Perceptually induced speech motor representations

C. Neufeld[1], R. Craioveanu[2], F. Rudzicz[3], W. Wong[4], and P. van Lieshout[1]

University of Toronto
[1]Speech-Language Pathology, [2]Linguistics, [3]Computer Science, [4]Electrical and Computer Engineering
christopher.neufeld@utoronto.ca

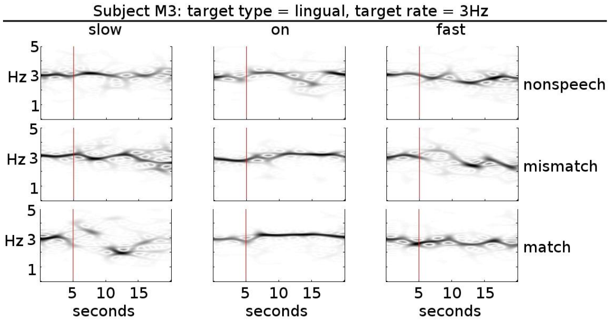
Recent neurological research has shown that motor areas associated with the movement of speech articulators are activated by the mere presentation of speech stimuli in both auditory and visual modalities (Fadiga et al., 2002; Pulvermüller et al., 2006; Watkins et al., 2003). Since the activity of these brain regions typically results in the overt movement of speech articulators, this invites us to question how perceptually induced speech-motor representations (PISR) differ from explicit motor plans. Listeners refrain from compulsively moving their articulators in response to speech stimuli, suggesting several possible explanations for the cortical motor activity observed in speech perception in the absence of explicit movement. Either PISR are a fundamentally different kind of motor representation from typical motor-plans, and the findings cited above simply reflect shared neural resources, rather than any cognitive overlap. Alternatively, the activation of speech-motor areas may reflect a priming effect, which is speech-specific, but not necessarily a detailed motor-plan. Or, finally, PISR may be fine-grained motor-plans which are filtered or damped by some antagonistic mechanism to prevent the involuntary movement of speech-articulators during speech perception. This latter hypothesis would provide a explanatory mechanism for convergence in speech as a form of entrainment: the speech motor plans of speakers are neurologically represented as such in listeners, and affect the fine-grained structure of the listener's motor plans when it is their turn to speak (Pickering & Garrod, 2004).

To distinguish these three possibilities, an experiment was devised which engages both explicit and tacit speech motor streams. Subjects are presented with prompts which instruct them either to produce two different **stimuli types** /ma/ or /na/ (labial or lingual) without vocalizing, to the beat of a visual metronome. There are three different **stimuli rates**: 2 Hz, 3 Hz, and 4.5 Hz. After several seconds of following the visual metronome, a rhythmic audio distractor is played over headphones. There are three **distractor rates**: a slower rhythm with respect to the visual metronome, on-beat or faster. Slow and fast rhythms are the target speed divided by, or multiplied by 1.5 respectively. Finally, there are three **distractor types**: nonspeech, matched and mismatched. Matched and mismatched distractors were created by recording the first author repeating /ba/ or /da/ at various speeds. Matched distractors have the same place of articulation as the target gesture (/ba/ in the case of /ma/, /da/ in the case of /na/), and mismatched distractors have the opposite place of articulation (/da/ in the case of /ma/). Speech distractors were automatically realigned to be exactly on the desired beat, and normalized for pitch and amplitude. Nonspeech distractors were created by superposing pure sine waves at the frequencies and amplitudes of F0-F3 of the speech distractors and multiplying the resulting complex wave by the amplitude envelope of speech distractors.

If PISR are fine-grained motor plans, we should expect that off-beat, matched distractors will be most disruptive to the accurate maintenance of the target frequency. For example, if a subject is attempting to produce a labial gesture at 3 Hz, and is suddenly presented with the audio of a labial speech gesture at 2 Hz, motor-areas responsible for driving the oscillations of the lips will be simultaneously representing 2 and 3 Hz, and this conflict should be detectable at the level of oral motor output. Mismatched off-beat speech distractors will be somewhat disruptive since there is some movement of lips and tongue for both labial and lingual gestures, but should not be as disruptive as matched off-beat speech distractors. Finally, nonspeech, off-beat distractors should be the least disruptive. If PISR are just a generic priming of the speech-motor system, organ matched and mismatched off-beat distractors should be equally distracting but more distracting than off-beat nonspeech, and speech on-beat speech distractors should be equally facilitative, but more facilitative than nonspeech. If PISR reflect shared neural, but not cognitive resources, we would predict that there should be no difference across distractor type conditions: off-beat distractors should all be equally distracting, regardless of type, and on-beat distractors should be equally facilitating, regardless of type.

