Atten Percept Psychophys (2016) 78:1203–1217
DOI 10.3758/s13414-016-1067-x

Natural fast speech is perceived as faster than linearly time-compressed speech

Eva Reinisch[1]

Published online: 9 February 2016

Abstract Listeners compensate for variation in speaking rate: In a fast context, a given sound is interpreted as longer than in a slow context. Experimental rate manipulations have been achieved either through linear compression or by using natural fast speech. However, in natural fast speech, segments are subject to processes such as reduction or deletion. If speaking rate is then defined as the number of segments per unit time, the question arises as to what impact such processes have on listeners' normalization for speaking rate. The present study tested the effect of sentence duration and fast-speech processes on rate normalization for a German vowel duration contrast. Results showed that a naturally produced short sentence containing segmental reductions and deletions led to the most "long" vowel responses whereas the long sentence with clearly articulated segments led to the fewest. This suggests that speaking rate is not merely calculated as the number of segments realized per unit time. Rather, listeners associate properties of natural fast speech with a higher speaking rate. This contrasts with earlier results and a second experiment in which perceived speaking rate was measured in an explicit task. Models of speech comprehension are evaluated with regard to the present findings.

Keywords Speech perception · Spoken word recognition · Speaking rate

In order to understand spoken language listeners have to overcome large amounts of variability in the speech signal. One of the most prominent sources of variation is speaking rate, because it may vary considerably even within the same speaker (Miller, Grosjean, & Lomanto, 1984; Quené, 2008). Speaking rate is often defined as articulation rate (i.e., excluding pauses; Crystal & House, 1990; throughout this paper, the term *speaking rate* will be used as a synonym to articulation rate) and measured in realized segments or syllables per unit time (e.g., per second). Taking this definition, a larger number of realized segments per unit time indicates a faster rate (Koreman, 2006). However, natural variation in speaking rate may be confounded with variation in clarity of articulation. Formal speech tends to be slow, with most or all intended segments clearly realized. Natural fast speech, however, often leads to articulatory undershoot (Lindblom, 1963, 1990; see below) and may include segmental reductions and deletions (see, e.g., Ernestus, 2014, for a recent overview). If, then, speaking rate is defined as the number of segments and syllables per unit time, the question arises as to what impact such fast-speech processes have on listeners' perception of speaking rate. How do segments that are produced noncanonically or even deleted contribute to such a segment count? Addressing this question is important, as experimental manipulations demonstrating how listeners cope with variability in speaking rate typically use one of two types of rate manipulation: (1) linear compression of normal-rate speech that keeps all segmental properties in place and shortens every segment to the same extent, and (2) natural fast speech that may be subject to fast-speech processes such as reductions and deletions. The present study addresses the possible impact of such fast-speech processes on rate normalization and compares their effect to normalization for linearly compressed fast speech.

Normalization for speaking rate means that listeners take into account that at a fast rate all segments shorten to some extent (Crystal & House, 1982, 1988) and compensate for this. Through this process, they deal with the large variability

✉ Eva Reinisch
evarei@phonetik.uni-muenchen.de

[1] Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Schellingstraße 3, 80799 Munich, Germany

in speaking rate during speech perception. That is, following a fast context sentence, a given sound is interpreted as longer than when it follows a slow context sentence (e.g., Ainsworth, 1972, 1974; Allen & Miller, 2001; Dilley & Pitt, 2010; Kidd, 1989; Miller, 1981, 1987; Miller & Dexter, 1988; Newman & Sawusch, 2009; Reinisch, Jesse, & McQueen, 2011; Reinisch & Sjerps, 2013; Summerfield, 1981). For example, the word-initial stop voicing contrast in English (e.g., /g/ vs. /k/) is mainly cued by temporal properties, namely duration of voice onset time (VOT). When a stop such as /g/ or /k/ is preceded by a fast carrier sentence, listeners report hearing /k/ (long VOT) more often than /g/ (short VOT). Following a slow sentence, more /g/ responses are reported (e.g., Newman & Sawusch, 2009). As a result, context information influences whether listeners hear words, for instance, as *goat* or *coat*. This information can aid speech perception, especially in the case of ambiguous phonemes (e.g., Newman & Sawusch, 2009; Reinisch et al. 2011; Sawusch & Newman, 2000) and even when other potential cues are available to the listener (Reinisch & Sjerps, 2013).

Studies of rate normalization in phonetic categorization have largely used one of two different methods to implement the rate manipulation: either a speaker is recorded at his or her natural rate and the sentence is manipulated by linear compression or expansion such as through PSOLA (e.g., Dilley & Pitt, 2010; Reinisch et al. 2011; Reinisch & Sjerps, 2013), or the speaker is asked to produce a given sentence at normal, fast, and slow rates (e.g., Kidd, 1989; Newman & Sawusch, 2009). Although both methods of obtaining stimuli at different overall durations have produced reliable effects of speaking rate context on phonetic categorization, little is known about possible differences in the magnitude of these effects. Differences could be expected if the sentences that were naturally spoken at fast versus slow rates differed in the presence of natural fast-speech processes; that is, if the sentence that has been spoken fast contained segmental reductions and deletions but the sentence spoken at a normal rate (that then would be linearly compressed) did not contain these processes (usually little information is given on the segmental properties of these fast vs. slow sentences).

Differences in the perception between natural fast speech and artificially compressed fast speech have been reported with regard to intelligibility (Janse, Nooteboom, & Quené, 2003). In natural fast speech, not all segments are compressed equally (Gay, 1978; Janse et al., 2003). Vowels tend to shorten relatively more than consonants, and unstressed syllables get shortened to a greater extent or are more likely to be deleted than stressed syllables. Janse et al. (2003) tested whether the perceptual system would be specifically tuned to this natural variation such that natural fast speech or speech that mimics the temporal relations of natural fast speech would be more intelligible than linearly compressed fast speech that had been spoken at a normal rate. However, this is not what they found. Rather, linearly compressed fast speech appeared to be most intelligible. Moreover, Adank and Janse (2009) showed that although listeners were able to adapt to and improve perception of linearly compressed speech as well as natural fast speech, transfer of improved understanding occurred only from linearly compressed speech to natural fast speech but not the other way around. Both studies hence suggest that listening to linearly compressed speech where all segments are realized as in normal-rate speech is “easier” than listening to natural fast speech. While these previous studies were mostly concerned with the timing of segments, the present study is additionally concerned with the number of realized segments per unit time.

Segmental reductions and deletions tend to occur even in moderately fast speech (Ernestus, 2014; Pluymaekers, Ernestus, & Baayen, 2005). While segmental deletion means that a segment is completely absent, segmental reduction means that a segment is realized, but not fully. An example here would be that a vowel is produced slightly centralized (i.e., somewhat more schwa-like). The occurrence of these two types of processes is correlated so that speech with more reductions also contains more deletions. Sometimes it is even difficult to distinguish the two: Browman and Goldstein (1990) provided an example of the phrase “perfect memory,” in which there is no audible trace of the /t/, but articulatory measures show a brief alveolar contact (which is, however, only released after the labial closure). From the point of view of the listener, this would be a deletion of the /t/, even though from an articulatory points of view it was “only” a reduction. That is, deletions are typically accompanied by segmental reductions, and sentences spoken at a very fast rate tend to also contain deletions.[1] Because, therefore, their exact roles are hard to disentangle, the present study will consider them as a combination of common speech processes in natural fast speech (see Vitela, Warner, & Lotto, 2013, for an attempt of assessing the role of segmental reductions alone). Importantly, the number of realized segments per unit time has been directly linked to the calculation of speaking rate, hence any differences between the normalization of clearly articulated, linearly time-compressed fast speech and natural fast speech would force us to reconsider how speaking rate is being calculated. Consider the phrase *He probably said . . .*, in which the word *probably* can be produced as the canonical three-syllable word or as the two-syllable form

[1] To illustrate, the sentences recorded for the present study that did not contain deletions were much longer than the tokens that did contain deletions – even when both were spoken at a fast rate.

prob'ly. Assuming an overall duration of about 1,200 ms for this phrase (which constitutes a typical carrier sentence), the two-syllable form would result in a speaking rate decrease of roughly 13 versus 11 segments, or four versus three syllables per second. That is, articulatory processes in natural fast speech may affect the magnitude of normalization for speaking rate relative to fast speech that has been created by linear compression, where all segments are realized as they would be at a normal or slow speaking rate.

