PERCEPTUO-MOTOR INTERACTIONS ACROSS AND WITHIN PHONEMIC CATEGORIES

Eugen Klein[1], Kevin D. Roon[2,3], Adamantios I. Gafos[1,3]

[1]Universität Potsdam, Germany
[2]CUNY Graduate Center, USA
[3]Haskins Laboratories, USA
euklein@uni-potsdam.de, kroon@gc.cuny.edu, gafos@uni-potsdam.de

ABSTRACT

Past studies have demonstrated that reaction times in producing CV syllables are modulated by audio distractors participants hear while preparing their responses. Galantucci et al. [5] showed that participants respond faster on trials with identical response-distractor pairs than when the response and distractor differ in voicing or articulator. The dynamical model of phonological planning developed in [9] attributes these differences in reaction times to phonological parameters of a distractor exciting or inhibiting the planning of a response. A so far untested prediction of the model is that within-category phonetic variation in the voicing parameter of the distractors still gives rise to congruency effects of the articulator. We report new results which show effects of response-distractor articulator congruency for distractors with varying voice onset times. These results extend previous findings by pursuing predictions concerning within-category variability of distractor stimuli.

Keywords: perceptuo-motor effects, reaction time, phonetic detail, phonological planning, voice onset time

1. INTRODUCTION

During a response-distractor task, participants repeatedly produce CV syllables as prompted by visual cues. On some trials, shortly before responding, participants are presented with distractor stimuli via headphones. Galantucci et al. [5] showed that reaction times (RTs) in identical response-distractor pairs (e.g., /ba/-/ba/) are faster than when the response and distractor differ in articulator (e.g., /ba/-/da/). These interactions between response and distractor have been dubbed "perceptuo-motor effects".

These results are accounted for by a dynamical model of phonological planning [9]. In this model, during the planning of a response, a set of phonological parameters needs to be set (e.g., tongue tip constriction, lower lip constriction, voicing), whereby the level of activation for each parameter evolves over time by receiving values from distractors and intended responses. In case of voicing, for instance, the relevant phonological parameter is voice onset time (VOT). In a dynamical model of phonological form, a voiced (/da/ or /ga/) and voiceless (/ta/ or /ka/) response correspond to two attractors or categories [4]. Within these regions, say, the voiceless, continuous VOT values of /ta/ and /ka/ are close enough that activation of one value increases neighbouring voiceless activation levels, via local excitation. Across the two response types, the voiced /da/ and the voiceless /ta/ categories occupy two regions in the VOT continuum that are sufficiently distant from each other so that activation of one results in suppression of the other, via lateral inhibition. We note that 'close enough' in our description of local excitation above is elaborated in the dynamical model in [9] by an interaction term parameterized for distance within the relevant phonetic space (here, VOT) and also for the slope of excitation as a function of distance (thus effecting more or less excitation, depending on distance). The mechanisms of local excitation for within-category VOTs and lateral inhibition for across-category VOTs are inherent to the formal model and derive from broad assumptions about functional cortical modeling.

Local excitation and lateral inhibition predict specific effects of distractors on responses. Hearing a distractor with a mismatched VOT (e.g., /da/-/ka/) should result in slower RTs than in matched distractor-response pairs (e.g., /ta/-/ka/), due to lateral inhibition between the distractor and response VOTs. The same applies when the mismatch is in terms of articulator. See results in [10] and extensions in [11] for these predictions.

In the results reviewed above, the distractor stimuli had fixed VOT values. In this study, we take up predictions which relax that assumption of invariant, specific VOTs. Thus, a so far untested prediction of the model is that distractor and response VOTs in phonemically identical pairs do not need to have identical phonetic values in order to excite each other. Consequently, speed-up in RTs for response-distractor pairs belonging to the same phonemic category (e.g., /ka/-/ka/ or /ta/-/ta) should be observed even when there is variability in the phonetic detail of the distractor stimuli. However, the degree of speed-up may be finely tuned by within-category differences, and any evidence to that effect would contribute information crucial to sharpening specifics of the model in [9]. Furthermore, phonetic variability in one phonological parameter, here in VOT, should not affect the inhibition effect for articulator across two phonemic categories, i.e., we should observe longer RTs for response-distractor pairs like /ka/-/ta/ or /ta/-/ka/.

We present below results from an experiment that tested these predictions.

2. METHOD

2.1. Participants

40 students (28 female, 12 male) from Potsdam University participated in the experiment. All participants were monolingual native speakers of German without history of speech and/or hearing disorders. The mean age was 23.7 years.

