Contextual confusability leads to targeted hyperarticulation

Esteban Buz (ebuz@bcs.rochester.edu)
Department of Brain and Cognitive Sciences, Meliora Hall, Box 270268
Rochester, NY 14627-0268

T. Florian Jaeger (fjaeger@bcs.rochester.edu)
Department of Brain and Cognitive Sciences, Meliora Hall, Box 270268
Rochester, NY 14627-0268

Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Department of Brain and Cognitive Sciences, Meliora Hall, Box 270268
Rochester, NY 14627-0268

Abstract

A central question in the field of language production is the extent to the speech production system is organized for robust communication. One view holds that speakers' decision to produce more or less clear signals or to speak faster or slower is primarily or even exclusively driven by the demands inherent to production planning. The opposing view holds that these demands are balanced against the goal to be understood. We investigate the degree of hyperarticulation in the presence of easily confusable minimal pair neighbors (e.g., saying *pill* when *bill* is contextually co-present and thus a plausible alternative). We directly test whether production difficulty alone can explain such hyperarticulation. The results argue against production-centered accounts. We also investigate how specific hyperarticulation is to the segment that contrasts the target against the contextually plausible alternative. Our evidence comes from a novel web-based speech recording paradigm.

Keywords: Psychology; Linguistics; Communication; Language understanding; Speech recognition; Human experimentation

Introduction

One of the central debates in the field of language production centers around the extent to which speech is designed for robust communication. For example, what determines how fast we talk and how clearly we articulate? Similarly, what determines speakers' lexical and structural decisions, such as whether they articulate optional words or not (e.g., the optional *that* in *I think (that) is true*)? One broadly held view states that the (implicit) decisions speakers make during language production are mostly or wholly dominated by the attentional and memory demands inherent to linguistic encoding (e.g., Arnold, 2008; Bard et al., 2000). Following the literature, we refer to this as the *production-centered* view.

This view is called into question by recent work on hyperarticulation. In a series of experiments Baese-Berk and Goldrick (2009) found that speakers hyperarticulate voiceless stop consonants of target words that have lexical neighbors that only differ from the target in voicing. For example *pill*, which has the minimal neighbor *bill*, which differs in voicing where as *pipe*, does not have a minimal neighbor, *bipe* (see also Kirov & Wilson, 2012; Schertz, 2013). Moreover, hyperarticulation of voiceless stop consonants increases when the minimal pair neighbor (i.e., *bill*) is contextually co-present (e.g., by presenting both words on the same screen, Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012). One interpretation of these findings (though not necessarily shared by the authors of the above studies) appeals to the fact that one common and important goal of speaking is communication (Jaeger, 2013; Lindblom, 1990, e.g.,). Just as task-relevant errors drive learning and behavior in non-linguistic motor planning (Wei & Körding, 2009, among others), preferences during language production are taken to be the consequence of implicit learning with the goal to reduce task-relevant error (Jaeger & Ferreira, 2013). This allows the systems underlying language production to strike a balance between production ease and successful information transfer. This *trade-off* account provides a straightforward explanation for the results of Baese-Berk and Goldrick (2009) and Kirov and Wilson (2012): the likelihood of successful information transfer will increase if more confusable words are produced with more distinguishable signals and if hyperarticulation is further increased when the word would be even more confusable in its current context. This interpretation seems to be supported by other studies finding that words with more phonological neighbors in the lexicon (words that differ from the target in only one phoneme) tend to be hyperarticulated compared to words with fewer phonological neighbors (e.g., Scarborough, 2010). These latter studies found words with a greater number of phonological neighbors are produced with longer vowel durations and vowels that are further from the center of the first and second formant vowel space (greater *vowel dispersion*), both results suggesting that speakers provide a more distinguishable signal for (*a priori*) more confusable words.

However, alternative interpretations of the above results have been advanced under the production-centered perspective (e.g., Baese-Berk & Goldrick, 2009; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl, Yao, & Johnson, 2012). According to this view, lexical or contextual presence of phonologically similar words increases production difficulty, which is reflected in hyperarticulation. For example Baese-Berk and Goldrick (2009) argue that competition between phonologically similar forms increases the difficulty of phonological encoding (see also the discussion in Kirov & Wilson, 2013). The idea that difficulty during the *plan-*