The present study follows Koreman (2006), who first investigated the impact of casual speech and the number of realized segments per unit time with regard to the *explicit* perception of speech tempo in two rate judgment tasks. Koreman selected a variety of carefully versus sloppily pronounced sentences from the Kiel Corpus of German spontaneous speech and paired them according to their *intended* speaking rate (i.e., as if all segments were fully realized) or their *realized* speaking rate (i.e., number of realized segments per second, where deletions equal fewer realized segments; note that spectral and temporal reductions in segments were not topic of the specific comparisons). Participants were asked to judge which of two sentences sounded faster and to rate the speaking rate in both sentences on a sliding scale, from *too fast* to t*oo slow.* Results showed that both intended and realized articulation rate correlated with listeners' explicit rate comparison judgments. In keeping with the example given above, the phrase *He probably said* . . . sounded faster when all five syllables and thirteen segments were realized than when it had the same overall duration but fewer syllables and segments realized (e.g., *He pro'bly said* . . only four syllables and eleven segments realized; effect of realized rate). However, the version with deletions still sounded faster than other utterances of the same duration involving the same number (11) of segments/syllables that were instead all realized (i.e., *He always says* . . .; effect of intended rate). In summary, segmental deletions influence the perception of speaking rate in an explicit task such that fewer realized segments as well as fewer intended segments are taken as a sign of a slower rate.

The present study focuses on the influence of fast-speech processes in natural fast speech on the *implicit* perception of speaking rate. To test whether and how natural fast speech versus linearly time-compressed speech affect the speech perception process, the present study uses the well-established effect of normalization for speaking rate in phonetic categorization. This task may be called "implicit," as participants will not be asked explicitly how fast they perceive an utterance to be. Rather, it will be tested whether the speaking rate of an utterance influences the perception of the following stimulus. The goal is to further inform psycholinguistic models of speech perception and to test whether perceived speaking rate depends on fast-speech processes, as this may impact the widely held assumption that speaking rate is calculated as the number of segments per unit time in implicit perception.

Two fundamentally different types of psycholinguistic models have been proposed to account for how listeners access mental representations of sounds and words. Simplifying for the sake of the argument, abstractionist models (e.g., Norris, 1994; Norris & McQueen, 2008) assume that listeners store one representation – typically, a canonical form of each word. Any variation due to speaker or speaking rate has to be "abstracted" prelexically to access word meaning. Note that this is the common way to describe listeners' reactions to variation in speaking rate. Notions such as "rate normalization," "compensation for speaking rate," and "coping with variability" suggest that speaking rate changes are a problem for the listener that have to be resolved before accessing the lexical representations. However, although "normalization" is the classical way of thinking of how to link the variable speech input to lexical representations, rate normalization can also be explained in terms of models with multiple representations for a given word.

Exemplar models (e.g., Goldinger, 1998) assume storage of each variant of a word, including, for example forms with deletions, items produced by different speakers, and at different speaking rates. Word forms are accessed by directly comparing the acoustic input to stored representations. That is, variability due to speaking rate does not have to be abstracted, but words spoken at a fast rate are mapped onto fast exemplars and tokens spoken at a slow rate are mapped onto slow exemplars (exemplars are "labeled" to have occurred in fast vs. slow speech; Pierrehumbert, 2001). Speaking rate is then assessed via the labels for rate of the best matching exemplars. Although it is commonly accepted that in their extreme assumptions neither fully abstractionist nor exemplar models can account for the majority of findings in the psycholinguistic literature, the nature of the best hybrid model has yet to be established (see, e.g., Ernestus, 2014, for a discussion with regard to processing reductions in casual speech).

Notably, a third class of models has recently been suggested to describe how listeners flexibly adapt to certain properties of the speech signal. These probabilistic models of speech perception, such as the belief-updating model of perceptual adaptation (Kleinschmidt & Jaeger, 2015) state that listeners track consistencies in the speech signal for a given situation, for instance, a specific speaker's idiosyncratic pronunciation of certain sounds, and create models of cue distributions for this given situation. These specifically adapted models of cue distributions will be reapplied in perception when the situation or the speaker is recognized again, hence facilitating perception and providing new starting points for further adaptation. Although the time scale of this adaptation remains widely unspecified, contextual speaking rate is likely a signal property for which cue distributions have been established (see, e.g., Reinisch, 2015; Sjerps & Reinisch, 2015).

Whether or not speech processes that are common in natural fast speech affect the implicit processing of speaking rate will help to further evaluate the different properties of the different speech perception models. Importantly the present study will contribute toward resolving the question whether speaking rate is indeed sufficiently explained by calculating the number of segments per unit time. A priori, two scenarios seem likely for how the processing of natural fast speech that typically includes segmental reductions and deletions may differ from the processing of linearly compressed normal-rate speech.

Under a first scenario, given the same sentence (i.e., the same intended words) and an identical overall duration, natural fast speech is perceived as slower than linearly compressed fast speech. This is because natural fast speech tends to contain segmental reductions and deletions. In contrast, linearly compressed speech tends to have all segments realized. If speaking rate were calculated from the number of realized segments per unit time or the relative speed of the articulators, we would expect the perception of slower speech tempo for a naturally produced fast than a linearly compressed normal-rate sentence. This scenario would be in line with the results for explicit speech perception (Koreman, 2006), which show that a higher number of realized segments leads to a higher perceived speaking rate. In terms of abstractionist speech perception models, it would suggest that speaking rate is calculated before abstraction during processing occurs. If rate were calculated only after abstraction, no difference between the natural fast and linearly compressed sentence would be expected. In terms of exemplar models, listeners would calculate rate by matching the signal onto stored exemplars, either full forms or forms that include reductions and deletions. Short forms and forms with a higher number of segments would be "labeled" as fast. Long forms and forms with fewer or less clearly articulated segments would be "labeled" slow. Probabilistic models of speech perception could also account for such an outcome. The belief updating model would state that a higher number of segments per unit time or faster perceived articulator speed would be directly associated with higher speaking rate and hence support the use of cue distributions for fast speech. These types of association would also be in line with accounts arguing for top-down prediction as a driving factor in perception (Clark, 2013; Farmer, Brown, & Tanenhaus, 2013, commentary).

Under a second scenario, given the same sentence (i.e., the same intended words) and an identical overall duration, natural fast speech that includes reductions and deletions is perceived as faster than linearly time-compressed fast speech with all segments present. This scenario would go against the traditional view of speaking rate being calculated by the number of realized segments/syllables per unit time. Abstractionist models would have a hard time explaining this outcome. However, it would be in line with exemplar models of speech perception in which stored fast sentences tend to include reductions and deletions. That is, not only shorter forms would be labeled fast but also forms that include properties that are typically found in fast speech, such as segmental reductions and deletions. Specifically, the information "short" and "includes fast-speech processes" would both contribute to the perceived speech tempo. In other words, Scenario 2 could be explained through listeners' "knowledge" or association that segmental deletions tend to occur in fast speech. Probabilistic models of speech perception or accounts arguing for top-down prediction could also account for such an outcome. In fact, such an outcome would strongly favor the involvement of top-down knowledge. This point will be taken up in the General Discussion.

