Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations

Esteban Buz[a,*], **Michael K. Tanenhaus**[a,b], and **T. Florian Jaeger**[a,b,c]

[a]Department of Brain and Cognitive Sciences, University of Rochester, United States

[b]Department of Linguistics, University of Rochester, United States

[c]Department of Computer Science, University of Rochester, United States

Abstract

We ask whether speakers can adapt their productions when feedback from their interlocutors suggests that previous productions were perceptually confusable. To address this question, we use a novel web-based task-oriented paradigm for speech recording, in which participants produce instructions towards a (simulated) partner with naturalistic response times. We manipulate (1) whether a target word with a voiceless plosive (e.g., *pill*) occurs in the presence of a voiced competitor (*bill*) or an unrelated word (*food*) and (2) whether or not the simulated partner occasionally misunderstands the target word. Speakers hyper-articulated the target word when a voiced competitor was present. Moreover, the size of the hyper-articulation effect was nearly doubled when partners occasionally misunderstood the instruction. A novel type of distributional analysis further suggests that hyper-articulation did not change the *target* of production, but rather reduced the probability of perceptually ambiguous or confusable productions. These results were obtained in the absence of explicit clarification requests, and persisted across words and over trials. Our findings suggest that speakers adapt their pronunciations based on the perceived communicative success of their previous productions in the current environment. We discuss why speakers make adaptive changes to their speech and what mechanisms might underlie speakers' ability to do so.

Introduction

Speech production is context sensitive. This is most obvious and best understood with regard to linguistic context. For example, how a sound is articulated and pronounced depends on the surrounding sounds (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) and its

*Corresponding author. ebuz@bcs.rochester.edu (E. Buz), mtan@bcs.rochester.edu (M.K. Tanenhaus), fjaeger@bcs.rochester.edu (T.F. Jaeger).

position in the larger linguistic structure (e.g., due to stress assignment and other prosodic factors, Klatt, 1976). Speech production is also sensitive to the broader non-linguistic context. This includes, for example, adjustments in how we talk due to the levels of acoustic noise in the local environment—speakers tend to talk louder when in a noisy environment (known as the Lombard effect, Lombard, 1911; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). It also includes adjustments based on whom we are talking to. For example, we sometimes revert to our home dialect when talking to friends or family from that region, or switch to less formal registers when talking to people we know, resulting in changes to speech rate and clarity of articulation, among other things (e.g., Bell, 1984; Finegan & Biber, 2001). Sensitivity to the socio-indexical context in which speech takes place goes beyond adjustments to interlocutors we know. Speakers also can adjust their pronunciations based on types of interlocutors. For example, speech directed at adults differs systematically from speech directed at infants (e.g., Kuhl et al., 1997; Pate & Goldwater, 2015) or pets (e.g., Burnham, Kitamura, & Vollmer-Conna, 2002). Similarly, speech directed at typical adult native interlocutors differs from speech directed at non-native interlocutors (e.g., Uther, Knoll, & Burnham, 2007) or audiences with impaired comprehension (e.g., "clear speech", speech directed at the hard of hearing, Picheny, Durlach, & Braida, 1986).

Examples like these illustrate that non-linguistic context can affect pronunciation. They also suggest that speech production is to some extent adaptive, allowing speakers to adjust their productions depending on their audience. These examples leave open, however, how *dynamic* such adjustments are. The present study begins to address this question. This question is both under-explored and of central importance to our understanding of the architecture underlying language production. In the longer-term, understanding adaptive processes holds the potential to shed light on the origin of socio-indexically conditioned registers, such as infant- and foreigner-directed speech. Adaptive processes may also be key to reconciling seemingly conflicting results in research on audience design (as proposed in Jaeger & Ferreira, 2013; for discussion see Jaeger & Buz, in press). Beyond contributing to these longer-term goals, our more immediate goal is to gain a better understanding of the nature of adaptation in speech. Specifically, we investigate whether and if so how speakers adapt their pronunciations—by *hyper-articulating* certain sounds—based on feedback from interlocutors about the communicative success of the speaker's previous productions. This then leads us to investigate hyper-articulation in such situations more closely. Precisely how do speakers adapt their articulations in response to feedback that suggests that their previous utterance was perceptually confusing? And what is the likely function or goal of this adaptive behavior?

We approach these questions in a novel web-based task-oriented *simulated partner* paradigm for speech recording. Participants provide instructions to a partner, who—unbeknownst to the participant—is simulated by a computer program. This allows us to control the timing and type of feedback that speakers received from their interlocutors, while maintaining ecological validity (as indexed by ratings reported below). Specifically, we manipulate what feedback participants receive on individual trials about whether their partner understood them. We then assess the degree to which participants hyper-articulate as a function of the perceived communicative success of their previous productions.

Before we describe our study in more detail, we briefly summarize previous work on the effect of interlocutor feedback on speakers' articulations and highlight how the present experiment contributes to this literature.

Previous work and how the present study contributes to it

Only a few studies have directly investigated the role of interlocutor feedback on subsequent productions. One line of research that is particularly relevant to the current goals has investigated the articulations of corrections following explicit clarification requests (Maniwa, Jongman, & Wade, 2009; Ohala, 1994; Oviatt, Levow, Moreton, & MacEachern, 1998; Oviatt, MacEachern, & Levow, 1998; Schertz, 2013; Stent, Huffman, & Brennan, 2008). For example, Schertz (2013, Study 1) recorded participants as they produced speech directed at what they believed to be an automatic speech recognition system. Target words either had voiced or voiceless plosive onsets (e.g., *pit*). On critical trials, the (simulated) automatic speech recognition system displayed a recognition error and requested clarification. Participants then had to repeat the same word. Schertz found that corrections were hyper-articulated (see also Maniwa et al., 2009; Oviatt, Levow, et al., 1998; Oviatt, MacEachern, et al., 1998; Schertz, 2013; Stent et al., 2008; for similar findings in response to a simulated human partner see Ohala, 1994).

Interestingly, the hyper-articulation observed in these studies was often targeted to the specific part of the production that seemed to have caused the misrecognition. For example, the (simulated) automatic speech recognition system in Schertz (2013) used more or less specific clarification prompts to indicate which part of participants' productions had likely caused the misrecognition. Sometimes clarification prompts were general ("???"). Other times, prompts contained specific guesses that deviated from the target (e.g., *pit*) in either voicing ("bit?"), place ("kit?"), or manner (e.g., "sit?"). Schertz found that voice onset times (VOT)—the primary cue to the English voicing distinction (e.g., *pit* vs. *bit*)—were hyper-articulated only following voicing-contrastive word prompts (e.g. "bit?") but not when participants saw general prompts ("???") or manner/place-contrastive prompts (e.g. "kit?" or "sit?"). Additionally, hyper-articulation after voicing-contrastive prompts was limited to VOTs: neither the overall amplitude nor overall word duration was hyper-articulated (see also de Jong, 2004; Maniwa et al., 2009; though see Ohala, 1994).

These results suggest that speakers can adapt productions of the same word immediately following an explicit request for clarification, and that they can do so in a targeted manner. Here we seek to contribute to this literature and to extend it in several ways. First, the majority of previous studies had participants produce words towards (simulated) automatic speech recognition systems (but see Ohala, 1994). There is evidence that speech directed at automatic speech recognition systems differs qualitatively from speech directed at human interlocutors (Oviatt, Levow, et al., 1998; Oviatt, MacEachern, et al., 1998; see also the discussion in Stent et al., 2008, p. 166). For this reason, the present paradigm employs a (simulated) human interlocutor.

Second, the studies summarized above employed specific requests for clarifications to elicit hyper-articulated productions. In those paradigms, participants typically are asked to produce the same word again immediately following the clarification request. This arguably

increases the likelihood that speakers use conscious repair strategies. While it is an open question whether such strategies originate in different mechanisms than more implicit adjustments to speech, the paradigm we present below aims to test adaptive speech behavior under more implicit conditions. To this end, we used a word naming paradigm that has previously been shown to elicit context-specific hyper-articulation (e.g., Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012). In this task, participants and their (confederate) partners sat in front of separate computer screens and saw the same three words. One target word was highlighted on the participant's screen who then verbally produced it so as to get their partner to click on the same word on their own screen.

Critical target words had voiceless plosive onsets (i.e., /*p*/ as in *pill*). In critical trials, a voiced onset competitor (*bill*) either was or was not displayed as one of the two distractor words. When the voiced competitor was present on the screen, participants hyper-articulated the VOT on the voiceless target compared to when the voiced competitor was not present (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012; for related results see also Buz, Jaeger, & Tanenhaus, 2014; Kang & Guion, 2008).[1] Paralleling the majority of studies on explicit clarification requests, these hyper-articulations seem to be targeted (Buz et al., 2014; Kirov & Wilson, 2012; Seyfarth et al., in press). For instance, Kirov and Wilson (2012), found that speakers exaggerate voiceless onset plosive VOTs when a voiced competitor was present but not when another type of minimal contrast competitor was present.

Here, we use this paradigm to test whether the type of feedback participants receive from their interlocutors affects the degree to which they hyper-articulate. This allows us to embed the manipulation of interest (whether a speaker sees that her production was understood or misunderstood) in a task-oriented language game. Feedback from the (simulated) partner came in the form of the partner clicking the target word or a competitor. Unlike in previous studies with explicit clarification requests, participants did not repeat the misunderstood word. Rather, the program continued to the next trial. Only after intervening filler trials would another critical trial with the same phonetic (voicing) contrast occur. This serves our goal of testing adaptive speech behavior under more implicit conditions: investigating hyper-articulations of the same contrast across words with several intervening filler trials, rather than repeated productions of the same word, further reduces the likelihood of conscious strategies.

Assessing adaptive adjustments to a phonetic contrasts across *different* words is also of theoretical interest. First, if dynamic hyper-articulation is observed under these circumstances, this would argue that the speech production system is substantially more adaptive than previously demonstrated and, we will argue, in 'smarter' ways than previously assumed. In the General Discussion, we expand on this idea and discuss a general framework that motivates the current work and, we hope, can help guide future work on adaptation. Second, as we describe in more detail below, our design also lets us test whether

---

[1]One recent study failed to find evidence that speakers hyper-articulate voicing contrasts on syllable-final plosives for words with voicing contrastive minimal pairs, e.g. no hyper-articulation for words like *sob* (minimal pair: *sop*) as compared to *knob* (**nop*) (Goldrick, Vaughn, & Murphy, 2013). After addressing confounds in the selection of stimuli, which reduced its statistical power, this study returned a null effect. A recent follow-up study on voicing contrasts in syllable-final sibilants, however, does find hyper-articulation in the presence of a voicing competitor (Seyfarth, Buz, & Jaeger, in press).

these hyper-articulations are *context-specific*: if interlocutor feedback affects subsequent productions of the same phonetic contrast, it could do so either across any type of context (e.g., hyper-articulation of subsequent voiceless plosives regardless of whether a voiced competitor is present or not) or only in the type of context in which the speaker previously experienced miscommunication.

Third, we will argue, the use of non-verbal feedback in particular allows us to evaluate a number of competing explanations for the effects we observe. This includes, in particular, explanations of our results in terms of competition processes during lexical planning (Arnold & Watson, 2015; Arnold, Kahn, & Pancani, 2012; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Baese-Berk & Goldrick, 2009; Goldrick et al., 2013; Kahn & Arnold, 2012; Kirov & Wilson, 2013; Lam & Watson, 2010; for review see Jaeger & Buz, in press), alignment to interlocutors (Giles, Coupland, & Coupland, 1991; Pickering & Garrod, 2004), self-monitoring (Levelt, 1999; Levelt, Roelofs, & Meyer, 1999), or the perception-production loop (Pierrehumbert, 2001, 2002; Wedel, 2006). We argue that a plausible explanation of our results requires that speakers are capable of monitoring the effect of their productions on their interlocutors and that this can result in hyper-articulation of subsequent productions.

Finally, we present additional analyses of the effect of voicing competitors and feedback on the *distribution* of articulations. These analyses can help to narrow down hypotheses about the function of hyper-articulation as well as the mechanisms underlying it. For example, one possibility is that hyper-articulation in the present paradigm leads to *exaggerated* targets for production. Another possibility is that speakers aim for the same articulatory, acoustic, or perceptual target (and thus, e.g., VOT value), regardless of whether or not they are hyper-articulating. In this latter scenario, increases in the average VOT produced by speakers could result from changes in the *shape or variance* of the overall distribution of VOTs. One specific hypothesis that prompted us to conduct this distributional analysis is that hyper-articulation in the present paradigm serves to reduce the probability of (unintentionally) producing perceptually confusable VOT values. This analysis thus also holds the potential to shed light on property of hyper-articulation in the presence of competitors that is, at first blush, puzzling. The magnitude of VOT hyper-articulation in the presence of voicing competitors, for example, is typically small (Baese-Berk & Goldrick, 2009; Buz et al., 2014; Kirov & Wilson, 2012). This has led some to question whether the function of such hyper-articulation originates in communicative goals (e.g., Goldrick, personal communication, and an anonymous reviewer). The distributional analysis we present below serves as a first step towards addressing this question.

Experiment

In an isolated word naming experiment, participants saw three words displayed on their screen and after a delay named a highlighted target word to their partner. A 2×3 design manipulated whether or not one of the two distractor words was an onset voicing contrastive minimal pair of the target (within-participants) and what feedback participants received from their partner after each utterance (between participants). Participants in the No Feedback group were informed at the end of each trial that their partner had made a response and to

prepare for the next trial. Participants in the Positive Feedback group were informed at the end of each trial that their partner had successfully selected the target. Participants in the Mixed Feedback group saw feedback similar to the Positive group but on a small number trials they were informed their partner had selected a different word.