2.4. Apparatus and stimuli

The distractor stimuli (/ta/ and /ka/) were spoken by a female native speaker of German. These had a VOT of 56 and 58 ms for /ta/ and /ka/, respectively. The duration of the aspiration for each syllable was adjusted in Praat [2]. Six versions of /ta/ and /ka/ were created such that the VOT ranged from 45 to 120 ms in 15 ms steps. The duration of each token was cut after 230 ms at zero amplitude and attenuated during the last 45 ms. The stimuli were presented using the Psychophysics Toolbox [6].

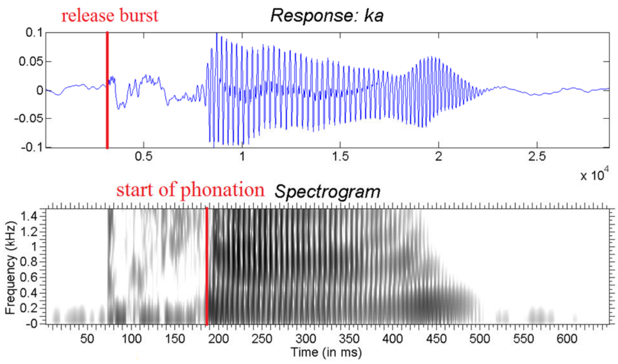
2.3. Procedure

Participants wore headphones while seated in front of a computer screen in a sound-attenuated booth. On each trial, a fixation cross appeared for 500 ms at the center of the screen followed by a visual cue (## or **) which was presented in gray on black background. Half of the participants were instructed to respond with /ka/ to ## and with /ta/ to **, and vice versa for the other half. Participants were instructed to respond as fast as possible once the cue appeared. The visual cue was kept on screen until a response was detected by the built-in microphone of the prompter computer, after which the visual cue disappeared and a new trial began after 800 ms. Shortly before producing a response but after the presentation of the visual cue, participants heard a distractor that matched or mismatched the intended response in articulator. The participants were told to ignore everything they would hear. The distractor stimuli were presented with 100 or 200 ms stimulus onset asynchrony (SOA) with respect to the onset of the visual cue. During the experiment each combination of a response and a distractor was presented 16 times at each SOA yielding 768 trials. Additionally, on 128 trials participants heard a non-linguistic tone distractor, and on 128 trials no distractor was presented. The total of 1024 trials was presented in 4 blocks yielding 256 trials per block for each participant. All trials were randomized within a block.

3. RESULTS

The total recording of 40 participants amounted to 40960 trials. Before the analysis, 5.4% of the trials were discarded due to either an erroneous response or initiation of response before the beginning of a distractor. The recordings were manually annotated by inspecting the spectrogram and waveform in MATLAB. Measurement points were taken at the burst release and at the start of phonation. Fig. 1 shows an example measurement. The start of phonation was marked at the start of modal voicing visible in the range of the fundamental frequency (F0) (cf. [3]). Based on these landmarks, RTs and response VOTs were computed. For any given trial, RT was defined as the interval in milliseconds between the presentation of the visual cue and the burst release of the prevocalic consonant. VOT was defined as the interval in milliseconds between the onset of voicing and the burst release [7].

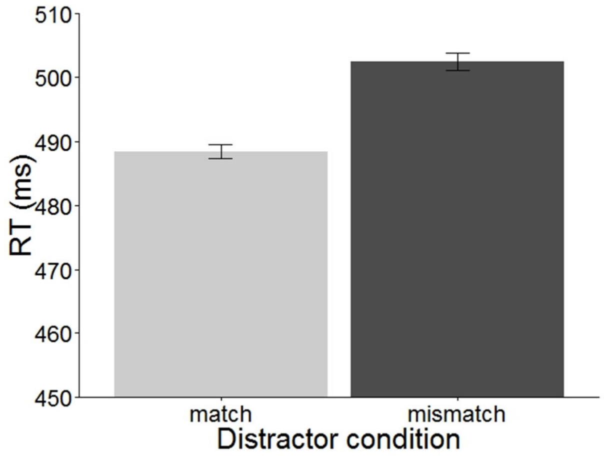
Figure 1: Waveform and spectrogram of participant's response with landmarks for release burst and start of the phonation.



3.1. Congruency effects across response syllables

Here, we consider the effect of articulator congruency between responses and distractors on participants' response RTs. RTs were logtransformed for the statistical analysis (however, all figures show RTs in milliseconds for convenience).