The present study set out to test whether and how processes in natural fast speech contribute to the perception of speaking rate. As laid out above, participants were not asked explicitly how fast they thought a sentence was. Instead, the perceived speaking rate was measured indirectly by asking participants to categorize a duration-based phonological contrast. Specifically, listeners categorized a German /a/–/a:/ duration continuum. In German, the /a/–/a:/ vowel contrast is described as a real duration contrast without consistent co-variation of spectral properties (Jessen, 1993; Pätzold & Simpson, 1997).[2] The /a/–/a:/ continuum appeared in a German minimal word pair at the end of a rate-manipulated carrier sentence. The faster this carrier sentence is perceived, the more "long" (/a:/) answers should be given. To be better able to judge the effect of speech processes typical to natural fast speech against an expected effect of rate due to overall context duration, the concept of speaking rate was split into two subcomponents: sentence duration (long/short) and fast-speech processes (present/absent). Sentence duration was defined such that "long" matched the duration of a naturally spoken normal-rate sentence and "short" matched a naturally spoken fast sentence including typical fast-speech processes such as reductions and deletions. That is, the first version of the sentence had a long overall duration with all segments articulated as is typical for natural normal-rate speech ("long/absent" condition). In the second version, the same sentence was at a short overall duration including reductions and deletions as produced in natural fast speech ("short/present" condition). That is, it was shorter, and contained fast-speech processes leading

[2] This becomes apparent in the acquisition of Dutch as a second language, where the lack of a spectral difference between lax and tense /a/ is one of the telltale signs of a German accent.

to a lower number of segments than the first version. For each of these sentences an additional version was created by artificial rate manipulation (i.e., linear compression and expansion). As a third version, the normal-rate sentence was linearly compressed to the short overall duration of the sentence that had been spoken fast ("short/absent" condition). This version was fast and had a higher articulation rate than the naturally spoken fast sentence, since more segments were realized in the same amount of time (and spoken more clearly). Finally, as a fourth version, the sentence that was spoken fast was linearly expanded to the overall duration of the long sentence (i.e., the normal-rate sentence; "long/present" condition). This sentence had the lowest articulation rate of all four conditions. All of these four conditions, except for the long one that was an expanded version of the naturally spoken fast sentence, can be found in studies on normalization for speaking rate in phonetic categorization. The fourth condition was included to complete the factorial design, as it is conceivable that speech processes found in natural fast speech may be perceived differently in fast speech (where they can be expected) than in slow speech, where they are unusual.

More specifically, the question was whether the different combinations of a long and short duration and the presence versus absence of fast-speech processes would result in rate effects of different magnitude as instantiated in shifts of the /a/–/a:/ category boundary. The faster a condition is perceived to be, the more /a:/ responses are expected. In a second experiment, the same sentences were subjected to an explicit rate comparison task similar to Koreman (2006) to allow for a better comparison of "implicit" effects of speaking rate on phonetic categorization and explicit judgments of perceived speaking rate.

Experiment 1

Method

Participants

Twelve native speakers of German from the student population of the University of Munich took part for a small monetary compensation. They reported no language, speech, or hearing impairment.

Materials

The German words *bannen* and *bahnen* ("to banish"–"to channel") that differ minimally in the /a/–/a:/ vowel duration contrast served as targets. Both words were recorded by a female native speaker of German at the end of the carrier sentence *Sie vermied in ihrem Text den Begriff {TARGET}* (literal translation: "She avoided in her text the term {TARGET}"). The sentence was recorded multiple times at a "normal" speaking rate in a clear speech style, as well as fast with the possibility for segmental deletions (for details about the carrier, see below). One token of the target word *bahnen* (i.e., containing the long vowel; 161 ms) was selected from the normal-rate items and excised from the sentence for further manipulation.

Using this selected token of *bahnen*, an /a/–/a:/ vowel continuum was created by manipulating the duration tier in PRAAT (Boersma & Weenink, 2009) and subsequent PSOLA resynthesis. Pretests on similar minimal word pair continua in a different study showed that vowel durations of 51 and 146 ms suffice for clear /a/ and /a:/ continuum endpoints, respectively. This duration range was split into 13 steps with a step size of 7.3 ms. Nine of these were chosen for the present experiment (i.e., steps 1, 3, 5, 6, 7, 8, 9, 11, and 13). The other four were dropped to reduce the overall number of steps and to allow for a larger number of repetitions per point. To avoid any influence of segment durations other than the vowel biasing listeners towards the base word *bahnen*, all segment durations were set to an average duration between the manipulated word's segments and a reference token of *bannen* (also selected from the normal-rate productions).

For the context manipulation, two tokens of the carrier sentence (*Sie vermied in ihrem Text den Begriff {TARGET}*) were selected: one spoken at a "normal" rate (long sentence, fast-speech processes absent), and one spoken fast with several segments reduced and deleted (short sentence, fast-speech processes present). Two trained phoneticians used broad IPA transcription to assess the number of segments produced in the chosen carrier sentences. Transcriptions were based on listening as well as visual inspection of the signals (spectrogram, oscillogram). The two transcribers agreed that in the clear version, 26 segments were realized (i.e., [zi: f mi:t n ir m tɛks d m begr f]), while in the version including deletions, only 20 segments were realized (i.e., [s f mit n i m tɛks tm g f]). Note that in sequences like *Text dem* the final /t/ in *Text* did not show a separate release in either version. Given the same closure duration in the natural fast and linearly time-compressed fast version (approximately 30 ms) the same number of segments was counted in both versions (here: one). Appendix A lists all segments of the two carrier sentences, including segment durations, and values of the first and second formants of all vowels in Hertz and Bark to assess spectral reductions in addition to the difference in number of realized segments. Because the spectral values showed a slight tendency toward centralization of the vowels in the naturally spoken fast sentence, it was further established that the long-term average spectra of the two versions of the sentence did not show any consistent, long-lasting differences above and below the first two formants. This precluded the impact of spectral contrast effects as discussed in Vitela et al. (2013). Most important for the present study was that duration is the main cue to the German /a/–/a:/ contrast and that the naturally produced fast sentence had undergone processes typical of naturally produced fast speech such that it contained fewer segments than the naturally produced normal-rate sentence.

Two additional tokens of the carrier sentence were created to construct a fully crossed design. To disentangle the presence or absence of natural fast-speech processes from speaking rate as instantiated by overall sentence duration (or the time within which segments could be counted) the concept of speaking rate will henceforth be referred to as two components: sentence duration and presence versus absence of fast-speech processes. That is, the factors Duration (long/short) and fast-speech Processes (present/absent) were fully crossed. Using PSOLA resynthesis, the original short/present sentence was expanded to the overall duration of the original normal-rate sentence, and the normal-rate sentence was (linearly) compressed to the overall duration of the naturally spoken fast sentence. Figure 1 illustrates this manipulation. In order to account for differences in artifacts due to the expansion and compression, the original sentences were also resynthesized (speeded up and slowed down again) such that all four versions of the sentence had undergone manipulation.