We tried to establish a task environment in which successful communication was perceived to be relevant. This is important because there is evidence that under some circumstances communicative behaviors may emerge only when speakers perceive that their utterances are relevant to their interlocutor. For example, some studies in which interlocutors' response timing suggests that they have access to information that makes audience design on the part of the speaker *unnecessary* (e.g., because the interlocutor is an *informed* confederate who already knows what the speaker will say) have not found evidence for audience design (Brown & Dell, 1987; Keysar, Barr, & Horton, 1998). Crucially, similar studies in which interlocutors were *uninformed* and thus exhibited "realistic" response timing have found effects of perspective taking (Lockridge & Brennan, 2002). In a recent review, Kuhlen and Brennan (2013) identify realistic timing of interlocutors' responses and backchannels as more important in eliciting audience design behavior than top-down indexical information about the interlocutor (i.e., whether an interlocutor is described as a confederate or as a naive participant).

We employed a simulated partner paradigm, in which participants were told they were talking to a human partner over the web. In reality the partner was simulated by the experimental software. Critically, the partner's response times were simulated based on naturalistic response time data from previous experiments (see below for details). This also allowed us to manipulate feedback across participants, while holding everything else constant including partner identity, response timing, and other partner behavior (e.g., changes in confederate behavior that result from repeatedly participating in the same experiment; cf. Heller, Grodner, & Tanenhaus, 2009).

Method

Participants—Sixty participants (25 female, 30 male, 5 declined to provide a gender; age range 18–62 years, mean = 30.18) were recruited using Amazon's Mechanical Turk (www.mturk.com). All participants were self-reported native speakers of American English (as confirmed by listening to their audio recordings). Speakers were randomly assigned to one of the three Feedback conditions, until each condition had been completed by 20 participants.

Materials—There were 36 critical, 54 filler, and six practice trials. Critical targets were monosyllabic words which began with a voiceless plosive consonant (/k, p, t/, e.g. *pill*) and had an onset voicing contrastive minimal pair (/g, b, d/, e.g. *bill*). 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 (e.g. target: *chime* and distractors *job* and *hop*). Critical targets were presented in one of two Voiced Competitor conditions. In Voiced Competitor Absent trials, critical targets were displayed with two phonologically unrelated monosyllabic words (e.g. *pill* and distractors *food* and *hair*). In

Competitor Present trials, critical targets were displayed with its onset voicing contrastive minimal pair and an unrelated monosyllabic word (e.g. target: *pill* and distractors *bill* and *hair*). All stimuli are listed in Table A.6.

We employed a Latin square design for the within-participant Competitor manipulation; each participant saw each critical target in either the Competitor Absent or Present condition. We generated one pseudo-randomized trial order for the two Latin square lists and used this and its reversed order to create a total of four experimental lists. Randomization was constrained such that one or two fillers occurred at the start, end, and between critical items, and no more than three critical items in the same condition occurred consequently (ignoring intervening filler). The three Feedback groups were presented these four lists, five participants per list per Feedback group.

Unlike previous studies on competitor effects in speech production (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012; Ohala, 1994; Schertz, 2013), we avoided repeated naming of the same target. Repeated mentions of words lead to reduced articulations (Arnold, 2008; Bard et al., 2000; Bell et al., 2009; Fowler & Housum, 1987; Lam & Watson, 2010). This effect might counteract or interact with the effect of interest, hyper-articulation of words in the presence of a minimal pair. In addition, avoiding repetition arguably reduces the potential for strategic effects or conscious strategies that are specific to the words presented in the experiment.

Procedure—The experiment was conducted online with Mechanical Turk using software to record speech over the web (Gruenstein, McGraw, & Badr, 2008). A short demonstration of the paradigm is available at https://www.hlp.rochester.edu/mturk/demos/simpartnerdemo/index.html.. Our lab has successfully used this web-based approach for spoken sentence elicitation (Jaeger & Grimshaw, 2013; Weatherholtz, Campbell-Kibler, & Jaeger, 2014) and speech perception (Kleinschmidt & Jaeger, 2012; Kleinschmidt, Raizada, & Jaeger, 2015). The present paper extends it to the study of phonetic production. Participants were instructed that they were taking part in an interactive communication experiment. The experiment began with a microphone test to ensure recording was possible. After that, participants read through a short task description that presented text and images describing their role and their partner's role. Participants were instructed that during each trial their partner would hear the participant but that due to technical limitations they would not hear their partner. After reading the study description and giving informed consent, our software displayed a waiting screen and participants were asked to wait for a partner to register through the same online interface. After a variable delay (approximately two minutes), participants were informed that they had been matched with a partner. In reality, the partner was simulated by our experimental software.

Each trial began with a short "re-sync" screen that illustrated, at variable timing (1100–2200 ms), establishment of a one-way audio connection to the (simulated) partner. Three words were then presented horizontally across the participant's screen along with a horizontal timer bar at the bottom of the screen. Participants were instructed to silently read all three words during this initial display. After a 1500 ms delay the target was highlighted with a

black box surrounding the word and the timer bar began to run down (following Kirov & Wilson, 2012).

Participants were instructed to name the cued target to their partner. To ensure prompt responses, participants were instructed that trials ended when their partner responded or after ten seconds (signified by the timer bar). This timing was chosen based on pilot results. After the pre-determined partner response time elapsed (see below), the trial bar stopped and a three second pause followed. During this pause, the screen for the No Feedback group stayed the same. The Positive Feedback group saw the target word turn green and a small icon appear above the target that indicated their partner's choice. The Mixed Feedback group saw an equivalent display for correct partner responses. However, for incorrect partner responses a red box and the partner icon appeared above a distractor word.

Participants in the Mixed Feedback group saw their partner make seven errors out of the 90 filler and critical trials (no errors occurred during the six practice trials). The first error occurred on the second filler trial (across lists this was the second or third trial following the practice trials). Another four errors occurred on Competitor Present trials within the first 26 trials. The two remaining errors also occurred during Competitor Present trials, approximately at trial 37 and trial 71. These later errors were included to maintain any assumptions about the partner's comprehension that would have developed in response to the first five errors.

The speech for each trial was recorded using the participant's own computer and microphone configuration. Each trial was saved to a server for analysis (Gruenstein et al., 2008). After the experimental trials were completed, each participant was presented with a post-test survey that collected demographic information and assessed participants' response to the study and simulated partner.

The simulated partner—We took a number of steps to make the simulated partner believable. Speakers are exquisitely sensitive to the timing of feedback from interlocutors. For example, unnatural timing can lead participants to not engage in perspective-taking or other communicative behaviors (for review see Kuhlen & Brennan, 2013). For this reason, we calculated realistic partner response times based on data from similar experiments with naive participants. Specifically, partner response times were modeled as the sum of three components. The first was the time it would take a speaker to initiate articulation once the target word had been indicated. We estimated these speech onset times from a picture naming study (Buz & Jaeger, 2015). The materials used by Buz and Jaeger resembled the current stimuli in that they were monosyllabic. Second, we considered the time it would take a listener to recognize and select the target word, measured from speech onset. We estimated these response times based on data from a 4-alternative forced choice picture selection study that— like the simulated participant in the present study—had participants click on the target (Clayards, Tanenhaus, Aslin, & Jacobs, 2008). Like the current experiment, Clayards and colleagues used monosyllabic target and distractor words. Finally, we considered that speech onset latencies and response times tend to become faster throughout experiments, due to practice effects. We estimated this speed-up for both speech onset latencies (against the picture naming experiment) and response times (against the 4-alternative forced choice

picture selection experiment) to obtain trial-by-trial simulated speech onset latencies and response times. These were summed to obtain the overall response time for each trial. Response times of the simulated partner ranged from approximately 3100 ms for the first practice trials and decreased to approximately 2800 ms over the course of the study (measured from the time the target was highlighted on the participant's screen). Initial piloting confirmed that these were response times felt natural (see below also).

To further increase the believability of the partner, participants were allowed to request an extended break of up to five minutes from their partner via a button press during the inter-trial interval. Critically, participants had to wait for partner approval to take these longer breaks. During these breaks, when participants wanted to continue the experiment, they had to alert their partner (again through a button press), who would respond with variable delay to the request to continue. On occasion, the (simulated) partner indicated readiness to resume the experiment prior to the participant requesting to continue the experiment. Our software limited participants to two long breaks (the simulated partner denied additional requests for breaks) to keep the experimental session within an hour's length.

Believability of the (simulated) partner—We assessed the believability of the simulated partner by asking participants a series of increasingly targeted questions about their partner's behavior. We chose this stepwise approach to reduce possible biases in participants' responses. Our assessment was scripted and automatic, further reducing biases that might arise from an interviewer (in face-to-face post-experimental interviews) or from non-systematic unscripted assessments. Questions were presented on separate screens, which prevented participants from returning to previous screens to change their responses. On the first screen, we asked participants to rate the their connection quality on a 1–7 scale (poor to good, mean = 6.5; SE = .1). On the second, we asked participants to rate two aspects of their partner's response time ("How fast would you say your partner responded?" and "How much delay in transmitting the sound between you and your partner would you estimate was there?"). Participants rated their partners as fairly fast responders (1–7 scale, slow to fast; mean = 6.3, SE = .1). When asked to rate the amount of audio transmission delay between them and their partner they rated the delay as low (1–7 scale, no delay to very delayed; mean = 1.6, SE = .1). 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 too slow.

On the third screen, we asked participants to comment on the experiment and on their partner ("Did you notice anything weird during the experiment?"). Two participants noted technical issues.[2] One participant commented that their partner's response times were very "consistent" (although—as outlined above—response times did, in fact, change throughout the experiment). One participant stated their partner occasionally misidentified voiceless for voiced segments. Nine participants explicitly stated that they did not believe they had a partner. Of these nine participants, one was in the Mixed Feedback group, three were in the No Feedback group and five were in the Positive Feedback Group. A χ^2–test of

---

[2]Our software defaulted to sending participants directly to the post experiment survey if any technical issues occurred. Participants were informed this was due to an inability to effectively reconnect them with their original partner or to a new partner.

independence found that Feedback groups did not significantly differ in the number of participants who stated they did not believe they had a partner ($\chi^2(2) = 3.14$, $p = .21$). These nine participants were excluded from further analysis (subsequent analyses indicated that these exclusions did not affect the conclusions drawn below).

In summary, the majority of participants ($N = 51$ out of 60, 85%) did not make any comments indicating that they found anything odd either about the experiment or their (simulated) partner. This suggest that the simulated partner was convincing for most participants. This is encouraging for future web-based studies within the simulated partner paradigm. All participants were fully debriefed about the actual study design at the end of the experiment.

Acoustic annotation and data exclusion—The speech onset latencies, voice onset times (VOT), and word durations of critical targets were manually annotated and measured using Praat (Boersma & Weenink, 2014) by seven research assistants trained by the first author. Annotators were blind to experimental conditions; annotations were checked by the first author. Speech onset latency was the time between the presentation of the target word cue and the onset of the target word. This onset was determined as the point of zero-amplitude on the waveform nearest the plosive release of the target word. 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 (this follows the procedure used by Baese-Berk & Goldrick, 2009). Word duration was measured as the time from word onset until when no visible speech was present in the waveform or spectrogram, which was confirmed acoustically. During annotation recordings were marked for disfluencies, mispronunciations, and noise or other recording issues that prevented annotation. Inter-annotator agreement was assessed by having annotators code one participant and Pearson's r was calculated over VOT values between each pair of annotators and averaged. Mean pairwise r was .9 (range .71–.99).

Aside from the nine participants excluded for not believing their simulated partner, no other participants were excluded. Tokens were excluded for disfluencies, mispronunciations, noise obscured word or vowel onsets, and recording issues (1% of all tokens).

Two participants in the Mixed Feedback group produced a subset of their productions with highly hyper-articulated articulations (e.g. *pill* as [pʰI.Iɬ or [pʰ (..) Iɬ]). This behavior was not observed in the other Feedback groups. These productions starkly differed from all other productions (one of these two speakers essentially produced bisyllabic utterances). Such strong hyper-articulation is consistent with our prediction that perceived miscommunications will affect hyper-articulation but may unduly bias our results. For these reasons, these hyper-articulated tokens were excluded from further analysis (1% of all tokens). Finally, tokens that constituted outliers in terms of their log speech onset latency, log word duration, or VOT (by-participant absolute z-score value > 2.5) were removed from further analysis (6% of all tokens). Subsequent analyses indicated that these exclusions did not affect the conclusions drawn below. This left 1679 data points from 51 participants for the remainder of the analysis.

Predictions of the effect of Feedback on speech—Of primary interest to the present study is the effect of Feedback. If the perceived communicative success of previous productions affects subsequent articulation, we expect Feedback to affect VOT production. Whether this hypothesis predicts a difference between the No and Positive Feedback condition depends on whether participants (on average) assume that their partners should understand them given the task constraints of the paradigm. Here, previous work is informative. Although not intended as a manipulation, two previous studies using the competitor paradigm differed in the type of feedback participants received. Baese-Berk and Goldrick (2009) provided no feedback to participants as to whether they had been understood. Participants only saw *when* their partner responded. Kirov and Wilson (2012) provided positive feedback after each trial. This makes these two studies very similar to our No Feedback and Positive Feedback groups, respectively. The two studies found comparable effects of competitors, with slightly smaller effects in Kirov and Wilson's study, compared to Baese-Berk and Goldrick's. We thus tentatively predict little difference between the No Feedback and Positive Feedback groups.

