Journal of Experimental Psychology:
Learning, Memory, and Cognition
2017, Vol. 43, No. 8, 1225–1238

© 2017 American Psychological Association
0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000381

How Our Own Speech Rate Influences Our Perception of Others

Hans Rutger Bosker
Max Planck Institute for Psycholinguistics and Radboud University

In conversation, our own speech and that of others follow each other in rapid succession. Effects of the surrounding context on speech perception are well documented but, despite the ubiquity of the sound of our own voice, it is unknown whether our own speech also influences our perception of other talkers. This study investigated context effects induced by our own speech through 6 experiments, specifically targeting rate normalization (i.e., perceiving phonetic segments relative to surrounding speech rate). Experiment 1 revealed that hearing prerecorded fast or slow context sentences altered the perception of ambiguous vowels, replicating earlier work. Experiment 2 demonstrated that talking at a fast or slow rate prior to target presentation also altered target perception, though the effect of preceding speech rate was reduced. Experiment 3 showed that silent talking (i.e., inner speech) at fast or slow rates did not modulate the perception of others, suggesting that the effect of self-produced speech rate in Experiment 2 arose through monitoring of the external speech signal. Experiment 4 demonstrated that, when participants were played back their own (fast/slow) speech, no reduction of the effect of preceding speech rate was observed, suggesting that the additional task of speech production may be responsible for the reduced effect in Experiment 2. Finally, Experiments 5 and 6 replicate Experiments 2 and 3 with new participant samples. Taken together, these results suggest that variation in speech production may induce variation in speech perception, thus carrying implications for our understanding of spoken communication in dialogue settings.

Keywords: speech rate normalization, self-monitoring, covert speech, phonetic convergence, speaking-induced suppression

Words seldom occur in isolation. Rather, spoken words are produced in rich acoustic contexts, including other speech from the same talker (e.g., the surrounding sentence), speech from other interlocutors (e.g., in conversational settings), and other nonspeech acoustic signals (e.g., music playing in the background). Research on speech perception has long recognized that the spectral and temporal properties of the acoustic context may influence speech perception (e.g., Ladefoged & Broadbent, 1957; Miller & Liberman, 1979). These context effects are contrastive: Manipulating the characteristics of a context sentence in one direction (e.g., lowering F2; increasing speech rate) will bias the perception of a subsequent target word in the other direction (e.g., perception of higher F2; longer syllable duration). For instance, the perception of an ambiguous Dutch vowel midway between short /ɑ/ and long /a:/ is biased toward perceiving long /a:/ when it is presented in a context sentence with a relatively fast speech rate (Bosker & Reinisch, 2015; Bosker, Reinisch, & Sjerps, 2016).

This contrastive effect of the surrounding speech rate, known as rate normalization, has been characterized as a general auditory process (Reinisch & Sjerps, 2013). Rate normalization has been reported for proximal contexts (i.e., adjacent phonemes; Summerfield, 1981), distal contexts (i.e., surrounding words; Dilley & Pitt, 2010; Reinisch, Jesse, & McQueen, 2011), and even for more global contexts (i.e., effects of the average speech rate calculated over an extended period of time; Baese-Berk et al., 2014). Moreover, rate normalization seems to generalize across different sound sources. That is, the rate of one speaker may affect the perception of another speaker (Newman & Sawusch, 2009; Sawusch & Newman, 2000). In fact, rate normalization appears even to be triggered by nonspeech contexts such as (fast and slow) tone sequences (Wade & Holt, 2005; but see Pitt, Szostak, & Dilley, 2016).

Several explanations have been proposed to account for rate normalization in speech perception. Gestural accounts of rate normalization (e.g., Fowler, 1990, 1991; Miller & Liberman, 1979) hold that rate normalization is the result of listeners retrieving the speaker's rate, and adjusting their perception of subsequent target words appropriately. These gestural accounts have been challenged by general auditory accounts of rate normalization. These accounts hold that rate normalization does not involve retrieval of a speaker's speech rate but rather involve general auditory principles. One example of a general auditory account

This article was published Online First January 16, 2017.

Hans Rutger Bosker, Max Planck Institute for Psycholinguistics, and Donders Institute for Brain, Cognition, and Behaviour, Radboud University.

This research was supported by a Gravitation grant from the Dutch Government to the Language in Interaction Consortium.

I thank Eleanor Drake and two anonymous reviewers for useful comments on an earlier version of this article. Thanks also to Antje Meyer for useful comments and suggestions, to Ronald Fischer and Johan Weustink for technical support, to all the student assistants who helped with testing participants and evaluating audio and video materials, and to Annelies van Wijngaarden, who coordinated their efforts and whose voice was recorded for the speech materials used in this study.

Parts of this study have been presented at Speech Prosody 2016, Boston University.

Correspondence concerning this article should be addressed to Hans Rutger Bosker, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, the Netherlands. E-mail: hansrutger.bosker@mpi.nl

involves durational contrast (Wade & Holt, 2005), and holds that listeners perceive duration cues relative to adjacent temporal cues. Another example, as proposed recently by Bosker (2016), involves neural entrainment with endogenous neural oscillations phase-locking to the rhythm of the speech signal (Peelle & Davis, 2012). The empirical finding that rate normalization seems to occur across different speech and nonspeech streams is in line with either of these general auditory accounts of rate normalization. At the same time, it challenges a gestural account of rate normalization since listeners cannot be assumed to retrieve the speaking rate of a series of tones.

If rate normalization indeed generalizes across different speech streams, then yet another source of context may be observed to influence speech perception, namely the sound of our own voice. In natural conversations, our own utterances and those of others follow each other in rapid succession. Interlocutors universally try to avoid overlapping talk and to minimize the silence between conversational turns. In fact, across a range of typologically diverse languages, turn transitions between speakers were found to have a fairly consistent duration of approximately 100 ms (Stivers et al., 2009). As such, the immediate context of an utterance spoken by our conversational partner includes speech that we produced ourselves moments earlier. Given the close temporal proximity of our own speech to that of others, our own speech rate may potentially induce rate normalization of the speech of others.

There is already some indication in the literature that our own speech rate alters our perception of others. Such studies typically investigate effects of listeners' habitual speech rate by means of explicit evaluative judgments. For instance, listeners with a habitually slow speech rate have been found to judge speech as faster than listeners with a relatively fast habitual speech rate (Schwab, 2011). However, this effect was only observed with slow and neutral rates, not with fast speech. Furthermore, Koreman (2006) failed to find any effect of listeners' own habitual speech rate on speech rate evaluation. One complicating factor in these studies is that explicit judgments of perceived speed do not always reflect the acoustic speech rate. For instance, acoustic measures of speed of articulation only explain 53% of the variance of perceived speed judgments (Bosker, Pinget, Quené, Sanders, & De Jong, 2013). Therefore, the present study targets more implicit effects of a preceding self-produced speech rate on rate normalization. That is, does talking at a fast (or slow) rate change one's perception of a subsequent utterance, spoken by another talker?

The present study was designed to test whether and how one's own speech rate might influence the perception of utterances produced by another talker. It includes four experiments (Experiments 1–4) using a within-participants design and two replication experiments (Experiments 5–6) using a between-participants design. All aimed at examining the effects of preceding slow or fast speech rate on the perception of the Dutch vowel contrast between /ɑ/ and /a:/. First, Experiment 1 aimed to replicate the standard finding of rate normalization. Participants heard manipulated target words, with vowels ambiguous between /ɑ/ and /a:/, embedded in fast and slow prerecorded context sentences. Their task was to indicate which sentence-final target word they heard, involving a two-alternative forced choice between two response options representing a Dutch minimal word pair differentiated only by the vowel (e.g., *staf*–*staaf*). Fast context sentences were expected to bias perception of the target vowel toward /a:/, and slow context sentences toward /ɑ/.

Experiment 2 extended this experimental design by investigating whether producing the context sentences at a fast and slow speech rate oneself (i.e., without the target word) would elicit similar rate normalization effects (i.e., self-induced rate normalization). Participants were explicitly instructed, using a visual cue, to speak at a particular fast or slow rate, after which the target words were automatically presented. If rate normalization indeed operates across different talker streams (Newman & Sawusch, 2009), we may find that talking at a fast rate oneself (prior to target word presentation) may bias the perception of the subsequent target word (produced by another talker) toward /a:/.

If self-induced rate normalization is indeed observed, one may question the mechanism underlying this effect. Findings of self-induced rate normalization may, perhaps most intuitively, stem from self-monitoring of the overt speech signal. When we speak, we typically hear the speech we produce, allowing us to monitor the external signal. Thus, self-perception of the external speech signal may be argued to account for potential effects of our own speech rate (e.g., through durational contrast or neural entrainment), similar to how rate normalization operates when we listen to speech produced by someone else.