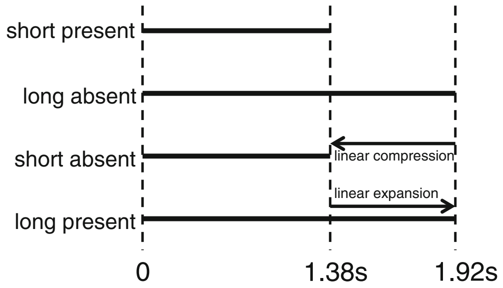

Fig. 1 Illustration of the four speaking rate conditions showing how speaking rate can be composed of duration (long vs. short sentence) and the presence versus absence of natural fast-speech processes. The arrows labeled "linear compression" and "linear expansion" illustrate the manipulation scheme used for the present experiments

Design and procedure

The four versions of the carrier sentence were combined with all nine selected steps of the *bannen–bahnen* continuum, resulting in a total of 36 sentences. Participants were seated in a sound-attenuated room and performed a phonetic categorization task. They listened to the sentences via headphones and indicated whether the last word in the sentence was *bannen* or *bahnen* by pressing the number keys 1 and 0 on a computer keyboard. Response options were displayed on the computer screen throughout the experiment with the layout of the words on the screen matching the sides of the response keys. Word-key assignments were counterbalanced across participants. The next trial started 700 ms after the participant's response. Each participant received a total of 252 trials, that is, 7 repetitions of each carrier-target combination. Every 63 trials, participants were allowed to take a break. The experiment was controlled by ePrime software (Psychology Software Tools, Inc.) and took approximately 25 minutes to complete.

Results

Results were analyzed using a linear mixed-effects model with the dichotomous dependent variable /a:/ responses (i.e., response /a:/ coded as 1, response /a/ coded as 0) and participant as a random factor, including random slopes for all within-participant fixed factors (see Barr, Levy, Scheepers & Tily, 2013). A logistic linking function was used. The model included three fixed factors and their interactions: Sentence Duration (short -> coded as −0.5, long -> coded as 0.5), fast-speech Processes (present -> −0.5, absent -> 0.5), and Continuum Step (centered on 0). All factors were contrast coded such that effects could be interpreted as main effects. The two outmost steps on both sides of the continuum were classified correctly with close to ceiling performance (>98 % and <2 % /a:/ responses), which shows that acoustically unambiguous steps are unlikely to be affected by acoustic context. Only responses to the middle five steps of the continuum were therefore analyzed. Figure 2 shows the categorization responses along the /a/–/a:/ continuum in the four conditions. Figure 3 aggregates the effects over the five middle continuum steps.

Results showed main effects of Continuum Step, ($b_{(\mathrm{Intercept})} = -0.61$, $SE = 0.39$, $z = -1.57$, $p = .12$; $b_{(\mathrm{step})} = 2.4$, $SE = 0.22$, $z = 11.19$, $p < .001$); Sentence Duration, ($b_{(\mathrm{duration})} = -1.18$, $SE = 0.25$, $z = -4.66$ $p < .001$); and fast-speech Processes, ($b_{(\mathrm{processes})} = -0.57$, $SE = 0.26$, $z = -2.20$, $p < .05$), without any of the interactions reaching significance, ($b_{(\mathrm{step*duration})} = -0.05$, $SE = 0.32$, $z = -0.16$, $p = .87$; $b_{(\mathrm{step*processes})} = -0.08$, $SE = 0.32$, $z = -0.25$, $p = .81$; $b_{(\mathrm{duration*processes})} = -0.43$, $SE = 0.4$, $z = -1.09$, $p = .28$; $b_{(\mathrm{step*duration*processes})} = 0.25$, $SE = 0.6$, $z = 0.42$, $p = .67$). The effect of Continuum Step indicates that the vowel duration manipulation resulted in the expected effect; that is, the longer the vowel, the more /a:/ responses were given. Critically, both Sentence Duration and fast-speech Processes influenced vowel perception: both short sentence duration and the presence of fast-speech processes lead to more /a:/-responses.

Discussion

Experiment 1 used the well-established speaking rate effect on phonetic categorization to test how duration and

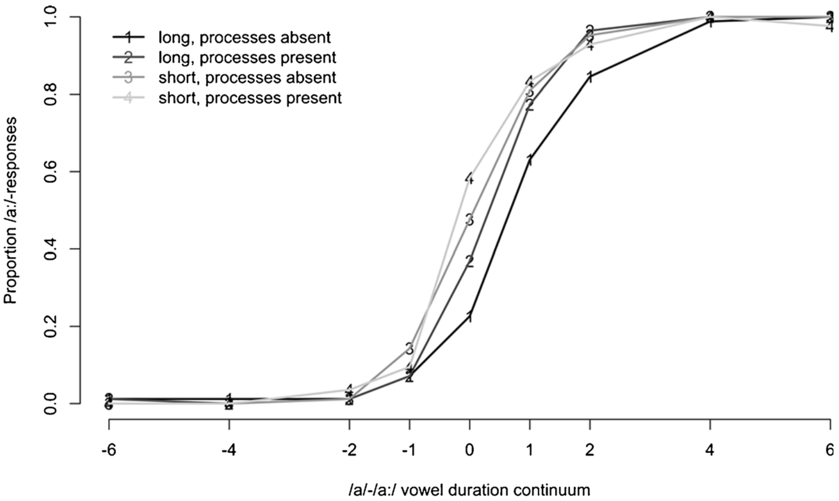

Fig. 2 Categorization function for the four conditions in Experiment 1. The two darker lines represent the long sentences; the lighter lines the short context. Lines with the point characters 1 and 3 represent the sentences without fast-speech processes, and the lines with the point characters 2 and 4 indicate sentences including fast-speech processes. Since participants identified points −6 and +6, and −4 and + 4 (on the x-axis) with close to ceiling performance, these were not included in the statistical analyses (see text for details)

natural fast-speech processes contribute to perceived speaking rate. Based on previous literature, it was expected that the faster listeners perceived the preceding carrier sentence to be, the more "long" responses should be given, here with regard to the perception of the German /a/–/a:/ contrast. Results showed that both manipulated subcomponents, Duration and presence versus absence of fast-speech Processes, exerted an effect. The short sentence that contained natural fast-speech processes led to the most /a:/ responses, whereas the long sentence in normal-rate speech without such processes led to the fewest /a:/ responses. The remaining two conditions (short/absent and long/present) showed intermediate effects (see Fig. 3). These results speak to two issues.

First, these findings inform the literature on normalization for speaking rate in phonetic categorization. There are currently two main approaches to manipulating speaking rate. Either a speaker is asked to produce carrier sentences fast versus slowly (e.g., Kidd, 1989; Newman & Sawusch, 2009) or utterances spoken at a "neutral" speaking rate are manipulated with linear compression and expansion (e.g., Dilley & Pitt, 2010; Reinisch et al. 2011; Reinisch & Sjerps, 2013). Although both types of rate manipulation have been shown to trigger the expected contrastive effects in the perception of duration contrasts, the present study was the first to directly compare the magnitude of the effects using the same sentence and target words in different conditions. Results suggest that naturally produced fast speech is perceived as faster than linearly compressed "normal-rate" speech. However, there was no interaction between Duration and fast-speech Processes. This suggests that although speech that contains fast-speech processes is perceived as faster than compressed or uncompressed normal-rate speech, the magnitude of the rate normalization effect (i.e., the comparison between the long vs. short sentence within each condition of the factor Processes) did not differ.

