ARTICULATORY CONSEQUENCES OF PREDICTION DURING COMPREHENSION

Eleanor Drake[1], Sonja Schaeffler[2], Martin Corley[1]

[1]PPLS, University of Edinburgh, [2]CASL, Queen Margaret University
E.K.E.Drake@sms.ed.ac.uk

ABSTRACT

It has been proposed that speech-motor activation observed during comprehension may, in part, reflect involvement of the speech-motor system in the top-down simulation of upcoming material [14]. In the current study we employed an automated approach to the analysis of ultrasound tongue imaging in order to investigate whether comprehension-elicited effects are observable at an articulatory-output level.

We investigated whether and how lexical predictions affect speech-motor output. Effects were found at a relatively early point during the pre-acoustic phase of articulation, and did not appear to be predicated upon the nature of the phonological-overlap between predicted and named items. In these respects effects related to comprehension-elicited predictions appear to differ in nature from those observed in production and perception experiments.

Keywords: Dynamic ultrasound tongue imaging, predictive coding, comprehension, production.

1. INTRODUCTION

Activation of neural speech motor areas and speech effectors is observed during speech listening and language comprehension [6][17][21]. Language comprehension is facilitated by prediction of upcoming material [1][7]. It has been suggested that speech motor activation during listening may reflect, in part, the top-down encoding of to-be-heard material via forward modelling [9][15].

It is thought that forward modelling involves the generation of efference copies, by which the agent predicts the sensory consequences of their own actions. The generation of efference copies is understood to be the mechanism underlying speech-induced auditory suppression (see [14]). It has been proposed that the generation of efference copies also occurs when people predict others' spoken output during speech listening [9][15].

Evidence that comprehension can involve the prediction of upcoming material comes largely from ERP studies (e.g., [1] [5] [7]). Phonological-form level prediction has been shown during reading comprehension [5]. Speech-motor area specific activity has been observed in response to violations of metrical predictions during speech comprehension [18]. It remains unclear, however, whether the generation of phonological-form expectations during reading comprehension is in any way related to somatotopic speech-motor activation observed during speech listening (e.g., [20]).

When competing phonological representations are activated in syllable-onset position during speech production, the consequences can be observed in both acoustic response times and in the speaker's articulatory patterns. During picture-naming tasks, onset-overlap between a picture-name and a heard word leads to reduced acoustic response latencies [13]. During error-elicitation tasks, competing onsets appear to lead to interference at an articulatory level, even in perceptually error-free productions [12][16]. During non-word reading aloud, simultaneous auditory presentation of a syllable with a competing onset leads to phoneme-specific articulatory interference [23]. If the lexical-predictions that arise during comprehension invoke production-associated phoneme-specific speech-motor activation, we would anticipate evidence in the articulatory output of the listener.

In the current study we elicited lexical predictions by presenting participants with auditory sentence-stems (from which the final, highly-predictable, word was omitted). At the end of each sentence-stem, participants were presented with an image to name, in order to allow us to examine the articulatory effects of predictions. Image names either fully matched the elicited lexical prediction (e.g., TAPE-tape) or overlapped with it in either syllable onset (e.g., TAPE-take) or rime (e.g., TAPE-cape) position. Articulation was recorded via dynamic ultrasound tongue imaging and analysed via an automated approach which allowed us to average spatio-temporal information across tokens, participants, and conditions. If comprehension-

elicited lexical predictions invoke speech-motor activation associated with phonological representations, we would expect articulation in the onset-overlap condition to be more similar to that in the full-overlap condition than is that in the rime-overlap (i.e. competing onset) condition. That is, we would expect to see an interference effect in articulation at word onset only when activated syllable onsets compete for articulatory realisation.

2. METHOD

2.1. Participants

Participants (7 female, 3 male) were monolingual speakers of English who reported normal visual and hearing acuity, and had no history of phonetic training or communication difficulty (age range = 19-27 years). All participants gave written informed consent in line with British Psychological Society guidelines. The study was granted ethical approval by Psychology Research Ethics Committee of the University of Edinburgh.