ning of a word results in slower and more detailed *articulation* of the word is seemingly supported by the observation that high frequency words tend to take less time to plan and tend to have shorter durations (e.g., Oldfield & Wingfield, 1965). Thus it seems that production-centered accounts provide a parsimonious explanation for both behavioral correlates of production difficulty (e.g., latencies) and articulation. However, despite the centrality for the claim that planning difficulty explains hyperarticulation (e.g., Bell et al., 2009; Gahl et al., 2012), we know of no study that directly tests this claim (and, in particular, not for effects of contextual confusability on articulation). Additionally, a recent comprehensive review found that more phonological neighbors do not always lead to increased difficulty (Sadat, Martin, Costa, & Alario, 2013). This calls into question production difficulty as an explanation of neighborhood density effects on articulation (like those obtained by, e.g., Scarborough, 2010). It remains unclear whether production-centered accounts can account for effects of *a priori* or contextual confusability on *articulation*. This is the primary question we seek to address here. Specifically we ask:

1. Does contextual confusability affect production difficulty?
2. Regardless of whether or not context affects productions, can differences in articulation be explained by planning?

To this end, we conducted an experiment similar to those reported in Baese-Berk and Goldrick (2009) and Kirov and Wilson (2012). Unlike those studies, we measured production latencies, which are a well accepted measure of the difficulty experienced during production planning (Oldfield & Wingfield, 1965). Production-centered accounts would predict that, to the extent that we replicate the contextual confusability effect on articulation, we should also observe an effect on production latencies and, at the least, we should observe that production latencies are a predictor of the degree of hyperarticulation (cf., Bell et al., 2009; Gahl et al., 2012). The opposing trade-off view, that language production is subject to not only planning demands inherent to production but also the goal of robust communication, predicts that hyperarticulation can be observed in the absence of production difficulty.

A secondary goal of this paper is to test the *specificity* articulation. One possibility is that speakers hyperarticulate all aspects of words presented with a minimal pair neighbor. An alternative possibility is that hyperarticulation is restricted to those aspects of the signal that contrast the target from its minimal pair are hyperarticulated (for preliminary evidence, see also Kirov & Wilson, 2012, discussed below in more detail). If hyperarticulation affects the whole word (e.g., increasing the duration of the word), this would also mean that previous findings of hyperarticulated VOTs (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012) are confounded: VOTs are known to be longer at slower speech rates (Kessinger & Blumstein, 1998). We hence compare the effect of contextual confusability on word durations and VOT and conduct additional analyses to address the possibility of a confound.

Study: Contextual confusability, planning and hyperarticulation

To collect speech data we adapted the paradigm used by Baese-Berk and Goldrick (2009) and Kirov and Wilson (2012). As in prior work, our critical stimuli were words with a voiceless stop onset that had a voiced stop onset minimal pair (e.g. critical target *pill* with minimal pair *bill*).

Method

Participants 10 participants (5 female; 5 male; aged $18-62$, mean $= 30.1$) were recruited using Amazon's Mechanical Turk (www.mturk.com). All participants were self-reported native speakers of American English.

Materials All materials were a subset of those used in Kirov and Wilson (2012), Study 2. There were 36 critical, 54 filler, and 6 practice target words. Critical targets began with a voiceless stop consonant (/k, p, t/) and had a voiced stop consonant minimal pair (/g, b, d/). Filler and practice targets were monosyllabic words that did not begin with /k, p, t, g, b, d/. Filler and practice targets were presented with two phonologically unrelated monosyllabic words. Critical targets were presented in one of two trial context conditions: with two phonologically unrelated monosyllabic words (competitor absent) or with its voiced minimal pair and an unrelated monosyllabic word (competitor present).

Procedure The experiment was conducted online with Mechanical Turk, using a novel procedure to record speech over the web. Participants were instructed that they were taking part in an interactive communication task. After reading the task description and giving informed consent they were asked to wait while a partner was found. After a variable delay, they were informed by our software that they had been matched to a partner. In reality, the partner was simulated by our software. We used this simulated partner approach to as closely match the procedure employed in experiments by Baese-Berk and Goldrick (2009) and Kirov and Wilson (2012), which were performed in the lab with a confederate partner.

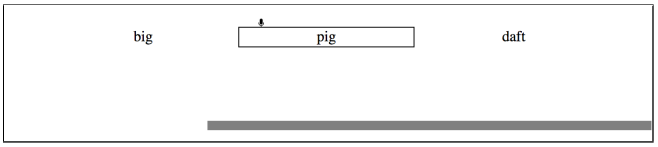

Figure 1: Participant screen with running timer.

