants using ultrasound imaging of the tongue. Onset consonant closures for /t, d, k, g/ in the contexts of /a/ and /æ/ were extracted from the video and traced using EdgeTrak software. Traces were analyzed using a mean minimum point-to-point distance measure between curves [12]. Curves were compared within allophonic group (e.g. [ka] in *cop* vs. [ka] in *com*) and across phoneme category (e.g. [ka] in *cop* vs. [ta] in *top*) to quantify similarity to the desired production and expected error. Perceived speech errors had greater within-group distance and smaller between-group distance, consistent with the production of the wrong target. Productions that deviate significantly from other productions within the group can be classified as gradient errors. By these measures both categorical errors and gradient errors were relatively rare.

Keywords: Speech error, ultrasound

1. INTRODUCTION

This study investigates errors on speech sounds (or phonemes) produced in speech. Speech errors themselves can be defined in reference to how the intended sound is realized, or how a listener perceives it. Mowrey & MacKay [8] stated that a production was erroneous if the output differed from that which the participant intended, however subtle, and no matter its effect upon the perception.

1.1. Speech error elicitation

Speech errors appear to follow common patterns [5]. Consonant errors will likely be similar to their targets (the phoneme the participant intends to produce), in fact, front and back consonants were more frequently interchanged with each other than with consonants having other places of articulation. An example of this would be a velar /k/ replacing an alveolar /t/. Errors predominantly occur in word-onset position and word-onset errors can be elicited in phrase form tongue twisters [10].

1.2. Instrumental studies of errors

In large part, the body of speech error studies had been based solely on listener perceptions. Recently studies have challenged this as a data collection procedure and demonstrated the need for more in depth analysis.

- Mowrey & MacKay [8] challenged the characterization of speech error data as phonemic through the use of EMG recording of muscle contraction during tongue twister productions.
- Frisch & Wright [2] used experimentally elicited speech errors with acoustic measures that compared perceptions from careful listening to the acoustic waveform. Looking at duration, amplitude, and percent of voicing they examined errors produced between fricatives /s/ and /z/.
- Pouplier & Goldstein [9] also addressed differences between production and perception using EMA data. The errors analyzed in this study demonstrated that less frequent elements are usually replaced by more frequent elements, but based on the errors analyzed, /k/ is more likely to replace /t/ than vice versa (Pouplier & Goldstein 2005).

All three studies conclude that there are some errors that transcend the current system of recording and certain combinations did appear that were outside of the realm of transcription using the accepted systems.

Stearns [11] related the perception of speech errors by naïve listeners to their observed production in ultrasound video based on measures of tongue dorsum height (indicative of a /k/ or /g/) and the elevation angle of the tongue tip (indicative of a /t/ or /d/). Differences in perception were expected between errors labeled as categorical or gradient but very little variation in perception was found; the sounds were heard on the basis of the closest typical sound [11].

McMillan, Corely & Lickley [6] used EPG to examine the articulation variation between words and their competitor in elicited speech errors based on contextual factors rather than category. McMillan & Corley [7] examined speech errors in tongue twister tasks using EPG, ultrasound and VOT. They determined that there was greater gradient variation in the articulation of a phoneme when it differed from its competitor by only one feature, and less gradient variation if a phoneme and its competitor differed by two features.

1.3. Ultrasound study of speech

In this project, ultrasound video of the tongue was used to record tongue twisters and analyze them for errors. Ultrasound has come into popular use because it is relatively inexpensive and does not require drastic individualization to the participant [1], also it is non-invasive and harmless [3]. Ultrasound is especially suited for research into tongue shape and position as a whole. The images of the oral cavity created by ultrasound are only of the soft tissue, this being the tongue from tip to root, neither bone nor air is imaged by ultrasound, as such, the hyoid, mandible, and sublingual cavity appear as shadows [3].

2. METHODS

2.1 Stimuli

There were 14 words arranged into four-word tongue twisters in two parts (Table 1). The baseline used twisters with all alveolar or velar onsets, to determine typical production. The experimental part used tongue twisters with alternating velar and alveolar onsets, to elicit errors.