If the perceived communicative success of previous productions affects subsequent articulation, we do, however, expect the Mixed Feedback groups to differ from the other two groups. Fig. 4 illustrates possible outcomes assuming that our paradigm replicates the Competitor effect found in previous work (Baese-Berk & Goldrick, 2009; Kang & Guion, 2008; Kirov & Wilson, 2012). We can distinguish between two outcomes. One possibility is that Mixed Feedback leads to additional hyper-articulation that is *not* context-specific and in this sense context-general. In this case, we should see two main effects of Competitor and Feedback, and no interaction between Feedback and Competitor (see Fig. 1, middle row). Alternatively, Feedback may lead to context-specific hyper-articulation. This would be reflected in main effects for Competitor and Feedback and an interaction between the two, such that the Competitor effect is larger in the Mixed Feedback group.

Finally, if we do not see an effect of Feedback, but replicate the Competitor effect (Fig. 1, top row), this would suggest that perceived communicative success does not affect speakers' subsequent productions or that speakers were not cooperative.

Main analysis: effects of Competitor and Feedback on speech—All analyses were conducted using mixed effect linear regression with fixed effects for Competitor and Feedback, and the maximal appropriate random effect structure (i.e., random by-participant intercepts and slopes for Competitor as well as random by-item intercepts and slopes for both Competitor and Feedback). Dependent measures and continuous independent measures were centered. Competitor was contrast coded $(+1, -1)$; Feedback was Helmert coded: the first contrast compares differences between the No and Positive Feedback groups $(-1, +1, 0)$; the second contrast compares differences between the Mixed Feedback group and the joint mean of the No and Positive Feedback groups $(-1, -1, +2)$. Significance and p-values are based on the χ^2-test of the change in deviance between the models with and without the predictor of interest.

Voiced Competitor and Feedback effects on VOT hyper-articulation—VOT means and standard deviations for each condition are summarized in Table 1. Fig. 2

visualizes VOTs across the Competitor and Feedback manipulations. Table 2 summarizes the mixed regression analysis of VOTs.

VOTs of target words were significantly longer in Competitor Present trials, compared to Competitor Absent trials ($\hat{\beta} = 4.1$; $t = 5.2$; $p < .01$). This effect was significantly larger in the Mixed Feedback group than in the other two groups ($\hat{\beta} = 1.1$; $t = 2.1$; $p < .05$). The effect of Competitor did not significantly differ between the Positive and No Feedback groups ($\hat{\beta} = .1$; $t = .1$; $p > .5$).

A simple effects analysis revealed that VOTs significantly differed between Competitor Present and Absent trials in all three Feedback groups ($ps < .05$) and that VOTs in Competitor Absent trials did not differ across Feedback groups (p's > .2).

Interpretation: We find that speakers hyper-articulate VOTs when a voiced competitor is present, replicating Baese-Berk and Goldrick (2009) and Kirov and Wilson (2012). The overall difference in VOT between trials with and without voiced competitors is 8.1 ms. This is similar to previous work that has investigated VOT hyper-articulation (5 ms Baese-Berk & Goldrick, 2009; 2 ms Kirov & Wilson, 2012; 9 ms Schertz, 2013). Critically, we find that speakers increase the degree to which they hyper-articulate VOT based the feedback they receive from their partners: Speakers in the Mixed Feedback group hyper-articulated VOTs at least twice as much as speakers in No Feedback and Positive Feedback groups (difference in VOTs between Voiced Competitor Present vs. Absent conditions, No Feedback: 4.9 ms; Positive Feedback: 6.4 ms; Mixed Feedback: 12.8 ms; see Table 1).

The simple effect analyses further suggest that speakers across all three Feedback groups had similar VOTs in the Competitor Absent trials. They only differed in how they produced VOTs in Competitor Present trials. This suggests that speakers in the Mixed Feedback group adjusted their VOT productions for precisely those critical trials that seemed to cause the most problems for their partner (recall that six out of seven partner errors in the Mixed Feedback condition occurred on critical trials with a voiced competitor). This makes the hyper-articulation observed here *context-specific*.

Before we discuss the broader implications of the present study, we present additional analyses of the *distribution* of VOTs that speakers produced in the different conditions. The aim of these analyses is to shed light on the function or goal of the observed hyper-articulations.

Why hyper-articulate VOT by a few milliseconds?—Our analyses, up to this point, have followed previous work by assessing mean differences in VOTs across our experimental manipulations of Competitor and Feedback. This approach reduces the dimensionality of the data, making it simpler to visualize and analyze. However, as we demonstrate here, it also risks missing important properties of the data. Identical differences in means can be caused by qualitatively rather different changes in the distributions. Hypotheses about the *cause* of hyper-articulation in the present study that are not distinguishable in terms of effects on mean VOTs might be distinguishable if we analyze effects on the overall distribution of VOTs.

Fig. 3 illustrates three possible scenarios that lead to the same effect of the Competitor on mean VOT, but differ in the effect of the Competitor on the distribution of VOTs. To facilitate comparison to the analyses we present below, the VOT distributions for the Competitor Absent condition (identical across all three panels of Fig. 3) are estimated under the assumption of normality from the mean and variance of the actual VOTs in the Competitor Absent condition. All three panels show an 8 ms increase in VOT from the Competitor Absent to the Competitor Present conditions, approximating the mean Competitor effect observed across all three Feedback groups. The three panels differ, however, in how this 8 ms difference comes about. In the left panel, only the mean is affected (this is essentially the assumption that an ANOVA analysis makes). In the middle panel, both the mean and the variance are affected. Finally, in the right panel, the Competitor Present condition also differs in its skewedness, making it non-normal.

In the left panel, all quantiles of the two distributions differ by 8 ms (e.g. the 5th quantile, mode, mean, and 95th quantile values all differ by 8 ms). In the middle panel the 5th quantile values differ by less than 8 ms, the mode and mean both differ by 8 ms, and the 95th quantile values differs by more than 8 ms. In the right panel, the 5th and 95th quantile values differ by more than 8 ms, the mean differs by 8 ms, and the mode differs by less than 8 ms.

Critically, determining how the distributions differ can shed light on *why* speakers hyper-articulate in the current paradigm. For example, an 8 ms increase in VOT, from 90 ms to 98 ms (cf. Table 1), is unlikely to serve a communicative purpose. While there is some inter-talker variance in the category boundary between voiced and voiceless plosives, the boundary typically lies around VOTs of 20 ms (for /p/) to 33 ms (for /k/) (estimated from speech analyses of reading data in Chodroff, Godfrey, Khudanpur, & Wilson, 2015). Hyper-articulating a sound that is far away from the category boundary by another 8 ms would therefore be unlikely to make it less confusable to the listener.[3]

We argue, however, that thinking about differences in mean VOTs in this way is likely to be misleading. Consider what a hypothetical rational or boundedly rational speaker with the goal to be understood should aim to do when producing a voiceless plosive in the context of a voiced competitor. The speaker wants to increase the probability that her productions are understood, while perhaps taking into account the articulatory effort that would be required. To a first approximation then, the speaker would try to avoid producing perceptually confusable VOTs, specifically, VOTs that are close to the category boundary. Critically though, articulation is implemented through a noisy motor system. Thus speakers do not have deterministic control over the actual VOT they produce. A rational articulatory planning system should take this into account (as has been proposed for motor planning, e.g., Harris & Wolpert, 1998; Trommershäuser, Maloney, & Landy, 2008; Todorov & Jordan, 2002; Wei & Körding, 2008; for articulation, see Houde & Nagarajan, 2011). The best a speaker who is willing to invest more effort into producing a specific VOT target can do is reduce the *probability* of producing perceptually confusable VOTs (cf. Jaeger & Ferreira, 2013).

---

[3]We thank an anonymous reviewer and Matt Goldrick for bringing this reasoning to our attention.

This can be achieved in at least two ways. If there are no other constraints on the target of articulation, the speaker could simply increase the target VOTs (i.e., the *mode* of the VOT distribution she is aiming for), shifting the VOT distribution for voiceless plosives to the 'right'. This corresponds to the left panel of Fig. 3. If, on the other hand, speakers prefer *not* to deviate from the target VOT or to limit deviation from the target VOT (e.g., because long VOTs require more effort or because VOTs that are prototypical for the intended category are preferred) then the most effective strategy for a rational speaker is to reduce the variance 'left' of the mode. This latter scenario corresponds to the right panel of Fig. 3. Both of these options would reduce the probability of perceptually confusable VOT values. For example, VOT differences of 8 ms at or near the category boundary strongly affect comprehension (cf. McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2002).

In short, whether or not an 8 ms increase in mean VOT could plausibly serve a communicative purpose depends on the underlying change in the distribution of VOTs. This motivates the distributional analyses we present next. We note that analyzing distributional changes in speech was not an initial focus of the current study. Thus, all the analyses presented below are post hoc. Nonetheless we believe they provide valuable insights that can advance our understanding of adaptation and guide future work.

Analyzing Voiced Competitor and Feedback effects on the distribution of VOTs—To understand how the distribution of speech changed in the current study we analyzed differences in the 5th and 95th quantile, mean, and mode values of VOTs across our manipulations of Competitor and Feedback. As there is no parametric test for differences in the first three of these values we conducted a non-parametric bootstrap. We split all productions into six bins based on our 2×3 experimental design and resampled each bin with replacement 2000 times. This method ignores the clustering of our data due to repeated measures from participants and items. This analysis is computationally simpler than resampling over clustered data. Importantly, the estimates obtained from the bootstrap analysis closely resemble the estimates from the linear mixed regression analysis reported above which does take into account the repeated measures structure of our data suggesting no undue bias in this un-clustered approach. For each sample we calculated the 5th and 95th quantile, the mean, and mode (the highest probability density in a Gaussian kernel density estimator). We report the median value of each statistic across all 2000 samples.

Fig. 4 visualizes the estimated density distributions of VOTs across the Competitor and Feedback conditions. Table 3 summarizes the estimated statistics for the Competitor effect for each Feedback group. Significance levels reflect the proportion of samples for which the difference across conditions was less than or equal to zero (e.g., in Table 3, $p < .05$ indicates that the statistic for Competitor Present trials was equal to, or less than, the statistic for Competitor Absent trials on fewer than 5% of the samples).

Mean VOTs were larger in Competitor Present trials than in Competitor Absent trials (p's < . 03), replicating the main effect of Competitor reported above. Similar or larger effects were also observed for the 5th quantiles in the Positive and Mixed Feedback groups (p's < .04), but not in the No Feedback group ($p = .56$). The 95th quantiles were numerically larger in

Competitor Present, compared to Competitor Absent, trials across all Feedback groups, but this difference was only significant in the Mixed Feedback group ($p < .01$). Interestingly, mode VOTs did *not* differ between Competitor Present and Competitor Absent trials (p's > .25; cf. the peaks of the distributions in Fig. 4).

Table 4 summarizes the statistics for the *difference* in the Competitor effect across Feedback groups (i.e., the interaction between Competitor and Feedback). There was a significant interaction between Competitor and Feedback such that the difference in mean VOT between Competitor Present and Competitor Absent trials was greater for speakers in the Mixed Feedback group than in the other two groups ($p = .01$), replicating the main analysis. The difference in 5th quantile differences was marginally larger for the Positive Feedback group compared to the No Feedback group ($p = .07$) and significantly larger for the Mixed feedback group as compared to the other two groups ($p = .04$). The difference in 95th quantile differences was marginally larger in the Mixed Feedback group compared to the other two groups ($p = .08$). Unsurprisingly given the uniform lack of an effect of Competitor on the mode of VOTs, there was no interaction between Competitor and Feedback on VOT modes (p's > .5).

Interpretation—The effects we observe for mean VOTs replicate our main analysis: VOTs were on average longer in the presence of a voiced competitor and this difference was larger for participants in the Mixed Feedback group, compared to the No and Positive Feedback groups. Critically, we further find evidence that speakers in the Positive and Mixed Feedback groups successfully reduced their probability of producing perceptually ambiguous or confusable VOTs.

This was particularly clear for speakers in the Mixed Feedback group, who successfully avoided producing short VOTs of 25–30 ms, reducing the probability of VOTs close to the category boundary (see Fig. 4). Interestingly, participants achieved this while holding the mode—and thus the likely target of their productions—almost constant.

General discussion

We find that speakers hyper-articulated VOTs of voiceless onset plosives in the presence of a voiced competitor. hyper-articulation was greater for participants who occasionally experienced misunderstanding, where their (simulated) partner clicked on the wrong word. The effect of interlocutor feedback on mean VOTs was driven by changes to the overall distribution of VOTs; participants in the Mixed Feedback group produced fewer VOTs that were close to the /p/-/b/ category boundary, thereby likely reducing the average perceptual confusability of their productions.

We first discuss our finding in relation to previous results, beginning with research on the role of interlocutor feedback on articulation. We then discuss the full set of results in the broader context of research on causes of pronunciation variation. This leads us to ask how effects of interlocutor feedback can be integrated into models of speech production. We discuss two possibilities—one in terms of adaptive or learning processes that sub-serve the goal of robust communication, the *adaptive speaker* framework, and one in terms of

competition during lexical planning. We close by identifying central questions the adaptive speaker framework raises for future research.