2.2. Materials

Experimental items were generated by combining two C syllable onsets (/t, k/) with six VC(C) syllable rimes in turn, producing 12 words (tape, take, toast, tone, tap, tan, cape, cake, coast, cone, cap, can). Each word was represented by a colour image selected from an online picture database. For each word, three high-cloze sentence-stems were generated that predicted that word. All sentence-stems ended in a vowel or semi-vowel, and were cut so as not to be phonetically informative as to the (predicted) final word. Picture-naming agreement and sentence-cloze likelihood were pre-tested via an online test completed by 10 native speakers of English who were not involved in the current study (mean item-naming agreement = 0.85, minimum sentence-cloze likelihood = 0.75). Sentence-stems were spoken by a female speaker of British English at a mean rate of 3.92 syllables per second (mean sentence stem duration = 3.10 seconds, range = 1.90 – 5.29 seconds).

2.3. Procedure

The experiment was run within the ultrasound tongue imaging suite at Queen Margaret University, and presented on a Dell XPS 1702 laptop running DMDX presentation software [8]. Item presentation order was automatically randomised within each block via the experimental software. Experiment design was fully within-participant and within-items.

During an initial familiarisation phase participants were presented with each image in turn accompanied by its name in written and spoken form. Once participants were able to name all pictures correctly they were fitted with an ultrasound probe stabilization helmet [2]. The ultrasound transducer was positioned to capture a mid-sagittal image of the tongue between the hyoid and mandible shadows (see [19] for equipment and procedure). Participants were then asked to name the pictures again in order to familiarise themselves with the experience of articulating whilst wearing the ultrasound equipment. Participants then began the experiment proper.

In the Control condition participants viewed a fixation point for 3.9 seconds and were then presented with a picture to name as quickly and accurately as possible. In the experimental conditions (Full Overlap, Onset Overlap, Rime Overlap) participants viewed a fixation point whilst listening to a sentence-stem. At the end of the sentence-stem a picture was presented for naming as quickly and accurately as possible. In the Full Overlap condition the predicted word and the picture name fully overlapped (e.g., TAKE-take), in the onset condition they overlapped in onset position (e.g., TAKE-tape), in the Rime Overlap they overlapped in rime position (e.g., TAKE-cake). Each picture was presented for naming once in each experimental condition in each of blocks two, three and four, and once in each Control block (blocks one and five).

2.4. Data Capture and Processing

Acoustic and ultrasound data were recorded via AAA software, Ultrasonix hardware, and a micro-convex probe (depth = 80mm, angle = 150˚, recording rate = 100fps; see [22]). Recording of each trial began at the point that the fixation point was presented and continued until the participant had completed picture-naming. Audio-visual data were exported from AAA in AVI format, and a synchronization check was performed in VirtualDub.

2.4.1. Audio processing

The exported data were manually tagged in Praat [4] to indicate the onset of picture presentation, the acoustic onset of the initial consonant and of the subsequent vowel, and the acoustic offset of the steady-state vowel. Acoustic landmark time-points were exported to .csv in order to be made available to the ultrasound video-processing software.

2.4.2. Video processing

Each frame of the ultrasound video comprised a 512 x 277 pixel grid. Pixel luminance varied between 000 (black) and 255 (white). Data tractability was improved by averaging pixel brightness over blocks of 8 x 8 contiguous pixels (for further information on method and rationale see [11]). Each frame was represented as a vector running from the bottom left of the screen to the top right, in which each 8 x 8 pixel block had a fixed position. The luminance of each pixel block was recorded in each vector. Pixel-block luminance varied from frame to frame in response to changes in the material imaged (i.e., tongue position). Frame vectors formed the input for subsequent analyses, allowing the automatic calculation and comparison of "Delta scores" (i.e., the Euclidean distances between individual or averaged vectors). Statistical analyses were performed on Delta scores.

3. ANALYSIS

3.1. Quality Metric

Ultrasound data is noisy and varies in articulatory informative-ness depending on participant anatomy-physiology and transducer positioning. We therefore used a data quality metric to geometrically weight the contribution of each participant's data to analyses. Taking a given CV onset, we calculated the average distance between articulations of the same CV onset (WITHIN); and the average distance between each CV exemplar and productions of different CVs (BETWEEN). The quality metric was calculated as BETWEEN/WITHIN. The quality metric indicated how well the Delta scores discriminated the tokens of a given word from tokens of all other words. This figure was entered as a weighting factor in all subsequent analyses (see [10]).

3.2. Statistical Modelling Approach

We used a mixed modelling approach, implemented in R via the lme4 package [2]. In line with common practice we report effects as significant where $|t| > 2$. In all analyses we modelled Delta scores as the outcome variable, and included Condition (Match, Onset Overlap, Rime Overlap) and Onset consonant (/k/, /t/) as fixed effects, and Participant and Picture-name (i.e. item) as random effects.

