Accepted for publication in Language, Cognition, & Neuroscience

(February 4, 2018)

Entrained theta oscillations guide perception of subsequent speech: behavioral evidence from rate normalization

Hans Rutger Bosker[ab][1], Oded Ghitza[cd]

[a]*Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH, Nijmegen, The Netherlands*

[b]*Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands*

[c]*Department of Biomedical Engineering, Hearing Research Center, Boston University, Boston, MA, USA*

[d]*Neuroscience Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany*

Word count: 9404

Figure count: 4

[1] Corresponding author. Tel.: +31 (0)24 3521 373.
E-mail address: HansRutger.Bosker@mpi.nl

Abstract

This psychoacoustic study provides behavioral evidence that neural entrainment in the theta range (3-9 Hz) causally shapes speech perception. Adopting the 'rate normalization' paradigm (presenting compressed carrier sentences followed by uncompressed target words), we show that uniform compression of a speech carrier to syllable rates *inside* the theta range influences perception of subsequent uncompressed targets, but compression *outside* theta range does not. However, the influence of carriers – compressed outside theta range – on target perception is salvaged when carriers are 'repackaged' to have a packet rate inside theta. This suggests that the brain can only successfully entrain to syllable/packet rates within theta range, with a causal influence on the perception of subsequent speech, in line with recent neuroimaging data. Thus, this study points to a central role for sustained theta entrainment in rate normalization and contributes to our understanding of the functional role of brain oscillations in speech perception.

Keywords: neural entrainment; theta oscillations; speech rate; rate normalization.

INTRODUCTION

Speech is a communicative signal with inherent slow amplitude modulations. These amplitude modulations fluctuate at a rate around 3-9 Hz, in various speech types and in different languages (Bosker & Cooke, in press; Ding et al., 2017; Krause & Braida, 2004; Varnet, Ortiz-Barajas, Erra, Gervain, & Lorenzi, 2017), driven primarily by the syllabic rate of speech. Recently, models of speech perception have pointed at the remarkable correspondence between the time scales of phonemic, syllabic, and phrasal linguistic units, on the one hand, and the periods of the gamma, theta, and delta oscillations in the brain, on the other (Ghitza, 2011; Poeppel, 2003). This correspondence has inspired recent hypotheses on the potential role of neuronal oscillations in speech perception (Ghitza, 2011, 2017; Ghitza & Greenberg, 2009; Giraud & Poeppel, 2012; Peelle & Davis, 2012; Poeppel, 2003). For instance, evidence has accumulated showing that the listening brain follows the syllabic speech rhythm by phase-locking endogenous theta oscillations (3-9 Hz) to the amplitude envelope of speech (Doelling, Arnal, Ghitza, & Poeppel, 2014; Gross et al., 2013; Luo & Poeppel, 2007; Peelle & Davis, 2012; Peelle, Gross, & Davis, 2013). This process, known as neural entrainment ('speech tracking'), has also been proposed to explain why low-frequency amplitude modulations in speech play a crucial role in perception (e.g., literature on locally time-reversed, interrupted, alternated, and filtered speech; Drullman, Festen, & Plomp, 1994a; Drullman, Festen, & Plomp, 1994b; Elliott & Theunissen, 2009; Ghitza, 2012; Peelle & Davis, 2012; Saberi & Perrott, 1999; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Ueda, Nakajima, Ellermeier, & Kattner, 2017). However, the functional role of this neural entrainment in speech perception remains a topic of debate: is entrainment causally involved in shaping successful speech perception (Riecke, Formisano, Sorger, Başkent, & Gaudrain, 2018; Zoefel, Archer-Boyd, & Davis, 2018) or is it merely a response-driven epiphenomenon of speech processing (Obleser, Herrmann, & Henry, 2012)? The present study will put forward psychoacoustic findings suggesting that, not only does neural entrainment to a particular syllable rate shape the decoding of *concurrent speech* (Experiment 1), but also that neural entrainment might persist when the entraining rhythm has ceased, influencing the perception of *subsequently presented words* (Experiments 2 and 3).

Neural entrainment shapes perception of *concurrent speech*

Evidence that neural entrainment shapes the decoding of *concurrent speech* (i.e., the speech presented during the time interval in which entrainment occurs) has been given by both behavioral and neuroimaging experiments. Behavioral evidence has come from studies using (highly) compressed (i.e., accelerated) speech, showing that the greater the compression, the more intelligibility is impaired (e.g., Dupoux & Green, 1997). Ghitza (2014) demonstrated that the intelligibility of compressed speech deteriorates particularly sharply when syllable rates exceed the upper frequency of the theta range (> 9 Hz; cf. Ghitza & Greenberg, 2009). He postulated that as long as the syllable rate is inside the theta frequency range, theta oscillations track the speech input, aligning neural theta cycles to the input syllabic rhythm[2]. When theta is in sync with the sensory syllable rate (i.e., in the 3-9 Hz range), intelligibility is high. However, when theta is out of sync, for instance, for syllable rates above 9 Hz, intelligibility drops. This account can also explain perceptual learning of compressed speech (i.e., better comprehension of compressed speech with greater exposure, even across talker changes; Adank & Janse, 2009; Dupoux & Green, 1997), since greater exposure would presumably allow for closer alignment of neural oscillations to the speech input.

