EXAMINING SPEECH PRODUCTION USING MASKED PRIMING

Chris Davis[1], *Jason Shaw*[1], *Mike Proctor*[2], *Donald Derrick*[3], *Stacey Sherwood*[1], *Jeesun Kim*[1]

[1] University of Western Sydney, The MARCS Institute, Australia, [2]Macquarie University, Australia,
[3]University of Canterbury, NZILBB, New Zealand
chris.davis@uws.edu.au, j.shaw@uws.edu.au, michael.proctor@mq.edu.au, donald.derrick@canterbury.ac.nz,
16115314@student.uws.edu.au, j.kim@uws.edu.au

ABSTRACT

The time to initiate naming a printed target word is reduced when preceded by an identical masked prime (match prime) or by one that has the same initial letter (onset prime) compared to an all letter different control. Masked priming has been examined using vocal response time but offers an opportunity to examine speech production dynamics before the onset of speech acoustics. We tracked tongue-dorsum, tongue-tip and lip motion from four participants pronouncing 19 targets in match, onset and unrelated control prime conditions. Control primes were selected so their articulation involved a different tongue gesture than the target. Prime influence was measured by tongue-dorsum height at gestural onset and peak velocity of the subsequent gesture. Results showed that relative to targets in the match condition, control targets had a significantly different tongue dorsum height and the peak velocity was greater when the subsequent gesture was achieved.

Keywords: speech production, masked priming, articulatory dynamics, articulation, EMA

1. INTRODUCTION

There is a long tradition of attempting to perturb the speech production system to understand its operation. For example, methods to experimentally induce speech errors often rely on sequences of overt speech (e.g., tongue twisters) or speech repetition (where repetition entrains specific segmental patterns and errors are induced by violating these). These overt articulatory tasks engage repetitive continuous production whereas other techniques induce perturbations through types of priming, in which entrainment occurs through silent reading (e.g., the SLIP paradigm, [1]). Our study used a masked priming paradigm [2] to examine how phonology activated from written words feeds into speech output. This approach has a number of attractive features: because it only uses a single prime, stimulus properties can be easily controlled; masking of the prime helps minimize strategic contributions, and the close proximity of primes and targets allows the phonological representations and articulation plans of these words to be activated nearly at once. This latter feature makes it a suitable method to examine whether a partially activated prime can influence target articulation.

In the masked priming paradigm ([3]), a clearly displayed upper case target is immediately preceded by a briefly displayed lower-case prime, which is immediately preceded by a clearly presented visual display (typically a series of # marks), which acts as a forward visual mask. The combined action of the forward and backward masks (the target word itself) makes the prime unlikely to be available for conscious report. When the response task involves naming a masked primed target, a specific masked priming effect has been identified (so called “onset priming effect” e.g., [2;4]). The onset effect refers to the reduction in the time to name (produce a vocal response) the clearly displayed target that has been preceded by a rapidly presented masked prime beginning with the same initial letter relative to when preceded by an unrelated control prime. Forster and Davis [2] called this an onset effect, although what they manipulated was only the initial letter of the primes and targets. The onset effect suggests that priming with the naming task involves the process of articulation as it is not found with other response tasks (such as the lexical decision task).

To our knowledge, this is the first time masked priming has been used to examine the speech production process per se. Measuring the dynamics of the tongue and lips allows for a more detailed examination of the production process than does simply measuring the time take to produce a vocalization. For instance, by tracking tongue motion it can be determined how different prime types affect the articulation of the target and whether there is any evidence that the prime has an influence on the way the target has been produced. This is an interesting issue as it bears on whether speech production can be influenced by features from more than a single phoneme at the same time (e.g., consistent with cascading activation).

Due to the novelty of this endeavour, our initial research strategy was to survey a range of possible effects by collecting a relatively large corpus of utterances from multiple speakers (two dialects of English) across differently composed primes and targets. In the current study we report data from a carefully selected sub-set of this corpus in which monosyllabic targets began with labials or coronals and the match, onset and control primes were carefully matched. (e.g., control primes for labial targets began with coronals and had the same rimes as the onset primes).

Given the nature of the target onset (mostly labials), we report the behavior of the tongue-dorsum (TD) at the onset of articulation. The control primes differ from target in both the onset consonant and the vowel. Our motivation for selecting the TD, corresponding to the vowel movement, (and not the tongue-tip, TT, corresponding to the consonantal onset) is that by comparing TD across control and match conditions, we are comparing controlled movements in different directions (see [5]). A problem with examining the TT is that since it is not under control in labials it is free to vary; as such, it would be harder to spot an intrusion against the high variability backdrop of an uncontrolled articulator. Also, the TD is a good choice to see coarticulation effects between planned and produced vowels [6]. Thus the aim was to compare height of the initial TD gesture when pronouncing a target in the control and match conditions to determine whether target articulation (tongue gestures) was influenced by the composition of the prime.