3.3. Location of articulation analysis

This analysis was performed on articulatory data acquired between -500ms and 0ms of the consonant acoustic burst. Data acquired during this period were collapsed to produce one average-luminance vector per token. For each item, a reference vector was generated by averaging across all Control productions of that item. This allowed us to calculate Delta scores which expressed the degree to which tokens produced in the experimental conditions (Full Overlap, Onset Overlap, Rime Overlap) differed from those produced in the Control condition.

Differences between individual articulations and mean control articulations were then modelled as the response variable in a linear mixed model (for details of predictor variables and model structure see above). Delta scores in the Full Overlap condition were significantly lower than those in the partial overlap conditions ($\beta = 1.687$, $t = 2.108$). Delta scores in the two partial overlap conditions (Onset and Rime) did not differ significantly ($\beta = 0.506$, $t = 0.364$). This indicates that, in line with our prediction, articulation in the Rime Overlap condition differed from that in the Full Overlap condition. However, contrary to our prediction, articulation in the Rime Overlap condition (where there was onset competition) did not differ from that in the Onset Overlap condition (where there was not onset competition).

In order to investigate whether tokens in the Rime Overlap condition exhibited traces of articulatory interference, we generated Delta scores that expressed the degree to which tokens produced in the experimental conditions differed from the Control reference vector for their onset competitor (i.e., rather than comparing tokens of "take" to the reference vector for "take" as above, we compared them to the reference vector for "cake"). These Delta scores did not differ by condition (in all cases $|t| < 1$): We did not observe a phoneme-specific articulatory interference effect.

3.4. Time-course analysis

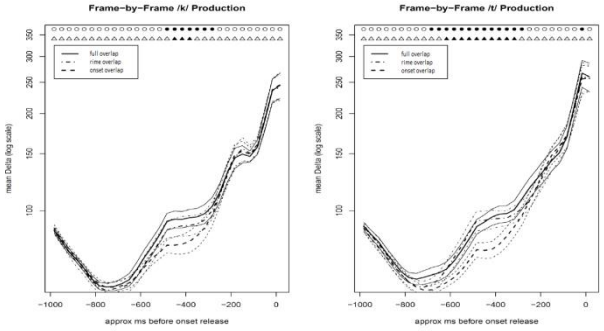

Figure 1: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation. Faint lines indicate 95% confidence intervals. Filled circles indicate inter-frame intervals where Delta in the rime overlap condition differs significantly from that in the full overlap condition. Filled triangles indicate inter-frame intervals where Delta in the onset overlap condition differs significantly from that in the full overlap condition.

This analysis was performed on all ultrasound frames acquired between -1000ms and 0ms of the onset consonant acoustic burst for each token (i.e., 31 frames per token). Within each token we calculated Delta scores for all inter-frame transitions over the time-course. Higher delta scores indicated greater frame-to-frame change associated with greater change in tongue configuration. Delta scores were automatically averaged and plotted by condition (Full, Onset, and Rime Overlap; see Fig. 1) and by onset-consonant (/k/, /t/).

We modelled Delta scores at each inter-frame transition via mixed-effects models comparing productions in the Onset and Rime Overlap conditions to those in the Full Overlap condition. We treated effects as significant only when they clustered across three or more consecutive time-points [9]. As illustrated in Fig. 1, effects of condition were found to be statistically significant (i.e., $|t| > 2$) for both consonant onsets (i.e., /k/ and /t/) between -500ms and -300ms. Effects were also significant for /t/ onset items in the time window -700ms to -500ms. Significant effects were observed over a longer time period in the Rime Overlap condition than in the Onset Overlap condition, but the two conditions pattern similarly, with consistently greater frame-to-frame movement in these conditions than in the Full overlap condition. This means that, as in the Location of Articulation analysis, we observed an articulatory effect of mismatch between the lexical prediction and the picture name. We did not find evidence that this effect was confined to situations in which there was onset competition between the predicted word and the picture name.

4. DISCUSSION

We reported a study in which we used an automated approach to the analysis of ultrasound tongue imaging data in order to investigate whether comprehension-elicited lexical predictions have articulatory consequences. Participants named pictures in one control condition and three experimental conditions. The experimental conditions differed with regard to the extent that the picture name overlapped with a predicted word at a phonological level. Lexical predictions were elicited by auditorily presenting participants with high-cloze sentence-stems (i.e. via comprehension). Of specific interest was whether comprehension-related predictions elicit cascade from a phonological to a motor-speech level as do representations activated during speech production and speech listening.