This oscillations-based account of compressed speech perception predicts that the maximum information transfer rate of syllables - the auditory channel capacity - is 9 syllables per second, the maximum speech tempo where the theta oscillations are still in sync with the speech input. Support for this prediction has been given in Ghitza (2014) by ‘repackaging’ compressed speech: dividing a time-compressed waveform into fragments, called packets, and delivering the packets at a prescribed packet delivery rate by inserting silent intervals in between packets (cf. Figure 1 below). While compressed speech with syllable rates higher than 9 Hz (i.e., outside theta range) was largely unintelligible,

[2] In order to be able to track the syllabic irregularities of spontaneous speech (e.g., a stressed syllable followed by a non-stressed syllable), the theta oscillator belongs to a special class of oscillators termed flexible oscillators (Ghitza, 2011). Such oscillators are different in important respects from autonomous, rigid oscillators (cf. Ghitza, 2011).

repackaging these compressed signals such that the packet delivery rate fell below 9 packets per second (i.e., allowing theta oscillations to be in sync with the input signal) restored intelligibility to a large extent. Note that the acoustics inside the packet is the compressed signal; hence the phonetic material available to the listener is identical in both (compressed and repackaged) conditions. Only the packet delivery rate is different between the compressed and repackaged speech conditions, indicating that auditory channel capacity is defined by information transfer rate, mostly independent of speech tempo[3].

Neuroimaging findings support the idea that successful neural entrainment shapes the decoding of concurrent speech. For instance, Doelling et al. (2014) recorded magnetoencephalography (MEG) while participants listened to speech from which the slow syllabic amplitude fluctuations had been removed (by filtering; cf. Ghitza, 2012). The authors observed that neural entrainment was reduced and intelligibility decreased, relative to control. However, artificially reinstating the temporal fluctuations by mixing in ‘clicks’ at syllabic intervals restored envelope-tracking activity (Doelling et al., 2014) and hence speech intelligibility (Ghitza, 2012).

Neural entrainment shapes perception of *subsequent speech*

It has recently been suggested that stimulus-induced entrainment may persist even when the driving stimulus has ceased (Lakatos et al., 2013; Spaak, de Lange, & Jensen, 2014). Influencing effects of this persisting neural rhythm on the perception of subsequent speech would be strong evidence for the notion of neural entrainment as a causal factor shaping perception (i.e., rather than being a response-driven epiphenomenon of language comprehension). That is, because the entraining rhythm has already ceased, the possibility of response-driven effects is excluded. For instance, Hickok, Farahbod, and Saberi (2015) presented listeners with entraining white noise modulated at 3 Hz, followed by

[3] It is important to note the distinction between the *interruption* of speech via Gating (G. A. Miller & Licklider, 1950), and the *insertion* of silent gaps via repackaging: interruption removes part of the speech signal, while repackaging maintains all speech information with artificial distribution of information in time, defined by the packet rate. While phonemic restoration from interrupted speech (Mattys, Brooks, & Cooke, 2009; Warren, 1970) has been attributed to informational masking, the insertion of gaps provides additional decoding time: a gradual change in gap duration should be viewed as tuning the packet rate in a search for a better synchronization between the input information flow and the capacity of the auditory channel; the optimal range of the packet rate is dictated by the properties of cortical theta (Ghitza, 2011, 2014).

stationary noise. Participants' task was to detect near-threshold pure tones that were embedded in the stationary noise signal in half of the trials. The authors observed a perceptual oscillation in participants' tone detection performance, matching the period of the entraining stimulus (i.e., 3 Hz) and persisting over several cycles. This suggests that neural entrainment may be sustained after rhythmic stimulation, consequently influencing subsequent auditory perception.

Most studies on *speech perception* addressing the issue of persisting entrainment and ensuing effects thereof have adopted the 'rate normalization' paradigm. That is, the perception of a target speech sound ambiguous between two members of a durational contrast (e.g., in English, short /b/ vs. long /w/; in Dutch, short /ɑ/ vs. long /a:/) may be biased towards the longer phoneme (i.e., /w/ in English; /a:/ in Dutch) if presented after a preceding sentence (hereafter: carrier) produced at a *faster speech rate* (Bosker, Reinisch, & Sjerps, 2017; Kidd, 1989; Pickett & Decker, 1960; Reinisch & Sjerps, 2013; Toscano & McMurray, 2015). This process, known as rate normalization, has been argued to involve general-auditory processes, since it occurs in human and non-human species (Dent, Brittan-Powell, Dooling, & Pierce, 1997), is induced by talker-incongruent contexts (Bosker, 2017b; Newman & Sawusch, 2009), and even by non-speech (Bosker, 2017a; Gordon, 1988; Wade & Holt, 2005); in contrast to other rate-dependent perceptual effects, such as the Lexical Rate Effect (Dilley & Pitt, 2010; Pitt, Szostak, & Dilley, 2016). Rate normalization has been proposed to be related to the cognitive principle of *durational contrast* (Diehl & Walsh, 1989). That is, the perception of ambiguous speech segments should be biased towards longer (shorter) percepts in the context of other shorter (longer) surrounding segments (Wade & Holt, 2005). Even though this principle accounts for the general-auditory nature of rate normalization effects, it is as yet unclear how this cognitive principle might be neurally implemented.