Table 1: Tongue twister stimuli.

Stimuli	
Baseline	Experimental
tap top Tom tab	top cap cop tab
gap gob gob gab	gap Don dam gob
cop cap cab com	Tom cab com tap
dam Don Don dab	dab gob gab Don
cap com cop cab	Don gap gob dam
gob gap gab gob	cab Tom tap com
top tab tap Tom	gob dab Don gab
Don dam dab Don	cap top tab cop

2.2 Participants

Six native English participants, undergraduate students at the University of South Florida, were paid for their participation. All reported no history of speech or hearing disorders.

2.2 Procedure

The participant was seated in a stabilizing framework with an acoustic standoff. The ultrasound probe was fixed under the chin relative to the stabilized head. The participant was asked to read stimulus at a normal rate of speech. Each stimulus was repeated 6 times for 96 target productions of each onset consonant.

2.3 Measurement

2.3.1. Tongue tracing

Ultrasound images of the stop closure (Figure 1) were traced using Edgetrak [4]. A tongue curve was output from Edgetrak as set of 100 data points. Figure 2 shows traces from /t, d/ onsets in twisters for two participants (/a/ context). Of note, there is a visible difference between participants 7 (left) and 8 (right) in the apparent stability of their productions.

Figure 1: Ultrasound image of stop closure.

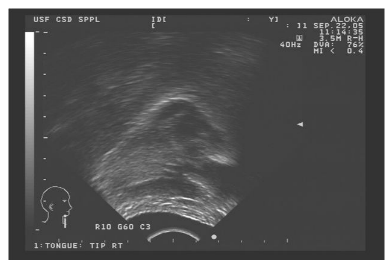
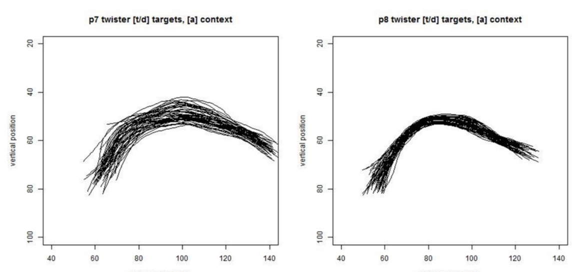