Previous work on interlocutor feedback

Previous research on interlocutor feedback has shown that immediately following requests for clarifications speakers make targeted changes to their subsequent articulations (Maniwa et al., 2009; Ohala, 1994; Oviatt, Levow, et al., 1998; Oviatt, MacEachern, et al., 1998; Schertz, 2013; Stent et al., 2008). Our results show that interlocutor feedback can induce subtle changes, specifically millisecond changes in VOT can be hyper-articulated (see also Kirov & Wilson, 2012; Schertz, 2013).[4] Taken together this body of work shows that hyper-articulation can be "a targeted and flexible adaptation rather than a generalized and stable mode of speaking" (Stent et al., 2008, p. 163).

Previous studies tested whether speakers can make targeted modifications to their speech signal when they are aware that (a) there has been a misrecognition such that the previous utterance had not been successful and (b) receive clear information about the likely source of the problem. Therefore, it could be argued that targeted hyper-articulation is restricted to situations where the speaker adopts a conscious strategy to explicitly correct misunderstood words.

Several aspects of our design were motivated by the goal to study interlocutor feedback under more implicit conditions. First, we did not elicit corrections or clarifications. Each trial ended with the partner selecting a word and participants had no opportunity to correct themselves. Second, because participants believed they were interacting with a human partner, they would not have used registers associated with computer-directed speech. Third, misrecognitions were rare: only seven of 90 trials in the Mixed Feedback condition ended in a misrecognition. Finally, the type of trials that most frequently led to misrecognitions (Competitor Present trials) occurred on only 18 of 90 trials, with on average 3.5 trials between Competitor Present trials.

Our exclusion criteria further contributed to our goal of testing hyper-articulation under more implicit conditions. We only included data from participants who seemed to believe that they were communicating with a human partner; none of the participants included in our analysis reported anything unusual about the partner's behavior. We further excluded all utterances that our annotators considered to be clearly due to conscious hyper-articulation ("P-H-ill"). This does not rule out the possibility that some participants became aware that there were trials with voiced competitors (or, in the Mixed Feedback condition, that these

---

[4]It is theoretically possible that the changes to VOT we observed are not targeted, but rather the side-effect of non-targeted hyper-articulation of entire words, thereby proportionally lengthening VOT among other phonetic variables. Space limitations prevent us from discussing this possibility in detail. We note, however, that previous studies have found the Competitor effect to be targeted (Kirov & Wilson, 2012). Additional analyses of the current data confirmed that the Competitor effect was clearly targeted to VOTs (though there also was some non-targeted hyper-articulation; see also Buz et al., 2014). In the Mixed Feedback condition, one participant exhibited clearly non-targeted hyper-articulation. When we excluded this participant, the enhanced VOT hyper-articulation in the Mixed Feedback condition compared to the No and Positive Feedback condition was marginal (.1 > ps > .05, depending on the specific statistical test). However, this exclusion reduced the power of the between-participant Competitor-by-Feedback interaction. More research will be needed to address this issue.

trials were more likely to cause misrecognition). It does, however, suggest that adaptive hyper-articulation is not limited to clarifications or corrections.

The present study also adds to previous work in that we observed hyper-articulation in the absence of immediate repetition of the same word. This suggests that interlocutor feedback can lead to hyper-articulation that is not word-specific, though possibly specific to a phonetic feature across words. This suggests that our effects of feedback are unlikely to be due to short-term effects, such as phonological priming. Moreover, our results suggest the longevity of feedback-induced hyper-articulation may depend on specific task constraints.

Finally, we found that feedback only affected the degree of hyper-articulation in the presence of a voiced competitor, rather than lengthening VOTs in both Competitor conditions. Recall that six of the seven times when participants in the Mixed Feedback condition saw their partner misunderstand them, the partner selected the voiced competitor (the remaining trial was a filler trial without a voiceless target or voiced competitor). Participants thus seem to hyper-articulate VOTs only in the context in which they previously perceived their interlocutor to misunderstand them. This context is not defined by specific words, but rather by pairings of specific types of words. It is possible, then, that these feedback effects are context-specific, although future work is necessary to confirm this. In summary, our results contribute to, and expand upon, previous findings that feedback from interlocutors can be an important factor in articulatory planning.

Although our study demonstrates hyper-articulation under more implicit conditions than previous studies, one might still argue that the results arise from a conscious strategy. Therefore, it is important to consider what it would mean for a conscious strategy to have a targeted effect on articulatory decisions that result in millisecond changes to the distribution of VOTs that are likely to reduce perceptual confusions for the listener.

Theorists agree that considerations of communicative success affect speaker's choices. Moreover, these considerations may differ in the extent to which the speaker is consciously aware of implementing them. The debate is about the degree to which these considerations affect linguistic and articulatory planning. Some find it useful to distinguish between automatic and strategic processes (for some examples in the literature see Horton & Keysar, 1996; Dell & Brown, 1991; Arnold, 2008; Bard et al., 2000; Pickering & Garrod, 2004; Shintel & Keysar, 2009). Automatic processes underlie normal production processes, whereas strategic processes represent special-purpose adjustments. Increasing the volume of one's utterances and choosing to speak more slowly and clearly (non-targeted, across-the-board hyper-articulation) could arguably be attributed to conscious decisions that have global effects on production. If, however, choices, even conscious ones, have targeted effects on low level processes and especially if those choices are adaptive, and generalize, then in our view, the automatic versus conscious distinction becomes much less useful. More research investigating targeted effects in production will be important for clarifying these issues.

Previous work on pronunciation variation

Previous research on pronunciation variation argues that speakers adapt their pronunciations as a function of alignment with interlocutors (e.g., Babel, 2010; Giles et al., 1991; Hay, 2000; Sanchez, Hay, & Nilson, 2015; possibly partly with social goals; for discussion, see Campbell-Kibler, 2010; Foulkes & Hay, 2015; Weatherholtz et al., 2014) or due to, for example, competition processes during lexical planning (Arnold et al., 2012; Baese-Berk & Goldrick, 2009; Kahn & Arnold, 2012; Watson, Buxó-Lugo, & Simmons, 2015; for discussion, see Buz & Jaeger, 2015; Jaeger & Buz, in press). The present results and the studies discussed in the previous section show that feedback from interlocutors also influences pronunciation variation. How might existing models of speech production accommodate these feedback effects? Alignment processes cannot be the source of the effects because participants in our study did not receive spoken input from their partner (see also, e.g., Maniwa et al., 2009; Ohala, 1994; Oviatt, Levow, et al., 1998; Schertz, 2013; Stent et al., 2008). The absence of spoken input also rules out an account based on a perception-production loop between interlocutors (as proposed in Pierrehumbert, 2001; Wedel, 2006). The idea behind the perception-production loop is that speakers productions are influenced by previously experienced exemplars stored in memory, which are assumed to contain rich sub-phonemic detail. Since only exemplars from successfully recognized words are assumed to be stored in memory, comprehension serves as a filter: on average, only exemplars that were sufficiently recognizable in the context in which they occurred will be reproduced through the perception-production loop between interlocutors (see also Guy, 1996; Ohala, 1989). Because spoken input to participants is required for the perception-production loop to explain hyper-articulation, it cannot explain our results. Thus, while the perception-production loop might play an important role in understanding how phonetic representations change over time, effects of interlocutor feedback suggest that additional mechanisms contribute to pronunciation variation. We now consider two mutually compatible candidates for such a mechanism. The first appeals to adaptive or learning processes, the second to processes inherent to the demands of lexical planning.

Why does interlocutor feedback affect subsequent productions?

The adaptive speaker—In the framework that motivated the present work, speakers monitor the effects of their productions on their interlocutors. They adapt their utterances based on interlocutor feedback with the goal of facilitating successful communication. Speakers hyper-articulate in trials with voiced competitors because they perceived their previous productions in similar contexts to be unsuccessful in that they led to misrecognitions.

This accounts for both the effects of Competitor and Feedback: speakers hyper-articulate utterances in contexts that are *a priori* more confusable and hyper-articulate more when they have additional evidence that they might be misunderstood. It also accounts for the results from our distributional analyses: once speakers have evidence that their previous productions were not sufficiently clear in the context of a voiced competitor, they invest additional effort to reduce the variance around the intended (perceptual or VOT) target; specifically, speakers should primarily aim to reduce variance along the task-relevant dimension (cf. similar proposals in models of motor adaptation Trommershäuser, Gepshtein, Maloney, Landy, &

Banks, 2005; Wei & Körding, 2008; for the role of goals in language processing and production, see Brown-Schmidt, Yoon, & Ryskin, 2015; Kuperberg & Jaeger, 2016). In the present case, an adaptive speaker should aim to reduce the probability of producing short, perceptually confusable VOTs. Therefore, the increase in very long VOTs is either an inevitable consequence of this goal (rather than being intended itself) or arises because speakers are sometimes willing to invest additional effort to produce hyper-articulated pronunciations.

How might feedback from interlocutors come to affect subsequent production? Previous research shows that speakers can dynamically adjust their pronunciations based on perception of their own speech. Perhaps the strongest evidence comes from perturbation studies (Houde & Jordan, 1998; Purcell & Munhall, 2006; Villacorta, Perkell, & Guenther, 2007). In these studies, participants produce words while wearing sound insulating head phones. Speakers productions are manipulated in real-time and played back through the head phones. For example, speakers might aim to produce the word *pen* but hear themselves say *pan* (e.g., Houde & Jordan, 1998; Purcell & Munhall, 2006). Speakers subsequently adapt their articulations, counter-acting the unintended shift they perceived in their own speech. Adaptations like these can be attributed to internal or external self-monitoring (see also Huettig & Hartsuiker, 2010; Jacobs, Yiu, Watson, & Dell, 2015) or—perhaps equivalently (cf. Pickering & Garrod, 2013, p. 9)—to changes to forward models underlying articulation (as proposed in Jaeger & Ferreira, 2013; for specific models, see Hickok, 2012; Houde & Nagarajan, 2011; Tourville & Guenther, 2011) or selection between different forward models (similar to recent proposals for speech perception, Kleinschmidt & Jaeger, 2015).

Effects of interlocutor feedback suggest that speaker's monitoring goes beyond encapsulated *self*-monitoring. Speakers also seem to integrate information from their interlocutors about the quality of their utterances. The mechanisms that implement this feedback into an adaptive change to subsequent behavior may, nonetheless, be the same as those that underlie adaptive responses in perturbation studies (cf. Jaeger & Ferreira, 2013). This would provide a straightforward way to integrate interlocutor feedback into existing models of speech production. Similar reasoning could be extended to possible effects of more implicit information about the communicative success of previous productions (such as hesitation or visible confusion of interlocutors, which speakers have been argued to be sensitive to; for review, see Jokinen, 2009). More generally, an adaptive speaker framework can accommodate the importance of communicative goals in understanding pronunciation variation (see also Brown-Schmidt & Tanenhaus, 2008; Fox Tree & Clark, 1997 Galati & Brennan, 2010; Jaeger, 2013; Kohler, 1990; Lindblom, 1990, among others; for review, see Jaeger & Buz, in press). Here, we have focused on adaptive or learning processes across productions, leaving open whether or not speakers continuously simulate the knowledge state of their interlocutors (see Brown-Schmidt et al., 2015; Horton & Gerrig, in press). Next, we discuss an alternative explanation of interlocutor feedback effects in terms of the dynamics of lexical planning.

Increased lexical competition due to visual feedback—Several recent papers argue that some, if not all, apparent effects of communicative demands on articulation may be attributable to processes inherent to lexical planning (e.g., Arnold & Watson, 2015;

Arnold et al., 2012; Bell et al., 2009; Kahn & Arnold, 2012; for review see, Jaeger & Buz, in press). Of specific relevance here, are *competition* accounts of pronunciation variation (Baese-Berk & Goldrick, 2009; Goldrick et al., 2013; Kirov & Wilson, 2013). According to these accounts, contextual co-presence of, for example, a voicing competitor increases competition during lexical planning. This assumption is supported by evidence that speakers avoid sequences of phonologically (onset) overlapping words (Jaeger, Furth, & Hilliard, 2012b). Moreover, producing these sequences is associated with increased rates of speech errors (O'Seaghdha & Marin, 2000; Sevald & Dell, 1994), longer production latencies, and slowed down speech rates (for review, see Jaeger, Furth, & Hilliard, 2012a). Increased competition is assumed to result in hyper-articulation, offering a potential explanation for Competitor effects, such as the ones observed here and in similar previous studies (Baese-Berk & Goldrick, 2009; Goldrick et al., 2013; Kirov & Wilson, 2013; see also Fox, Reilly, & Blumstein, 2015; Fricke, Baese-Berk, & Goldrick, 2016; Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011).

Could competition also account for the effect of interlocutor feedback, and specifically, (a) the increased hyper-articulation in the Mixed Feedback condition as well as (b) the distributional nature of this hyper-articulation? With regard to (a), we need to explain that Mixed Feedback only increased hyper-articulation in the presence of a voiced competitor, rather than lengthening of VOTs in all Competitor conditions. Recall that misrecognition was signaled by highlighting the wrong answer with a red box. This could have increased the amount of attention that participants gave to non-targets. For example, participants were instructed to read all three words during the preview phase of each trial. But (1) not all participants might have initially done so; and (2) highlighting a non-target word led those participants to read all three words on subsequent trials. Participants would experience increased competition between the target and the voiced competitor on those trials causing them to hyper-articulate. Because non-target words were only highlighted in the Mixed Feedback group this would both explain why increased hyper-articulation only occurred in that group (e.g., as opposed to the Positive Feedback group) and why it only affected the Competitor Present condition. Without further assumptions, an account along these lines predicts the same pattern of results if misrecognition in the Mixed Feedback condition always occurred during filler trials (which did not contain voicing contrasts). This prediction needs to be tested in future work.