One potential neurobiologically plausible mechanism involves sustained neural entrainment to the preceding sentence. Bosker (2017a) showed that anisochronous ('non-rhythmic') fast and slow carriers do not trigger rate normalization, suggesting that rate normalization is induced by the periodicity in the carrier. This is in line with an MEG experiment (Kösem et al., 2017), where native Dutch participants were presented with slow and fast carriers (amplitude modulations around 3 and

5.5 Hz, respectively), followed by (uncompressed) target words containing vowels ambiguous between short /ɑ/ and long /a:/. Behavioral vowel identification responses revealed a consistent rate normalization effect, with more long /a:/ responses after fast carriers. The MEG data showed that, during the carrier, neural theta oscillations efficiently tracked the dynamics of the slow and fast speech. Moreover, during the target word time window, theta oscillations in right superior temporal cortex were observed that corresponded in frequency to the preceding syllable rate, revealing persisting entrainment in the target window. In fact, the extent to which these theta oscillations persisted into the target window correlated with the observed behavioral biases: the more evidence for sustained entrainment in the target window, the greater the behavioral rate normalization effect (Kösem et al., 2017).

These findings suggest that neural oscillations actively shape speech perception (Bosker, 2017a; Bosker & Kösem, 2017; Peelle & Davis, 2012). Entrained theta oscillations would be thought to impose periodic phases of neuronal excitation and inhibition, thus sampling the input signal at the appropriate temporal granularity (Bosker, 2017a). In other words, entrainment at a higher theta frequency (e.g., 8 Hz) would raise the cortical 'sampling frequency'. If entrained neural rhythms persist after rhythmic stimulation has ended, this higher cortical 'sampling frequency' then also affects the perception of subsequent speech. That is, if a fast sentence is suddenly followed by an ambiguous target vowel (e.g., ambiguous between short /ɑ/ and long /a:/ in Dutch), entrainment to the fast sentence would lead to 'oversampling' the target vowel, inducing overestimation of the target's duration (i.e., more long /a:/ responses). Similarly, if the same target word is presented after a slow carrier, this would induce cortical 'undersampling', with the target's duration being underestimated (i.e., fewer long /a:/ responses).

Contextual effects of syllable rates inside and outside theta range

Oscillations-based accounts of rate normalization (Bosker, 2017a; Kösem et al., 2017; Peelle & Davis, 2012) would predict that, whenever theta oscillations are in sync with the syllable rate of the carrier (i.e., ranging from 3-9 Hz), they will consistently influence the cortical 'sampling frequency', with sustained effects influencing the perceived duration of following target segments. However, when carriers are compressed to syllable rates above 9 Hz, theta oscillations would be out of sync with the syllabic input rate, thus removing contextual influences on subsequent target perception. That is, these oscillations-based models predict an upper limit to the syllable rates that induce rate normalization: increasing syllable rates to up to nine syllables per second should bias perception of following ambiguous targets more and more towards longer percepts; however, no consistent contextual influences should be observed for syllable rates exceeding the theta range.

Interestingly, all studies on rate normalization in the literature so far have tested the effects of syllable rates inside the theta range of 3-9 Hz. As far as we could establish, the fastest syllable rates tested had a syllable rate of 8 Hz (e.g., Newman & Sawusch, 2009). As such, it is as yet unknown whether there is an upper limit to the contextual syllable rates that might induce rate normalization effects.

In this study, we tested the effect of rate normalization with contextual syllable rates beyond capacity. In Experiment 1, Ghitza (2014) was replicated with Dutch materials. In Experiment 2, Dutch participants were presented with time-compressed carriers followed by uncompressed target words containing vowels ambiguous between Dutch /ɑ/ and /a:/. The carriers were linearly compressed by various compression factors *κ*, resulting in some syllable rates within theta range (i.e., below 9 Hz) and some outside theta range (i.e., above 9 Hz). Instead of a gradual increase in long /a:/ responses with larger compression factors, oscillations-based models actually predict a gradual increase in the percentage of long /a:/ responses for carriers *with syllable rates up to 9 Hz*. However, when the carrier has a syllable rate above 9 Hz, cortical theta is assumed to be out of sync with the syllabic input rate, presumably eliminating any potential contextual effect on subsequent speech perception.

Restoring rate normalization through repackaging

If Experiment 2 would show that rate normalization effects are only triggered by syllable rates within theta range, then this would provide behavioral support for oscillations-based models of speeded speech comprehension. In turn, such a finding would predict that repackaging compressed speech (Ghitza, 2014; Ghitza & Greenberg, 2009) might be able to restore rate normalization effects. For instance, compressed speech with a syllable rate of 15 Hz would not be predicted to bias the perception of subsequent Dutch target vowels towards /a:/, because the syllable rate is outside theta. However, if this compressed speech signal would be repackaged such that the packet delivery rate would fall below 9 Hz, rate normalization effects should be restored.