Nevertheless, speakers do not only monitor their overt speech (i.e., after speech initiation), but also their inner speech (i.e., prior to speech initiation; cf. Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Lœvenbruck, 2014). This inner speech has been claimed to involve auditory forward models that are used to anticipate the sensory outcome of motor commands (e.g., Pickering & Garrod, 2013; Tian & Poeppel, 2010). As such, inner speech appears to be auditory in nature (i.e., sharing neural infrastructure with overt speech perception; Perrone-Bertolotti et al., 2014; Tian & Poeppel, 2010), to include a phonological level (Pickering & Garrod, 2013), and to be able to affect overt speech perception (Sams, Möttönen, & Sihvonen, 2005; Sato, Troille, Ménard, Cathiard, & Gracco, 2013). Inner speech seems to be a temporal signal (Dell & Oppenheim, 2015) including some specification of speech rate (Netsell, Ashley, & Bakker, 2010). If inner speech is indeed specified for speech rate, then inner speech alone may already elicit rate normalization. If so, findings of self-induced rate normalization may also be attributed to self-monitoring of the internal signal, rather than only to the monitoring of the external speech signal.

In order to disentangle the differential contributions of production mechanisms (i.e., monitoring of the internal signal) and perception mechanisms (i.e., monitoring of the external speech signal), Experiment 3 investigated potential effects of covert speech production at fast and slow rates (control over the rate at which inner speech is produced has been previously reported; e.g., Netsell et al., 2010; Shergill et al., 2002, 2003). If self-induced rate normalization is exclusively the result of self-perception (i.e., monitoring of the external speech signal), we would expect this effect to disappear when speech is produced covertly (i.e., as inner speech in the mind, without audible sound or articulatory movements). However, if effects of our own speech rate are due to the prosodic properties of the forward models involved in speech production, then covert speech production alone may be sufficient to elicit rate normalization—without any overt speech being present.

Another way of disentangling the contribution of perception versus production mechanisms is by comparing rate normalization induced during speech production (as in Experiment 2) to rate normalization induced by listening to recordings of your own voice. Since the temporal characteristics of the speech signals are identical in both situations (self-production vs. self-perception), any potential difference in the size of the rate normalization effect may be uniquely attributed to mechanisms involved in speech production. Therefore, in Experiment 4, participants were invited back to the lab to listen to the recordings of their own voice, talking fast or slow, recorded in Experiment 2. In this way, the contribution of the task of speech production (Experiment 2) can be separated from the contribution of self-perception (Experiment 4).

Finally, two replication experiments were run to test whether the obtained results in Experiments 2 and 3 could be substantiated in two new participant samples. Experiment 5 was identical to Experiment 2, testing effects of overt production of fast and slow speech rates, with a new sample of participants who had not taken part in any of the other experiments. Similarly, Experiment 6 was identical to Experiment 3, testing covert production of fast and slow speech, also with new participants who had not taken part in Experiments 1–5.

Experiment 1: Perception

Experiment 1 was designed to replicate the typical finding in studies on rate normalization (Bosker & Reinisch, 2015; Bosker et al., 2016; Reinisch, 2016a), namely that increasing the speech rate of a context sentence biases the perception of a subsequent target toward longer segments.

Method

Participants. Native Dutch participants (N = 45, 11 male, M_{age} = 27 years) with normal hearing were recruited from the participant pool of the Max Planck Institute. All participated with informed consent as approved by the ethics committee of the Social Sciences Department of Radboud University (Project code ECSW2014-1003–196). In Experiments 1–4, a within-participants design was adopted, with the same sample of participants taking part in each experiment, thus reducing error variance. Experiments 5–6 report two additional experiments, replicating Experiments 2 and 3, but using a between-participants design (i.e., new participant samples). The decision to use a within-participants design for Experiments 1–4 was motivated by (a) the need to familiarize participants with the words of the context sentence, the fast and slow speech rates, and the timing of the onset and offset of the context sentence relative to the target word, if they were to correctly produce the sentences themselves in Experiment 2; and (b) the need to familiarize participants with the task of speech production in Experiment 2 if they were to be expected to correctly perform the considerably more difficult task of covert speech production in Experiment 3. The chronological order of the experiments was fixed: Participants first took part in Experiment 1, then Experiment 2, and then Experiment 3 (all on the same day). Experiment 4 was run several weeks later to allow for the annotation and preparation of participants' self-produced context sentences.

Design and materials. A female native speaker of Dutch was recorded producing the following sentence: *Freek ging het hok eerst in en toen weer uit en zei dus het woord . . . [target]* ("Freek first went into the hut and then out again and then said the word . . . [target]"). This sentence did not favor any of the target words semantically and did not contain any /ɑ/ or /a:/ vowels. The sentence was produced multiple times at the speaker's habitual rate, ending in monosyllabic target words that either had the short vowel /ɑ/ or the long vowel /a:/. Six minimal target pairs were used: *zat–zaad* ("sat"–"seed"), *Stan–staan* ("Stan"– "stand"), *dat–daad* ("that"–"deed"), *stad–staat* ("city"–"state"), *staf–staaf* ("staff"–"bar"), and *zak–zaak* ("bag"–"shop").

From these recordings, context sentences were excised that included all speech up to target onset. One clear token (without silent pauses) near the speaker's median rate was selected and its intensity was scaled to 70 dB. This context sentence was then linearly compressed/expanded into one slow (ratio = 1.33; total duration = 4,055 ms) and one fast version (ratio = 0.75; total duration = 2,512 ms) using Pitch Synchronous Overlap and Add (PSOLA) with Praat software (Boersma & Weenink, 2016).

Target words were also excised from the recordings and one long vowel /a:/ was selected for manipulation (originating from the word *staat*). Because the Dutch /ɑ/–/a:/ contrast is cued by both spectral and temporal characteristics, a two-dimensional continuum was created from this one vowel token, comprising seven duration values and seven F2 values, all falling within the speaker's natural range. Spectral manipulations were based on Burg's linear predictive coding method (implemented in Praat), with the source and filter models estimated automatically from the selected vowel. The formant values in the filter models were inspected and adjusted to result in a constant F1 value (739 Hz, ambiguous between /ɑ/ and /a:/) and one of seven desired F2 values (1300–1600 Hz in steps of 50 Hz). Then, the source and filter models were recombined and the new vowels were adjusted to have the same overall amplitude as the original vowel. Based on these spectrally manipulated vowels, duration continua (110–170 ms in steps of 10 ms) were created using PSOLA. Durations of onset and coda consonants were equalized in duration (150 ms and 200 ms, respectively). Finally, the 49 vowel tokens were combined with the onset and coda consonants from the six target pairs, after which the intensity of each target token was scaled to 65 dB.

These target tokens were presented in isolation (i.e., without any preceding context sentences) to 26 native Dutch listeners in a categorization pretest (two-alternative forced choice). None of these participants took part in the other experiments. They indicated whether they heard the word with the short vowel /ɑ/ or the long vowel /a:/. Based on these categorization data, three vowel tokens with different F2 values but identical duration (140 ms) were selected for the following experiments, each sampling a different point from the categorization curve: Token 1, F2 = 1,300 Hz, 27% /a:/ categorization; Token 2, F2 = 1,450 Hz, 48% /a:/ categorization; and Token 3, F2 = 1,550 Hz, 67% /a:/ categorization. Only these three vowel tokens were used in the following experiments.

Finally, all target words were combined with the two (fast and slow) context sentences, with a silent interval of 75 ms in between, adding up to a total of 108 items (2 context rates × 6 target pairs × 3 vowel tokens = 36 unique stimuli, each stimulus presented three times).

Procedure. Stimulus presentation was controlled by Presentation software (version 16.5; Neurobehavioral Systems, Albany, CA). Speech stimuli were presented to half of the participants in a fixed random order, with the reversed order presented to the other half.

For purposes of comparability across experiments, visual displays were identical across all experiments (see Figure 1). Each trial started with a screen showing a (horizontal) hourglass running empty in 5 s (from right to left). Above the hourglass, the rate of the context sentence was displayed (*SNEL* "FAST" vs. *TRAAG* "SLOW"). Moreover, a mark on the hourglass indicated the time of context sentence onset: early in the case of slow contexts (945 ms after hourglass onset), late in the case of fast contexts (2,488 ms after hourglass onset). The hourglass always ran empty at context sentence offset, after which the target word followed. Participants in Experiment 1 did not receive specific instructions with respect to the hourglass, since it was irrelevant for the task in Experiment 1 (but essential in the other experiments).

At target offset, the screen was replaced by two response options and participants were instructed to indicate what sentence–final target word they had heard: *dat* or *daad*, *zak* or *zaak*, and so forth. The position of words (left or right) was counterbalanced across participants, who gave their response by pressing "1" for the word on the left side of the screen, and "0" for the word on the right side of the screen. If participants did not respond within 5 s, a missing response was recorded and the next trial was presented.

Results

Categorization data, calculated as the proportion of long vowel responses (percentage /a:/) of Experiment 1, are represented in Figure 2. The figure shows that participants reported more long vowels when the target vowel had a higher F2. Moreover, the difference between the two lines suggests that hearing a preceding context with a fast speech rate (solid line) biased listeners' perception toward /a:/. Visual inspection of the separate categorization curves of the six different target pairs did not reveal substantial variation across items.

A generalized linear mixed model (GLMM; Quené & Van den Bergh, 2008) with a logistic linking function as implemented in the lme4 library, version 1.0.5 (Bates, Maechler, Bolker, & Walker, 2015) in R (R Development Core Team, 2012), tested the binomial responses (0 = /ɑ/; 1 = /a:/) collected in Experiment 1 for fixed effects of Vowel F2 (continuous predictor, scaled and centered around the mean), Rate Condition (categorical predictor, intercept is slow), and their interaction, with crossed random effects of Participants and Items. Only by-participant random slopes for Rate Condition were included because models with more complex random effects structures failed to converge. Note that Vowel F2 was included as a continuous predictor, considering the linear nature of the underlying construct (namely, vowels' second formant frequencies), thus taking the relative distance between different measurement points into account.