Articulatory data was collected from 15 subjects with a 12-channel 3D Electromagnetic articulograph (EMA) with a sampling rate of 200 Hz. Gestures were derived from raw EMA data by calculating the Euclidean distance between the tongue-tip and the nose, for lingual targets, and the distance between upper and lower lip, for labial targets. Figure 1 shows some data from one subject. Each panel is a time-frequency representation of the trajectory of the tongue-tip with a target frequency of 3 Hz.

Figure 1: Time-frequency representations of speech gestures.



Each column of panels represents a distractor rate, and each row a distractor type. The vertical lines indicate the onset of the auditory distractor. Disorder in the signal can be observed where the time-frequency representation is diffuse and has no strong single peak frequency, or where the peak frequency changes rapidly over short time spans. It can be seen that on-beat distractors facilitate the maintenance of a 3 Hz rhythm. However, the signal appears slightly more disordered for the nonspeech distractors than for speech in the on-beat condition, and the matched speech distractor shows the greatest degree of facilitation: a highly ordered signal with almost all its energy at 3 Hz. The most extreme instance of distraction occurs with the slow matched distractor. After the onset of the auditory distractor, the signal becomes disordered, with more energy in higher frequencies. At around 10 seconds, the signal is orderly once more, but most of its energy is around 2 Hz: the distractor frequency. From 15-20 s, the signal becomes highly disordered again, without a strong peak in any frequency band. By contrast, for mismatch and nonspeech, the slow auditory distractors do not appear to interfere with the maintenance of a 3 Hz rhythm to nearly so dramatic an extent. The signal is more disordered than for on-beat distractors, but only marginally so, and the signal is generally concentrated around 3 Hz.

To quantify the disorder of these time series, 2^{14}-point spectrograms were calculated from the gestural time series using a 1.25 second Hamming window. At each time step, the frequency bin with the highest amount of energy was calculated, producing a peak-frequency track. The absolute value of the derivative of this frequency track was used as a measure of disorder: for disordered signals, the peak-frequency changes rapidly (such as can be seen in the bottom left panel of Figure 1, and for ordered signals, the peak frequency changes slowly or not at all (such as can be seen in the bottom middle panel).

A 4-way ANOVA (stimuli type × stimuli rate × distractor type × distractor rate) was calculated, and showed significant main effects for all independent variables ($p < 0.05$). Post-hoc Tukey tests showed that matched distractors caused significantly more disordered gestures than either mismatched or nonspeech distractors, with no significant differences between the latter. Post-hoc tests also showed that fast distractors caused significantly more disorder than slow distractors, both of which cased significantly more disorder than on-beat distractors. The 4.5 Hz target signals were significantly more disordered than either the 3 Hz target signals or 2 Hz distractor signals. Finally, lingual gestures were significantly more disordered than labial disorders.

These preliminary observations suggest that the paradigm presented here is a productive research tool. The fact that off-beat distractors disrupt the accurate maintenance of a rhythmic speech gesture indicates that asynchronous auditory input does impact speech-motor output. Thus, there is support for hypothesis that PISR are organ-specific motor plans, since matched distractors cause significantly more short-term volatility in peak frequency than either mismatched or nonspeech distractors, while mismatched distractors are not significantly different from one another, arguing against the priming hypothesis.

References

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.

Pickering, M. J., & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioural and Brain Sciences* 27, 169–226.

Pulvermüller, F., Huss, M., Kherif, F., Martin, F. M. d. P., Hauk, O., & Shtyrov, Y. 2006. Motor cortex maps articulatory features of speech sounds. *Proc. Nat. Acad. Sci.* 103, 7865–7870.

Watkins, K. E., Strafella, A. P., & Paus, T. 2003. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994.