Experiment 3 was designed to test the effect of repackaged speech on the perception of subsequent speech. Oscillations-based models would predict that compressed speech with a syllable rate above 9 Hz would not induce rate normalization, whereas repackaging this same acoustic signal such that the packet delivery rate would fall below 9 Hz would induce rate normalization (similar to compressed speech with syllable rates below 9 Hz).

EXPERIMENT 1

Experiment 1 served as an attempt to replicate the findings from Ghitza (2014) in Dutch, testing the *intelligibility* of Dutch speech recordings compressed with various compression factors κ such that resulting syllable rates would fall either within (i.e., below 9 Hz) or outside theta range (above 9 Hz). Moreover, it tested whether repackaging speech such that the resulting packet delivery rate falls within theta would restore intelligibility of highly compressed speech, as reported in Ghitza (2014). If Experiment 1, targeting overall intelligibility, succeeds in replicating the findings from Ghitza (2014), this allows us to use the Dutch speech materials for the subsequent experiments, targeting rate normalization.

Method

Participants. Native Dutch participants (N = 18) with normal hearing were recruited from the Max Planck Institute’s participant pool. They gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196). Data from one participant were excluded for failing to understand the experimental task. Data from another participant were excluded because she requested to stop the experiment, leaving data from 16 participants (12 females, 4 males; M_{age} = 21) for analysis.

Design and materials. Sentence materials were generated after Ghitza (2014). A male native speaker of Dutch (first author) was recorded, fluently producing 100 digit strings, followed by Dutch target words. Experiment 1 only made use of the recorded digit strings; description of the target words is provided in the method of Experiment 2.

Each digit string was comprised of 7 digits (e.g., *twee een vijf, vier zes vijf drie;* "two one five, four six five three"), approximately 2 seconds long, with an average syllable rate around 3 Hz (*SD* = 0.2). The speaker uttered the strings as a phone number, using a cluster of 3 digits followed by a cluster of 4 digits, without combining digits (e.g., no "sixteen", "three times five", etc.). All digits were used (0, 1, 2, 3, 4, 5, 6, 7, 9), except *acht* /ɑxt/ "eight", because it contained the critical vowel used in the target words in Experiment 2 and 3.

Digit strings were time-compressed using PSOLA in Praat (Boersma & Weenink, 2016). This manipulation alters a signal's duration while preserving spectral properties (e.g., original formant patterns and fundamental frequency contours are maintained; cf. Figure 1). The time-compression factor κ (i.e., the factor by which the original duration of an utterance was compressed; e.g., $\kappa = 5$ involves compressing speech to 20% of its original duration) was chosen such that $\kappa \in \{1, 2, 5\}$, resulting in three different experimental conditions. Signals with $\kappa = 1$ and signals with $\kappa = 2$ had an average syllable rate of around 3 Hz and 6 Hz, respectively. As such, their syllable rates fell below 9 Hz, generally assumed to be the upper limit of cortical theta and the maximum reliable information transfer rate through the auditory channel (i.e., auditory channel capacity; Ghitza, 2014). Signals with $\kappa = 5$ had an average syllable rate of around 15 Hz (i.e., above 9 Hz), falling outside cortical theta and auditory channel capacity.

[INSERT Figure 1 ABOUT HERE]

A fourth repackaged condition involved repackaging the $\kappa = 5$ condition (see Figure 1). Intervals of

66 ms were excised from the signal with $\kappa = 5$ and spaced apart by 100 ms using Praat. This manipulation resulted in a repackaged condition that contained acoustic-phonetic material that was identical to the $\kappa = 5$ condition, only with a significantly lower packet delivery rate of 6 Hz (i.e., within the cortical theta range). Finally, low-level speech-shaped noise was added to all digit strings (SNR = 20 dB; not shown in Figure 1).

Procedure. Stimulus presentation was controlled by Presentation software (v16.5; Neurobehavioral Systems, Albany, CA, USA). Each participant was presented with the four speech conditions ($\kappa = 1$; $\kappa = 2$; $\kappa = 5$; repackaged), with 80 digit strings per condition, chosen at random from the total set of 100.

Each trial started with a fixation cross presented on screen. After 500 ms, the auditory stimulus was played. At stimulus offset, the screen was replaced with a response screen. Participants were instructed to enter *only the last four digits* of the stimulus they had heard (using digits, e.g., “4653”), and hit Enter to proceed (i.e., self-paced). Requesting only the last four digits reduced the bias of memory load on error patterns and provided an opportunity for the presumed (cortical) theta oscillator to entrain to the input rhythm of the first three digits prior to the occurrence of the final four digits. Participants were instructed to always enter four digits, in the exact order in which they were spoken, encouraging participants to guess if they had missed any digit. After a response had been recorded, a blank screen was presented for 500 ms, and the next trial was initiated.