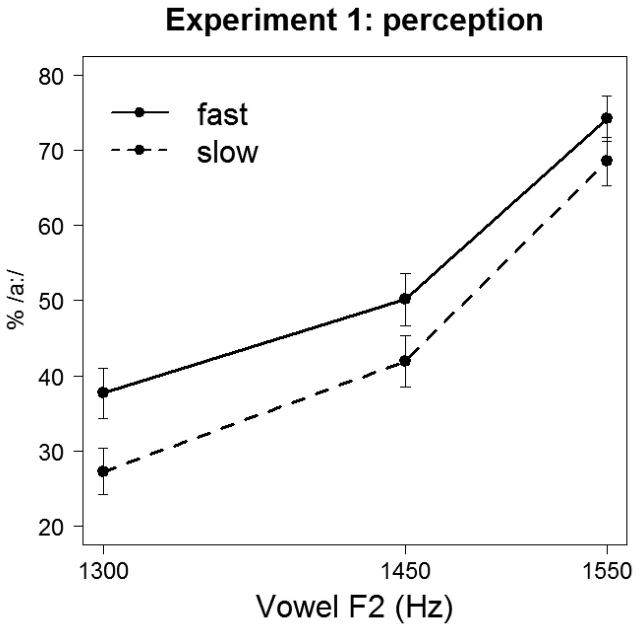
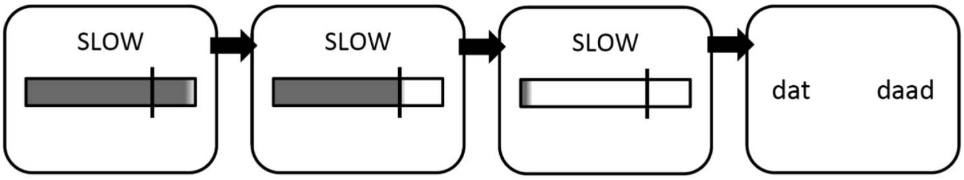


Figure 2. Average categorization data (in percentage of /a:/ responses) of Experiment 1 (perception), split by rate condition (error bars enclose 1.96 × *SE* on either side; 95% confidence intervals).

This GLMM, referred to as Model 1, revealed a significant effect of Vowel F2 ($\beta = 0.979$, $z = 17.643$, $p < 0.001$): The higher the vowel's F2, the higher the proportion of /a:/ responses. Also, a significant effect of Rate Condition was found ($\beta = 0.505$,



After preview (500 ms), the hour glass starts running empty from right to left. Total running time: 5 sec.

At the mark on the hourglass:
Expt 1: onset *recorded* context sentence
Expt 2: onset *spoken* context sentence
Expt 3: onset *silent* context sentence
Expt 4: onset *recorded self-produced* context sentence

At hourglass offset, the (*recorded/spoken/silent*) context sentence ends and immediately afterwards the target word is automatically presented.

At target offset, a response screen appears. Participants respond which sentence-final target word they heard by pressing 1 (left) or 0 (right).

Figure 1. Graphical representation of the visual display used in all experiments. Expt = experiment.

$z = 6.439$, $p < 0.001$): There was a higher proportion of /a:/ responses in the fast condition. No interaction between the two predictors was observed.

Discussion

The results of Experiment 1 demonstrate that the perception of the target vowels was influenced by the speech rate of the preceding context sentence. Thus, earlier findings of rate normalization in the literature are replicated with the present materials.

Experiment 2: Production

Experiment 2, building on Experiment 1, aimed to test whether producing fast and slow context sentences oneself may also bias perception of subsequent target vowels toward /a:/.

Method

Participants. The same participants who took part in Experiment 1 also participated in Experiment 2 ($N = 45$). Experiment 2 immediately followed Experiment 1 so that participants were familiar with the words of the context sentence, the two different speech rates, and the timing of the context sentence relative to the onset of the target word.

Procedure. Experiment 2 was identical to Experiment 1 except that participants were instructed to produce the context sentences themselves—but not the sentence–final target word. The rate at which the context sentence was to be produced could be gleaned from the rate displayed above the hourglass (*SNEL* "FAST" vs. *TRAAG* "SLOW"). Participants were instructed to imitate the rates from Experiment 1 as much as possible and to start speaking at the time point indicated by the mark on the hourglass and to finish when the hourglass ran empty. When the hourglass ran empty, the prerecorded target words from Experiment 1 were automatically presented and participants indicated by button press what target word they heard. The words of the sentence were not displayed on the screen, but had to be recited from memory. To remind participants of the exact wording, the words of the sentence were displayed on the screen after every sixth trial, but disappeared again for the next trial. Audio recordings were made of all participants' utterances.

Results

Overt speech. The self-produced context sentences from Experiment 2 were evaluated for accuracy (correct number of syllables, no hesitations, etc.) and inaccurate productions were excluded from analyses (fast trials: $n = 338$; slow trials: $n = 246$; 12% in total). Manual annotations of the total duration of the self-produced context sentences revealed that speakers indeed produced shorter sentence durations in fast trials ($M = 2{,}620$ ms, $SD = 234$ ms) than in slow trials ($M = 3{,}724$ ms, $SD = 289$ ms), $t(3539) = -124$, $p < 0.001$, although there was considerable variation within the two rate conditions (see Figure 3). Participants' timing of their speech offset relative to the onset of the target word varied ($M = 157$ ms, $SD = 270$ ms). Potential effects of participants' timing on their vowel categorization data are examined below.

Categorization data. The categorization data of Experiment 2, represented in Figure 4, look similar to the data of Experiment 1. Again, it seems that participants reported more long /a:/ vowels when the target vowel had a higher F2, and importantly, there seems to be a difference in target categorization after fast versus slow speech production. Nevertheless, the distance between the two lines representing the two rate conditions seems to be somewhat smaller.

A GLMM, in structure identical to the previous model (i.e., logistic linking function, fixed effects of Vowel F2, Rate Condition, and their interaction, with crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition), tested the binomial responses collected in Experiment 2. This Model 2 revealed a significant effect of Vowel F2 ($\beta = 1.045$, $z = 16.223$, $p < 0.001$) and of Rate Condition ($\beta = 0.187$, $z = 2.161$, $p = 0.030$), indicating a higher proportion of long /a:/ responses in the fast condition. Finally, no interaction between the two predictors was observed. These results suggest that self-produced context sentences in the fast rate condition biased the perception of subsequent targets toward longer vowels.

Note that the estimate of the Rate Condition effect in Experiment 2 ($\beta = 0.187$) is considerably smaller than the estimate of the Rate Condition effect in Experiment 1 ($\beta = 0.505$). In order to compare the effect of Rate Condition in the two experiments, the data sets from Experiment 1 and 2 were combined. These combined data were analyzed by Model 3 which was identical to Model 2, only including one additional categorical predictor Experiment (intercept is Experiment 2). This Model 3 indeed revealed an interaction between Rate Condition and Experiment ($\beta = 0.363$, $z = 3.399$, $p < 0.001$), demonstrating that the effect of preceding speech rate in Experiment 2 was reduced relative to the effect of preceding speech rate in Experiment 1.

Note that because participants in Experiment 2 produced the context sentences themselves, the time between the offset of (self-produced) contexts and the onset of (automatically presented)

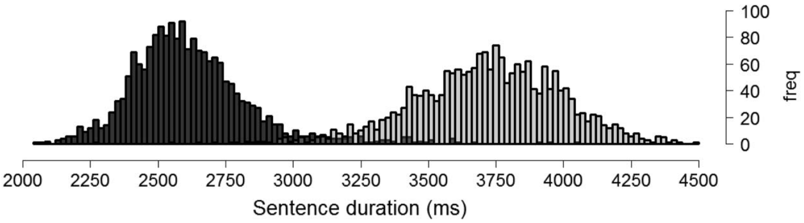


Figure 3. Two distributions of sentence durations from the fast condition (dark gray) and the slow condition (lighter gray). Only data from Experiment 2 (overt speech production) are plotted. Freq = frequency.