The second, more important finding of the present experiment is the effect of natural fast-speech processes itself. This effect showed the opposite pattern to the explicit rate judgments reported in Koreman (2006). In the present experiment, the sentence that had been spoken fast and included segmental reductions and deletions was perceived as *faster* than the same

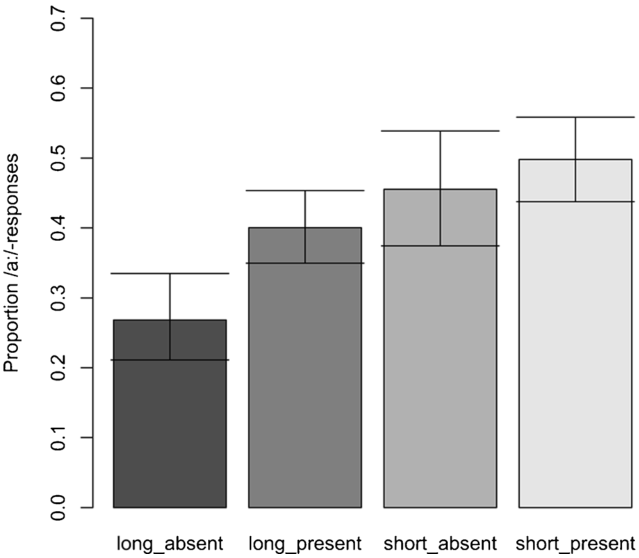

Fig. 3 Proportion /a:/ responses pooled over the five middle steps of the vowel duration continuum (hence the scale to .7 rather than 1), shown for the four different combinations of Duration and fast-speech Processes. The higher the bars, the "faster" the condition was perceived. Error bars show one standard error and were calculated in logistic space (as were the analyses), but for the purpose of plotting they were transformed back to the proportion scale

sentence of the same duration with all segments realized (as in normal-rate speech). That is, the second scenario as described in the introduction appears to be confirmed. The implications of this result for speech processing will be taken up in the General Discussion. However, before jumping to conclusions, the four sentence conditions of the present experiment were subjected to an additional rate comparison task, similar to that conducted by Koreman (2006). This additional task will allow the results of Experiment 1 to be set in relation to explicit rate comparison judgments.

Experiment 2

Method

Participants

Twelve participants who matched the same selection criteria but did not take part in Experiment 1 participated for a small monetary compensation.

Material

Following Koreman (2006), a rate comparison task was constructed by slightly adapting the materials used in Experiment 1. Listeners were presented with two versions of the carrier sentence and had to decide which of the sentences sounded faster or slower respectively. The target words used in Experiment 1 were not included in the present experiment. Given that in Experiment 1 only four versions of the same sentence were used and the difference in overall duration between long and short was rather large (i.e., about 540 ms) three additional intermediate durations were created for both types of sentences (i.e., with natural fast-speech processes present and absent). The duration manipulation was again implemented using PSOLA by linear compression of the long, normal-rate sentence and linear expansion of the short sentence that had been produced fast. Experiment 2 hence used five tokens of the sentence with natural fast-speech processes present and five tokens of the sentence with fast-speech processes absent.

Design

Sentences were paired in three different ways. First, sentences of the same duration but with fast-speech processes present versus absent were paired as the “same duration-different processes“ condition. If the results are consistent with those for the implicit perception of rate found in Experiment 1, the sentence containing fast-speech processes should be chosen as the faster one in this condition. A “different duration-same processes“ condition included sentences of the same type (i.e., either fast-speech processes present vs. absent) but with different overall durations. If “1” represents the shortest and “5” the longest of the five possible durations, the following duration pairs were used: 1–3, 2–4, 3–5, and 1–5. Using these pairs allowed us to see whether participants were able to do the task. If so, they should choose the physically shorter stimulus as the one with the faster rate, and responses should be more consistent for the easy (1–5) pair than for the physically less distinct pairs (1–3, 2–4, and 3–5). Finally, in the “different duration-different processes“ condition the two sentences were grouped into duration pairs (1–3, 2–4, 3–5, and 1–5), which also differed in the presence versus absence of fast-speech processes. If, as in the implicit task, participants perceive stimuli with segmental reductions and deletions as faster than the sentences without such processes, they should choose the physically shorter stimulus as the faster one more often if the shorter stimulus is also the one that contains fast-speech processes. Taking all possible combinations and orders (first/second) into account, a total of 42 sentence pairs was used (i.e., five pairs for the same duration–different processes condition × 2 orders = 10, 4 pairs for the different duration-same processes × 2 orders for duration × 2 types of sentences (fast-speech processes present vs. absent) = 16, 4 pairs for the different duration-different processes condition × 2 orders for duration × 2 orders for speech processes present vs. absent first = 16, 10 + 16 + 16 = 42).

Procedure

Participants were seated in a sound-attenuated room and were presented with the sentence pairs over headphones at a comfortable listening level. Sentences in a pair were separated by 500 ms. Half of the participants had to decide whether the first or the second sentence sounded *faster* while the other half decided whether the first or the second sentence sounded *slower*. They pressed the left arrow key on the computer keyboard for answering “first“ and the right arrow key for answering “second.“ Participants were encouraged to answer as quickly as possible, but only after the end of the second sentence, which was indicated by the appearance of a cross in the middle of the computer screen. Each sentence pair was presented four times in a separate random order with the restriction that all sentence pairs were presented once before being repeated. Every 42 trials, participants were allowed to take a break.

Note that this design includes a few adjustments relative to the task used by Koreman (2006). In the present study, the same sentence was used in all conditions, so no “same duration-same processes“ condition was included. This had two advantages: first, across conditions listeners had to compare only physically different stimuli, and second, there was no need for a “same“ answer to be allowed. This prevented participants from retreating to answering “same“ when they were unsure about their decision.

Results

Results were analyzed separately for the three different conditions (“same duration–different processes,” “different duration-different processes,” and “different duration–same processes”). Results will be plotted and reported such that the dependent variable is response “first sentence is faster” (i.e., *first faster* coded as 1, *second faster* coded as 0). That is, for those participants whose task it was to determine the slower sounding sentence, data were recoded to match “first faster.” In addition, following Koreman (2006), the different–duration pair data were further recoded such that the first sentence in the pair was the one with the shorter duration. These recoded data from all three conditions were analyzed with mixed effects models, with “first faster” as the dependent variable and fast-speech Processes (present coded as −0.5 and absent coded as 0.5) as a fixed factor. For the data in which the duration of the pair differed (“different duration–different processes,” and “different duration–same processes”), two additional factors were entered: duration Difference (*easy* coded as 0.5 and *difficult* coded as −0.5,) and an interaction between fast-speech Processes and Difference. There are two things to note about the coding. First, a positive significant intercept will indicate the use of duration such that listeners report “first faster“ if the first sentence is shorter. Second, the factor fast-speech Processes pertains to the first/faster sentence of a pair. A similar effect as in Experiment 1 (i.e., sentences containing fast-speech processes sound fast) would therefore be reflected in a negative regression weight of fast-speech Processes, since the level “present” was coded as −0.5, and a negative regression weight then means that more “positive”, first is faster responses result from this type of speech. Participant was entered as a random factor with random slopes for within-participant fixed factors. A logistic linking function was used.

Results for the *same duration–different processes* condition showed that in the absence of durational differences between the sentences, listeners had a bias toward responding first faster, ($b_{(\text{intercept})} = 0.39$, $SE = 0.1$, $z = 3.98$, $p < .001$), but there was no effect of the presence versus absence of fast-speech processes, ($b_{(\text{processes})} = -0.09$, $SE = 0.38$, $z = -0.24$, $p = .81$; $M_{(absent)} = 0.58$, $M_{(present)} = 0.61$). That is, listeners were unable to differentiate perceived speaking rate based on whether or not the sentence contained speech processes that are typical for fast speech (i.e., segmental reductions and deletions).

Results for the *different duration–same processes* condition are plotted in Fig. 4, left panel. Since the data for this condition were recoded such that the first sentence was the faster one, listeners’ strong preference to respond “first faster” shows that they based their responses on duration differences, ($b_{(\text{intercept})} = 3.13$, $SE = 0.32$, $z = 9.76$, $p < .001$). As expected, this preference was larger if the duration difference between the stimuli was large (1–5 pairs) than when it was small, other pairs; ($b_{(\text{difference})} = 1.84$, $SE = 0.62$, $z = 2.98$, $p < .005$). However, neither the effect of fast-speech Processes, ($b_{(\text{processes})} = -0.40$, $SE = 0.57$, $z = -0.71$, $p = .48$), nor the interaction of Processes and Difference, ($b_{(\text{processes*difference})} = -0.98$, $SE = 1.04$, $z = -0.95$, $p = .34$), were significant, suggesting a similar use of duration for fully articulated sentences and sentences including segmental reductions and deletions.