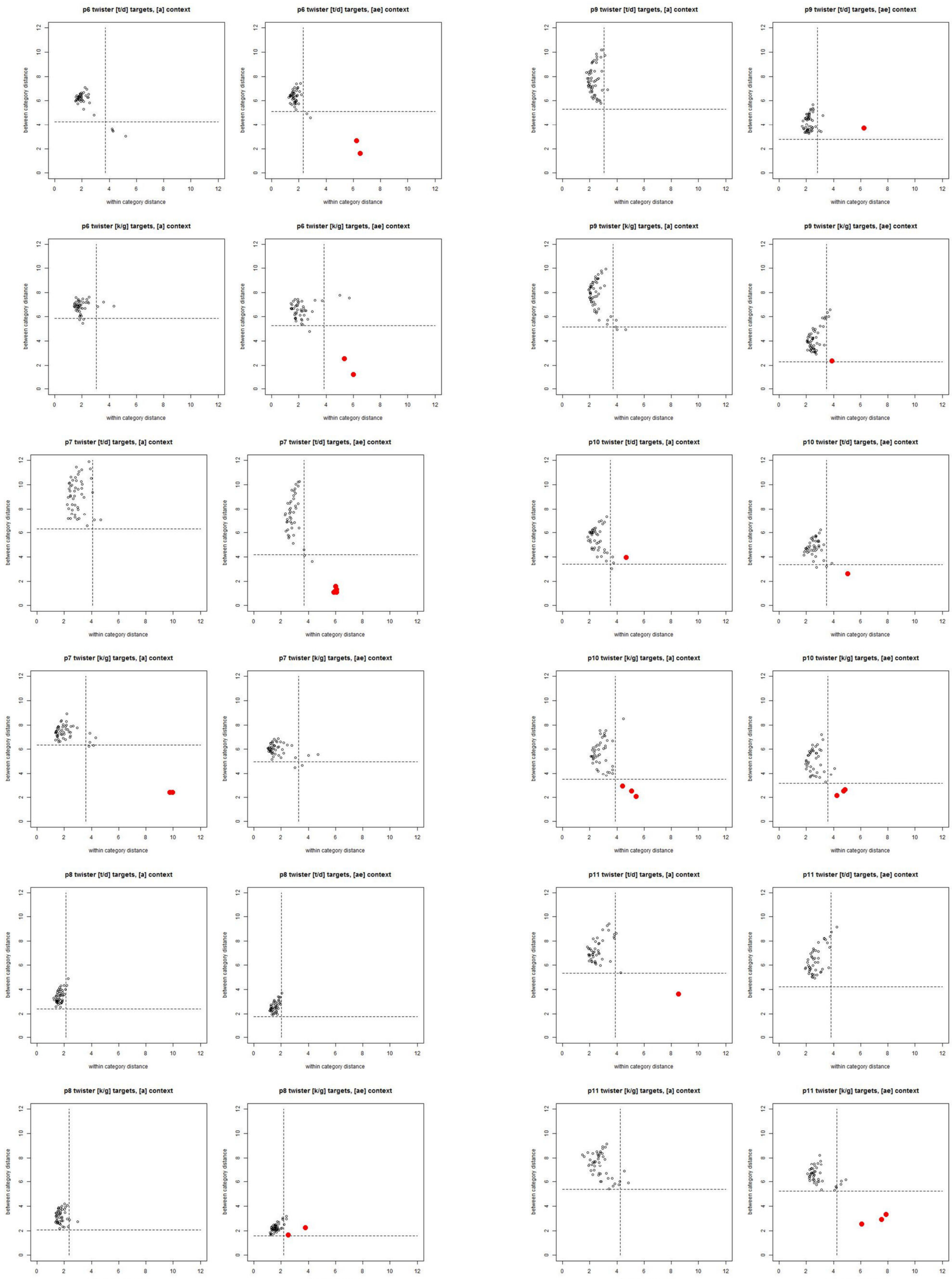
Figure 2: Edgetrak traces of twister /t, d/ for two participants, /a/ context.



2.3.2. Curve-to-curve analysis

Given a set of two curves, the mean curve-to-curve distance was determined using a point-to-point analysis. Each point on Curve A was compared to every other point on Curve B. The smallest distance value for each of the 100 points on Curve A was determined. The mean of these 100 values was taken. This value is the curve-to-curve distance between Curve A and Curve B.

Figure 3: Curve to curve distance for production organized by each consonant place and vowel context, compared to same place (within, x-axis) and other place (between, y-axis). Red points indicate perceived speech errors.



For each participant, curve-to-curve distance was computed for onsets at the same place of articulation (alveolar, velar) in the same vowel context (/a/, /æ/) in the baseline stimuli as a measure of articulatory stability or accuracy.

Curve-to-curve distance was also computed across places of articulation within the same vowel context in tongue twister stimuli as a measure of the degree to which a production resulted in articulation similar to the desired elicited error.

3. RESULTS

3.1. Measure of individual productions

Figure 3 shows each production in the experimental stimuli by each participant as a single point according to the distance within and between place of articulation group (by vowel context). Graphs are grouped by participant, place of articulation, and vowel context. Larger solid markers are used for productions perceived as errors (typically in the lower right quadrant). Dotted lines indicate 2 SD from the mean for the within group distance (same place of articulation, vertical) and the between group distance (alternative place of articulation, horizontal). These graphs show individual differences in variation and perceived speech error rate across the six participants.