Results

Trials with responses with less than four digits (n = 53; <1%) were excluded from analyses. Proportion correct scores were calculated as the proportion of digits in a given string that were correctly registered in the correct position (i.e., for the digit string “215 4652”, the response “4652” received a proportion correct value of 1.0; the response “4352”, 0.75; the response “6452”, 0.5; etc.), and derived percentage correct scores are presented in Figure 2.

[INSERT Figure 2 ABOUT HERE]

Proportion correct scores were entered into a Generalized Linear Mixed Model (GLMM; Quené & Van den Bergh, 2008) with a logistic linking function, as implemented in the lme4 library (Bates, Maechler, Bolker, & Walker, 2015) in R (R Development Core Team, 2012) , with weights specified as the maximum number of correct digits per trial (i.e., 4). Condition was entered as predictor (categorical variable, dummy coded with the κ = 1 condition mapped onto the intercept), with Participant and Digit String entered as random factors with by-participant and by-digit string random slopes for Condition (Barr, Levy, Scheepers, & Tily, 2013).

This model revealed significant differences between the κ = 1 and the κ = 5 condition (β = -9.291, *SE* = 0.688, *z* = -13.499, $p < 0.001$; lower accuracy in κ = 5 vs. κ = 1), between the κ = 1 and the repackaged condition (β = -7.807, *SE* = 0.699, *z* = -11.157, $p < 0.001$; lower accuracy in repackaged vs. κ = 1), and between the κ = 1 and the κ = 2 condition (β = -3.997, *SE* = 0.691, *z* = -5.781, $p <$ 0.001; slightly lower accuracy in κ = 2 vs. κ = 1).

The categorical predictor Condition can only compare conditions to its intercept (set to κ = 1). In order to also gain insight into comparisons between other conditions, a mathematically equivalent GLMM was built, including a re-leveled Condition predictor, this time mapping the repackaged condition onto the intercept. This analysis revealed significant differences between the repackaged and the κ = 5 condition (β = -1.484, *SE* = 0.122, *z* = -12.147, $p < 0.001$; lower accuracy in κ = 5 vs. repackaged), and between the repackaged and the κ = 2 condition (β = 3.811, *SE* = 0.285, *z* = -12.147, $p < 0.001$; higher accuracy in κ = 2 vs. repackaged).

Interim discussion

Experiment 1 replicated the findings from Ghitza (2014) using Dutch materials. Speech intelligibility was high for syllable rates within the theta range (i.e., κ = 1 and κ = 2 with syllable rates below 9 Hz) but deteriorated sharply for syllable rates outside the theta range (i.e., κ = 5 with syllable rates above 9 Hz). Moreover, when maximally compressed speech (i.e., κ = 5) was repackaged such that the resulting packet delivery rate falls within theta range, the intelligibility of the speech was restored to a large degree. Thus, Experiment 1 validated the use of these materials for subsequent experiments.

EXPERIMENT 2

Experiment 2 was designed to test whether there is an upper limit to the contextual syllable rates that may induce rate normalization. Instead of a gradual increase in rate normalization effects as the compression factor (hence the syllable rate) is raised, oscillations-based models predict that rate normalization effects should not be induced by syllable rates above 9 Hz, since cortical theta would then be assumed to be out of sync with the syllabic input rate.

Method

Participants. Native Dutch participants (N = 21; 17 females, 4 males; M_{age} = 22), that had not participated in Experiment 1, with normal hearing were recruited from the Max Planck Institute's participant pool. They gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196).

Design and materials. A set of 10 digit strings (out of the total set of a hundred) were adopted for use in Experiment 2. These digit strings were compressed (PSOLA in Praat) using various time-compression factors κ, such that $\kappa \in \{1, 2, 3, 4, 5\}$, resulting in five different experimental conditions (see Figure 1). Signals with $\kappa \in \{1, 2, 3\}$ had an average syllable rate around 3, 6, and 9 Hz, respectively, falling just below the upper limit of cortical theta and the maximum reliable information transfer rate through the auditory channel (i.e., auditory channel capacity; Ghitza, 2014). Signals with $\kappa > 3$ had a syllable rate above 9 Hz, falling outside cortical theta and auditory channel capacity.

The male native Dutch speaker that had been recorded for Experiment 1 had also produced Dutch target words following each digit string. These target words were proper names selected from four different minimal pairs containing either the short vowel /ɑ/ or the long vowel /a:/: *Ad* - *Aad,* /ɑt - a:t/; *Bas* - *Baas,* /bɑs - ba:s/; *Dan* - *Daan,* /dɑn - da:n/; *Mart* - *Maart,* /mɑrt - ma:rt/. The /ɑ/-/a:/ vowel contrast in Dutch is cued by both spectral (lower formant values for /ɑ/, higher formant values for /a:/)

and temporal cues (shorter duration for /ɑ/, longer duration for /a:/; Escudero, Benders, & Lipski, 2009). Therefore, one natural /a:/ vowel token was selected, taken from the word *Baas*, that had a duration that fell in between the speaker's typical /ɑ/ and /a:/ durations (120 ms; i.e., in duration ambiguous between /ɑ/ and /a:/). This temporally ambiguous /a:/ vowel was manipulated to also be spectrally ambiguous between /ɑ/ and /a:/ using Burg's LPC method in Praat. Source and filter models were estimated automatically from the selected vowel. The formant values in the filter models were inspected and adjusted (*F1* = 820 Hz; *F2* = 1150 Hz) to fall in between the speaker's /ɑ/ and /a:/ formant values. Recombination of the source and adjusted filter model resulted in a vowel that was ambiguous between /ɑ/ and /a:/ in both its temporal and spectral properties (corroborated by pretesting). Finally, the vowel was spliced into fixed consonantal frames (/ʔ_t/; /b_s/; /d_n/; /m_rt/) to create target words.