Each trial began with a short "re-sync" screen that illustrated, at variable timing, establishment of a connection to the (simulated) partner. Then three words were presented horizontally across the participant's screen along with a horizontal timer bar at the bottom of the screen (see Figure 1). Words were presented for 1500 msec, after which the target was outlined with a black box and paired microphone icon and the timer bar began to shorten. Participants were told to utter the cued target to their simulated partner. To avoid overly slow re-

sponses, trials were timed, with the timer bar counting down to 10 seconds. Participants were instructed that trials ended after 10 seconds or whenever their partner answered by clicking a word. Participants did not receive feedback about the (simulated) partner's choice, but the timer bar stopped and the trial ended.

Several steps were taken to increase the believability of the (simulated) partner. A simulated connection screen showed various stages of the connection being established with the participant's and simulated partner's computers. Participants were allowed a short post-trial break of 30 seconds and were allowed to "request" a longer 5 minute break from their (simulated) partner that could be ended early. Our software would respond with some variability to these requests with a naturalistic delay; our software limited participants to 2 long breaks. To simulate realistic partner response times we estimated participant speech onset time and partner mouse click response times (from speech onset) as a function of log trial number using data from an unrelated single picture naming experiment and another unrelated spoken word recognition 4AFC picture selection experiment. This resulted in simulated response times decreasing during the experiment at a rate that resembled natural behavior in this type of task.

The speech for each trial was recorded individually using the participant's own computer and microphone configuration and saved to a server for analysis (Gruenstein, McGraw, & Badr, 2008). After the experimental list was completed each participant was presented with a post-test survey that collected demographic information.

Believability of the paradigm In an effort to assess the the believability of the simulated partner we asked participants a series of increasingly targeted questions about their partner's behavior. Questions were presented on subsequent screens, with no option to return to previous screens. First, we asked participants to rate the their connection quality on a 1 to 7 scale (poor to good, mean $= 6.3$; se $= 0.2$).

Second, we asked participants to rate various aspects of their partner's response time. Participants rated their partners as fairly fast responders (mean $= 5.7$, se $= 0.3$; 1 to 7 scale, slow to fast). When asked to rate the amount of audio transmission delay between them and their partner they rated the delay as low (mean $= 2.3$, se $= 0.3$; 1 to 7 scale, no delay to very delayed). We asked participants to note how many times their partner 1) failed to respond in time, 2) responded prior to the participant finishing, and 3) responded prior to the participant starting to speak. One participant noted that their partner exhibited all three behaviors once. A different participant noted that their partner responded before they finished speaking on one trial (we informed participants in the instruction that this might happen as both participant and partner are under a time limit to finish the experiment). All other participants stated their partner did none of these behaviors. This suggests that the partner response times that we programmed were sufficiently natural to be neither too fast (e.g., responses before speech initiation), nor to slow.

Third, we asked participants to note any oddities in the experiment and in their partner (e.g. "Did you noticed anything weird during the experiment?"). One participant commented that their partner's response times were very consistent. One participant explicitly stated that they did not believe they had a partner. That is, prior to any more specific information, most participants did not seem to consider their (simulated) partner sufficiently odd to comment on.

Fourth, we told participants that in our study we randomly paired participants with a real person or a computer. We then asked participants to rate how human-like their partner acted (1 to 7 scale, computer to human-like). Predictably, the two participants who did not believe our setup stated their partner was computer like (ratings of 1 and 2). The remainder of the participants gave higher ratings (mean $= 3.4$, se $= 0.4$).

Finally, we told participants that they indeed had been (randomly) paired with a computer, rather than a human, and asked about the believability of their simulated partner. Participants rated our cover story as fairly believable (mean $= 5.3$, se $= 0.5$; 1 to 7 scale, not believable to very believable). Now being informed that their partner was in fact not human, participants were split when asked if they felt like they were interacting with a person: 5 ratings of 3 or less and 5 ratings of 6 or more (1 to 7, didn't feel real to felt real). Overall, these results suggested that the simulated partner was sufficiently convincing for most participants; only after participants were told that their partner was in fact not a person did their ratings drop. The final part of the survey also solicited comments on if and how they might have realized their parter was not real. Six participants noted that they felt their parter was too consistent in their response times (a detail we plan to modify for future studies). All participants were debriefed after the experiment.