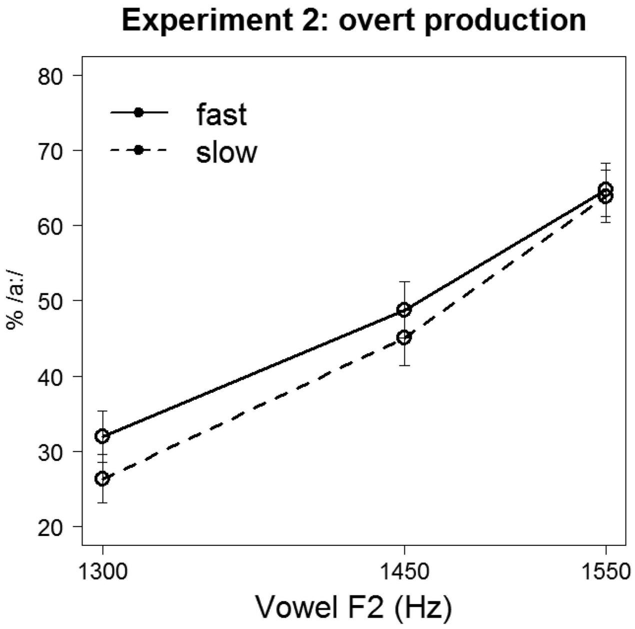


Figure 4. Average categorization data (in percentage of /a:/ responses) of Experiment 2 (overt production), split by rate condition (error bars enclose 1.96 × *SE* on either side; 95% confidence intervals).

targets varied ($M = 157$ ms, $SD = 270$ ms). In contrast, there was no timing variability in the prerecorded materials of Experiment 1 (fixed time interval of 75 ms). It has previously been argued that variability in the time window between context sentence and target may influence rate normalization effects (Newman & Sawusch, 1996; Sawusch & Newman, 2000). Therefore, the effect of timing variability (absent in Experiment 1, present in Experiment 2) on the size of the effect of preceding speech rate was investigated as potential explanation for the reduced effect of preceding speech rate in Experiment 2. Model 4 was identical to Model 2, only including one additional continuous predictor Latency, testing for effects of the time between context sentence offset and target word onset. This Model 4, however, did not reveal any effect of (or interactions with) Latency: The time between context sentence offset and target word onset could not be found to consistently influence vowel categorization.

Discussion

Experiment 2 extends our understanding of rate normalization by showing that one's own speech rate may change one's perception of another talker's utterance. When participants produced fast speech prior to target presentation, they perceived the target vowel durations as longer than when they were talking at a slow rate.

The combined analysis of the data from Experiment 1 and 2 revealed that rate normalization elicited by speech production is reduced relative to rate normalization elicited by speech perception. In-depth analyses of the data suggest that this reduction could not be attributed to larger variability in the time interval between context sentence offset and target word onset in Experiment 2.

However, another potential explanation for the reduced effect of preceding speech rate in Experiment 2 may be related to rate variability within the two rate conditions. As shown in Figure 3, there was considerable speech rate variation within the two rate conditions in the overt speech of Experiment 2. In contrast, there was no variation in speech rates (within the two rate conditions) in Experiment 1. The larger variability in the self-produced context sentences from Experiment 2 may potentially explain the reduced effect of Rate Condition in Experiment 2. This potential account of the reduced effect of preceding speech rate in Experiment 2 will be returned to in Experiment 4.

Experiment 3: Covert Production

Experiment 3 tested what mechanism may account for the self-induced rate normalization found in Experiment 2. First, self-induced rate normalization may be explained by perception mechanisms through self-monitoring of the external (self-produced) speech signal. Thus, self-induced rate normalization would operate similarly as "typical" rate normalization induced by speech from another talker. Alternatively, self-monitoring of the internal signal, suggested by some to involve forward models of the speech to be produced (e.g., Pickering & Garrod, 2013; Tian & Poeppel, 2010), may already be sufficient to elicit rate normalization. In order to disentangle the contributions of these potential mechanisms, Experiment 3 tested whether covert speech production at fast and slow rates may also induce rate normalization.

Method

Participants. The same participants that took part in Experiment 1 and 2 also participated in Experiment 3 ($N = 45$). Experiment 3 immediately followed Experiment 2 so that participants were already familiar with the experimental task.

Procedure. The task in Experiment 3 was identical to the task in Experiment 2, except that participants were now instructed to produce the context sentences covertly. Specifically, they were told to produce the context sentences "in their heads," without any audible speech or any articulatory movements, which was assessed by means of audio and video recordings. Since post hoc accuracy assessment of covert signals is impossible, participants themselves indicated after each trial whether they had succeeded in producing the sentence correctly (i.e., correct words, correct rate, no pauses or "uhm"s, etc.) by pressing "Y" or "N."

Results

Covert speech. To assess participants' task compliance, video and audio recordings of the time period in which participants were expected to covertly produce the context sentences were inspected. Trials with audible speech or visible articulatory movements were excluded from analyses (fast trials: $n = 94$; slow trials: $n = 94$; 4% in total). Also, trials in which the participant had reported to have failed to produce the sentence correctly were excluded from analyses (fast trials: $n = 338$; slow trials: $n = 246$; 13% in total).

Categorization data. The data from Experiment 3, calculated as the proportion of /a:/ responses, are represented in Figure 5. This figure suggests that participants were only sensitive to the spectral characteristics (F2) of the target vowels, without any difference between the two rate conditions. Model 5, a GLMM testing for effects of Vowel F2, Rate Condition, and their interac-

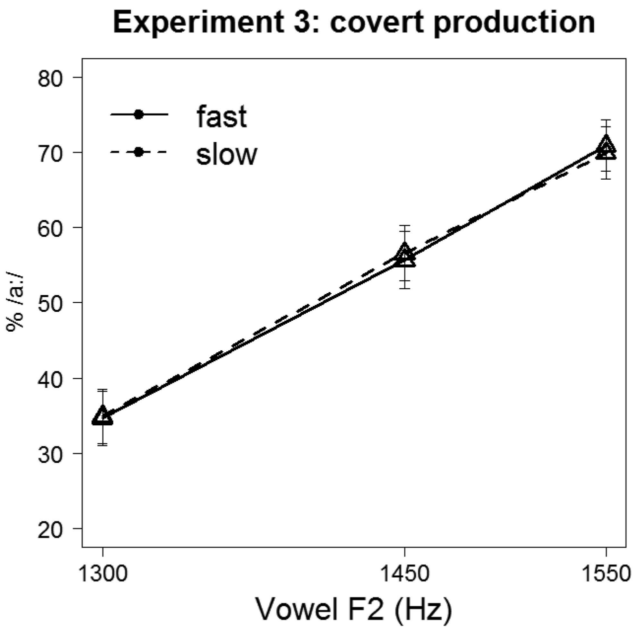


Figure 5. Average categorization data (in percentage of /a:/ responses) of Experiment 3 (covert production), split by rate condition (error bars enclose 1.96 × *SE* on either side; 95% confidence intervals).

tion (with crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition), only found an effect of Vowel F2 ($\beta = 1.047$, $z = 15.931$, $p < 0.001$). No effect of Rate Condition was observed, nor an interaction between Vowel F2 and Rate Condition.

The data from all experiments so far were entered into another analysis to be able to test for differences between experiments in the effect of preceding speech rate on categorization. Model 6 tested the combined dataset on the predictors Vowel F2, Rate Condition, and Experiment (categorical predictor with three levels; intercept is Experiment 3), and all their interactions. This model included crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition.

Model 6 found an effect of Vowel F2 ($\beta = 0.888$, $z = 15.399$, $p < 0.001$) but no effect of Rate Condition (nor an interaction between Vowel F2 and Rate Condition). Note that Experiment 3 was mapped onto the intercept of the predictor experiment so the absence of an effect of preceding speech rate is in line with the previous analysis (Model 5). Interactions between Rate Condition and the other two experiments (Rate Condition and Experiment 1: $\beta = 0.584$, $z = 5.360$, $p < 0.001$; and Rate Condition and Experiment 2: $\beta = 0.217$, $z = 1.942$, $p = 0.05$) suggest that the effect of preceding speech rate was significantly stronger in Experiments 1 and 2, relative to Experiment 3 (where no effect of preceding speech rate was observed).

Discussion

Experiment 3 failed to find evidence for contextual effects of covertly produced speech rate on the perception of overt target speech. Note, however, that it is impossible to assess participants' accuracy of covert speech production or participants' task compliance. Therefore, caution should be taken not to disregard the contribution of production processes to the observed effect of self-produced speech rate in Experiment 2, especially given earlier evidence for influences of covert speech production on overt speech perception (Sams et al., 2005; Sato et al., 2013). Nonetheless, given the current results, it is most likely that the effects of our own speech rate found in Experiment 2 arose primarily through monitoring of the external speech signal (i.e., similar to how rate normalization induced by another talker's speech rate operates).

Finally, let us return to the observation in Experiment 2 that the effect of (self-produced) preceding speech rate seemed to be reduced (relative to the effect of perceived speech rate in Experiment 1). It was argued that this reduction may potentially be attributed to larger rate variability within the two rate conditions in Experiment 2 (cf. Figure 3), compared to Experiment 1. In order to test whether this acoustic variation is to be held responsible for the reduced effect of preceding speech rate in Experiment 2, participants were invited back to the lab to passively listen to the recordings of their own voice from Experiment 2. In such a situation, the temporal characteristics of the speech materials are identical to those in Experiment 2, meaning that any potential difference in the size of the rate normalization effect cannot be attributed to the speech signal.

Experiment 4: Self-Perception

In Experiment 4, participants were invited back to the lab to passively listen to the speech they had produced themselves in Experiment 2. This guaranteed that the temporal characteristics of the speech in Experiment 2 and 4 were identical, allowing for proper comparison between the effect of preceding speech rate in Experiment 2 (self-production) and Experiment 4 (self-perception).

Method

Participants. The same participants who had taken part in the previous experiments were invited back to the lab for Experiment 4. Unfortunately, only data from 23 participants could be obtained; other participants were unable to come back to the lab. The experimental sessions of Experiment 4 were run several weeks after the previous experiments to allow for the annotation and preparation of participants' self-produced context sentences.