3.2. Correlations between measures

Table 2 shows aggregate measures for each participant.[1] Measures include:

- Mean curve-to-curve distance in mm for baseline productions of the same place of articulation in the same vowel context (BB_Stability). Calculation of this value excluded all productions heard in error or that fell more than 2 standard deviations from the mean of its category. It is a measure of the average distance between productions at baseline within a single participant.
- Mean curve-to-curve distance in mm for twister productions of the same place of articulation in the same vowel context (TB_Stability). Calculation of this value excluded all productions heard in error or that fell more than 2 standard deviations from the mean of its category. It is a measure of the average distance between non-error productions in the twisters within a single participant.
- The rate of occurrence of the productions that fall outside of 2 standard deviations of the mean for their production category that were not heard as errors (TG_Rate). This is the rate of gradient error production expressed as a percentage.
- The rate of occurrence of productions that were heard in error within each place and vowel expressed as a percentage (TE_Rate).

Table 2: Aggregate participant measures.

	BB_Stability	TB_Stability	TG_Rate	TE_Rate
p6	2.14	1.87	5.6%	2.1%
p7	3.17	2.26	7.0%	3.5%
p8	1.56	1.61	4.0%	1.0%
p9	2.48	2.35	5.4%	0.9%
p10	2.19	2.59	4.3%	4.3%
p11	2.31	2.65	4.8%	2.1%

Figure 4: Correlations between baseline stability and twister stability, gradient error rate, and perceived error rate.

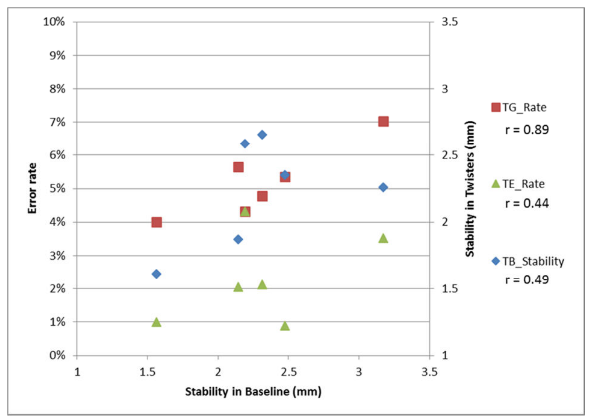


Figure 4 shows measures of twister stability, gradient error rate, and perceived error rate by baseline stability. In addition, correlations between baseline stability and the other measures are given. Across participants, baseline stability provides a moderate to large correlation with twister stability, gradient error rate, and perceived error rate.

Informally, individual differences in stability in tongue position (as seen visually in Figure 2) in the production of tongue twisters predicts both gradient and perceived error rate. Participants with less baseline stability in their productions had less stable tongue twister productions, more gradient errors with significant deviations from the norm, and more perceived errors in their production of tongue twisters. It would appear, therefore, that some level of individual difference in the basic repeatability of consistent speech gestures has an impact on behavioural measures of speech production performance under difficult conditions, such as in tongue twisters.

4. REFERENCES

[1] Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*, 19(6-7), 619-633.

[2] Frisch, S.A., & Wright, R. (2002). The phonetics of phonological speech errors: an acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30, 139-162.

[3] Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association*, 32(2), 113-121.

[4] Li, M., Kambhamettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics & Phonetics*, 19(6/7), 545.

[5] MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8(3), 323-350.

[6] McMillan, C. T., Corley, M., & Lickley, R. J. (2009) Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, 24:1, 44-66.

[7] McMillan, C. T., & Corley, M. (2010) Cascading influences on the production of speech: Evidence from articulation. *Cognition,* 117, 243-60.

[8] Mowrey, R., & MacKay, I. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88(3), 1229 – 1312.

[9] Pouplier, M., & Goldstein, L. (2005). Asymmetries in the perception of speech production errors*. Journal of Phonetics*, 33(1), 47-75.

[10] Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In Keller, E. & Gopnik, M. (Eds), *Motor and sensory processes of language* (pp. 17-51). Hillsdale, NJ: Lawrence Erlbaum Associates

[11] Stearns, A. (2006). *Production and perception of place of articulation errors*. Unpublished Master's Thesis, University of South Florida, Tampa, FL.

[12] Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /ɑ/. *Journal of Phonetics*, 37(2), 248-256.

---

[1] Participants are labelled p6 through p11 in a manner consistent with internal lab records. Participants p1 through p5 were recording using the same procedures but analyzed differently [11].