Some of our results are, however, problematic for the competition account. First, it is unclear how competition could explain the *distribution* of VOT productions (point b above). Speakers did not change their target VOT; neither did they uniformly shift their VOT distribution towards longer VOTs. Rather, speakers seemed to avoid short VOTs. The adaptive speaker framework provides a straightforward explanation for this pattern. It is unclear how it is predicted by competition during lexical planning. More generally, competition based accounts have been criticized for not specifying *how* competition affects the coordination of articulatory gestures (for discussion, see Buz & Jaeger, 2015; Goldrick & Chu, 2014; Goldrick et al., 2013; Pouplier & Goldstein, 2010, 2014; Watson et al., 2015). Clarifying the link between lexical planning and articulation is a critical issue for future work on competition accounts. For example, does competition always lead to *lengthening*? This would seem to be incompatible with findings that contextual co-presence of a minimal

pair competitor can lead to *shortening* of context-relevant phonetic features to increase the perceptual contrast to a competitor (Seyfarth et al., in press).

Aspects of our data in the Mixed Feedback condition are also inconsistent with competition. For example, increased competition typically increases the rate of speech errors (Goldrick, Folk, & Rapp, 2010; O'Seaghdha & Marin, 2000; Sevald & Dell, 1994). However, we did not find an increase in speech errors in the Mixed Feedback condition. Participants in the present study made a total of four speech errors—none of them in the Mixed Feedback condition. Neither did we find any evidence for increased competition on planning times (contrary to the explicit predictions of some competition accounts Kirov & Wilson, 2013; for more discussion see Buz & Jaeger, 2015). Speech onset latencies did not correlate with VOTs ($\hat{\beta} = .4$; $t = .1$; $p > .1$; for summary mean and standard deviations see Table 5). It is possible, though, that our speech error and latency measures might be too coarse-grained to find evidence for competition. Moreover, speakers had relatively long preview of target words (1500 ms). This might have weakened the degree to which planning difficulty is reflected in our latency measure (see also Munson, 2007). In sum, the competition account can account for some aspects of our results. However, unlike the adaptive speaker account, it faces challenges in accounting for the specific effects of interlocutor feedback on the distribution of VOTs.

Implications of the adaptive speaker framework

We proposed that the adaptive speaker framework accounts for the range of effects we reported, including the effect of Feedback on the distribution of VOTs. We favor this approach for two reasons. First, as mentioned earlier, it is naturally realized by classes of models that have proved successful and informative in other domains of motor planning. Second, the framework raises a number of interesting questions, and promising directions for future research. We focus next on two of these questions.

When should hyper-articulation be targeted?—From our perspective, questions about causal attribution are among the most challenging and exciting questions raised by an adaptive speaker approach embedded within a communicative (cooperative speaker) framework. Put most generally, the speaker, when faced with evidence that the listener has misunderstood her intended message, will have uncertainty about the source of the miscommunication. Within an adaptive speaker framework, the inferences that the speaker draws should affect when and how the speaker should adapt her utterances. We first focus on the specific case of targeted hyper-articulation of VOTs. We then briefly discuss the more general issue and connect it with related work on adaptation in other domains of language processing.

When a speaker in our study says *pill* to their partner and then sees the partner click on *bill* instead, she must determine the appropriate response. In order to do so, a cooperative speaker would need to determine the *cause* of the observed miscommunication. This process is subject to considerable uncertainty. If a speaker wishes to make a targeted correction (i.e. hyper-articulating VOT), then detecting (perhaps implicitly) that the partner's response differs from the intended target in only one phonetic feature (the voicing of the onset

plosive) is a necessary condition. However, it is not a sufficient condition. Even a cooperative speaker who has successfully made this inference will still have uncertainty about the cause of the miscommunication. The miscommunication might have resulted from an attentional lapse. It might also have resulted from a technical error, such as a failure in transmitting the sound file from the speaker to the partner or high levels of noise that make it difficult to recognize the word with any degree of certainty. In each of these cases, the partner's response would be random and should not be interpreted as informative about what the partner perceived. Hyper-articulation of later productions is not warranted under these circumstances. In effect, the speaker would be making an adjustment that might be effortful and yet not reduce the likelihood of subsequent miscommunication.

Miscommunication might, however, have resulted from environmental noise (including noise in the web-based transmission of the sound files, which the speaker cannot directly observe) that specifically affected the perception of the voiceless onset plosive. If noise is assumed to persist over subsequent trials, this would make targeted hyper-articulation a rational response. Miscommunication might also have resulted because the partner expected the speaker to use voicing cues differently (e.g., because the partner has a different dialect or language background, cf. work on the perception of English voicing by native Korean listeners, Schertz, Cho, Lotto, & Warner, 2015). In some such cases, targeted hyper-articulation might help, as long as it exaggerates phonetic dimensions that are relevant to the partner's voicing classification.

In sum, whether hyper-articulation is the most cooperative response to a *pill-bill* confusion depends on what is likely to have caused the miscommunication. Speakers have uncertainty about these causes, so that even a perfectly cooperative rational speaker would not be expected to always produce targeted hyper-articulation or hyper-articulate at all (for causal uncertainty in speech perception and evidence that listeners take it into account, see Kraljic, Samuel, & Brennan, 2008). Future research can address this question by manipulating the likelihood of different causes for misrecognition.

From adaptation to speech register?—A related question focuses on the relation between adaptive changes to production in a specific environment (such as the present experiment) and speech registers, such as clear speech, infant-, and foreigner-directed speech (e.g., Kuhl et al., 1997; Picheny et al., 1986; Uther et al., 2007).

It is conceivable that at least some of these speech registers partially originate in adaptive behavior like that observed in the current experiment. For example, while it is possible that adaptive changes to production are short-lived, it is also possible that speakers *remember* these changes. This would allow speakers to reuse previously adapted behaviors in subsequently encountered identical or similar situations. This would resemble similar proposals made for speech perception (Kleinschmidt & Jaeger, 2015; Weatherholtz, Seifeldin, Kleinschmidt, Kurumada, & Jaeger, submitted for publication). More generally, research on language comprehension suggests that comprehenders store rich-context specific representations in memory (e.g., Brown-Schmidt et al., 2015; Goldinger, 1998; Horton & Gerrig, 2005; Johnson, 1997; Pickering & Garrod, 2013; Pierrehumbert, 2001; Wedel, 2006). For example, a growing body of evidence demonstrates that phonetic, lexical,

syntactic and even higher level aspects of language usage depend not just on what is spoken but who spoke it and where (e.g., Arnold, Kam, & Tanenhaus, 2007; Hanulíková, Alphen, Goch, & Weber, 2012; Kleinschmidt & Jaeger, 2015; Kurumada, 2013; Niedzielski, 1999; Staum Casasanto, 2008; Strand, 1999; Walker & Hay, 2011; for positions and reviews, see Brown-Schmidt et al., 2015; Foulkes & Hay, 2015; Horton & Gerrig, in press; Weatherholtz & Jaeger, in press). Speakers might draw on this or similar implicit knowledge, which could be implemented in terms of situation-specific forward models (Jaeger & Ferreira, 2013). The same type of model might underlie speech registers as the ones discussed above (see also Dell & Brown, 1991).

Adaptation in response to communicative goals might be one factor that contributes to the development of these register, but it is unlikely to be the only one. For example, the function of specific characteristics of speech registers, such an infant- or foreigner-directed speech, is still under debate (for a recent review for infant-directed speech, see Martin et al., 2015).

Conclusions

To the best of our knowledge, the present study is the first to investigate fine-grained durational measures of speech in recordings elicited over the web. Our findings are consistent, both qualitatively and quantitatively, with previous lab-based studies suggesting that the current approach is feasible and reliable (at least for durational/temporal based measures; for similar results from a web-based study on word-final sibilants, see also Seyfarth et al., in press). In order to study how feedback from interlocutors dynamically affects speakers' subsequent productions, we combined this novel recording method with a simulated partner paradigm with naturalistic partner response timing. Participants were successfully led to believe that they were talking to an interlocutor over the web.

Replicating previous work, we found that speakers hyper-articulate their pronunciations in context where they would *a priori* be more perceptually confusable. Critically, the extent to which hyper-articulation occurred seems to be a function of the perceived communicative success of previous productions. This suggests that speakers monitor the effect of their productions on their interlocutors and, if necessary, can adapt their speech based on their perceived communicative success of similar previously produced words. We further found that speakers hyper-articulated in such a way that they avoided perceptually confusable pronunciations while keeping the articulatory target constant. We propose that such adaptive processes play an important role in understanding how speakers dynamically manage the planning demands associated with linguistic encoding, balancing them against their communicative goals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix A. Experimental stimuli

Table A.6.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2015.12.009.

References

Arnold JE. Reference production: Production-internal and addressee-oriented processes. Language and Cognitive Processes. 2008; 23:495–527. http://dx.doi.org/10.1080/01690960801920099.

Arnold JE, Kahn JM, Pancani GC. Audience design affects acoustic reduction via production facilitation. Psychonomic Bulletin & Review. 2012; 19:505–512. http://dx.doi.org/10.3758/s13423-012-0233-y. [PubMed: 22419403]

Arnold JE, Kam CLH, Tanenhaus MK. If you say thee uh you are describing something hard: The online attribution of disfluency during reference comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33:914–930. http://dx.doi.org/10.1037/0278-7393.33.5.914.

Arnold JE, Watson DG. Synthesising meaning and processing approaches to prosody: Performance matters. Language, Cognition and Neuroscience. 2015; 30:88–102. http://dx.doi.org/10.1080/01690965.2013.840733.

Babel M. Dialect divergence and convergence in New Zealand English. Language in Society. 2010; 39:437–456. http://dx.doi.org/10.1017/S0047404510000400.

Baese-Berk MM, Goldrick M. Mechanisms of interaction in speech production. Language and Cognitive Processes. 2009; 24:527–554. http://dx.doi.org/10.1080/01690960802299378. [PubMed: 19946622]

Bard EG, Anderson AH, Sotillo C, Aylett MP, Doherty-Sneddon G, Newlands A. Controlling the intelligibility of referring expressions in dialogue. Journal of Memory and Language. 2000; 42:1–22. http://dx.doi.org/10.1006/jmla.1999.2667.

Bell A. Language style as audience design. Language in Society. 1984; 13:145–204. http://dx.doi.org/10.1017/S004740450001037X.

Bell A, Brenier JM, Gregory M, Girand C, Jurafsky DS. Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language. 2009; 60:92–111. http://dx.doi.org/10.1016/j.jml.2008.06.003.

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer, version 5.4.08. 2014. Retrieved from http://www.praat.org/

Brown PM, Dell GS. Adapting production to comprehension: The explicit mention of instruments. Cognitive Psychology. 1987; 19:441–472. http://dx.doi.org/10.1016/0010-0285(87)90015-6.

Brown-Schmidt S, Tanenhaus MK. Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. Cognitive Science. 2008; 32:643–684. http://dx.doi.org/10.1080/03640210802066816. [PubMed: 19890480]

Brown-Schmidt, S.; Yoon, SO.; Ryskin, RA. People as contexts in conversation. In: Ross, B., editor. The psychology of learning and motivation. Academic Press; 2015. p. 59-99.http://dx.doi.org/10.1016/bs.plm.2014.09.003

Burnham D, Kitamura C, Vollmer-Conna U. What's new, pussycat? On talking to babies and animals. Science. 2002; 296:1435. http://dx.doi.org/10.1126/science.1069587. [PubMed: 12029126]

Buz E, Jaeger TF. The (in)dependence of articulation and lexical planning during isolated word production. Language, Cognition and Neuroscience. (Advance online publication). 2015 http://dx.doi.org/10.1080/23273798.2015.1105984.

Buz, E.; Jaeger, TF.; Tanenhaus, MK. Contextual confusability leads to targeted hyper-articulation. In: Bello, P.; Guarini, M.; McShane, M.; Scassellati, B., editors. Proceedings of the 36th annual meeting of the cognitive science society. Quebec City, Canada: Cognitive Science Society; 2014. p. 1970-1975.

Campbell-Kibler K. The sociolinguistic variant as a carrier of social meaning. Language Variation and Change. 2010; 22:423–441. http://dx.doi.org/10.1017/S0954394510000177.

Chodroff, E.; Godfrey, J.; Khudanpur, S.; Wilson, C. The Scottish Consortium for ICPhS 2015. Glasgow, UK: The University of Glasgow; 2015. Structured variability in acoustic realization: A corpus study of voice onset time in American English stops. Proceedings of the 18th international congress of phonetic sciences

Clayards M, Tanenhaus MK, Aslin RN, Jacobs RA. Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 2008; 108:804–809. http://dx.doi.org/10.1016/j.cognition.2008.04.004. [PubMed: 18582855]

Dell, GS.; Brown, PM. Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In: Napoli, DJ.; Kegl, JA., editors. Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman. Hillsdale, NJ: Lawrence Erlbaum; 1991. p. 105-130.

Finegan, E.; Biber, D. Register variation and social dialect variation: The register axiom. In: Eckert, P.; Rickford, JR., editors. Style and sociolinguistic variation. Cambridge: Cambridge University Press; 2001. p. 235-267.

Foulkes, P.; Hay, JB. The emergence of sociophonetic structure. In: MacWhinney, B.; O'Grady, W., editors. The handbook of language emergence. Oxford, UK: Blackwell Publishers; 2015. p. 292-313.

Fowler CA, Housum J. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. Journal of Memory and Language. 1987; 26:489–504. http://dx.doi.org/10.1016/0749-596X(87)90136-7.