Finally, data for the *different duration–different processes* condition are shown in Fig. 4, right Panel. Again, a strong preference to respond “first faster” was found, which reflects listeners’ use of duration differences in making speaking rate judgments, ($b_{(\text{intercept})} = 2.85$, $SE = 0.32$, $z = 9.01$, $p < .001$). However, even though in this condition the presence versus absence of fast-speech processes differed in sentences of a pair, fast-speech Processes did not show a significant effect, ($b_{(\text{processes})} = -0.48$, $SE = 0.5$, $z = -0.96$, $p = .34$). Again, pairs with a larger duration difference (1–5) led to more ”first faster” responses in line with the difference than pairs with small duration differences, ($b_{(\text{difference})} = 1.64$, $SE = 0.47$, $z = 3.47$, $p < .001$). Since there was a marginal interaction of Processes and Difference, ($b_{(\text{processes*difference})} = -1.67$, $SE = 0.92$, $z = -1.82$, $p = .07$), an additional analysis excluding the duration pair with the largest difference (i.e., 1–5) was performed, since the large duration difference may have overshadowed any effect of fast-speech Processes. But this analysis showed similar results, ($b_{(\text{intercept})} = 1.85$, $SE = 0.25$, $z = 7.35$, $p < .001$; $b_{(\text{processes})} = -0.32$, $SE = 0.29$, $z = -1.11$, $p = .27$). In sum, results for this explicit rate comparison task show that listeners based their decisions only on the overall duration of the sentence and if the presence versus absence of fast-speech processes had an effect it was too small to be detected in an explicit task.

Discussion

Experiment 2 set out to compare the results from the implicit speech perception task in Experiment 1 to explicit judgments about the perceived speaking rate of the carrier sentences. The results from the sentence pairs including durational differences showed that participants could readily perform the task of judging which sentence was faster. Participants indicated well above chance that the shorter sentence was the faster one, and did so more consistently if the duration difference was larger. Importantly, for the smaller duration differences, the overall performance was not quite at ceiling (i.e., around 88 %), which would have left room for variables other than duration to influence responses. In other words, the duration differences were small enough to leave some uncertainty about which stimulus was faster, which could have allowed the presence versus absence of fast-speech processes to exert

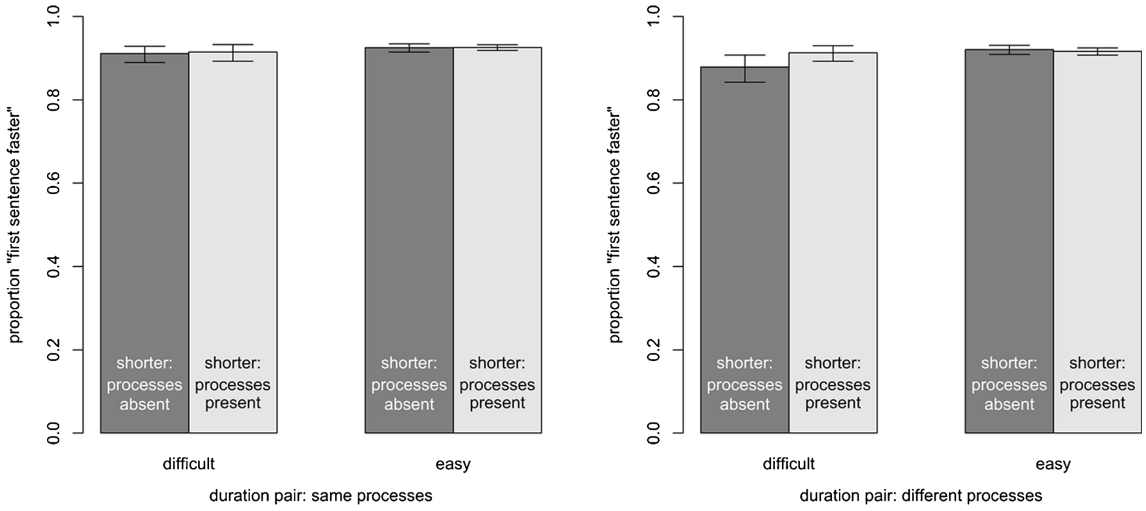

Fig. 4 Proportion "first faster responses" for the explicit rate comparison task in Experiment 2. The left panel shows results for the "different duration–same processes condition" and the right panel shows results for the "different duration–different processes condition." Note that the data were recoded such that first sentence indicates the faster one. The labels "shorter: processes absent" (dark bars) and "shorter: processes present" (light bars) indicate whether or not the first, that is, shorter sentence of the pair contained fast-speech processes. Error bars show one standard error and were calculated in logistic space (as were the analyses) but for the purpose of plotting they were transformed back to the proportion scale

an influence. In addition, even in the "same duration–different processes condition," where the difference in the presence versus absence of fast-speech processes was the only cue to suggest differences in rate, listeners appeared not make use of this information.

Given that listeners could perform the task in general, the main question can be addressed. This was whether the effect of Speech Processes found for the implicit task in Experiment 1 could be replicated in the explicit judgments of this second experiment. That is, would sentences that include speech processes that are typically found in natural fast speech be judged as faster than sentences with all segments (or at least a larger number of segments) clearly realized, or would results pattern with the explicit rate judgments in Koreman (2006)? It turns out, in a direct comparison of the same sentences with and without fast-speech processes, listeners did not take this difference into account. If durational cues were not available, they responded at random (with an overall bias toward responding "first faster"), otherwise all judgments were based on the duration difference.

An alternative account on the apparent lack of an effect of fast-speech processes in the explicit task could be taken if speaking rate was not only calculated as the number of segments per unit time, but listeners additionally took transition speed or articulator speed into account. That is, segmental reductions and deletions lead to shallower transitions and reduced articulator speed relative to the fully articulated sentence. If this measure was perceived in a trading relation with rate perception as a segment count then the two types of measures may indeed cancel the effects in the same duration-different processes condition. Moreover, the duration difference in the different duration–different processes condition – despite the addition of smaller duration differences than in the implicit task – could have merely overpowered an effect of fast-speech Processes that underlyingly would have been there. Note that this interpretation is different from the account that fast-speech processes including a reduced number of segments did not influence perception at all. It suggests that effects of fast-speech processes simply cannot be measured in an explicit task. However, no matter how speaking rate is being calculated, the results of Experiment 2 do not match those from the implicit task in Experiment 1 or the explicit rate judgment task in Koreman (2006).

General discussion

An ambiguous vowel is perceived as relatively longer following a fast context sentence than following a slow context sentence. This effect of normalization for speaking rate was used to assess whether and to what extent speech processes common to natural fast speech such as segmental reductions and deletions contribute to the perception of speaking rate. Results showed that overall sentence duration as well as the presence versus absence of fast-speech processes both influenced listeners' categorization of a German minimal word pair differing in the /a/–/a:/ duration contrast. More /a:/ responses were given in the short than the long context, and more /a:/ responses were given following a sentence that was spoken fast and contained segmental reductions and deletions as compared to a sentence that had been produced at a normal rate and was linearly compressed to the same duration as the naturally spoken fast sentence. One objective of this study was to test whether the number of realized segments per unit time is sufficient to calculate speaking rate to explain its effects on speech perception. However, the sentences including fast-speech processes were perceived as faster than the sentences without those processes, despite the lower number of realized segments within the same amount of time. Results thus match the second scenario as described in the Introduction. They suggest that the number of realized segments per unit time or perceived articulator speed may not be a sufficient definition of speaking rate in implicit rate perception. Rather listeners appeared to interpret the presence of processes typical to natural fast speech as a sign of fast speech despite the lower number of realized segments.