Figure 2: Mean RT (ms) for response-distractor pairs with matching and mismatching articulator pooled across two SOAs and two response syllables, /ka/ and /ta/.



The dataset of each participant was inspected for outliers, resulting in discarding a further 0.1% of the data. 28987 data points entered the analysis. The results by condition are shown in Fig. 2. Using the lme4 package [1] for R [8], a linear mixed-effects model was fitted including the distractor condition as a fixed effect and RT as the dependent variable. Further, the model included the fixed effects of response syllable, the log-RT of the previous trial and SOA. Random intercepts were modeled for trial number and for every participant with random slopes for response syllable and distractor condition.

RTs in the mismatch condition were significantly longer than in the match condition (t = 7.085, p < 0.001). The effect of the response syllable was also significant with overall longer RTs for /ta/ responses (t = 3.617, p < 0.001). The RT on the previous trial was a significant predictor of RT (t = 61.092, p < 0.001). The effect of SOA was not significant. The novel aspect of this result is that perceptuo-motor congruency effects of articulator were obtained even in the presence of within-category phonetic variation within the voicing of the distractor stimuli.

3.2. Congruency effects within response syllables

In this section, we take a closer look at the effect of articulator congruency within each distractor condition. The distractor stimuli were grouped relative to the mean VOT of each participant. Participant mean VOT was defined as the mean VOT of all responses for that participant within a phonemic category (/ta/, /ka/). For any given participant, each distractor was categorized as having either a shorter or longer VOT relative to the participant mean VOT. Shorter and longer relative VOTs were binned into VOT steps of approximately 15 ms, so that, for instance, VOT step of 1 means that all distractors in this bin had a VOT which was up to 15 ms longer than the participant's mean VOT for that specific syllable, and VOT step of –1 means that all distractors in this bin had a VOT which was up to 15 ms shorter than the participant's mean VOT for that specific syllable, and so on for the other VOT steps. This procedure was done separately for /ka/ and /ta/ responses. VOT steps containing less than 500 observations (–5, –4, and 6) where discarded to get a sufficient number of observations across all VOT steps (1.9% of the data were discarded). Fig. 3 shows the RTs by VOT step within the distractor conditions of matching and mismatching articulator, collapsed across response syllable and SOA.

Figure 3: Mean RT (ms) for distractor conditions with matching and mismatching articulator pooled across two SOAs. RTs are binned by VOT step (see text for definition).

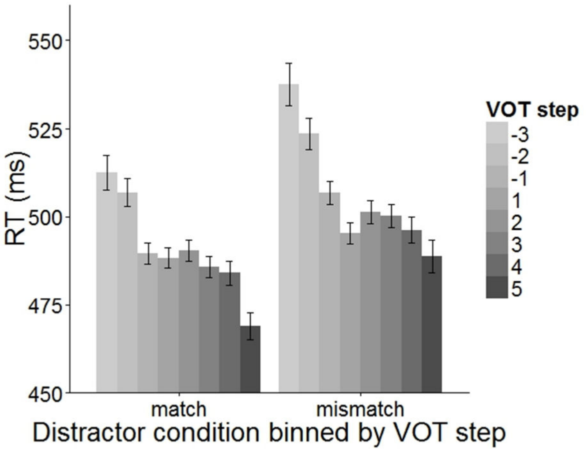


Fig. 3 appears to suggest a relation between VOT step and RTs where shorter VOT steps tend to have longer RTs. To assess this relation statistically, we fitted a model which included the distractor condition (match, mismatch) and an interaction term between distractor condition and VOT step (coded as continuous predictor) as fixed effects. The model further included the same random effects as the model described in section 3.1. 28198 data points entered the analysis.

Once more, the distractor condition was a significant predictor of RTs, with longer RTs for the mismatch condition (t = 5.610, p < 0.001). The interaction between the distractor condition and the

VOT step was not significant either for the match or the mismatch condition.

4. DISCUSSION AND CONCLUSION

The experimental results reported here replicate those of previous response-distractor studies, showing an effect on RTs based on articulator (in)congruency between a response and an acoustic distractor. As in the study by Galantucci et al. [5], RTs on trials where the distractor differed from the response in articulator (/ka/-/ta/ and /ta/-/ka/) were significantly longer than RTs on trials where distractor and response matched in articulator (/ka/-/ka/ and /ta/-/ta/). The results reported here extend that finding, showing that this effect of articulator congruency is obtained despite within-category variation in VOT, since the distractor VOT was varied for voiceless stimuli with both congruent and incongruent articulator stimuli. Previous studies [10] that have obtained an articulator effect did so using distractor stimuli with single VOT values that were fixed throughout the experiment. The clear effect of articulator congruency and its independence from distractor voicing are consistent with the model of phonological planning that has been proposed in [9] to account for RT modulations in this task.