Fox NP, Reilly M, Blumstein SE. Phonological neighborhood competition affects spoken word production irrespective of sentential context. Journal of Memory and Language. 2015; 83:97–117. http://dx.doi.org/10.1016/j.jml.2015.04.002. [PubMed: 26124538]

Fox Tree JE, Clark HH. Pronouncing "the" as "thee" to signal problems in speaking. Cognition. 1997; 62:151–167. http://dx.doi.org/10.1016/S0010-0277(96)00781-0. [PubMed: 9141905]

Fricke M, Baese-Berk MM, Goldrick M. Dimensions of similarity in the mental lexicon. Language, Cognition and Neuroscience (Advance online publication). 2016 http://dx.doi.org/10.1080/23273798.2015.1130234.

Galati A, Brennan SE. Attenuating information in spoken communication: For the speaker, or for the addressee? Journal of Memory and Language. 2010; 62:35–51. http://dx.doi.org/10.1016/j.jml.2009.09.002.

Giles, H.; Coupland, N.; Coupland, J. Accomodation theory: Communication, context, and consequence. In: Giles, H.; Coupland, N.; Coupland, J., editors. Contexts of accomodation: Developments in applied sociolinguistics. Cambridge, UK: Cambridge University Press; 1991. p. 1-68.

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. http://dx.doi.org/10.1037/0033-295X.105.2.251. [PubMed: 9577239]

Goldrick M, Chu K. Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010) . Language, Cognition and Neuroscience. 2014; 29:452–458. http://dx.doi.org/10.1080/01690965.2013.807347.

Goldrick M, Folk JR, Rapp B. Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. Journal of Memory and Language. 2010; 62:113–134. http://dx.doi.org/10.1016/j.jml.2009.11.008. [PubMed: 20161591]

Goldrick M, Vaughn C, Murphy A. The effects of lexical neighbors on stop consonant articulation. The Journal of the Acoustical Society of America. 2013; 134:EL172–EL177. http://dx.doi.org/10.1121/1.4812821. [PubMed: 23927221]

Gruenstein A, McGraw I, Badr I. The WAMI toolkit for developing, deploying and evaluating web-accessible multimodal interfaces. 10th International conference on multimodal interfaces. 2008

Guy, GR. Form and function in linguistic variation. In Towards a social science of language: Papers in honor of William Labov. In: Guy, GR.; Feagin, C.; Schiffrin, D.; Baugh, J., editors. Variation and change in language and society. Vol. 1. Amsterdam: Benjamin Publishing Company; 1996. p. 221-252.

Hanulíková A, Alphen PMV, Goch MMV, Weber A. When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. Journal of Cognitive Neuroscience. 2012; 24:878–887. http://dx.doi.org/10.1162/jocn_a_00103. [PubMed: 21812565]

Harris CM, Wolpert DM. Signal-dependent noise determines motor planning. Nature. 1998; 394:780–784. http://dx.doi.org/10.1038/29528. [PubMed: 9723616]

Hay JB. Functions of humor in the conversations of men and women. Journal of Pragmatics. 2000; 32:709–742. http://dx.doi.org/10.1016/S0378-2166(99)00069-7.

Heller, D.; Grodner, D.; Tanenhaus, MK. The real-time use of information about common ground in restricting domains of reference. In: Sauerland, U.; Yatsushiro, K.; Breheny, R., editors. Semantics and pragmatics: From experiment to theory. London, UK: Palgrave Macmillan; 2009.

Hickok G. Computational neuroanatomy of speech production. Nature reviews neuroscience. 2012; 13:135–145. http://dx.doi.org/10.1038/nrn3158.

Horton WS, Gerrig RJ. Conversational common ground and memory processes in language production. Discourse Processes. 2005; 40:1–35. http://dx.doi.org/10.1207/s15326950dp4001_1.

Horton WS, Gerrig RJ. Revisiting the memory-based processing approach to common ground. Topics in Cognitive Science. in press

Horton WS, Keysar B. When do speakers take into account common ground? Cognition. 1996; 59:91–117. http://dx.doi.org/10.1016/0010-0277(96)81418-1. [PubMed: 8857472]

Houde JF, Jordan MI. Sensorimotor adaptation in speech production. Science. 1998; 279:1213–1216. http://dx.doi.org/10.1126/science.279.5354.1213. [PubMed: 9469813]

Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in Human Neuroscience. 2011; 5:1–14. http://dx.doi.org/10.3389/fnhum.2011.00082. [PubMed: 21283556]

Huettig F, Hartsuiker RJ. Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. Language and Cognitive Processes. 2010; 25:347–374. http://dx.doi.org/10.1080/01690960903046926.

Jacobs CL, Yiu LK, Watson DG, Dell GS. Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. Journal of Memory and Language. 2015; 84:37–48. http://dx.doi.org/10.1016/j.jml.2015.05.004. [PubMed: 26089592]

Jaeger TF. Production preferences cannot be understood without reference to communication. Frontiers in Psychology. 2013; 4:1–4. http://dx.doi.org/10.3389/fpsyg.2013.00230. [PubMed: 23382719]

Jaeger, TF.; Buz, E. Signal reduction and linguistic encoding. In: Fernández, EM.; Cairns, HS., editors. Handbook of psycholinguistics. Wiley-Blackwell; in press

Jaeger TF, Ferreira VS. Seeking predictions from a predictive framework. The Behavioral and Brain Sciences. 2013; 36:359–360. http://dx.doi.org/10.1017/S0140525X12002762. [PubMed: 23789872]

Jaeger TF, Furth K, Hilliard C. Incremental phonological encoding during unscripted sentence production. Frontiers in Psychology. 2012a; 3:1–22. http://dx.doi.org/10.3389/fpsyg.2012.00481. [PubMed: 22279440]

Jaeger TF, Furth K, Hilliard C. Phonological overlap affects lexical selection during sentence production. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012b; 38:1439–1449. http://dx.doi.org/10.1037/a0027862.

Jaeger, TF.; Grimshaw, J. 19th Architecture and mechanisms for language processing. France: Marseille; 2013. Information density affects both production and grammatical constraints.

Johnson, K. Speech perception without speaker normalization: An exemplar model. In: Johnson, K.; Mullennix, J., editors. Talker variability in speech processing. San Diego: Academic Press; 1997. p. 145-165.

Jokinen, K. Gaze and gesture activity in communication. In Universal access in human–computer interaction. In: Stephanidis, C., editor. Intelligent and ubiquitous interaction environments. Vol. 5615. Berlin: Springer; 2009. p. 537-546.http://dx.doi.org/10.1007/978-3-642-02710-9_60

de Jong KJ. Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. Journal of Phonetics. 2004; 32:493–516. http://dx.doi.org/10.1016/j.wocn.2004.05.002.

Kahn JM, Arnold JE. A processing-centered look at the contribution of givenness to durational reduction. Journal of Memory and Language. 2012; 67:311–325. http://dx.doi.org/10.1016/j.jml.2012.07.002.

Kang K-H, Guion SG. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. The Journal of the Acoustical Society of America. 2008; 124:3909–3917. http://dx.doi.org/10.1121/1.2988292. [PubMed: 19206816]

Keysar B, Barr DJ, Horton WS. The egocentric basis of language use: Insights from a processing approach. Current Directions in Psychological Science. 1998; 7:46–50. http://dx.doi.org/10.1111/1467-8721.ep13175613.

Kirov, C.; Wilson, C. Proceedings of the 34th annual conference of the cognitive science society. Cognitive Science Society; 2012. The specificity of online variation in speech production; p. 587-592.

Kirov, C.; Wilson, C. Proceedings of the 35th annual conference of the cognitive science society. Cognitive Science Society; 2013. Bayesian speech production: Evidence from latency and hyper-articulation; p. 788-793.

Klatt DH. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. The Journal of the Acoustical Society of America. 1976; 59:1208–1221. http://dx.doi.org/10.1121/1.380986. [PubMed: 956516]

Kleinschmidt, DF.; Jaeger, TF. A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. In: Miyake, N.; Peebles, D.; Cooper, RP., editors. Proceedings of the 34th annual conference of the cognitive science society. Austin, TX: Cognitive Science Society; 2012. p. 599-604.

Kleinschmidt DF, Jaeger TF. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. Psychological Review. 2015; 122:148–203. http://dx.doi.org/10.1037/a0038695. [PubMed: 25844873]

Kleinschmidt, DF.; Raizada, R.; Jaeger, TF. Supervised and unsupervised learning in phonetic adaptation. In: Dale, R.; Jennings, C.; Maglio, P.; Matlock, T.; Noelle, D.; Warlaumont, A., et al., editors. Proceedings of the 37th annual meeting of the cognitive science society. Austin, TX: Cognitive Science Society; 2015. p. 1129-1134.

Kohler, KJ. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In: Hardcastle, WJ.; Marchal, A., editors. Speech production and speech modeling. Boston/London: Kluwer; 1990. p. 69-92.

Kraljic T, Samuel AG, Brennan SE. First impressions and last resorts: How listeners adjust to speaker variability. Psychological Science. 2008; 19:332–338. http://dx.doi.org/10.1111/j.1467-9280.2008.02090.x. [PubMed: 18399885]

Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, Lacerda F. Cross-language analysis of phonetic units in language addressed to infants. Science. 1997; 277:684–686. http://dx.doi.org/10.1126/science.277.5326.684. [PubMed: 9235890]

Kuhlen AK, Brennan SE. Language in dialogue: When confederates might be hazardous to your data. Psychonomic Bulletin & Review. 2013; 20:54–72. http://dx.doi.org/10.3758/s13423-012-0341-8. [PubMed: 23188738]

Kuperberg GR, Jaeger TF. What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience. 2016; 31:32–59. http://dx.doi.org/10.1080/23273798.2015.1102299.

Kurumada, C. Doctoral thesis. Stanford University; 2013. Navigating variability in the linguistic signal: Learning to interpret contrastive prosody.

Lam TQ, Watson DG. Repetition is easy: Why repeated referents have reduced prominence. Memory & Cognition. 2010; 38:1137–1146. http://dx.doi.org/10.3758/MC.38.8.1137. [PubMed: 21156876]

Levelt, WJM. Producing spoken language: A blueprint of the speaker. In: Brown, C.; Hagoort, P., editors. The neurocognition of language. Oxford Press; 1999. p. 83-122.

Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral and Brain Sciences. 1999; 22:1–38. [PubMed: 11301520]

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review. 1967; 74:431–461. http://dx.doi.org/10.1037/h0020279. [PubMed: 4170865]

Lindblom, BEF. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, WJ.; Marchal, A., editors. Speech production and speech modelling. Dordrecht: Springer Netherlands; 1990. p. 403-439. http://dx.doi.org/10.1007/978-94-009-2037-8_16

Lockridge CB, Brennan SE. Addressees' needs influence speakers' early syntactic choices. Psychonomic Bulletin & Review. 2002; 9:550–557. http://dx.doi.org/10.3758/BF03196312. [PubMed: 12412896]

Lombard E. Le signe de l'elevation de la voix. Ann Maladies Oreille, Larynx, Nez, Pharynx. 1911; 37:101–119.

Maniwa K, Jongman A, Wade T. Acoustic characteristics of clearly spoken English fricatives. The Journal of the Acoustical Society of America. 2009; 125:3962–3973. http://dx.doi.org/10.1121/1.2990715. [PubMed: 19507978]

Martin A, Schatz T, Versteegh M, Miyazawa K, Mazuka R, Dupoux E, et al. Mothers speak less clearly to infants than to adults: A comprehensive test of the hyper-articulation hypothesis. Psychological Science. 2015; 26:341–347. http://dx.doi.org/10.1177/0956797614562453. [PubMed: 25630443]

McMurray B, Aslin RN, Tanenhaus MK, Spivey MJ, Subik D. Gradient sensitivity to within-category variation in words and syllables. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34:1609–1631. http://dx.doi.org/10.1037/a0011747. [PubMed: 19045996]

McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. Cognition. 2002; 86:B33–B42. http://dx.doi.org/10.1016/S0010-0277(02)00157-9. [PubMed: 12435537]

Munson, B. Lexical access, lexical representation, and vowel production. In: Cole, J.; Hualde, JI., editors. Laboratory phonology. Vol. 9. Berlin: Mouton de Gruyter; 2007. p. 201-228.

Niedzielski N. The effect of social information on the perception of sociolinguistic variables. Journal of Language and Social Psychology. 1999; 18:62–85. http://dx.doi.org/10.1177/0261927X99018001005.

Ohala, JJ. Sound change is drawn from a pool of synchronic variation. In: Breivik, LE.; Jahr, EH., editors. Language change: Contributions to the study of its causes. Berlin: Mouton de Gruyter; 1989. p. 173-198.

Ohala JJ. Acoustic study of clear speech: A test of the contrastive hypothesis. In Proceedings of the international symposium on prosody. 1994:75–89).

O'Seaghdha PG, Marin JW. Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. Journal of Experimental Psychology: Human Perception and Performance. 2000; 26:57–73. http://dx.doi.org/10.1037/0096-1523.26.1.57. [PubMed: 10696605]

Oviatt S, Levow G-A, Moreton E, MacEachern M. Modeling global and focal hyper-articulation during human-computer error resolution. The Journal of the Acoustical Society of America. 1998; 104:3080–3098. http://dx.doi.org/10.1121/1.423888. [PubMed: 9821350]

Oviatt S, MacEachern M, Levow G-A. Predicting hyperarticulate speech during human-computer error resolution. Speech Communication. 1998; 24:87–110. http://dx.doi.org/10.1016/S0167-6393(98)00005-3.

