Phonetic distortion as a window to investigate motor engagement in speech perception: Articulatory evidence from a Stimulus-Response Compatibility paradigm

Takayuki Nagamine[1*], Zhouyiting Zhao[1], Zihui Wang[1], Chris Carignan[1], Adamantios Gafos[2], Patti Adank[1]
[1]Speech, Hearing and Phonetic Sciences, University College London (*takayuki.nagamine@ucl.ac.uk)
[2]Department of Linguistics, University of Potsdam

Background: This study presents our approach to studying the perception-production link in speech using the Stimulus-Response Compatibility (SRC) paradigm. In a typical speech SRC task, participants are prompted to produce a target syllable while being presented with either congruent or incongruent distractors. Responses tend to be slower in incongruent trials, reflecting a competition between perception-driven and goal-driven motor plans. The short response-distractor time lag in the SRC task design makes it suited to study the motor system engagement upon speech perception during speech planning [1].

Previous research demonstrates that participants' responses exhibit a 'trace' of the distractor stimulus in the incongruent condition, compared to the baseline production, both acoustically (e.g., an increased F2 for incongruent /aba/) and articulatorily (e.g., intrusion of tongue tip raising for incongruent /k/) [2, 3]. The exact nature of this 'phonetic distortion' phenomenon, however, remains inconclusive, due to a lack of articulatory evidence and possible alternative interpretations in the previously used paradigm. To address these issues, we combine the SRC paradigm with electromagnetic articulography (EMA).

Methods: Ten L1 British English speakers (six female, M_{age} = 27.4, *SD* = 10.63) produced [ɹa] or [va] upon seeing the prompt symbols ('£' or '&'). The symbols are superimposed around the lip area on the audiovisual distractor stimuli (a female British English speaker saying [ɹa] or [va]). Three levels of delay were added in symbol presentation timing (stimulus-onset asynchrony; SOA). Responses were recorded both acoustically (44,100 Hz) and articulatorily using NDI Vox-EMA (400 Hz). A total of 288 utterances were elicited per speaker, and analysis focuses on 2,638 veridical utterances ([ɹa]: *n* = 1,320, [va]: *n* = 1,318).

Reaction time (RT) is defined as the time lag between the symbol presentation and (1) the acoustic onset of each utterance or (2) the vertical displacement peak of the primary articulators: tongue tip (TT) for [ɹa] and lower lip (LL) for [va]. Articulatory analysis considers the movements of both TT and LL for [ɹa] and [va].

Results: Acoustic and articulatory RT analyses demonstrate a clear congruency effect for [ɹa] but not for [va].

Despite this, the time-varying TT displacement for [va] shows a slight between-condition difference. This is especially evident at SOA3 (indicated with the circle in Figure 1) when the symbol is displayed temporarily closest to the visual lip movement in the distractor. Functional Principal Component Analysis and Bayesian linear mixed-effect modelling (PC1 score (z-normalised per speaker) ~ congruency*SOA) shows a credible congruency effect on TT trajectory shape for [va] (β = -0.14 [-0.19, -0.09]).

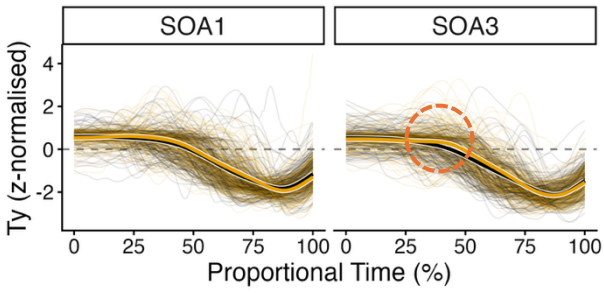

Figure 1: Time-varying TT vertical displacement (Ty) for the [va] response at SOA1 and 3. Time is expressed proportionally, between the symbol presentation (0%) and the utterance offset (100%). By-condition smooths are superimposed on raw EMA signals (low-pass filtered with a 10 Hz cut-off). Colours indicate compatibility (yellow: incongruent, black: congruent).

Discussion: The incongruent [va] responses may demonstrate an automatic and involuntary TT involvement, due likely to the activation of TT motor commands by perceiving the distractor [ɹa]. The lack of congruency effect in the RT could result from the independent involvement of the tongue and lips for [va], enabling articulatory adjustment without any consequences on RT. We will further complement data from ongoing experiments and discuss the findings in the light of multimodality in speech perception.

References: [1] Roon, K. D., & Gafos, A. I. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language, 89*, 222–243. https://doi.org/10.1016/j.jml.2016.01.005: [2] Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*(1), 66–75. https://doi.org/10.1007/s00221-005-0008-z: [3] 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. https://doi.org/10.1073/pnas.0904774107