We excluded the two participants who did not believe the interlocutor from the analyses reported below (all results hold if these participants are included). Interestingly, exclusion increased the context effect on VOT reported below by 25%, suggesting that the believability of communicative partners affects articulation (cf. Lockridge & Brennan, 2002, for a similar finding for lexical and syntactic planning).

Acoustic analysis Speech onset latency, VOT and word duration were manually annotated and measured using Praat (Boersma & Weenink, 2014). Speech onset latency was the time between target word cue presentation and the onset of speech. Word onset was defined as the point of zero-amplitude on the waveform nearest the stop consonant release. VOT was defined as the time between word and vowel onset. Vowel onset was defined as the point of zero-amplitude on the waveform nearest the onset of periodicity. Word duration was measured as the time from word onset to when no visible speech signal was present in the waveform or spectrogram. Word durations were log-transformed for analysis.

All participants followed the task (e.g. not uttering non-target words or uttering multiple words) so no further participants were excluded. Following participant exclusions,

tokens were excluded for disfluencies, mispronunciations, background noise obscuring word and or vowel onsets, or recording issues (0% of all tokens). Finally, latency, logged word duration and VOT outliers (absolute z-score value $>$ 2.5) were removed by participant (0.06% of all tokens). All results reported below hold with or without exclusions.

Results

We first assessed whether the context manipulation (competitor present vs. absent) affected VOT and word durations of critical targets. Following that, we present a mediation analysis, assessing the effect of the context manipulation on VOT *while controlling for effects of word duration*. Following this we assessed whether the context manipulation affected speech onset latencies. Further we assessed whether speech onset latencies affected VOT *while controlling for effects of context and word duration*. All analyses were conducted using a mixed effect linear regression (maximal RE structure) with fixed effects for context condition (competitor present or absent, ANOVA-coded). Following standard procedure, no random slopes were added for the covariates in the mediation analysis (adding these slopes, did not change the results). Significance was assessed by comparing model fit without a predictor of interest to model fit with that predictor of interest.

If participants hyperarticulate contextually confusing words, we would expect exaggerated VOTs when the competitor is present. Further, if hyperarticulation is specific to the feature that increases the relevant contextual contrast (in this case, VOT), we should not observe effects of context on word duration—or, at least, none that cannot be reduced to changes in VOT. Finally, if the inherent demands on production planning, rather than a bias for robust communication, underlie whatever effects are observed for VOT and word duration, these effects should be reducible to a (possibly non-linear) function of speech onset latencies.

Context effect on articulation The only articulatory measure significantly affected by our design manipulation (context) was VOT: VOTs were on average 9.1 msecs longer when the competitor was present on the screen compared to when it was absent ($\hat{\beta} = 4.6$; $t = 3.4$; $p < .01$). Total word duration did not differ significantly across contexts ($\hat{\beta} = 0.003$; $t = 0.7$; $p > .4$). This replicates results of previous lab-based studies (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012). Figure 2 shows VOT difference across contexts. These results suggest that participants hyperarticulated VOTs of contextually confusable words and this hyperarticulation was restricted to VOTs, rather than the entire word.

To examine to what extent differences in VOT are driven by differences in word duration (Kessinger & Blumstein, 1998), we also conducted model comparisons between a model predicting VOT by context and a model predicting VOT by context and word duration. Word duration significantly improved model fit ($\chi^2(1) = 36.7$; $p < .01$). Longer word durations predicted longer VOTs ($\hat{\beta} = 133.3$; $t = 6.5$; $p < .01$). However, the effect of context remained significant ($\hat{\beta} = 4.3$; $t = 3.7$; $p < .01$). Table 1 summarizes both the main analysis of VOT (analysis 1) and the follow-up analysis controlling for log-transformed word duration (analysis 2).

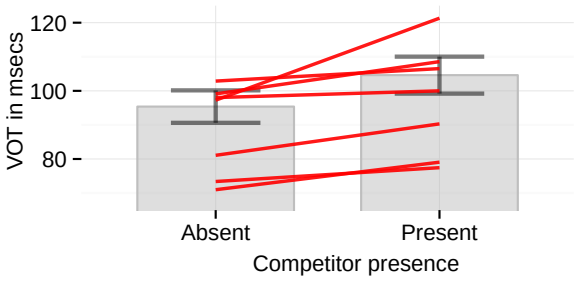

Figure 2: Voice onset timing (VOT) by condition aggregating within participants (lines) and across participants (bars). Error bars indicate $\pm$ 1SE after aggregating over participants.