Pate JK, Goldwater S. Talkers account for listener and channel characteristics to communicate efficiently. Journal of Memory and Language. 2015; 78:1–17. http://dx.doi.org/10.1016/j.jml.2014.10.003.

Peramunage D, Blumstein SE, Myers EB, Goldrick M, Baese-Berk MM. Phonological neighborhood effects in spoken word production: An fMRI study. Journal of Cognitive Neuroscience. 2011; 23:593–603. http://dx.doi.org/10.1162/jocn.2010.21489. [PubMed: 20350185]

Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing II. Journal of Speech Language and Hearing Research. 1986; 29:434–446. http://dx.doi.org/10.1044/jshr.2904.434.

Pickering MJ, Garrod S. Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences. 2004; 27:169–190. http://dx.doi.org/10.1017/S0140525X04000056. [PubMed: 15595235]

Pickering MJ, Garrod S. An integrated theory of language production and comprehension. Behavioral and Brain Sciences. 2013; 36:329–347. http://dx.doi.org/10.1017/S0140525X12001495. [PubMed: 23789620]

Pierrehumbert, JB. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J.; Hopper, P., editors. Frequency and the emergence of linguistic structure. Amsterdam: John Benjamins Publishing Company; 2001. p. 137-157.

Pierrehumbert, JB. Word-specific phonetics. In: Gussenhoven, C.; Warner, N., editors. Laboratory phonology. Vol. 7. Berlin: Mouton de Gruyter; 2002. p. 101-139.

Pouplier M, Goldstein L. Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. Language and Cognitive Processes. 2010; 25:616–649. http://dx.doi.org/10.1080/01690960903395380. [PubMed: 25009365]

Pouplier M, Goldstein L. The relationship between planning and execution is more than duration: Response to Goldrick & Chu. Language, Cognition and Neuroscience. 2014; 29:1097–1099. http://dx.doi.org/10.1080/01690965.2013.834063.

Purcell DW, Munhall KG. Compensation following real-time manipulation of formants in isolated vowels. The Journal of the Acoustical Society of America. 2006; 119:2288–2297. http://dx.doi.org/10.1121/1.2173514. [PubMed: 16642842]

Sanchez K, Hay JB, Nilson E. Contextual activation of Australia can affect New Zealanders' vowel productions. Journal of Phonetics. 2015; 48:76–95. http://dx.doi.org/10.1016/j.wocn.2014.10.004.

Schertz JL. Exaggeration of featural contrasts in clarifications of misheard speech in English. Journal of Phonetics. 2013; 41:249–263. http://dx.doi.org/10.1016/j.wocn.2013.03.007.

Schertz JL, Cho T, Lotto AJ, Warner N. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. Journal of Phonetics. 2015; 52:183–204. http://dx.doi.org/10.1016/j.wocn.2015.07.003. [PubMed: 26644630]

Sevald CA, Dell GS. The sequential cuing effect in speech production. Cognition. 1994; 53:91–127. http://dx.doi.org/10.1016/0749-596X(90)90011-N. [PubMed: 7805353]

Seyfarth S, Buz E, Jaeger TF. Dynamic hyperarticulation of coda voicing contrasts. Journal of the Acoustical Society of America. in press

Shintel H, Keysar B. Less is more: A minimalist account of joint action in communication. Topics in Cognitive Science. 2009; 1:260–273. http://dx.doi.org/10.1111/j.1756-8765.2009.01018.x. [PubMed: 25164932]

Staum Casasanto, L. Does social information influence sentence processing?. In: Love, BC.; McRae, K.; Sloutsky, VM., editors. Proceedings of the 30th annual meeting of the cognitive science society. Austin, TX: Cognitive Science Society; 2008. p. 799-804.

Stent AJ, Huffman MK, Brennan SE. Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. Speech Communication. 2008; 50:163–178. http://dx.doi.org/10.1016/.specom.2007.07.005.

Strand EA. Uncovering the role of gender stereotypes in speech perception. Journal of Language and Social Psychology. 1999; 18:86–100. http://dx.doi.org/10.1177/0261927X99018001006.

Todorov E, Jordan MI. Optimal feedback control as a theory of motor coordination. Nature Neuroscience. 2002; 5:1226–1235. http://dx.doi.org/10.1038/nn963. [PubMed: 12404008]

Tourville JA, Guenther FH. The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes. 2011; 26:952–981. http://dx.doi.org/10.1080/01690960903498424. [PubMed: 23667281]

Trommershäuser J, Gepshtein S, Maloney LT, Landy MS, Banks MS. Optimal compensation for changes in task-relevant movement variability. The Journal of Neuroscience. 2005; 25:7169–7178. http://dx.doi.org/10.1523/JNEUROSCI.1906-05.2005. [PubMed: 16079399]

Trommershäuser J, Maloney LT, Landy MS. Decision making, movement planning and statistical decision theory. Trends in Cognitive Sciences. 2008; 12:291–297. http://dx.doi.org/10.1016/j.tics.2008.04.010. [PubMed: 18614390]

Uther M, Knoll MA, Burnham D. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. Speech Communication. 2007; 49:2–7. http://dx.doi.org/10.1016/j.specom.2006.10.003.

Van Summers W, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA. Effects of noise on speech production: Acoustic and perceptual analyses. Journal of the Acoustical Society of America. 1988; 84:917–928. http://dx.doi.org/10.1121/1.396660. [PubMed: 3183209]

Villacorta VM, Perkell JS, Guenther FH. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. The Journal of the Acoustical Society of America. 2007; 122:2306–2319. http://dx.doi.org/10.1121/1.2773966. [PubMed: 17902866]

Walker AJ, Hay JB. Congruence between 'word age' and 'voice age' facilitates lexical access. Laboratory Phonology. 2011; 2:219–237. http://dx.doi.org/10.1515/labphon.2011.007.

Watson, DG.; Buxó-Lugo, A.; Simmons, DC. The effect of phonological encoding on word duration: Selection takes time. In: Gibson, E.; Frazier, L., editors. Explicit and implicit prosody in sentence processing. Switzerland: Springer International Publishing; 2015. p. 85-98.http://dx.doi.org/10.1007/978-3-319-12961-7_5

Weatherholtz K, Campbell-Kibler K, Jaeger TF. Socially-mediated syntactic alignment. Language Variation and Change. 2014; 26:387–420. http://dx.doi.org/10.1017/S0954394514000155.

Weatherholtz K, Jaeger TF. Speech perception and generalization across talkers and accents. Oxford Research Encyclopedia in Linguistics. in press

Weatherholtz K, Seifeldin M, Kleinschmidt DF, Kurumada C, Jaeger TF. (submitted for publication) Language processing as probabilistic inference under uncertainty based on social-indexical knowledge. Language and Linguistics Compass.

Wedel A. Exemplar models, evolution and language change. The Linguistic Review. 2006; 23:247–274. http://dx.doi.org/10.1515/TLR.2006.010.