Effects of prediction were observed in articulatory data acquired prior to the acoustic onset of picture naming. When data were collapsed over the 500ms preceding acoustic onset, articulations were more similar to the control productions when there was full overlap between the predicted word and the picture name than where there was only partial overlap. This indicates that lexical predictions elicited via comprehension can have articulatory consequences.

The findings of previous error-elicitation studies led us to predict that if motor-speech activation during comprehension reflects the representation of upcoming material at an abstract gestural level we would observe interference from a competing representation at an articulatory level. We therefore investigated whether tokens produced in the Rime Overlap condition, where the lexical prediction would activate a competing onset representation, were more similar to articulations of the competing word than were those in the Onset Overlap. We did not find evidence to support this interpretation.

We used a time-course analysis to further investigate the nature of the effects on articulation of comprehension-elicited predictions. This analysis approach revealed that effects are seen only at a relatively early stage during the pre-acoustic phase. During this early time-window, frame to frame change is greater in conditions where the picture name does not match the predicted word. However, the degree and pattern of frame to frame change did not appear to differ according to whether there was phonological conflict at word onset: Articulation in the Onset Overlap condition (in which there was no conflict at word onset) differed from articulation when there was a full-word match but not from articulation when there was conflict at word onset. This suggests that articulatory effects may arise at a whole-word level. It should be noted, however, that picture names were all monosyllabic so it is not possible to distinguish between effects arising at a whole syllable level and those arising at a whole word level.

The suggestion that articulatory effects arise at a whole word level is compatible with the early time-window at which effects are observed, and with the failure to find evidence of phonological interference effects. The apparent non-specificity of the articulatory effects observed raises the possibility that the articulatory consequences of the lexical predictions reflect general conflict monitoring and resolution processes (possibly an articulatory correlate of neural error-related negativity).

5. ACKNOWLEDGEMENTS

We thank Alan Wrench (Articulate Instruments) and Steve Cowen (Queen Margaret University) for ongoing technical advice and support.

6. REFERENCES

[1] Altmann, G., Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

[2] Articulate Instruments Ltd. (2008). *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd.

[3] Bates, D., Maechler, M., Dai, B. (2008). Ime4: Linear mixed-effects models using S4 classes (R package version 0.999375-27). Retrieved from: *http://www. r-proiect. org*.

[4] Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

[5] DeLong, K. A., Urbach, T. P., Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117-1121.

[6] Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15(*2), 399-402.

[7] Federmeier, K. D. (2007). "Thinking ahead: The role and roots of prediction in language comprehension". *Psychophysiology*, *44(*4), 491-505.

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

[9] Garrod, S., Gambi, C., Pickering, M. J. (2014). Prediction at all levels: forward model predictions can enhance comprehension. *Language, Cognition and Neuroscience*, *29*(1), 46-48.

[10] Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., Galán, L. (2010). False discovery rate and permutation test: an evaluation in ERP data analysis. *Statistics in medicine*, *29*(1), 63-74.

[11] Mardia, K. V. (1978). Some properties of classical multi-dimensional scaling. *Communications in Statistics-Theory and Methods*, *7*(13), 1233-1241.

[12] McMillan, C. T., Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117(*3), 243-260.

[13] Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146.

[14] Niziolek, C. A., Nagarajan, S. S., Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience*, *33*(41), 16110-16116.

[15] Pickering, M. J., Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in cognitive sciences*, *11*(3), 105-110.

[16] Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*, *50*(3), 311-341.

[17] Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865-7870.

[18] Rothermich, K., Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *Neuroimage*, *70*, 89-100.

[19] Scobbie, J. M., Wrench, A. A., van der Linden, M. (2008). "Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement". In *Proceedings of the 8th International Seminar on Speech Production* (pp. 373-376).

[20] Watkins, K., Paus, T. (2004). "Modulation of motor excitability during speech perception: the role of Broca's area". *Journal of Cognitive Neuroscience*, *16*(6), 978-987.

[21] Wilson, S. M., Saygin, A. P., Sereno, M. I., Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech

[22] Wrench, A. A., Scobbie, J. M. (2008). High-speed Cineloop Ultrasound vs. Video Ultrasound Tongue Imaging: Comparison of Front and Back Lingual Gesture Location and Relative Timing. In Proceedings of the Eighth International Seminar on Speech Production (ISSP).

[23] Yuen, I., Davis, M. H., Brysbaert, M., Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, *107*(2), 592-597.