The pattern of results in section 3.2 hinted at a relation between RTs and distractor VOT relative to the mean participant VOT, shown in Fig. 3 as VOT step: RTs seem to increase at the lowest VOT steps. However, our statistical analysis failed to confirm significance of this relation. The absence of an effect of VOT step may have been due to lack of sufficient statistical power since the data in the lowest VOT step bins (–3, –2) were substantially sparser compared to the rest of VOT step bins. This is because for a response to be in the lowest VOT step bins implies that the participant's mean VOT had to be substantially (about 30 to 45 ms) higher than the distractor VOT. Apparently, our participant population did not happen to include a sufficient number of participants with that high participant mean VOT. Note that data sparsity could not have been circumvented by using distractors with even lower VOT values; the lowest distractor VOT was set to 45 ms. Lower than that would bring the intended /ka/ or /ta/ distractor in the region of the corresponding voiced-initial syllable.

Despite the absence of significance in the relation between RTs and distractor VOT in our current dataset, such a relation, if valid, deserves attention because of its potential to inform model development. One possible explanation for slower RTs on trials with distractors with short VOTs is that distractors (/ka/ and /ta/) were misperceived by participants with high mean VOT as starting with the corresponding voiced consonants (/ga/ and /da/). That would result in lateral inhibition and thus slower RTs since on these trials the distractor's perceived voicing parameter did not match that of the intended response (cf. Experiment 1 in [10] where this result was obtained with distractor-response pairs that mismatched in voicing, i.e., /ta/-/da/ and /ka/-/ga/). A second potential explanation for the relation between VOT step and RT is that distractors with short VOTs induced less local excitation for participants with high mean VOT compared to distractors with longer VOTs. Recall that in the model in [9], excitation obtains between two VOT values when these values are 'close enough'. It is conceivable that this notion is not symmetric. In other words, distractor VOT values shorter than the planned (same phonemic category) VOT response may not excite that response to the same extent as distractor VOT values longer than the planned VOT response. Specifically, speaking to the relation between VOT step and RT seen in Fig. 3, this relation would be predicted if the degree of local excitation falls as the distractor VOT approaches the voiced-voiceless category boundary.

In sum, over and above the main predictions, our data suggest the presence of a relation between VOT and RT which, however, our statistical analysis failed to establish firmly. Further experiments and modelling work is required to address this relation.

5. ACKNOWLEDGEMENTS

AIG gratefully acknowledges support by ERC AdG 249440.

6. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4

[2] Boersma, P., Weenik, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 28 December 2014 from http://www.praat.org/.

[3] Francis, A. L., Ciocca, V., Yu, J. M. 2003. Accuracy and variability of acoustic measures of voicing onset. *J. Acoust. Soc. Am.* 113(2), 1025–1032.

[4] Gafos, A. I. 2006. Dynamics in grammar. In: Goldstein, L. M., Whalen, D. H., Best, C. T. (eds.), *Laboratory Phonology 8*. New York: Mouton de Gruyter, 51–79.

[5] Galantucci, B., Fowler, C. A., Goldstein, L. 2009. Perceptuomotor compatibility effects in speech. *Atten. Percept. Psychophys.* 71(5), 1138–1149.

[6] Kleiner, M., Brainard, D., Pelli, D. 2007. What's new in Psychtoolbox-3? *Perception* 36. ECVP Abstract Supplement.

[7] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.

[8] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

[9] Roon, K., Gafos, A. 2013. A dynamical model of the speech perception-production link. In: Knauff, M., Pauen, M., Sebanz, N., Wachsmuth, I. (eds), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin TX: Cognitive Science Society, 1241–1246.

[10] Roon, K., Gafos, A. 2015. Perceptuo-motor effects of response-distractor compatibility in speech: beyond phonemic identity. *Psychon. Bull. Rev.* 22(1), 242–250.

[11] Roon, K., Klein, E., Gafos, A. 2014. Distractor effects on response times in fricative production. In: Fuchs, S., Grice, M., Hermes, A., Lancia, L., Mücke, D. (eds), *Proceedings of the 10th International Seminar on Speech Production (ISSP)*, 356–359.