Note that only using a single ambiguous vowel would have made the experiment needlessly difficult for participants, whose task was to categorize the vowel in the target word (i.e., negatively affecting participants' motivation). Therefore, filler trials were included that contained clear (i.e., unambiguous) /ɑ/ and /a:/ vowels.

Procedure. Each participant was presented with the 10 digit strings in all 5 compression conditions, followed by any of the 4 ambiguous minimal pairs (*n* = 200; or followed by any of the 8 unambiguous target words in filler trials; *n* = 400).

Each trial started with a fixation cross presented on screen. After 500 ms, a digit string was played, followed by a 100 ms silent interval and a target stimulus. At target offset, the fixation cross was replaced by a screen with two response options, one word on the left, another on the right (position of /ɑ/-/a:/ words counter-balanced across participants). Participants entered their response as to which of the two words they had heard (*Bas* or *Baas*, etc.) by pressing "1" for the option on the left, or "0" for the option on the right. After their response (or timeout after 4 seconds), the screen was replaced by an empty screen for 500 ms, after which the next trial was initiated.

Results

Trials with missing categorization responses (*n* = 6; <1%) were excluded from analyses. Categorization data, calculated as the percentage of long /aː/ responses (% /aː/), are presented in Figure 3, and were analyzed by a GLMM with a logistic linking function. The dependent variable was response /aː/ (coded as 1) or /ɑ/ (coded 0). Condition was entered as predictor (categorical variable, dummy coded with the $\kappa = 1$ condition mapped onto the intercept), with Participant and Digit String entered as random factors with by-participant and by-digit string random slopes for Condition (Barr et al., 2013).

[INSERT Figure 3 ABOUT HERE]

This model revealed significant differences between the $\kappa = 1$ and the $\kappa = 2$ condition ($\beta = 0.500$, $SE = 0.121$, $z = 4.107$, $p < 0.001$; higher percentage of /aː/ responses in $\kappa = 2$); and between the $\kappa = 1$ and the $\kappa = 3$ condition ($\beta = 0.353$, $SE = 0.136$, $z = 2.601$, $p = 0.009$; higher percentage of /aː/ responses in $\kappa = 3$). However, no differences were observed between the $\kappa = 1$, $\kappa = 4$, and $\kappa = 5$ conditions.

A mathematically equivalent GLMM with a re-leveled Condition predictor, this time mapping the $\kappa = 3$ condition onto the intercept, revealed additional significant differences between the $\kappa = 3$ and the $\kappa = 4$ condition ($\beta = -0.286$, $SE = 0.112$, $z = -2.552$, $p = 0.011$; lower percentage of /aː/ responses in $\kappa = 4$); and between the $\kappa = 3$ and the $\kappa = 5$ condition ($\beta = -0.326$, $SE = 0.117$, $z = -2.778$, $p = 0.005$; lower percentage of /aː/ responses in $\kappa = 5$).

Interim discussion

Experiment 2 was designed to test whether there is an upper limit to the syllable rates that elicit rate normalization effects on following ambiguous target words. We observed that compressing a naturally produced carrier by a factor of 2 induced a higher percentage of /aː/ responses for following ambiguous target words (relative to the uncompressed signal), replicating earlier findings in the literature (Bosker, 2017b; Bosker et al., 2017; Reinisch, 2016b; Reinisch & Sjerps, 2013).

A novel finding of Experiment 2 is that when carriers are compressed to have syllable rates outside the theta range (i.e., the κ = 4 and κ = 5 conditions), no difference in target word categorization is observed compared to the uncompressed condition (i.e., κ = 1). This finding supports an oscillations-based mechanism underlying rate normalization: only when theta oscillations can optimally track the rate of the carrier do we find effects on the perception of subsequent target words.

Note that the categorization data in the κ = 3 condition fall in between the results for the κ = 2 and the κ = {1, 4, 5} conditions. This observation may be a consequence of the fact that the average syllable rate in the κ = 3 condition is around 9 Hz, just at the upper border of the theta range. Recalling the biophysical nature of the neuronal theta, its frequency range is not precise and the 9 Hz limit should be considered as an estimated mean. Hence, a plausible explanation for this finding may be related to individual differences between participants in how successful their theta oscillations were in tracking the speech input at syllable rates near capacity.