Procedure. Experiment 4 used the self-produced speech materials from Experiment 2. All characteristics of the auditory stimuli of Experiment 2 were maintained (e.g., onset of context sentences, onset of target words, order of presentation), thus replicating the exact situation of Experiment 2. To control for loudness, all self-produced context sentences were scaled to 70 dB and manipulated targets to 65 dB, similar to Experiment 1. Participants were told that they would hear their own (fast and slow) context sentences, followed by a target word which they had to categorize. Thus, the task was similar to Experiment 1 (passive listening) but used the participant-specific self-produced materials from Experiment 2.

Results

The categorization data of Experiment 4, represented in Figure 6, look similar to the data of Experiment 1. As expected, participants reported more long /a:/ vowels when the target vowel had a higher F2. Similar to the previous experiments, again there would seem to be a difference in target categorization for the two rate conditions. Moreover, the distance between the two lines representing the two rate conditions seems to be similar to that observed in Experiment 1.

The data from Experiment 4 were analyzed using a GLMM with a similar structure as the model used for Experiment 2 (i.e., logistic linking function, fixed effects of Vowel F2, Rate Condition, and their interaction, with crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition). This Model 7 revealed a significant effect of Vowel F2 ($\beta = 1.088$, $z = 12.375$, $p < 0.001$) and also of Rate Condition ($\beta = 0.508$, $z = 3.595$, $p < 0.001$), indicating a higher proportion of /a:/ responses in the fast condition.

In order to compare the different sizes of the effects of preceding speech rate in the different experiments, the data from Experiments 1, 2, and 4 were combined into a larger dataset. A new GLMM, Model 8, was built that included effects of Vowel F2, Rate Condition, Experiment (categorical predictor with three levels, intercept is Experiment 4), and all their interactions. This model, again, included crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition. This analysis found effects of Vowel F2 ($\beta = 1.067$, $z = 12.737$, $p < 0.001$), and importantly, of Rate Condition ($\beta = 0.511$, $z = 4.126$, $p < 0.001$). Since Experiment 4 was mapped onto the intercept of the predictor experiment, the presence of an effect of preceding speech rate is in line with the previous analysis (Model 7). Furthermore, only one interaction was observed, namely between Rate Condition and Experiment 2 ($\beta = -0.344$, $z = -2.451$, $p = 0.014$), indicating that the effect of preceding speech rate in Experiment 2 was reduced relative to Experiment 4. No interaction between Rate Condition and Experiment 1 was found, suggesting that there was no evidence for the effect of preceding speech rate in Experiment 1 to be significantly different from that in Experiment 4.

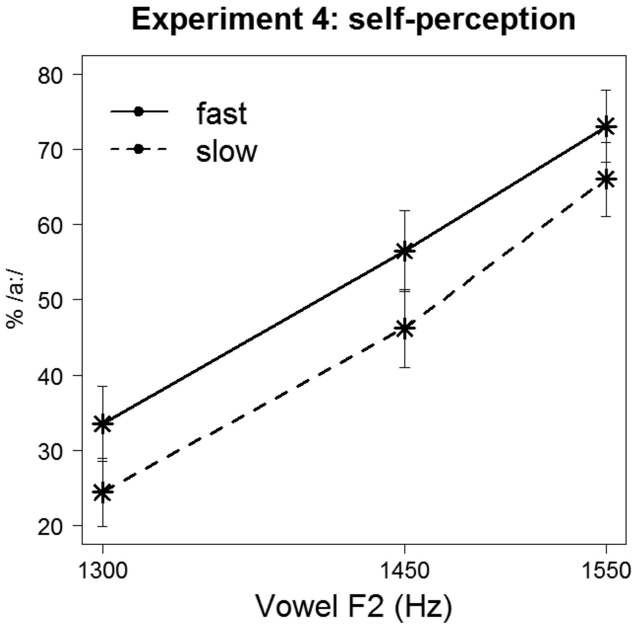


Figure 6. Average categorization data (in percentage of /a:/ responses) of Experiment 4 (self-perception), split by rate condition (error bars enclose 1.96 × *SE* on either side; 95% confidence intervals).

Discussion

In Experiment 4, participants passively listened to the speech they had produced earlier in Experiment 2. Despite the fact that the temporal characteristics of the speech presented in Experiment 4 were identical to those in Experiment 2, results showed that the effect of preceding speech rate was larger in Experiment 4 than in Experiment 2. This would suggest that the additional task of speech production—which uniquely differentiated participants' tasks in Experiment 2 from Experiment 4—may be responsible for the reduced effect of preceding speech rate observed in Experiment 2.

However, the stronger effect of preceding speech rate in Experiment 4 (compared to Experiment 2) may also be explained by individual differences in producing fast and slow speech. Perhaps Experiment 4 included precisely those participants that could very successfully produce very slow and very fast speech in Experiment 2. If so, then one would indeed expect a stronger effect in Experiment 4 relative to Experiment 2 simply because the speech materials in Experiment 4 contained greater separation between (really) slow and (really) fast speech. However, inspection of the produced context sentences used in Experiment 4 (i.e., of those participants from Experiment 2 who would also participate in Experiment 4) revealed that their sentence durations were comparable to the speech produced by the total sample of participants (fast trials: $M_{\text{subset}} = 2{,}603$ ms, $SD_{\text{subset}} = 264$ ms, $M_{\text{all}} = 2{,}620$ ms, $SD_{\text{all}} = 234$ ms; slow trials: $M_{\text{subset}} = 3{,}695$ ms, $SD_{\text{subset}} = 303$ ms, $M_{\text{all}} = 3{,}724$ ms, $SD_{\text{all}} = 289$ ms; cf. Figure 7), challenging this alternative explanation.

Finally, another alternative explanation may be related to the chronological order in which participants took part in the different experiments. In order to familiarize participants with the words of the context sentence, the two different speech rates, and the timing of the context sentence relative to the target word, the first experiments of the present study adopted a within-participants design, reducing error variance. That is, in a single experimental session, the same participants first participated in Experiment 1, then in Experiment 2, and then in Experiment 3. The disadvantage of this particular design is that it cannot exclude the possibility that the experimental order, either through fatigue or familiarity with the experimental stimuli, reduced the effect of preceding speech rate observed in Experiment 2 and eliminated the effect altogether in Experiment 3. In comparison, Experiment 4 was run several weeks later, meaning that participants came into the lab refreshed, possibly explaining the larger effect of preceding speech rate in Experiment 4.

In order to investigate whether the observed reduction (in Experiment 2) and the observed elimination (in Experiment 3) of the effect of preceding speech rate is due to experimental order, two new experiments were conducted. These new experiments adopted

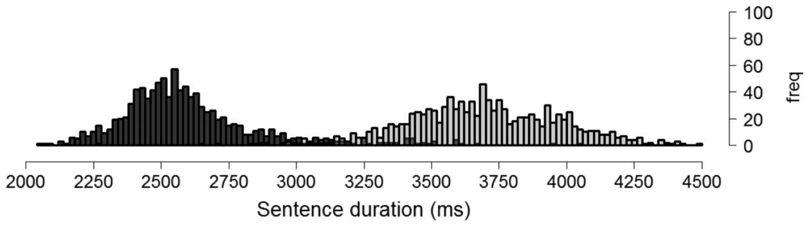


Figure 7. Two distributions of sentence durations from the fast condition (dark gray) and the slow condition (lighter gray). This bar plot only gives the data from those participants in Experiment 2 who would also participate in Experiment 4. Freq = frequency.

a between-participants design: Two new participant samples were recruited, one for each experiment. Each new participant either took part in Experiment 5 or in Experiment 6, and had not participated in any of the previous four experiments. Participants in Experiment 5 received the overt production task (identical to Experiment 2) and participants in Experiment 6 received the covert production task (identical to Experiment 3).

Experiment 5: Overt Production

Method

Participants. A sample of 25 native Dutch participants with normal hearing were recruited from the participant pool of the Max Planck Institute to take part in Experiment 5. All participated with informed consent as approved by the ethics committee of the Social Sciences Department of Radboud University (Project code ECSW2014-1003–196). None of these participants had participated in any of the other experiments in this study. Data from 5 participants were excluded for reasons of technical failures or noncompliance, leaving data from 20 participants (3 males, M_{age} = 22 years) for analysis.

Procedure. The entire procedure of Experiment 5 (the task, stimulus materials, visual screens, etc.) was identical to that of Experiment 2. Participants were to overtly produce the context sentence "*Freek ging het hok eerst in en toen weer uit en zei dus het woord* . . ." at a fast rate and a slow rate, as indicated by a horizontal hourglass, after which the target words were played automatically. Audio recordings were made of their performance.

Because participants in Experiment 5 had not participated in any of the other experiments, they received eight "perception" practice trials where they just listened to the fast and slow context sentences, as in Experiment 1. Thus they were familiarized with the fast and slow rate, the words of the sentence, and the timing of the onset and offset of the context sentence relative to the target word. After another six "production" practice trials (where they could practice self-production of the context sentence), the actual experimental session started. In order to account for the greater error variance due to the between-participants design, the number of items in Experiment 5 was doubled relative to Experiment 2 (i.e., more trial repetitions; 216 items in total).