2. METHOD

2.1. Participants

The current study was based on the data from four Australian speakers (two females) selected from a larger corpus: 5 Australian (2 female; average age 29.6, range 20 – 42) and 5 North American (3 female, average age 43.4, range 31-61) speakers.

2.2. Equipment

An NDI Wave (100Hz sampling rate) electromagnetic articulograph system was used to capture non-line-of-sight 3D motion. This system captures sensor positions based on the principles of two-dimensional magnetometers (i.e., a varying strength signal induced in a sensor by means of an alternating electromagnetic field generated by a transmitter). The reported accuracy (RMS) of the Wave is: static positional 0.6 mm; static angular accuracy 0.2 degrees; dynamic positional accuracy 1.5 mm; and dynamic angular accuracy 0.6^0 ([7;8]).

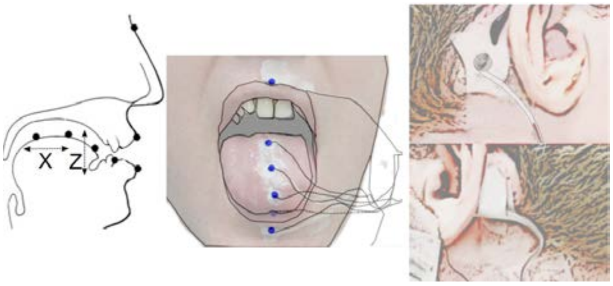
2.3. Stimuli

The full corpus consisted of 63 target words presented three times so that each target would appear in each of three priming conditions (match prime, e.g., tame-TAME, onset prime, till-TAME and control, fill-TAME). The onset and control prime always had the same rhyme. Target onsets consisted of 27 labials, 21 coronals and 15 fricatives, had a mean log HAL frequency (see [9]) of 9.21 (SD 1.68). The current study reports data from 19 Targets (14 bilabials and 5 coronals). For the bilabial targets, control primes were selected where pronouncing the initial prime segment would involve a coronal TT gesture and the following vowel differed in height, e.g., target "BAKE" (/bæɪk/, near low front vowel) with control prime "tend" (/tend/, high mid front vowel). For the coronal targets, the control primes were selected to differ in vowel height/frontedness (e.g., target "TEEN" /tiːn/, prime "walk" /woːk/). The targets had a mean log HAL frequency 8.74 and the primes, onset, 9.20 and control, 9.56.

2.4. Sensor placement

To track tongue movements three 5D sensors were attached to the tongue mid-sagittally, at the tongue tip (TT), tongue body (TB) and tongue Dorsum (TD) (see Figure 1).

Figure 1: Position of the nine 5D Wave sensors. The centre panel shows that in addition to surgical glue (epiglu), Ketac (dental glue) was used to hold the tongue sensors in place. The X and Z labels illustrate horizontal and vertical (height) displacement respectively.



An additional sensor was affixed to participant's lower jaw at the gum line in order to measure jaw motion; one sensor was placed on the upper and one on the lower lip to record lip movements. Three 5D sensors were used to track head motion, these were placed on the participant's nose bridge and the right and left mastoids.

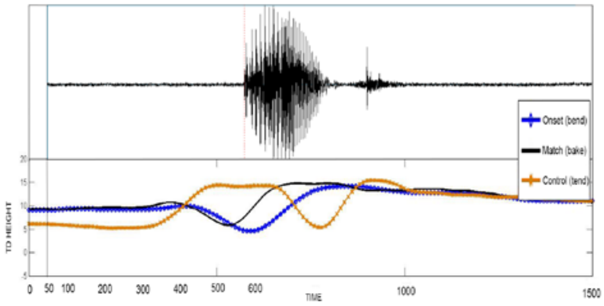
2.5. Data processing

Item presentation timing and vocal recording was accomplished using the DMDX software [10] with synchronization of item display times achieved by cross-correlating the Wave and DMDX vocal recordings and using the record clock on time option in DMDX. Kinematic analyses were carried out using a Matlab based multi-channel visualization application, MVIEW [11], which dynamically displays sensor motion. The onset of the tongue tip gesture was determined semi-automatically by applying a 20% threshold on the tangential peak velocity of the TT sensor.

2.6. Measures

Measurement of the acoustic onset times were conducted in Praat. The onset of the articulatory movement was determined by using a 20% threshold of local peak velocity of the velocity signal of the tongue dorsum (TD). Figure 2 shows an example of TD trajectories for the three prime conditions (where there is a delay in the 'correct' TD gesture for the control condition relative to the others).