EXPERIMENT 3

Experiment 3 was designed to test whether repackaged compressed speech may induce rate normalization. Since repackaging of heavily compressed speech can lower the packet delivery rate to below 9 Hz (i.e., within theta range), oscillations-based models predict that repackaging might restore rate normalization effects.

Method

Participants. Native Dutch participants (*N* = 29) with normal hearing, that had not participated in the previous experiments, were recruited from the Max Planck Institute’s participant pool. They gave informed consent as approved by the Ethics Committee of the Social Sciences department of Radboud University (project code: ECSW2014-1003-196). Data from 6 participants were excluded for reasons of technical errors (*n* = 1), illness (*n* = 1), fatigue (*n* = 2) or non-compliance with the experimental task (*n* = 2), leaving data from 23 participants (19 females, 4 males; M_{age} = 21) for analysis.

Design and materials. Experiment 3 combined the four speech conditions from Experiment 1 (κ = 1; κ = 2; κ = 5; repackaged) with the procedure from Experiment 2. That is, the 10 digit strings used in Experiment 2 were time-compressed and repackaged, using the method described in Experiment 1, and presented together with the target words, described in Experiment 2.

Procedure. The experimental procedure in Experiment 3 mirrored the procedure in Experiment 2. Each participant was presented with the 10 digit strings in all 4 conditions, followed by any of the 4 ambiguous minimal pairs (*n* = 160; or by any of the 8 unambiguous target words in filler trials; *n* = 320). Again, participants' task was to indicate which target word they had heard (*Bas* or *Baas*, etc.).

Results

Trials with missing categorization responses (*n* = 3; <1%) were excluded from analyses. Categorization data, calculated as the percentage of long /a:/ responses (% /a:/), are presented in Figure 4, and were analyzed by a GLMM with identical structure as the one built for Experiment 2.

[INSERT Figure 4 ABOUT HERE]

This GLMM revealed significant differences between the κ = 1 and the κ = 2 condition (β = 0.398, *SE* = 0.101, *z* = 3.927, $p < 0.001$; higher percentage of /a:/ responses in κ = 2); and between the κ = 1 and the repackaged condition (β = 0.272, *SE* = 0.100, *z* = 2.694, *p* = 0.007; higher percentage of /a:/ responses in repackaged condition). No difference was found between the κ = 1 and κ = 5 condition ($p > 0.7$).

A mathematically equivalent GLMM with a re-leveled Condition predictor, this time mapping the repackaged condition onto the intercept, revealed an additional significant difference between the repackaged condition and the κ = 5 condition (β = -0.245, *SE* = 0.100, *z* = -2.440, *p* = 0.015; lower percentage of /a:/ responses in κ = 5). However, no difference was found between the repackaged condition and the κ = 2 condition ($p > 0.2$).

Interim discussion

First, results from Experiment 3 replicate the findings for the κ = 1, κ = 2, and κ = 5 conditions from Experiment 2. That is, only when syllable rates were inside the theta range, rate normalization was observed.

The novel finding of Experiment 3 is that when the κ = 5 condition was repackaged such that its packet delivery rate was set to be around 6 Hz (inside theta range), rate normalization was restored. In fact, the categorization responses from the repackaged condition were comparable to those from the κ = 2 condition, with a syllable rate comparable to the packet rate of the repackaged condition (around 6 Hz). Note that the acoustics inside the packets in the compressed and repackaged condition were identical (i.e., time-compressed by κ = 5); nevertheless, very different contextual effects were observed on target word perception. Thus, Experiment 3, together with the replicated findings from Experiment 2, supports an oscillations-based account of rate normalization, with a central role for sustained entrainment of oscillations in the theta range in rate normalization.

GENERAL DISCUSSION

The present study contributes psychoacoustic data to the ongoing debate about the functional role of entrained neural oscillations in speech comprehension: does entrainment to a speech rhythm actively shape the decoding of the spoken input signal (i.e., a causal factor) or is it merely a manifestation of modulated evoked responses to the driving auditory stimulus (i.e., a consequence)? Previous studies have contributed to this debate by providing empirical evidence that neural entrainment guides the decoding of *concurrent speech*: when the natural speech rhythm is disrupted, entrainment is reduced (Doelling et al., 2014), and intelligibility suffers (Ghitza, 2012, 2014; Ghitza & Greenberg, 2009). A demonstration that entrainment influences the perception of *subsequent speech* (i.e., after the driving stimulus has ceased) – thus allowing the exclusion of potential evoked effects from the driving stimulus – could provide further evidence for the causal influence of neural entrainment on speech comprehension. Our study was concerned with providing such evidence by means of psychoacoustic experimentation.

Experiment 1 showed that compressing Dutch speech such that syllable rates fall outside the theta range (i.e., > 9 Hz) greatly harms intelligibility, in line with Ahissar et al. (2001) who showed that MEG neural responses track the temporal envelope of compressed speech as long as it is intelligible. This behavior is explained by the TEMPO model in Ghitza (2011), suggesting that the decline in intelligibility with speech speed is dictated by cortical theta (Ghitza, 2014)[4]. If the highly compressed speech signal is 'repackaged' (i.e., delivering compressed speech packets at a lower rate by inserting silent intervals in between packets; see Figure 1) such that the packet delivery rate falls within theta range, intelligibility is much enhanced. Together, these outcomes suggest that entrainment of theta oscillations supports the decoding of *concurrent speech*, extending earlier work in English (Ghitza, 2014) to a new language: Dutch.