Results

Overt speech. Similar to Experiment 2, inaccurate productions (hesitations, incorrect words, etc.) were excluded from analyses (fast trials: n = 215; slow trials: n = 147; 8% in total). Manual annotations of the total duration of the self-produced context sentences revealed that speakers indeed produced shorter sentence durations in fast trials (M = 2,454 ms, SD = 213 ms) than in slow trials (M = 3,810 ms, SD = 358 ms), $t(4314) = -151$, $p < 0.001$). Participants' timing of their speech offset relative to the onset of the target word varied (M = 164 ms, SD = 250 ms).

Categorization data. The categorization data of Experiment 5, represented in Figure 8, look similar to the data of Experiment 2. Again, it seems that participants reported more long vowels when the target vowel had a higher F2, and importantly, there also seems to be a difference in target categorization after fast versus slow speech production. Again, however, the distance between the two lines representing the two rate conditions seems to be somewhat smaller compared to Experiment 1 (as was also observed for Experiment 2).

A GLMM with logistic linking function tested the binomial responses (0 = /ɑ/; 1 = /a:/) collected in Experiment 5. This Model 9 included fixed effects of Vowel F2 (continuous predictor, scaled and centered around the mean), Rate Condition (categorical predictor;

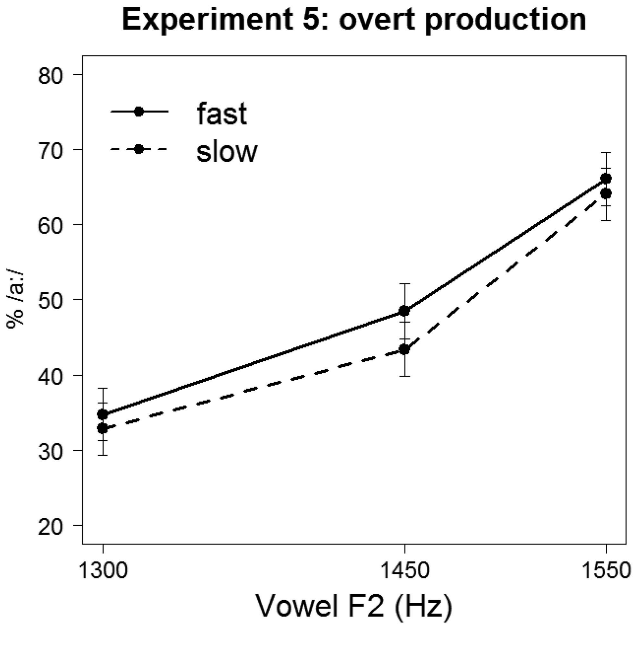


Figure 8. Average categorization data (in percentage of /a:/ responses) of Experiment 5 (overt production), split by rate condition (error bars enclose 1.96 × *SE* on either side; 95% confidence intervals).

intercept is slow), and their interaction. It also included crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition. Model 9 revealed a significant effect of Vowel F2 ($\beta = 0.986$, $z = 14.741$, $p < 0.001$) and also of Rate Condition ($\beta = 0.194$, $z = 2.023$, $p = 0.043$), indicating a higher proportion of long vowel responses in the fast condition. Finally, no interaction between the two predictors was observed. These results suggest that, in Experiment 5, producing context sentences at a fast rate biased the perception of subsequent targets toward long vowels.

Participants in Experiment 5 (between-participants design) received twice as many trials as participants in Experiment 2 (within-participant design) in order to increase statistical power. To examine whether there was any effect of fatigue or familiarity with the stimuli within Experiment 5, Model 9 was extended to include the predictor Split Half (categorical predictor, trials in the first half of the experimental session coded as 0, trials in the second half coded as 1), together with interactions with all other predictors. This extended model did not reveal an interaction between Rate Condition and Split Half, suggesting that there was no variation in the effect of preceding speech rate within Experiment 5.

Another GLMM compared the effect of preceding speech rate in Experiment 5 (overt production) to the effect observed in Experiment 1 (perception). Model 10 tested the combined data from both experiments on the predictors Vowel F2, Rate Condition, and Experiment (categorical predictor; intercept is Experiment 5), and all their interactions. It also included crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition. Model 10 revealed a statistically significant interaction between Rate Condition and Experiment ($\beta = 0.310$, $z = 2.610$, $p = 0.009$), suggesting that the effect of preceding speech rate was significantly larger in Experiment 1 (relative to Experiment 5). A similar analysis comparing Experiment 5 to Experiment 2 revealed no interaction between Rate Condition and Experiment, revealing that there was no evidence for different effects of preceding speech rate across Experiments 2 and 5.

Discussion

Experiment 5 replicates the findings from Experiment 2 by showing that one's own speech rate may change one's perception of another talker's utterance. Also, the combined analysis of Experiment 1 and Experiment 5 shows that rate normalization elicited by speech production is reduced relative to rate normalization elicited by speech perception. Because Experiment 5 tested a new participant sample, this reduction in the effect of preceding speech rate cannot be attributed to fatigue or familiarity effects. Moreover, it is unlikely that the reduction in the effect of preceding speech rate was due to fatigue or familiarity effects within Experiment 5, since no difference was found between the effect of preceding speech rate in the first half versus the second half of the experiment.

Experiment 6: Covert Production

Method

Participants. A sample of 23 native Dutch participants with normal hearing were recruited from the participant pool of the Max Planck Institute to take part in Experiment 6. All participated with informed consent as approved by the ethics committee of the Social Sciences Department of Radboud University (Project code ECSW2014-1003–196). None of these participants had participated in any of the other experiments reported in this study. Data from 3 participants were excluded for reasons of technical failures or noncompliance, leaving data from 20 participants (5 males, $M_{\text{age}} = 24$ years) for analysis.

Procedure. The entire procedure of Experiment 6 (the task, stimulus materials, visual screens, etc.) was identical to that of Experiment 3. Participants were to covertly produce the context sentence "*Freek ging het hok eerst in en toen weer uit en zei dus het woord* . . ." at a fast rate and a slow rate, as indicated by the horizontal hourglass, after which the target words were played automatically. Instructions were to produce the sentence "in your head" without audible sound or articulatory movements. Similar to Experiment 5, participants first received eight "perception" practice trials where they just listened to the fast and slow context sentences (cf. Experiment 1). Thus they were familiarized with the fast and slow rates, the words of the sentence, and the timing of the onset and offset of the context sentence relative to the target word. After another six "covert production" practice trials (where participants could practice covert production of the context sentence), the actual experimental session started. In line with Experiment 5, error variance due to the between-participants design was reduced by increasing the number of items in Experiment 6 (216 items in total).

Results

Covert speech. To assess participants' task compliance, video and audio recordings of the time period in which participants were expected to covertly produce the context sentences were inspected. Trials with audible speech or visible articulatory movements were excluded from analyses (fast trials: $n = 97$; slow trials: $n = 112$; 5% in total). Also, trials in which the participant had reported to have failed to produce the sentence correctly were excluded from analyses (fast trials: $n = 170$; slow trials: $n = 137$; 8% in total).

Categorization data. The data from Experiment 6, calculated as the proportion of /a:/ responses, are represented in Figure 9. This figure suggests that participants were only sensitive to the spectral characteristics (F2) of the target vowels, without any difference between the two rate conditions. Model 11, a GLMM testing for effects of Vowel F2 and Rate Condition, and their interaction (crossed random effects of Participants and Items, and by-participant random slopes for Rate Condition), only found an effect of Vowel F2 ($\beta = 1.449$, $z = 17.073$, $p < 0.001$). No effect of Rate Condition was observed, nor an interaction between Vowel F2 and Rate Condition.

Similar to the analyses in Experiment 5, it was examined whether there was any effect of fatigue or familiarity with the stimuli within Experiment 6. Model 11 was extended to include the predictor Split Half (categorical predictor, trials in the first half of the experimental session coded as 0, trials in the second half coded as 1), together with interactions with all other predictors. This extended model did not reveal an interaction between Rate Condition and Split Half, suggesting that there was no order effect within Experiment 6.

In order to show that the effect of preceding speech rate was eliminated in Experiment 6, Model 12 compared the data from

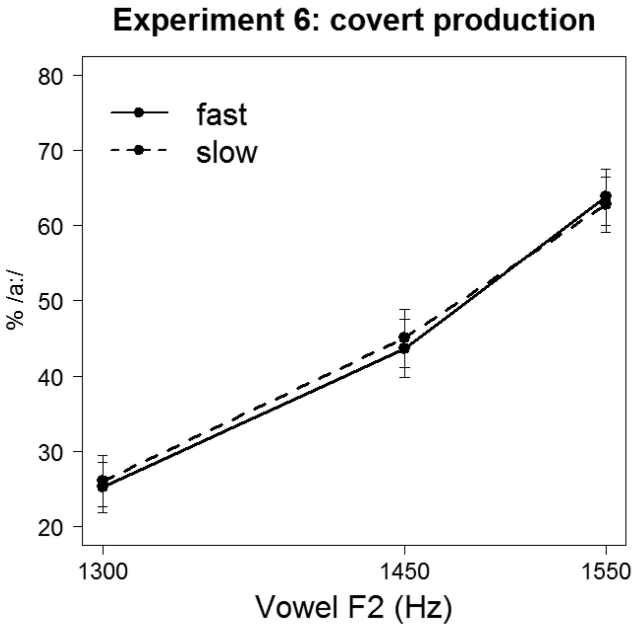


Figure 9. Average categorization data (in percentage of /a:/ responses) of Experiment 6 (covert production), split by rate condition (error bars enclose 1.96 × *SE* on either side; 95% confidence intervals).