Exemplar models of lexical representation (e.g., Goldinger, 1998) can account for such a finding since they are based on the assumption that linguistic categories consist of a collection of detailed instances/exemplars in memory (i.e., one for each form spoken by a specific speaker, at a specific rate, etc.). A particular stimulus then refers to the exemplars with the closest match to the input and listeners can draw on their "knowledge" how natural fast speech tends to sound. Specifically, the properties "short duration" and "includes processes typical for fast speech" (such as reductions or deletions) would both contribute to the perception of a fast rate.

An analogy of the present findings can be drawn with effects of long-term experience/expectations on normalization for spectral characteristics of utterances. For example, Strand and Johnson (1996) demonstrated that listeners categorize fricative continua differently depending on the perception of the voice as male or female. More specifically, listeners relied on their prior experience that female voices tend to be higher and have higher resonance frequencies than male voices and assumed accordingly that the /s/–/ / boundary would be at higher spectral center frequencies for female than male speakers. Similar effects have been demonstrated with regard to vowel perception depending on the expected vowel spaces of male and female speakers (Johnson, Strand & D'Imperio, 1999). Talker normalization emerges from the extraction of memorized correlations between talker representation and the linguistic exemplars. Since speaking rate is also considered a part of the representation of exemplars, the association of fast-speech processes, such as segmental reductions or deletions with a fast speaking rate, could be explained in a similar fashion. However, it has to be noted that exemplar models of speech perception cannot explain all effects of naturally fast casual speech (Ernestus, 2014).

Another class of perception models that can account for effects of experience with certain listening situations and associations such as "fast speech is likely to contain reductions and deletions" are probabilistic models of perception, for instance, the Belief Updating Model of perceptual adaptation (Kleinschmidt & Jaeger, 2015). The belief-updating model suggests that whenever listeners recognize consistencies in the speech signal for a given situation, they will track these situation- or speaker-specific distributions of acoustic cues. These specifically adapted models of cue distributions will be reapplied in perception when the situation or the speaker is recognized again. Although not explicitly stated in the description of the model, structural correlations such as "*natural fast speech tends to contain segmental reductions and deletions*" are likely candidates for "situational" cooccurrences that could be tracked in order to facilitate recognition.

So what does this associative relation between the occurrence of fast-speech processes and listeners' knowledge that they tend to occur in fast rather than normal-rate or slow speech add to our understanding of how speaking rate is being processed? Previous research suggests that normalization for speaking rate is an early perceptual or even general auditory process: normalization for speaking rate occurs across speakers (Green, Tomiak & Kuhl, 1997; Newman & Sawusch, 2009). This has been used to argue that rate normalization takes place before other early perceptual processes, such as stream segregation (i.e., the perceptual separation of voices), occur. Rate normalization can be elicited by non-speech contexts (e.g., Diehl & Walsh, 1989; Wade & Holt, 2005), and it has been found in non-human perception (e.g., Welch, Sawusch, & Dent, 2009). More recently, studies using eye tracking have shown that listeners use speaking rate online during speech perception to interpret upcoming sounds as soon as they become available in the acoustic signal (Reinisch et al. 2011, Reinisch & Sjerps, 2013). All this evidence on the earliness and generality of rate normalization may seem to be counter to the present findings if the explanation is that prior knowledge (that fast speech tends to contain reductions and deletions) influences the effects. Abstractionist models of speech perception that argue for such early, prelexical rate normalization appear unable to explain the present results.

However, this apparent discrepancy can be reconciled in two ways. First, by allowing speaking rate to influence processing at more than one point during processing. In fact, speaking rate may be similar to spectral contrast effects. Spectral contrast effects have been shown to arise early during processing occurring between the periphery of the auditory system (Summerfield, Haggard, Foster, & Gray, 1984; Wilson, 1970), up to language-specific levels of speech perception (Sjerps, Mitterer & McQueen, 2011, 2012; Viswanathan, Fowler, & Magnuson, 2009). The example presented above, that spectral contrasts are interpreted differently dependent on whether they are perceived to originate from a male versus female speaker, appears at yet higher processing levels. This may indeed mirror the relation of previous findings on the earliness of speaking rate normalization and the present findings that segmental deletions tend to be associated with fast speech – *especially* in implicit processing.

Note that at least one other study did show an influence of an external factor on the perception of speech tempo. Bosker and Reinisch (2015) showed that carrier sentences that were spoken by a nonnative speaker led to more "long vowel" responses than when the carrier was spoken by a native speaker. That is, despite an overall match in long-term spectral properties, duration, and number of realized segments, nonnative speech was perceived as faster than native speech in an implicit rate normalization task. This is because nonnative speech is typically slower than native speech, and listeners tend to take this expectation into account. Therefore, presenting native and nonnative speech at the same overall duration made the nonnative material sound faster.

However, Bosker and Reinisch also propose an alternative explanation to their results that would be a second way to reconcile our results with a low-level abstractionist normalization account. Bosker and Reinisch argued that accented sentences may be harder to process than native sentences and this increase in cognitive effort may speed up time perception (see Block, Hancock, & Zakay, 2010). This explanation has gained support by a set of studies suggesting that increased cognitive load as instantiated in a dual-task experiment indeed appears to speed up time perception. Bosker, Reinisch, and Sjerps (2016) asked listeners to categorize a vowel duration continuum following fast and slow context sentences during which participants had to complete an easy versus difficult visual search task. Results showed that the context was (implicitly) perceived as faster if the visual search was difficult than when it was easy. Similar arguments could be brought forward with regard to the present study. In addition, they may rescue an account of abstract representation. If fast-speech processes such as segmental deletions are seen as a form of signal degradation (as would be assumed in abstract models of speech processing), the effect of fast-speech processes could be explained by the increased cognitive load caused by reductions and deletions. The only drawback for the explanation through signal degradation is that, in the present study, trouble with understanding should not have been an issue since listeners were presented with the same sentence in four different versions over and over again. However, the same was true for the Bosker et al. (2016) study.

The other finding of the present study is the contrast between the results from the implicit rate normalization task in Experiment 1 and a task requiring explicit rate comparison judgments (Experiment 2). In Experiment 2, listeners were presented with pairs of two different versions of the carrier sentence and had to decide which version sounded faster or slower, respectively. Although most of the duration differences between sentences were smaller (i.e., half the size) than in Experiment 1, listeners based their judgments exclusively on duration. Moreover, even in the "same duration–different processes" condition, the presence versus absence of fast-speech processes was not taken into account. Note that dissociations between implicit and explicit processing have been shown in other cognitive domains such as visual processing (e.g., Vorberg, Mattler, Heinecke, Schmidt, & Scharzbach, 2003). However, the results of Experiment 2 were not in line with Koreman's (2006) results for explicit rate judgments, either. Here differences in experimental setup are the most likely explanation: Koreman (2006) used a large number of semantically differentiated sentences from a corpus of spontaneous speech, whereas a simpler setup was used in the present study. This was because the main focus of the present study was to compare perceived speaking rate for natural (fast) speech and fast linearly time-compressed clear speech in an implicit task (Experiment 1). Given that previous studies addressing these types of effects used simple designs (i.e., one or a few context sentences presented over and over again), a similarly tightly controlled set-up seemed like a good starting point for Experiment 1 and had to be maintained for the explicit task in Experiment 2.