Wei K, Körding KP. Relevance of error: What drives motor adaptation? Journal of Neurophysiology. 2008; 101:655–664. http://dx.doi.org/10.1152/jn.90545.2008. [PubMed: 19019979]
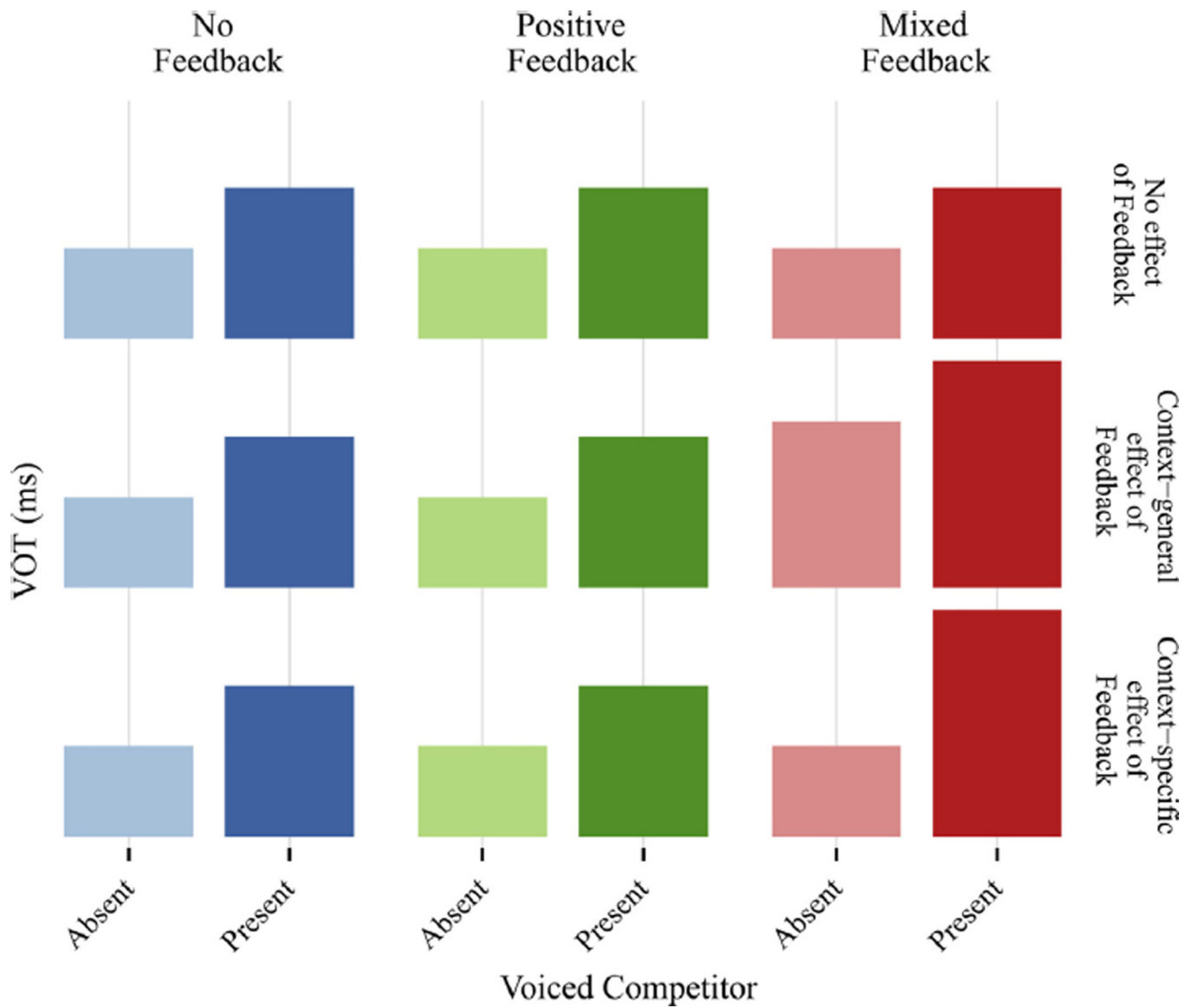
Fig. 1.
Possible changes in VOT based on Competitor and Feedback. The top panel illustrates no effect of Feedback, the middle panel illustrates a main effect of Feedback such that speakers in the Mixed Feedback group lengthen VOT across Competitor conditions. The lower panel illustrates an interaction between Feedback and Competitor.
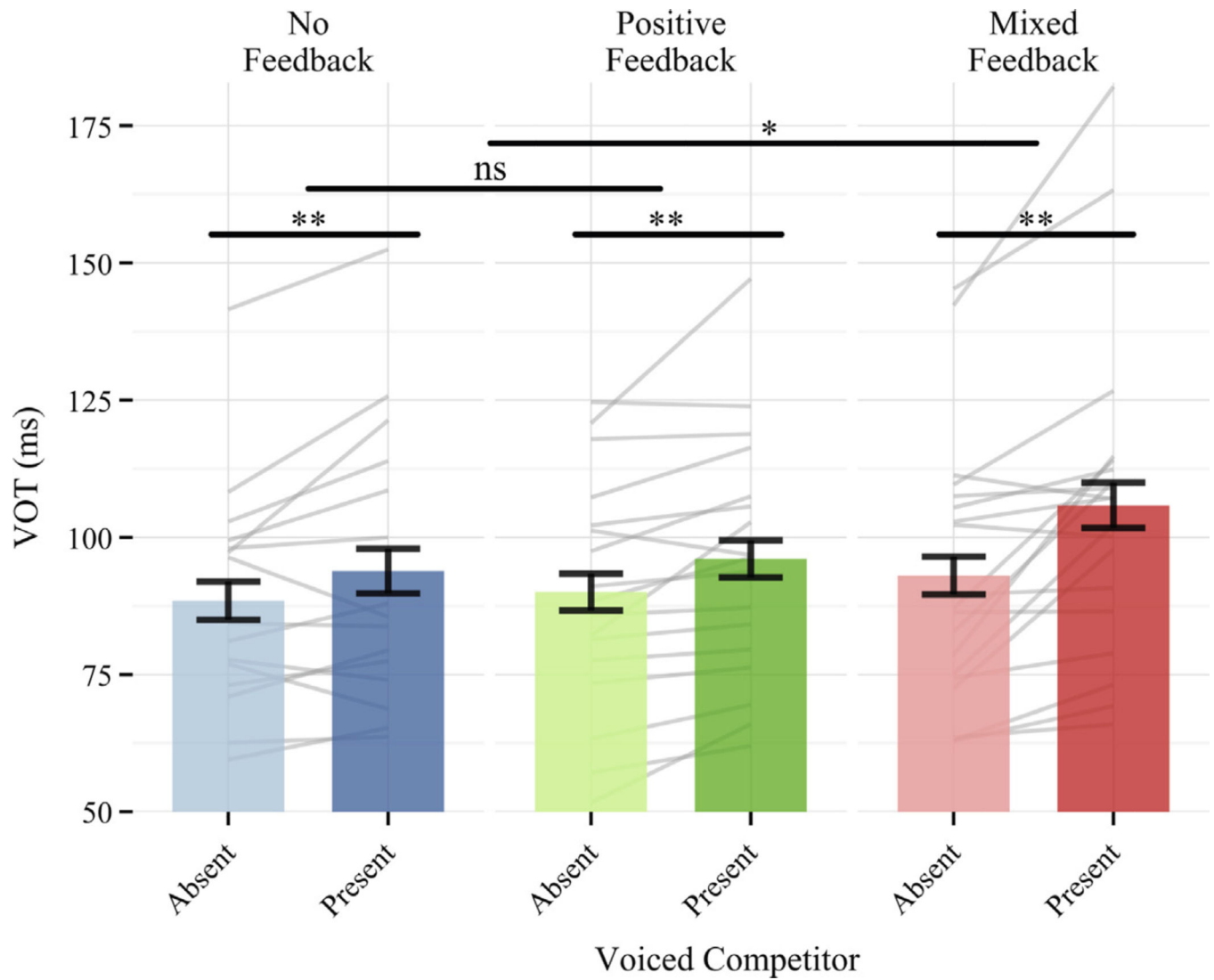
Fig. 2.
VOT by condition aggregating within participants (lines) and across participants (bars). Error bars indicate ± 1SE after aggregating over participants. Thick lines indicate significant simple effects of Voiced Competitor and the interaction of this effect with Feedback conditions (Helmert-coded; see text for details).
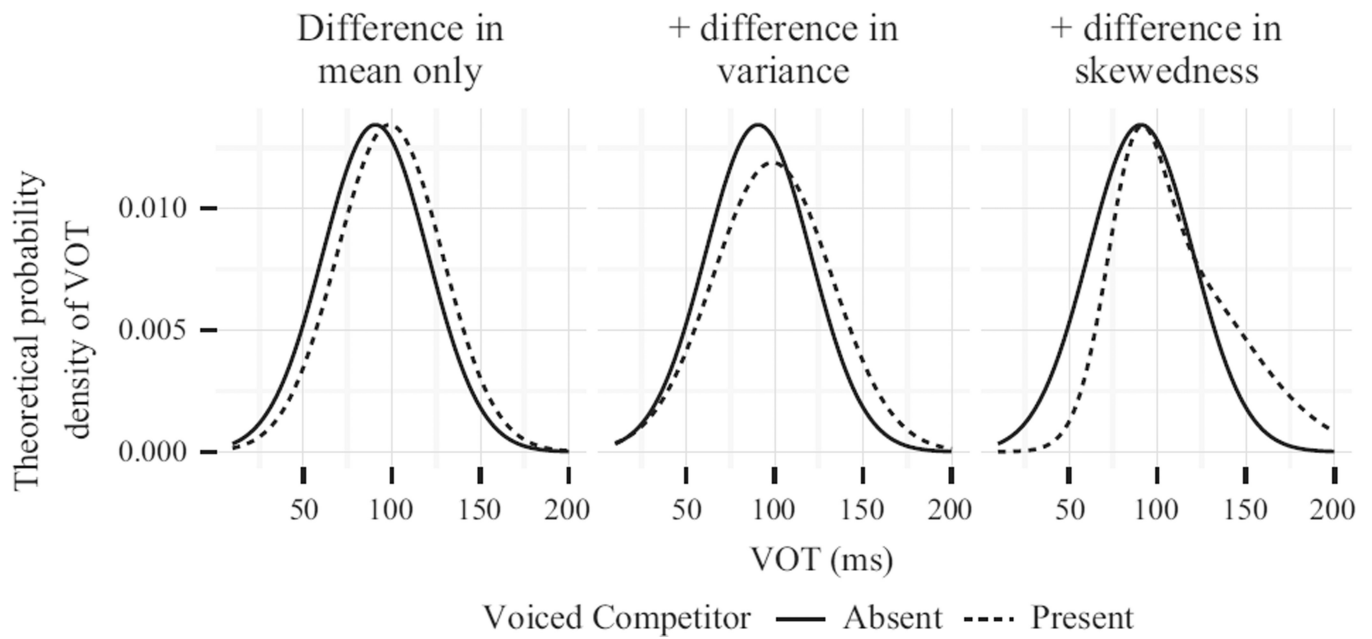
Fig. 3.
Different hypothetical scenarios that lead to an 8 ms difference in VOTs between the Voiced
Competitor Absent and Present conditions. In the left panel, only the mean differs between
the Voiced Competitor Absent and Present conditions. In the middle panel, both the mean
and the variances differ. In the right panel, the mean, variance, and skewedness differ. For
further details, see text.
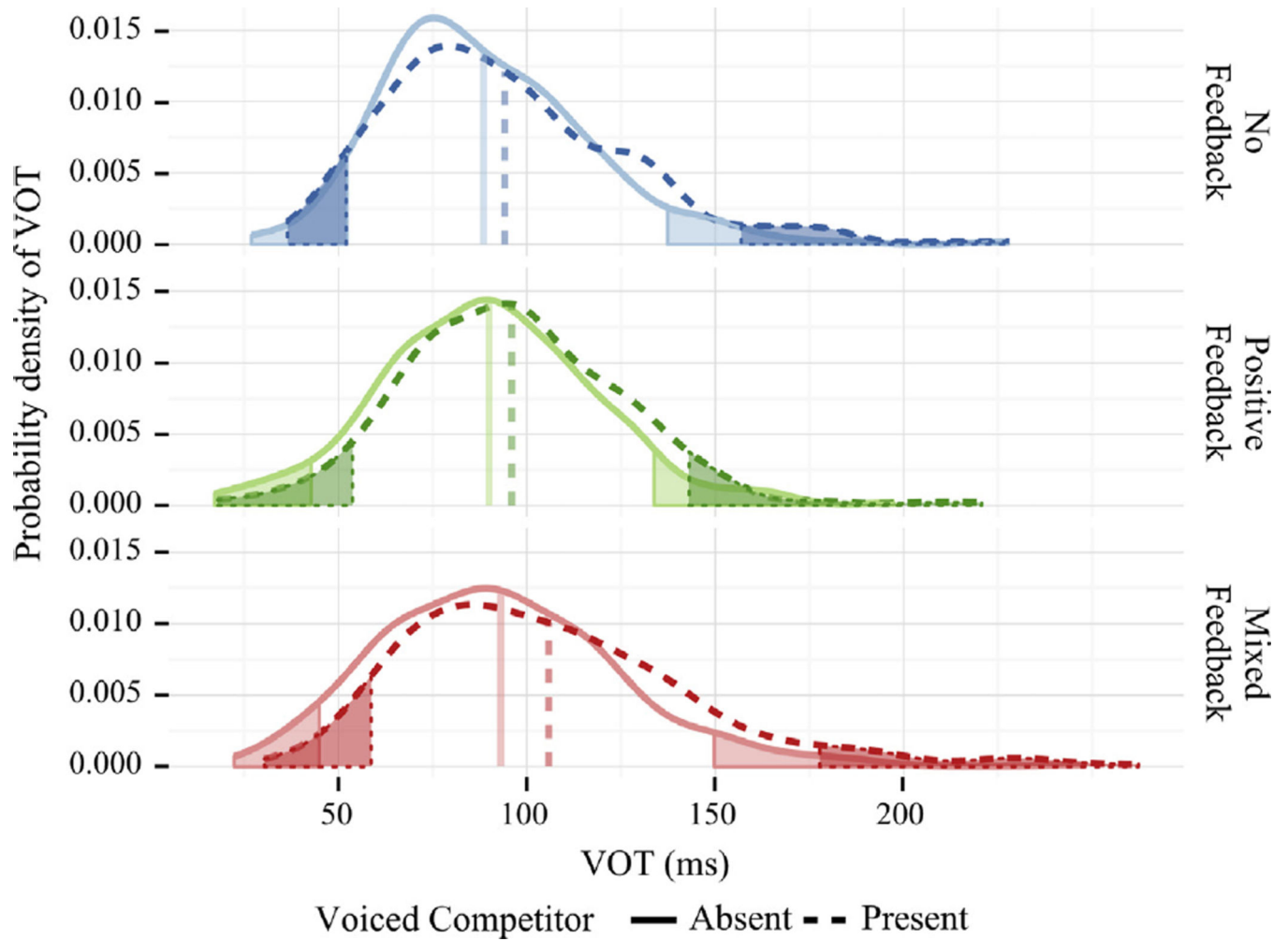
Fig. 4.
Probability density of VOT distributions across Voiced Competitor and Feedback manipulations. Also plotted is the non-parametric bootstrap estimated shift of the 5th and 95th quantiles and the mean.

Table 1

Mean and standard deviation (sd) of VOTs across conditions.

Feedback	VOT					Difference
	Voiced Competitor					
	Present		**Absent**			
	Mean	**sd**	**Mean**	**sd**		
No	93.95	32	89.03	28		4.92
Positive	96.21	29	89.74	28		6.47
Mixed	105.74	38	92.88	32		12.86
All	98.81	34	90.68	30		8.13

Table 2

Model summary of Voiced Competitor and Feedback effects on VOTs. The first number for each predictor is the coefficient estimate, the second number is the value of the t-statistic. Significance is indicated by stars.

Predictor	$\hat{\beta}$	*t*-value
Intercept	−.3	−.1
Voiced Competitor Present vs Absent	4.0	5.2**
Positive vs No Feedback (PvN)	.8	.2
Mixed vs other Feedback (MvO)	2.4	1.1
Voiced Competitor X PvN	.1	.1
Voiced Competitor X MvO	1.1	2.1*

Note:

+$p < .1$;

*$p < .05$;

**$p < .01$.

Table 3

Descriptive statistics for the VOT distributions plotted in Fig. 4. Statistics are estimated using non-parametric bootstrap resampling of each distribution.

Statistic	Feedback	Voiced Competitor effect (Present–Absent)	*p*-value
5th quantile	No	−.28	.56
	Positive	9.25	.04 *
	Mixed	13.59	.00 **
Mode	No	3.97	.25
	Positive	4.45	.27
	Mixed	−1.64	.58
Mean	No	4.79	.03 *
	Positive	6.34	.00 **
	Mixed	12.90	.00 **
95th quantile	No	13.39	.15
	Positive	9.63	.13
	Mixed	28.48	.00 **

Note:

$^{+}p < .1$;

$^{*}p < .05$;

$^{**}p < .01$.

Table 4

Descriptive statistics of differences in the VOT distributions across Feedback groups plotted in Fig. 4. Statistics are estimated using non-parametric bootstrap resampling of each distribution.

Statistic	Comparison	Present-Absent	
		Difference between groups in Competitor effect	*p*-value
5th quantile	Positive-No	9.98	.07[+]
	Mixed-Others	9.16	.04[*]
Mode	Positive-No	−.11	.50
	Mixed-Others	−5.61	.68
Mean	Positive-No	1.54	.34
	Mixed-Others	7.17	.01[**]
95th quantile	Positive-No	−5.14	.64
	Mixed-Others	17.97	.08[+]

Note:

[+] $p < .1$;

[*] $p < .05$;

[**] $p < .01$.

Table 5

Mean and standard deviation (sd) of speech onset latencies across conditions. Subsequent analyses using the same methodology as used to analyze VOT (see Methods section in main text) found no significant effect of Competitor, Feedback nor their interaction on latencies (p's > .2). We note, however, that speech onset latency estimates are based on individual participant computer hardware, possibly making these measures too coarse-grained to find reliable differences.

Feedback	Speech onset latencies				
	Voiced Competitor				
	Present		Absent		Difference
	Mean	sd	Mean	sd	
No	801.4	213	805.3	371	−3.9
Positive	814.6	322	787.1	357	27.5
Mixed	918.0	293	871.5	257	46.5
All	846.2	287	823.5	330	22.7

Table A.6

Critical stimuli. Voiced minimal pair and non-minimal pair distractors are shown as *minimal pair/non minimal pair*.

Condition	Target	Distractor 1	Distractor 2	Target Onset
item	cap	gap/dim	wolf	k
item	kilt	guilt/yacht	toast	k
item	cuff	guff/noon	reed	k
item	cab	gab/rice	surf	k
item	kit	git/bean	lace	k
item	cot	got/wig	haze	k
item	cape	gape/thieve	yard	k
item	cob	gob/shed	save	k
item	code	goad/chip	nun	k
item	core	gore/want	type	k
item	curl	girl/soft	nest	k
item	pike	bike/fog	dart	p
item	pig	big/heart	daft	p
item	punk	bunk/ship	lard	p
item	punch	bunch/cool	wool	p
item	peat	beat/wren	thug	p
item	palm	balm/hag	nose	p
item	pack	back/wrist	guild	p
item	pun	bun/golf	tar	p
item	pill	bill/food	hair	p
item	pore	bore/mile	shelf	p
item	putt	butt/land	zone	p
item	pad	bad/van	norm	p
item	teem	deem/board	mast	t
item	tent	dent/worm	barn	t
item	tab	dab/shoal	wedge	t
item	tart	dart/gel	mouse	t
item	tomb	doom/whale	male	t
item	ton	done/mouth	hill	t
item	tip	dip/moon	hood	t
item	tame	dame/sign	bench	t
item	tuck	duck/beard	wit	t
item	tote	dote/shaft	rug	t
item	taunt	daunt/mood	guile	t
item	tore	door/half	goose	t
item	teal	deal/yap	goof	t