Experiment 6 to the data from Experiment 1. Model 12 tested the combined dataset on the predictors Vowel F2, Rate Condition, and Experiment (categorical predictor, intercept is Experiment 6), and all their interactions (crossed random effects of Participants and Items, by-participant random slopes for Rate Condition).

Model 12 found an effect of Vowel F2 ($\beta = 1.425$, $z = 17.265$, $p < 0.001$) but no effect of Rate Condition. Note that Experiment 6 was mapped onto the intercept of the predictor experiment so the absence of an effect of preceding speech rate is in line with the previous analysis (Model 11). The interaction between Rate Condition and Experiment ($\beta = 0.497$, $z = 3.726$, $p < 0.001$) demonstrates that the effect of preceding speech rate was significantly larger in Experiment 1, compared to Experiment 6 (where no effect of preceding speech rate was observed). A similar analysis comparing Experiment 6 to Experiment 3 revealed no interaction between Rate Condition and Experiment, revealing that there was no evidence for differences between Experiments 3 and 6.

Discussion

Experiment 6 demonstrated that when participants covertly produce fast and slow context sentences, the rate of covert production has no effect on subsequent target categorization. This finding is in line with the outcomes of Experiment 3.

To conclude, Experiment 5 and Experiment 6 replicate the findings from Experiment 2 and Experiment 3 (respectively) with a between-participants design. Given this outcome, the reduced effect of preceding speech rate in Experiments 5 and 6 cannot be attributed to fatigue or familiarity effects, since both experiments recruited new samples of participants who had not participated in any of the other experiments in this study. Moreover, it is unlikely that the reduced effects of preceding speech rate in Experiments 5 and 6 were due to fatigue or familiarity effects within these experiments, since no difference were found when comparing the first half to the second half of the experiment.

General Discussion

This study tested whether characteristics of one's own voice (here: speech rate) may influence one's perception of other talkers. Experiment 1 replicated standard findings from earlier rate normalization studies: Hearing a fast context sentence biased perception toward long target vowels. Experiment 2, together with Experiment 5, extended this effect to self-produced speech: Talking at a fast rate was also found to bias perception of subsequent target words toward long vowels. And Experiment 3, together with Experiment 6, indicated that covert speech production at a faster rate did not bias perception toward long vowels.

Note that the absence of an effect of preceding speech rate in Experiments 3 and 6 does not necessarily entail that inner speech is underspecified for speech rate. In fact, several studies support the claim that inner speech is a temporal signal (Anderson, 1982; Dell & Oppenheim, 2015; Dell & Repka, 1992; Mackay, 1981; Weber & Castleman, 1970). Moreover, inner speech seems to be produced at a similar rate as overt speech (Netsell et al., 2010). Finally, one should be careful not to draw conclusions from null results since they do not easily lend themselves for proper interpretation. The findings of Experiments 3 and 6 (only) suggest that the effect of one's own speech rate in Experiment 2 most likely arose through self-monitoring of the external speech signal. That is, the overt speech signal seems to be necessary to elicit rate normalization (cf. a similar dissociation between covert and overt speech in eliciting repetition reduction; Jacobs, Yiu, Watson, & Dell, 2015).

This observation may be interpreted as arguing against gestural accounts of rate normalization (e.g., Fowler, 1990, 1991; Miller & Liberman, 1979). These accounts hold that rate normalization is the result of listeners retrieving the speaker's speech rate. During covert speech production at different rates, the intended rate is available to the speaker since it is under his or her own control. Therefore, gestural accounts would hypothesize that this covert rate should also influence the perception of subsequent target words. This, however, was not found to be the case, challenging a gestural interpretation of rate normalization findings.

The present finding that the overt speech signal plays a central role in rate normalization rather supports general auditory accounts of rate normalization, such as durational contrast (Wade & Holt, 2005) and neural entrainment (Bosker, 2016; Peelle & Davis, 2012). For instance, it may be argued that self-perception of self-produced speech induces durational contrast in a similar way as perception of speech from other talkers does. In the same vein, one may argue that neural oscillators phase-lock to self-produced sensory signals as much as to external signals. The present data do not discriminate between different general auditory accounts of rate normalization; further (neuroimaging) investigations are required for that purpose.

Even though Experiment 2 showed that one's own speech rate influences one's perception of another talker's utterance, the effect of (self-produced) speech rate in Experiment 2 was found to be significantly reduced when compared to a situation where participants passively listened to their own fast and slow speech (in

Experiment 4). Further analyses revealed that the reduced effect of preceding speech rate could not be attributed to the within-participant design used in Experiments 1–4 (since the same patterns were observed in Experiments 5–6 using a between-participants design) or fatigue/familiarity effects within Experiments 5–6 (no order effects were observed). Another possibility might be that participants in the production experiments (Experiments 2, 3, 5, and 6) experienced increased cognitive load (relative to Experiments 1 and 4) due to the fact that they were required to memorize and recite the context sentence, and monitor their own speech for speech errors to be reported after each trial. However, a recent study by Bosker et al. (2016) has shown that rate normalization effects are not modulated (cf. Experiments 2 and 5), let alone eliminated (cf. Experiments 3 and 6), by increasing cognitive load (through a secondary visual search task in Bosker et al., 2016). Thus, the findings of Bosker et al. (2016) cast doubt on an account that attributes the reduced effect of preceding speech rate, observed in the current study, solely to cognitive load.

Interestingly, comparing Experiment 2 (producing fast and slow speech) to Experiment 4 (listening passively to one's own fast and slow speech) reveals that the difference in the size of the effect of preceding speech rate cannot be attributed to the speech materials used, since the temporal characteristics of the speech materials were identical in Experiments 2 and 4. Instead, this difference suggests that the processes involved in speech production may be responsible for the reduced effect of preceding speech rate observed in Experiment 2.

One potential explanation for the reduced effect of preceding speech rate during self-production may be found in the neurocognitive literature. This literature has established that the neural response to perception during production (i.e., hearing one's own voice while speaking) differs from the neural response to perception without production (i.e., passive listening to recordings of your own voice). In particular, activity in the auditory cortex in response to self-produced speech is attenuated relative to hearing tape-recorded speech (known as speaking-induced suppression; Houde, Nagarajan, Sekihara, & Merzenich, 2002). This attenuation has been attributed to internal forward models that simulate the sensory consequences of speech motor actions (Houde & Nagarajan, 2011). Moreover, auditory responses during speech production are not only significantly inhibited, but have also been found to be slightly delayed (Numminen & Curio, 1999).

One may interpret the present findings in light of this neurocognitive literature. For instance, one may speculate that the processing of one's own speech rate during speech production (cf. Experiment 2) is attenuated relative to the processing of one's own speech rate during passive listening (cf. Experiment 4). As such, speaking-induced suppression may attenuate listeners' sensitivity to their own speech rate during speech production, thus reducing the influence of this signal on subsequent target words. Further investigations, involving electrophysiological and neuroimaging methods, may shed light on the cognitive and neural mechanisms behind the reduced effect of self-produced speech rate.

Regardless of this reduced effect size, the present study is rather unique in finding effects of our own voice on speech perception. Even though we are repeatedly exposed to the sound of our own voice, earlier studies have only provided equivocal evidence for effects of listeners' own speech rate on explicit rate judgments (Koreman, 2006; Schwab, 2011). These studies concerned effects of the listeners' habitual speech rate. By contrast, the current experiments tested more local effects of self-produced fast and slow context sentences, and showed that one's own speech may indeed influence the perception of a following utterance, spoken by another talker. However, since only local effects of self-produced speech rate were tested, the current data do not tell us whether habitually slow speakers will perceive the same speech signal differently from habitually fast speakers. Especially given recent evidence for the tracking of habitual speech rate as a speaker-specific property (Reinisch, 2016b), this remains an intriguing question for further experimentation.

The finding that talking at a fast pace changes our perception of a subsequent utterance carries implications for our understanding of speech perception, and communication in dialogue. The ubiquity of the sound of our own voice implies that it forms a considerable part of the context in which speech from other speakers occurs. Moreover, speech rate varies considerably both between individuals (Jacewicz, Fox, & Wei, 2010; Quené, 2008) and within a given talker, for instance, depending on age (Quené, 2013), the length of utterances (Quené, 2008), conversational register, emotion, and so forth Given this large-scale variation, the fact that our own speech rate production bears consequences for speech perception may be seen as a substantial source of variation in speech comprehension and word recognition.