Figure 2: Example TD gestures for the three prime types (the three curves superimposed here). Top panel acoustic signal (from the control condition); Lower panel vertical position of the TD sensor for the three priming conditions. Time (ms) begins from the onset of the prime.



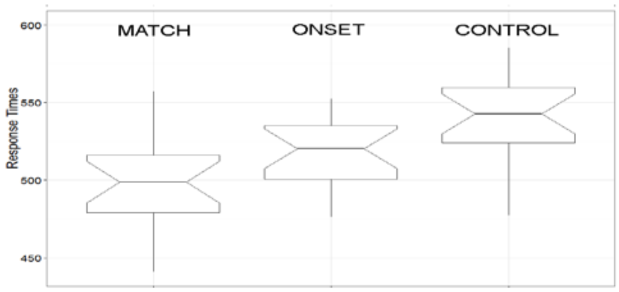
2.7. Procedure

NDI wave sensors were taped to the participant's nasion, left mastoid, and right mastoid, and glued to the midsagittal line of the tongue tip, tongue back as far back as comfortable for the participant, and mid-way on the tongue body. Sensors were also glued to the gum just under the inner lower left incisor, and the midsagittal line of the upper and lower lip next to the vermillion border. Participants were seated at approximately 1.5 meters from a CRT video monitor that displayed each item. Each item consisted of three stimuli. The first was a forward mask consisting of a row of four hash marks (duration 500 ms). This was immediately followed by the prime (duration 50 ms), which was presented in lowercase letters. The prime was in turn immediately followed by the target (duration 500 ms), which was presented in uppercase letters. Each stimulus was centered in the viewing screen and was superimposed on the preceding stimulus. Items were presented on a computer-controlled video display using DMDX software and the timing of the display was synchronized with the video raster. Participants were asked to pronounce aloud the word presented in uppercase letters. No mention was made of the number of stimuli that would be presented on each trial. The vocal response of the target was recorded for off-line analysis. The experiment consisted of three blocks of the same 63 target words (presented in random order).

3. RESULTS

We begin with determining if there was a priming effect in naming response times and then examine the correlation between vocal and gestural onset time. After this we report on the different TD height between the match and control conditions to determine how the prime type affected articulation. Mean naming response times (SE) are shown in Figure 3.

Figure 3: Mean response time to initiate naming the target as a function of prime condition.



A linear mixed model (LMM) analysis (random-slopes and random-intercepts) using Kenward-Roger approximation for degrees of freedom indicated that effect of priming was significant, F(1,2.86) = 11.043, p = 0.048 (a repeated measures item-based ANOVA also indicated the overall difference was significant, F2(2,36) = 9.06, $p < 0.05$; $\eta_p^2 = 0.34$). There was a priming effect between the match and control conditions, LMM (as above), F(1,2.67) = 12.233, p = 0.047.

We next examined the relationship between the timing of the initial tongue gesture (TD in the vertical Z axis, see Figure 1) and the time to initiate vocalization. When the prime and target matched (match condition), the correlation was r = 0.76. For the control condition the correlation was r = 0.61. The difference in correlations using Zou's (2007) confidence interval was significant (as the 95%

confidence interval did not include the null hypothesis, 0) [12].

A possible reason for the lower correlation in the control condition between the onset time of the TD gesture and the time to begin a correct vocalization (acoustic onset of the target word) is that in the control condition participants initiated a tongue gesture based on the control prime rather than the target. The gesture initiated by the control prime would subsequently need to be to overcome in order to correctly pronounce the target. To examine this, we first tested the difference in the times of the TD gesture in the match and control conditions and then between the lower-lip gesture (for coronal targets TT gesture) in the match and control conditions. We used a LMM (random slopes, intercepts) to contrast TD gestural onset times for the match and control conditions. This difference was not significant, F(1, 2.9576) = 3.92, p = 0.14. We then conducted the same analysis for the match vs control conditions for the offset of the LL (TT) gesture. Here it turned out that there was a significant difference, F(1, 2.73) = 12.086, p = 0.046.

If the participant's initial articulatory gesture was influenced by the makeup of the control primes, then given how their onsets differed from the targets (vowel height), it would be expected that there will be a difference in the initial gesture for TD in the Z axis (TD_Z). Of course in determining this, the direction of the predicted interference differences needs to be taken into account (i.e., if the control prime vowel is higher than the target then it would be expected that TD_Z at gesture onset would be higher in the control than in the match condition, the opposite would follow if the control prime had a lower vowel).