Effect of context on planning difficulty We found no difference in speech onset latencies across contexts ($\chi^2(1) = 0.04$; $p > .6$). Log transforming latency did not change this result. This suggests that the visual co-presence of a minimal pair neighbor does not result in planning difficulties.

While context did not affect planning difficulty, planning difficulty may still affect articulation. We conducted model comparisons between a model predicting VOT by context and a model predicting VOT by context and latency. Speech onset latency did not significantly improve model fit ($\chi^2(1) = 1.6$; $p > .6$). Because word duration does predict VOT we additionally tested if latency improved model fit after controlling for word duration. Speech onset latency did not significantly improve model fit ($\chi^2(1) = 1.4$; $p > .5$). Table 1 summarizes the two follow-up analysis controlling for latency (analysis 3) and both duration and latency (analysis 4). We further tested for non-linear effects of latency on VOT, with and without word duration as a covariate, by testing for the addition of restricted cubic spline transformations of latency (with 3, 4, or 5 knots). In neither case, did the non-linear transformations of latency significantly improved model fit ($\chi^2(4)$s< 5.1; ps$> .2$)

In sum, then, we did not find any evidence that context affected production planning. Neither did any of our analysis reveal evidence in support of the hypothesis that production difficulty causes the observed effects on VOT. Our results do not, of course, rule out the possibility that production difficulty does not affect articulation in other situations. They do, however, argue against an explanation of the current VOT results (hyperarticulation of contextually confusable words) in terms of only production difficulty.

Discussion

We evaluated two opposing explanations for why speakers choose to articulate a word with more or less signal. According to the production-centered account, hyperarticulation is caused by production difficulty (e.g., Bell et al., 2009; Gahl

Table 1: Coefficients (and SEs) of context effect on VOT while controlling for possible confounds.

	Dependent variable: VOT			
	(1)	(2)	(3)	(4)
Intercept	−0.5	−0.3	−0.5	−0.4
	(7.6)	(6.0)	(7.5)	(5.9)
Competitor	4.6***	4.3***	4.5***	4.2***
	(1.3)	(1.2)	(1.3)	(1.2)
Log word duration		133.3***		133.1***
		(20.7)		(20.7)
Latency			−0.005	−0.004
			(0.004)	(0.003)
Note:				*p<0.1; **p<0.05; ***p<0.01

et al., 2012). The opposing trade-off account holds that hyperarticulation serves to facilitate robust recognition of the target word (Jaeger, 2013; Lindblom, 1990). We focused on the effect of a minimal neighbor on VOTs, replicating previous results that VOTs where hyperarticulated when a voice contrastive minimal pair was co-present during production (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012, 2013). We also conducted analyses that eliminated a confound in previous studies, that VOTs are correlated with overall word duration (Kessinger & Blumstein, 1998), and found that hyperarticulation was specific to the distinguishing dimension. Crucially the production-centered account, but not the trade-off account, predicts that planning difficulty should account for increased VOTs. Using speech onset latency as a measure of planning difficulty we found no evidence that copresence of a minimal pair increased planning difficulty. We also found no evidence that latencies modulate VOTs, suggesting that planning difficulty is not the only factor underlying articulation.

We further find no evidence that contextually driven VOT lengthening is the result of overall word lengthening. This suggests that the hyperarticulation caused by contextual confusability is quite specific. Speakers are modifying fine grain aspects of their productions and do so in a way that suggests they are producing words that are perceptually further than contextually relevant competitors. Two recent studies corroborate this finding. Kirov and Wilson (2012) found that speakers hyperarticulated VOT of words like *pill* when a voiced minimal pair neighbor (e.g. *bill*) was contextually present but not a vowel or coda contrastive minimal pair (e.g. *pull* or *pick*). Using a different paradigm Schertz (2013) replicated this finding and extended it to duration contrastive vowels (e.g. hyperarticulation of /i/ vs /ɪ/).

