SEPARATING NORMAL VARIATION IN MOVEMENT AMPLITUDES FROM GRADIENT SPEECH ERRORS

Anneke Slis[1], and Pascal van Lieshout[1,2,3,4]
[1]Oral Dynamics Lab, Department of Speech-Language Pathology, University of Toronto, anneke.slis@utoronto.ca;
[2]Toronto Rehab, Toronto, Ontario, Canada
[3]Department of Psychology, University of Toronto Mississauga, Mississauga, Ontario, Canada
[4]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada

1. INTRODUCTION

The results presented in this paper are part of a larger study that looks at articulatory constraints and its influence on the occurrence of gestural intrusion and reduction errors, as described in a study by Goldstein, Pouplier, Chen, Saltzman, and Byrd[1].

Until recently, speech errors were mostly explained in abstract models of speech production[2], in which the feature or phoneme is the smallest unit of speech. The idea of the phoneme as the smallest unit of speech, however, is challenged by evidence from recent kinematic studies[13]. In these studies, speech errors, when studied at the level of articulatory gestures, frequently resulted into sub-phonemic speech errors. Most errors consisted of the activation of a gesture, not supposed to be involved in that particular sound production, accompanying the production of the actual target gesture. These unintended activations were often gradual in nature. In order to define a gradual error, Goldstein et al. compared the activation level of a gesture, for example tongue dorsum (TD) in alternating trials (e.g. *cop top*) with the mean activation level of the TD in non-alternating trials (see figure 1).

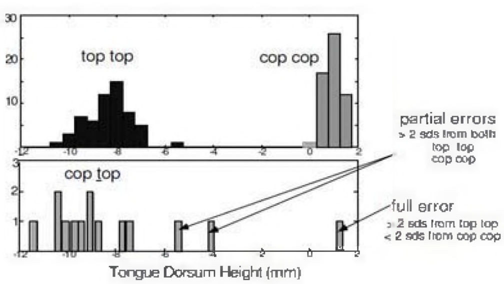

Figure 1. Comparison between amplitudes of TD gestures in alternating and non-alternating trials. From: Pouplier[3], p. 37.

They did indicate that this method introduced the risk of comparing speech sequences with different coarticulatory properties, which could affect the analysis. This issue was reinforced in a recent study where it was found that higher production rates could be achieved in alternating trials, tapa, compared to non alternating trials, like tata, because of less energy consumption for the jaw[4]. This indeed suggests a different production mechanism for alternating trials compared to non alternating trials.

Another aspect that can make it difficult to separate normal amplitude variation from intrusion errors is that the amount of amplitude variability of the articulators differs across types of constrictions and contexts[56], so different production constraints could play a role[6]. Finally, the articulators can be activated as a consequence of active control in forming a constriction gesture or they can move as a passive consequence because of the activation of another gesture[5], in which the amount of passive movement is dependent on the context. Taken all this into account, it is important to determine what the co-articulatory properties of a C1VC C2VC sequence are and how these differ from non-alternating trials C1VC C1VC. This information will help to establish what an intrusion error is, because it allows us to determine the potential contributions of biomechanical constraints, coarticulatory influences, and normal variation[7]. The current study is a first attempt to investigate to what extent changes in movement amplitudes in alternating trials can be related to above mentioned factors. As a first approach, we will use correlations to determine the (in) dependence of gestures and their contributing articulators[589].

2. METHOD

Speech errors were invoked by a repetitive speech task, comparable to the study described in Goldstein et al[1]. Two speech rates were employed, normal and fast, which were individually determined for each participant and controlled by metronome presentation. Movement data were recorded with the 3D EMA system[10]. The raw movement amplitudes were normalized such that the maximum amplitude of a constriction per trial was set to 100 % and the minimum constriction per trial was set to 0 %. This way it is possible to compare across trials and speakers. Maxima for TT and TD during each segment were automatically determined using a peak-picking algorithm. In this paper, data are presented for the bisyllable *topcop* in the normal speaking rate condition for 8 participants. For each separate trial, correlations were calculated between the normalized amplitude of the target gesture and that of the co-occurring non-target gesture in the bisyllables *top cop*, *top top* and *cop cop*. Furthermore, correlations between normalized amplitudes of the same gestures at the target and non-target location were used to determine co-articulatory influences across syllable positions. Finally, correlations are reported for the tongue tip and tongue dorsum after subtracting the contribution of the jaw to assess the latter influence on these measures. It was hypothesized that if the correlation between target and non-target is high, TT and TD are not moving independently from each other and biomechanical constraints play a role[58]. Moreover, it was hypothesized that if the gestures were highly dependent due to co-articulatory constraints, high correlations for the same gesture in target and non-target positions should be found across words.

3. RESULTS AND DISCUSSION

The correlation values are shown in table 1a and b. No consistent significant positive correlations were found between target and non target gestures in the same syllable, except for one participant, C9.