In fact, production may even hurt perception in a situation of spoken communication between interlocutors with highly divergent speech rates, with Talker A interpreting the speech of Talker B relative to his or her own divergent speech rate. Of course, top-down information, such as semantic context, may help to avoid misinterpretation of the spoken signal. However, in the absence of such information, comprehension, and hence communication, would be facilitated if interlocutors converged in their speech rates, thus minimizing the interference from their own speech rate. This reasoning is relevant for the study of phonetic convergence, the phenomenon that interlocutors tend to align on phonetic and prosodic features of their speech, such as speech rate (Bell, Gustafson, & Heldner, 2003; Finlayson, Lickley, & Corley, 2012; Giles, Coupland, & Coupland, 1991; Jungers & Hupp, 2009; but see conflicting evidence in the work of Pardo, 2010; Pardo, Gibbons, Suppes, & Krauss, 2012; Pardo et al., 2013; Pardo, Jay, & Krauss, 2010). While the benefits of phonetic convergence have consistently been sought in the social domain (reducing social distance and facilitating social integration, approval, and conformity; Giles et al., 1991; Natale, 1975; Pardo et al., 2010), the present study would suggest a novel function, namely to serve speech comprehension at a phonetic level. This view is in line with findings that people tend to prefer speakers who talk at a rate similar to their own (Street, Brady, & Putman, 1983), and with findings that phonetic convergence promotes comprehensibility (Berger & Roloff, 1980; Giles & Powesland, 1975). Therefore, phonetic convergence on speech rate may not only provide social advantages but may also reduce adverse effects of one's own (divergent) speech rate on the comprehension of one's interlocutor.

Conclusion

This study demonstrated that one's own speech rate can influence the perception of speech produced by another talker. This effect of one's own voice was shown to operate most likely through self-monitoring of the external speech signal. The effect of self-produced speech rate was found to be reduced relative to hearing another talker's speech rate, and an explanation in terms of speaking-induced suppression was formulated. Since temporal characteristics of our own voice may affect our perception of others, dialogic communication may be facilitated when talkers converge toward their interlocutor's speech rate.

References

Anderson, R. E. (1982). Speech imagery is not always faster than visual imagery. *Memory & Cognition, 10,* 371–380. http://dx.doi.org/10.3758/BF03202429

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25,* 1546–1553. http://dx.doi.org/10.1177/0956797614533705

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human–computer interaction. In M. J. Sole, D. Recasens, & J. Romero (Eds.), In *Proceedings of the 15th International Congress of Phonetic Sciences 2003 (ICPhS-15), Barcelona* (Vol. 3, pp. 833–836). Paris, France: International Phonetic Association.

Berger, C. R., & Roloff, M. E. (1980). Social cognition, self-awareness, and interpersonal communication. In B. Dervin & M. J. Voight (Eds.), *Progress in communication sciences* (Vol. 2, pp. 1–49). Norwood, NJ: Ablex.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program], Version 5.4.12. Retrieved from http://www.praat.org/

Bosker, H. R. (2016). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*. Advance online publication. http://dx.doi.org/10.3758/s13414-016-1206-4

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T. J. M., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30,* 159–175. http://dx.doi.org/10.1177/0265532212455394

Bosker, H. R., & Reinisch, E. (2015). Normalization for speechrate in native and nonnative speech. In M. Wolters, J. Livingstone, B. B. R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences 2015 (ICPhS XVIII), Glasgow*. Paris, France: International Phonetic Association.

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (in press). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*. http://dx.doi.org/10.1016/j.jml.2016.12.002

Dell, G. S., & Oppenheim, G. M. (2015). Insights for speech production planning from errors in inner speech. In M. Redford (Ed.), *The handbook of speech production* (pp. 404–418). Boston, MA: Wiley-Blackwell. http://dx.doi.org/10.1002/9781118584156.ch18

Dell, G. S., & Repka, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error* (pp. 237–262). Boston, MA: Springer. http://dx.doi.org/10.1007/978-1-4899-1164-3_10

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664–1670. http://dx.doi.org/10.1177/0956797610384743

Finlayson, I., Lickley, R. J., & Corley, M. (2012). Convergence of speech rate: Interactive alignment beyond representation. In D. Bradley, E. Fernández, and J. D. Fodor (Eds.), In *Proceedings of the 25th CUNY Conference on Human Sentence Processing* (p. 24). New York, NY: City University of New York (CUNY).

Fowler, C. A. (1990). Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America, 88,* 1236–1249. http://dx.doi.org/10.1121/1.399701

Fowler, C. A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America, 89,* 2910–2915. http://dx.doi.org/10.1121/1.400729

Giles, H., Coupland, N., & Coupland, I. (Eds.). (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511663673

Giles, H., & Powesland, P. F. (1975). *Speech style and social evaluation*. New York, NY: Academic Press.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience, 5,* 82.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience, 14,* 1125–1138. http://dx.doi.org/10.1162/089892902760807140

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America, 128,* 839–850. http://dx.doi.org/10.1121/1.3459842

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language, 84,* 37–48. http://dx.doi.org/10.1016/j.jml.2015.05.004

Jungers, M. K., & Hupp, J. M. (2009). Speech priming: Evidence for rate persistence in unscripted speech. *Language and Cognitive Processes, 24,* 611–624. http://dx.doi.org/10.1080/01690960802602241

Koreman, J. (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America, 119,* 582–596. http://dx.doi.org/10.1121/1.2133436

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104. http://dx.doi.org/10.1121/1.1908694

Mackay, D. G. (1981). The problem of rehearsal or mental practice. *Journal of Motor Behavior, 13,* 274–285. http://dx.doi.org/10.1080/00222895.1981.10735253

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457–465. http://dx.doi.org/10.3758/BF03213823

Natale, M. (1975). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills, 40,* 827–830. http://dx.doi.org/10.2466/pms.1975.40.3.827

Netsell, R., Ashley, E., & Bakker, K. (2010). The inner speech of persons who stutter. Poster presentation at the International Conference on Motor Speech 2010, Savannah, Georgia. Retrieved from http://www.madonna.org/file_download/a723dc53-17cf-4822-8e91-49c78c41c7c3

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics, 58,* 540–560. http://dx.doi.org/10.3758/BF03213089

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics, 37,* 46–65. http://dx.doi.org/10.1016/j.wocn.2008.09.001

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters, 272,* 29–32. http://dx.doi.org/10.1016/S0304-3940(99)00573-X

Pardo, J. S. (2010). Expressing oneself in conversational interaction. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identity* (pp. 183–196). New York, NY: Psychology Press.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40,* 190–197. http://dx.doi.org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013). Influence of role-switching on phonetic convergence in conversation. *Discourse Processes, 50,* 276–300. http://dx.doi.org/10.1080/0163853X.2013.778168

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics, 72,* 2254–2264. http://dx.doi.org/10.3758/BF03196699

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3,* 320. http://dx.doi.org/10.3389/fpsyg.2012.00320

Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciu, M., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research, 261,* 220–239. http://dx.doi.org/10.1016/j.bbr.2013.12.034

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36,* 329–347. http://dx.doi.org/10.1017/S0140525X12001495

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention, Perception, & Psychophysics, 78,* 334–345. http://dx.doi.org/10.3758/s13414-015-0981-7

Quené, H. (2008). Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *Journal of the Acoustical Society of America, 123,* 1104–1113. http://dx.doi.org/10.1121/1.2821762

Quené, H. (2013). Longitudinal trends in speech tempo: The case of Queen Beatrix. *Journal of the Acoustical Society of America, 133*(6), EL452–EL457. http://dx.doi.org/10.1121/1.4802892

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59,* 413–425. http://dx.doi.org/10.1016/j.jml.2008.02.002

R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reinisch, E. (2016a). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics, 78,* 1203–1217. http://dx.doi.org/10.3758/s13414-016-1067-x

Reinisch, E. (2016b). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics, 37,* 1397–1415. http://dx.doi.org/10.1017/S0142716415000612

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 978–996. http://dx.doi.org/10.1037/a0021923

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics, 41,* 101–116. http://dx.doi.org/10.1016/j.wocn.2013.01.002

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23*(2–3), 429–435. http://dx.doi.org/10.1016/j.cogbrainres.2004.11.006

Sato, M., Troille, E., Ménard, L., Cathiard, M.-A., & Gracco, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research, 227,* 275–288. http://dx.doi.org/10.1007/s00221-013-3510-8

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate: II. Effects of signal discontinuities. *Perception & Psychophysics, 62,* 285–300. http://dx.doi.org/10.3758/BF03205549

Schwab, S. (2011). Relationship between speech rate perceived and produced by the listener. *Phonetica, 68,* 243–255. http://dx.doi.org/10.1159/000335578

Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Jr., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping, 16,* 219–227. http://dx.doi.org/10.1002/hbm.10046

Shergill, S. S., Brammer, M. J., Fukuda, R., Williams, S. C., Murray, R. M., & McGuire, P. K. (2003). Engagement of brain areas implicated in processing inner speech in people with auditory hallucinations. *British Journal of Psychiatry, 182,* 525–531. http://dx.doi.org/10.1192/bjp.182.6.525

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 10587–10592. http://dx.doi.org/10.1073/pnas.0903616106

Street, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. *Journal of Language and Social Psychology, 2,* 37–56. http://dx.doi.org/10.1177/0261927X8300200103

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095. http://dx.doi.org/10.1037/0096-1523.7.5.1074

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology, 1,* 166. http://dx.doi.org/10.3389/fpsyg.2010.00166

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics, 67,* 939–950. http://dx.doi.org/10.3758/BF03193621

Weber, R. J., & Castleman, J. (1970). The time it takes to imagine. *Perception & Psychophysics, 8,* 165–168. http://dx.doi.org/10.3758/BF03210196

Received July 28, 2016
Revision received October 26, 2016
Accepted November 20, 2016 ■