Our finding of contextually specific hyperarticulation rather than generic hyperarticulation highlights a potential confound in earlier work. Many studies have used duration of individual segments (e.g., vowel duration) as well as whole word duration to argue for or against the trade-off view of speaker behavior. For example, the longer duration of words in denser phonological neighborhoods may be seen as "for the listener" in that lengthening may aid comprehension (Scarborough, 2010). Alternatively, evidence that these same words are actually produced with shorter durations has been argued as evidence that speaker behavior is not, in part, driven by communicative goals (Gahl et al., 2012). However, we find that if speakers are hyperarticulating to aid their listeners they do so in a contextually specific way. It is an open question if these findings are applicable to other features of articulation, such as vowel production. Contextually specific changes in duration suggests contextually specific vowel changes. The upshot is that a measure such as vowel dispersion (Gahl et al., 2012; Scarborough, 2010), which tracks overall changes in vowel space rather than specific movements in vowel space (distance from contextually contrastive vowels), may be too coarse grained a measure of if and how speakers hyperarticulate to increase utterance intelligibility. Rather, it suggests that the best place to look for evidence for or against the trade-off view is using measures that are specific to the stimuli in question (e.g. our use of VOT and stimuli which differ in VOT).

One recent finding is potentially in conflict with this hypothesis. Schertz (2013), using target words with vowels that contrasted with contextually relevant alternatives found no evidence that speakers shifted target vowels away from their co-present minimal pair. It is possible that speakers are only able to vary certain aspects of production, such as VOT, but not others, such as vowel formants. Another possibility is that the vowel space of most languages—or at least English—is too densely populated. That is, speakers might not be able to increase the intelligibility of vowels by moving articulation away from a contextually present competitor without inevitably increasing confusability with another competitor (see Schertz, 2013). This put vowels in contrast with phonetic features like VOT, which (in English) can be safely exaggerated. Further work is warranted on this point.

Future directions

To the best of our knowledge, this is the first study to use a web-based paradigm to investigate speech production. Despite the fact that the context effect on VOTs was in the order of only 9.1 msec (as in previous work), our paradigm reliably replicated the effect. The results were robust and independent of all exclusion criteria. This suggests that, at least for durational/temporal acoustic variables (such as VOT and word durations, but also, e.g., vowel durations), web-based recordings can achieve the accuracy required for speech production research.

The web-based paradigm introduced here has several advantages. Large numbers of participants can easily be recruited within a day (e.g., in another ongoing study, we recorded over 300 speakers in a few days) and at lower costs. One particular advantage specific to the question

under discussion here—to what extent articulation reflects a trade-off between production ease and by a bias for robust communication—is that it allows direct control over the amount of feedback provided to the speaker. While research within the paradigm of speech perturbation (e.g., Houde & Jordan, 1998), has shed light on the role of self-monitoring during articulation, little is known to what extent speakers can integrate feedback form their interlocutors (whether implicit or explicit, verbal or visual, etc.) to change subsequent articulations. In the current experiment, we intentionally removed any form of feedback about the success of communication. The only feedback speakers received was when their interlocutor responded (the timer stopped and the next trial began). This was done to address the possibility that the confederates in Baese-Berk and Goldrick (2009) and Kirov and Wilson (2012) subconsciously provided feedback to the speaker (e.g., through facial expressions), thereby confounding *a priori* articulation preferences with those that result from learning based on interlocutor feedback (cf. Lockridge & Brennan, 2002). In ongoing work, we are using the same paradigm to investigate how feedback from interlocutors affects subsequent productions.

Acknowledgments

This work was partially funded by NSF CAREER Award (NSF IIS-1150028) to TFJ and an NIH training Grant at the University of Rochester (#T32 DC000035).

References

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527.

Baese-Berk, M. & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, *24*(4), 527–554.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M. P., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806.

Gruenstein, A., McGraw, I., & Badr, I. (2008). The WAMI toolkit for developing, deploying and evaluating web-accessible multimodal interfaces. In *10th international conference on multimodal interfaces*.

Houde, J. F. & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(1213), 1213–1216.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, *4*, 230.

Jaeger, T. F. & Ferreira, V. S. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, *36*(4), 359–360.

Kessinger, R. H. & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, *26*(2), 117–128.

Kirov, C. & Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 587–592). Austin, TX: Cognitive Science Society.

Kirov, C. & Wilson, C. (2013). Bayesian speech production: Evidence from latency and hyperarticulation. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 788–793). Austin, TX: Cognitive Science Society.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Kluwer Academic Publishers.

Lockridge, C. B. & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*(3), 550–7.

Oldfield, R. C. & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2013). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68C*, 33–58.

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kuehnert, M. Imperio, & N. Vallee (Eds.), *Laboratory phonology 10* (Vol. 10, pp. 557–586). Berlin; New York: De Gruyter Mouton.

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, *41*(3-4), 249–263.

Wei, K. & Körding, K. P. (2009). Relevance of error: what drives motor adaptation? *Journal of Neurophysiology*, *101*, 655–664.