1a

Top Cop										
		With jaw				Without jaw				
	Beats per minute	TT TD top	TD TT cop	TT top TT cop	TD cop TD top	TT TD top	TD TT cop	TT top TT cop	TD cop TD top	n
C6	80	0.01	0.10	0.54	-0.30	0.09	0.8	-0.2	-0.1	16
C7	92	-0.28	-0.06	-0.07	-0.29	-0.23	-0.16	-0.04	0.14	17
C8	90	0.34	0.47	0.02	0.10	-0.2	0.08	-0.05	0.4	17
C9	113	-0.15	0.66	-0.18	0.40	-0.04	0.21	-0.44	0.17	16
C10	90	-0.6	0.03	-0.04	-0.56	-0.59	-0.1	-0.55	-0.48	18
C11	100	-0.44	-0.55	-0.27	-0.44	0.18	-0.29	-0.58	0.03	17
C14	90	0.20	0.40	0.13	[illegible]	0.25	0.2	-0.03	-0.35	16
C19	90	0.3	0.45	-0.46	-0.46	-0.74	0.25	-0.12	0.05	17

1b

top top							cop cop				
		With jaw		Without jaw			With jaw		Without jaw		
	Beats per minute	TT TD top1	TT TD top2	TT TD top1	TT TD top2	n	TD TT cop1	TT TD cop2	TD TT cop1	TD TT cop2	n
C6	80	0.04	-0.24	0.02	0.28	17	-0.3	-0.14	-0.22	0.22	17
C7	92	0.75	0.37	-0.3	0.2	17	-0.01	0.47	-	-	17
C8	90	0.47	0.71	-0.18	0.01	16	0.77	-0.04	0.33	0.29	16
C9	113	0.25	0.18	0.55	0.36	17	-0.23	-0.12	-0.02	0.08	17
C10	90	0.33	-0.30	-0.58	-0.37	17	0.59	0.49	-	-	15
C11	100	0.08	0.34	0.12	-0.08	17	0.23	0.41	0.32	0.73	16
C14	90	-0.3	0.69	-0.35	-0.05	17	-	-	-	-	-
C19	90	0.78	0.88	-0.08	0.44	18	0.01	0.38	0.18	0.65	17

Table1a & b. Correlations for TT and TD in both alternating (top) and non-alternating (bottom) trials. The last column shows the number of repetitions in a trial. The marked boxes are significant at $p < 0.01$.

Figure 2 shows movement data for subjects C9 and C6 (who had very low correlations). The significant correlation between TT and TD in the word cop for C9 may indicate a passive movement of the TT during TD, due to the shared organ, the tongue root[1]. However, after subtracting the jaw component, correlations in most cases decreased, suggesting a common influence of the jaw underlying the tongue movements during /k/[8] especially for C9. In general, correlations between TD and TT were low and non-significant, suggesting that biomechanical constraints of the tongue tip and tongue dorsum are not a major factor in our study. This is also in accordance with Jackson & Singampalli[8] who showed similar low correlations between TD and TT. They did find a high correlation between jaw and TT, which would fit our findings for jaw-corrected tongue movements. One exception is C6, who showed a very high correlation in the no-jaw condition for TD and TT in the word *cop*.

With respect to correlations across syllables, no consistent correlations were found between the target gesture and the same gesture in the non target position. This means that the activation of a gesture in the target position doesn't influence the activation of this gesture in a non target position and consequently doesn't contribute to co-articulatory influences. Finally, there are large individual differences, as can be observed in table 1 and figure 2, and it is safe to assume that speakers use individual control strategies that may or may not result in a higher correlation between TT and TD.

In the non-alternating trials more significant correlations between TT and TD were found. The values for the non-alternating trials in the word *top* were slightly higher. However, the values didn't show a consistent pattern, ruling out a biomechanical or co-articulatory constraint. Again, the jaw plays an important role in the higher correlation values. Once the jaw is left out, most of the significant values become insignificant. The fact that non alternating trials show a different pattern than the alternating trials means that these two conditions cannot be compared. Therefore, the use of non-alternating trials for the estimation of intrusion or reduction errors in alternating trials can affect the analysis and may cause an inaccurate determination of normal variation.

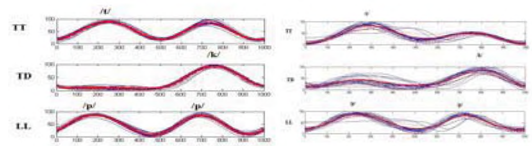

Figure 2. C9 (left) and C6 (right). The y-axis shows normalized amplitude and the x-axis shows normalized time.

REFERENCES

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

[2] Levelt W. (1989). Speaking. From intention to articulation. MIT Press, Cambridge.

[3] Pouplier, M. (2003). Units of phonological encoding: Empirical evidence. Unpublished dissertation. Yale University.

[4] Rochet-Capellan, A. and Schwartz, J.L. (2007). An articulator basis for the Labial-to-Coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/. *Journal of the Acoustical Society of America, 121*(6), 3740-3754.

[5] Browman, C. P. (1994). Lip aperture and consonant releases. *Phonological Structure and Phonetic Form. Papers in Laboratory Phonology III,* Cambridge: Cambridge University Press, 331-353.

[6] Koenig, L.L., Lucero, J.C. & Lofqvist, A. (2003). Studying articulatory variability using functional data analysis. *Proceedings of the 15th International Congress of Phonetic Sciences.* Barcelona, Universitat Auto`noma de Barcelona.

[7] Frisch, S.A. (2007). Walking the tightrope between cognition and articulation: The state of the art in the phonetics of speech errors. In Schutze, C.T. & Ferreira, V.S. (Eds.), Proceedings of the LSA Institute Workshop. MIT Working Papers in Linguistics, 53, The State of the Art in Speech Error Research (pp. 155-171), MITWPL.

[8] Jackson. P.J.B., & Singampalli, V. (2009). Statistical identification of critical articulators in the production of speech. *Speech Communication*, 51(8), 695-710.

[9] Stone, M., Epstein, M. A., & Iskarous, K. (2004). Functional segments in tongue movement. *Clinical Linguistics & Phonetics, 18,* 507-521.

[10] Yunusova, Y., Green, J., & Mefferd, A. (2009). Accuracy Assessment for AG500, Electromagnetic Articulograph. *Journal of Speech, Language, and Hearing Research*, 52(2), 547-555.

ACKNOWLEDGEMENTS

We would like to thank Dr. Aravind Namasivayam for his assistance in running the EMA sessions.