In summary, the present study provides a first step into exploring the implicit perception of speaking rate by looking at effects of natural fast speech versus linearly time-compressed speech. At first glance, the results may seem surprising: the presence of natural fast-speech processes such as segmental reductions and deletions leads to the perception of faster speech. Although this is counter the traditional assumption that speaking rate is calculated as the number of segments per unit time (note that segmental deletions imply a lower number of realized segments and reductions tend to be related to articulator speed), all reviewed speech perception models can account for the effect. While exemplar models and probabilistic models of perception can deal with the effect via direct associations between fast speech and the presence of fast-speech processes, abstractionist models would assume increased cognitive load due to more difficult mapping between signal and representations. Together with other recent studies on external influences on the implicit perception of speaking rate (Bosker & Reinisch, 2015; Bosker et al., 2016), evidence appears to accumulate that normalization for speaking rate may in fact be explained at multiple levels of processing.

Acknowledgments This research was funded by a grant of the Alexander von Humboldt Foundation to the author. I would like to thank Holger Mitterer, Mary Stevens, and Hans-Rutger Bosker for comments on a previous version of the paper.

Appendix A

The table below shows the broad IPA transcriptions of the two versions of the context sentence, that is, without fast-speech processes (measures in the table on the left) and with fast-speech processes (measures in the table in the middle). As discussed in the main text, the two types of the sentence differed in the number of transcribed segments, and shared segments differed in their respective duration. Since vowels are especially prone to spectral reductions due to articulatory undershoot, spectral measures of the first and second formants are given in Hertz and the perceptual scale Bark. Measures were taken at the midpoints of the respective time intervals. Differences (on the left of the table) suggest that the vowels in the version of the sentence that had been spoken fast are, as would be expected, somewhat more centralized. The rather large difference in F2 for the schwa marked with * may result from the presence of an adjacent /r/ in the normal rate/ linearly compressed sentence that was completely absent in the natural-fast version.

natural normal rate speech							natural fast speech						spectral differences			
IPA	natural duration (sec)	compressed duration (sec)	F1 (Hz)	F2 (Hz)	F1 (bark)	F2 (bark)	IPA	duration (sec)	F1 (Hz)	F2 (Hz)	F1 (bark)	F2 (bark)	F1 (Hz)	F2 (Hz)	F1 (bark)	F2 (bark)
z	0.161	0.116					**s**	0.053								
i:	0.078	0.056	275	2678	2.88	14.86		0.057	281	2360	2.94	14.01	6	318	0.06	0.85
f	0.083	0.059					**f**	0.067								
	0.081	0.058	699	1722	6.54	11.91		0.037	502	1531	4.98	11.14	197	191	1.56	0.71
m	0.103	0.074					**m**	0.090								
i:	0.114	0.082	323	2784	3.35	15.13	**i**	0.039	379	2689	3.88	14.89	56	95	0.53	0.24
t	0.066	0.047					**t**	0.126								
	0.051	0.037	413	2372	4.19	14.04										
n	0.074	0.053					**n**	0.076								
i	0.103	0.073	425	2624	4.30	14.72	**i**	0.060	429	2625	4.34	14.73	4	1	0.04	0.01
r	0.038	0.027														
	0.029	0.021	543	1405	5.32	10.59		0.055	462	2070	4.63	13.13	81	665*	0.69	2.54
m	0.088	0.064					**m**	0.108								
t	0.142	0.102					**t**	0.123								
ɛ	0.040	0.029	472	2456	4.72	14.28	**ɛ**	0.030	442	2429	4.45	14.20	30	27	0.27	0.08
k	0.091	0.065					**k**	0.073								
s	0.040	0.029					**s**	0.030								
d	0.082	0.059					**t**	0.066								
	0.027	0.019	370	2410	3.80	14.15										
m	0.096	0.069					**m**	0.126								
b	0.025	0.018														
ɛ	0.034	0.025	390	2017	3.98	12.95										
g	0.077	0.055					**g**	0.062								
r	0.062	0.044														
	0.043	0.031	336	1309	3.48	10.14		0.082	277	1263	2.90	9.91	59	46	0.58	0.23
f	0.089	0.064					**f**	0.020								

References

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *Journal of the Acoustical Society of America, 126,* 2649–2659.

Ainsworth, W. (1972). Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America, 51,* 648–651.

Ainsworth, W. (1974). The influence of precursive sequences on the perception of synthesized vowels. *Language & Speech, 17,* 103–109.

Allen, J., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics, 63,* 798–810.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta- analytic review. *Acta Psychologica, 134,* 330–343.

Boersma, P., & Weenink, D. (2009). *PRAAT, doing phonetics by computer (Version 5.1) [Computer program]*. www.praat.org.

Bosker, H. R., & Reinisch, E. (2015). Normalization for speechrate in native and nonnative speech. *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, UK.*

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2016). Listening under cognitive load makes a sentence sound fast. *Paper presented at the Workshop on Speech Processing in Realistic Environments (SPIRE),* Groningen, The Netherlands.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 341–376). Cambridge: Cambridge University Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–253.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America, 72,* 705–716.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America, 83,* 1553–1573.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups. *Journal of the Acoustical Society of America, 88,* 101–112.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *The Journal of the Acoustical Society of America, 85,* 2154–2164.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664–1670.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua, 142,* 27–41.

Farmer, T., Brown, M., & Tanenhaus, M. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences, 36,* 211–212.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America, 63,* 223–230.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics, 59,* 675–692.

Janse, E., Nooteboom, S., & Quené, H. (2003). Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. *Speech Communication, 41,* 287–301.

Jessen, M. (1993). Stress conditions on vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory, 8,* 1–27.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory visual integration of talker gender in vowel perception. *Journal of Phonetics, 27,* 359–384.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology. Human Perception and Performance, 15,* 736–748.

Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. *Psychological Review, 122,* 148–203.

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.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35,* 1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling*. Dordrecht: Kluwer Academic.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39–74). Hillsdale: Erlbaum.

Miller, J. L. (1987). Rate-dependent processing in speech perception. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 3, pp. 119–157). London, England: Erlbaum.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology. Human Perception and Performance, 14,* 369–378.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41,* 215–225.

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.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395.

Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in read speech. In A. P. Simpson, K. J. Kohler, & T. Rettstadt (Eds.), *The Kiel corpus of read/spontaneous speech—Acoustic database, processing tools and analysis results, AIPUK 32* (pp. 215–247). Germany: IPDS.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America, 18,* 2561–2569.

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.

Reinisch, E. (2015). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics*. doi: 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.

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.

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

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics, 73,* 1195–1215.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2012). Hemispheric differences in the effects of context on vowel perception. *Brain and Language, 120,* 401–405.

Sjerps, M. J., & Reinisch, E. (2015). Divide and conquer: how perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception. *Journal of Experimental Psychology. Human Perception and Performance, 41,* 710–722.

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology: Results of the Third KONVENS Conference* (pp. 14–26). Berlin: Mouton de Gruyter.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology. Human Perception and Performance, 7,* 1074–1095.

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics, 35,* 203–213.

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review, 16,* 74–79.

Vitela, A., Warner, N., & Lotto, A. (2013). Perceptual compensation for differences in speaking style. *Frontiers in Psychology, 4,* 399.

Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Scharzbach, J. (2003). Different time courses for visual perception and action priming. *Proceedings of the National Academy of Sciences, 100,* 6275–6280.

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics, 2005*(67), 939–950.

Welch, T. E., Sawusch, J. R., & Dent, M. L. (2009). Effects of syllable-final segment duration on the identification of synthetic speech continua by birds and humans. *Journal of the Acoustical Society of America, 126,* 2779–2787.

Wilson, J. P. (1970). An auditory after-image. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 303–318). Leiden: Sijthoff.