For the items where there was clear difference expected due to differences in vowel height and where the priming effect was greater or equal to zero (n = 22), the magnitude of the difference in TD_Z at gesture onset between the match and control (taking account of the signed direction of predicted difference, M = 3.34 mm) was significantly different from zero (one sample test) t(21) = 6.26, $p < 0.05$. Given this interference from the control prime, we also tested whether the peak velocity of the lower lip (LL) gesture for the bilabial targets (TT for the coronal ones) differed between the control and match prime conditions. It did, the magnitude of the difference in peak velocity (M = 1.33 cm/sec) was significantly different from zero (one sample test) t(21) = 2.55, $p < 0.05$. It should be noted that the greater peak velocity of the LL (or TT for the coronal targets) found in the control condition does not allow it to compensate for the delayed gestural onsets, as indicated in a delayed naming time

4. DISCUSSION

To date, masked priming naming studies have exclusively used vocal onset (naming) time as the response measure. Such data have motivated models [13] whereby reading the prime activates phonemes that reach threshold to drive articulation. On this view, the response time cost shown when the initial letter of the prime and target mismatch (the traditional view of the onset effect) is due to competition at the phonological encoding stage that once resolved has no impact on the dynamics of speech articulation. Similar mechanisms are posited in traditional staged models of speech errors where simultaneously active segments compete, but this ends once a selected sound sequence is output to the phonetic processing stage.

Recent phonetic studies of speech errors induced by tongue twisters or laboratory techniques (e.g., repetitive speech, the slip technique, etc.) have shown that errors can involve simultaneously articulated speech gestures and that such can lead to speech that may not necessarily be recognized as errorful [14]. This result suggests that minor degrees of gesture co-activation and partial intrusion may be commonplace in everyday speech and, potentially, an important component of natural variation. A recent study [15] has shown that the time to begin naming a nonword target (e.g., BAF) was reduced by a masked prime that shared all features except voicing with the first phoneme of the target (e.g., piz) compared to a control (e.g., suz). Mousikou et al. [15] interpreted this in terms of the activation or inhibition of featural representations, however the current results suggest that the effect could be due to gestural intrusion from the control prime.

5. CONCLUSION

The data show that a masked prime can have an early influence on articulator position that has carry-on effects on articulatory kinematics, as in the increased peak velocity in articulating the target (i.e., masked control primes lead to intrusive gestures). This suggests articulatory processes can be triggered by partially activated phonology, an interpretation consistent with a cascaded approach [16]. We propose that the identity of the vowel is driving prime effects on TD displacement. This is significant, since vowel priming effects are not reliably obtained from naming latencies. The current study has established that combining masked priming and 3D electromagnetic articulography may serve as a useful tool for the study of how phonology from near overlapping inputs trigger articulatory gestures before acoustic onset.

6. REFERENCES

[1] Motley, M. T., & Baars, B. J. (1976). Semantic bias effects on the outcomes of verbal slips. Cognition, 4(2), 177-187.

[2] Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. Journal of Memory and Language, 30(1), 1-25

[3] Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. Journal of experimental psychology: Learning, Memory, and Cognition, 10(4), 680.

[4] Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. Psychonomic Bulletin & Review, 7(1), 133-141.

[5] Iskarous, K., Fowler, C. A., & Whalen, D. H. (2010). Locus equations are an acoustic expression of articulator synergy. The Journal of the Acoustical Society of America, 128(4), 2021-2032.

[6] Whalen, D. H. (1990). Coarticulation is largely planned 7/3. Journal of Phonetics, 18, 3-35.

[7] www.ndigital.com/msci/wp-content/uploads/sites/11/2014/02/Wave_4page_Brochure_Nov2010_email.pdf

[8] Berry, J. J. (2011). Accuracy of the NDI Wave speech research system. Journal of Speech, Language, and Hearing Research, 54(5), 1295-1301.

[9] Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. Behavior Research Methods, 39, 445-459.

[10] Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. Behavior Research Methods, Instruments, & Computers, 35(1), 116-124.

[11] M. K. Tiede. MVIEW: Multi-channel visualization application for displaying dynamic sensor movements. Haskins.

[12] Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. Psychological methods, 12(4), 399.

[13] Mousikou, P., Coltheart, M., Finkbeiner, M., & Saunders, S. (2010). Can the dual-route cascaded computational model of reading offer a valid account of the masked onset priming effect?. The Quarterly Journal of Experimental Psychology, 63(5), 984-1003.

[14] Pouplier, M. (2003). The dynamics of error. In Proceedings of the 15th International Congress of Phonetic Sciences (pp. 2245-2248).

[15] Mousikou, P., Roon, K. D., & Rastle, K. (2014, December 22). Masked Primes Activate Feature Representations in Reading Aloud. Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication. http://dx.doi.org/10.1037/xlm0000072.

[16] Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 719–750.