Experiment 2 and 3 targeted potential effects of sustained entrainment on *subsequent speech* using the rate normalization paradigm: participants heard digit strings compressed by various compression factors, followed by (uncompressed) minimal pair target words ambiguous between containing the short vowel /ɑ/ vs. the long vowel /a:/ (e.g., *Dan - Daan*, /dɑn - da:n/). Experiment 2 demonstrated that compression of the digit strings biased listeners towards reporting more long /a:/ responses, corroborating previous studies on rate normalization (e.g., Bosker, 2017b; Newman & Sawusch, 2009; Reinisch & Sjerps, 2013; Toscano & McMurray, 2015). However, this only applied for those compressed speech conditions with syllable rates safely within theta range. In fact, the perception of target words preceded by compressed digit strings with syllable rates outside the theta range (i.e., well

[4] It is important to mention here a recent EEG-based study (Pefkou, Arnal, Fontolan, & Giraud, 2017) showing that brain-wave oscillations follow the rhythm of *comprehensible* fast speech, with speeds up to 14 syllables/sec. Does this finding contradict the 9 syllables/sec limit suggested by Ghitza (2014)? Interpreting their data Pefkou et al. (2017) suggested two distinct functional processes, a bottom-up theta driven and a top-down beta driven process, concurrently at play during speech perception. And so, even when the theta oscillator ceases to accurately track the input, thus preventing the bottom-up process from extracting important syllabic information from the sensory input, a top-down process helps to obtain the missing information. Hence, the brain-wave oscillations above 9 Hz – beyond the biophysical range of theta – reflect beta feedback. In relation to our study, note that the effectiveness of a top-down process is determined by the amount of contextual information in the speech stream. In Pefkou et al. (2017), participants listened to short stories with high semantic context, feeding beta oscillations and thus the brain-wave activity above 9 Hz. In contrast, our study (as well as Ghitza, 2014) used random digit strings that do not allow much opportunity for top-down processing (random sequences of digits; relatively short; no coherence between trials), hence involving mainly the bottom-up theta driven function.

above 9 Hz) was observed to be comparable to the baseline uncompressed condition. Thus, the present study is the first to demonstrate that there is an upper limit to the syllable rates that induce rate normalization, namely a limit around 9 Hz.

One study that may seem to be at odds with this suggestion is Wade and Holt (2005), who reported that tone sequences with a presentation rate of 25 Hz also elicited rate normalization. Note, however, that this study used non-speech carriers to elicit normalization effects on the perception of consonantal target segments (i.e., short /b/ vs. long /w/ in English). It could be argued that the perception of vowels (relatively long; syllable nuclei) is governed by slow theta oscillations (corresponding in modulation rate), whereas the perception of consonants (relatively short; higher modulation rate; syllable onsets and codas) would be sensitive to faster gamma oscillations in the 25-40 Hz range (Giraud & Poeppel, 2012). This, however, remains speculative and future studies may examine potentially differential normalization of vowels and consonants.

Experiment 3 also used the rate normalization paradigm but this time participants were presented with repackaged digit strings, followed by the ambiguous target words. Oscillations-based models predict that repackaging highly compressed speech would restore cortical speech tracking, with consequences for the temporal sampling of subsequent speech. Results indeed showed that, whereas highly compressed speech (with syllable rates > 9 Hz) did not have an influence on listeners' target vowel perception (replicating Experiment 2), repackaged compressed speech – with packet delivery rates inside theta range – did bias listeners to report more long /a:/ vowels.

Taken together, the present experiments point to a central role for sustained theta entrainment in rate normalization. Neural theta oscillations are suggested to entrain to the rhythmic properties of a driving speech stimulus (Giraud & Poeppel, 2012; Peelle & Davis, 2012), imposing an appropriate sampling regime onto the incoming sensory stimulus (cf. Experiment 1; Ghitza, 2012, 2014). These oscillations may persist for several cycles after stimulation has ceased (Hickok et al., 2015; Kösem et al., 2017; Lakatos et al., 2013; Spaak et al., 2014), allowing for the possibility that a preceding rhythm influences the perception of subsequent speech by imposing its own cortical 'sampling frequency'. For instance, fast rhythms would induce 'oversampling' of the following speech content, with listeners overestimating the duration of subsequent spoken segments (Bosker, 2017a; Bosker & Kösem, 2017). When syllable rates fall outside the theta range (i.e., > 9 Hz), theta oscillations cannot effectively entrain to the incoming speech signal, and consequently the sustained effect of these theta oscillations on subsequent perception is reduced (cf. Experiment 2). However, when the information transfer rate is tuned back to within the theta range (via repackaging), theta entrainment is restored, improving the decoding of the concurrent speech (intelligibility enhancement; cf. Experiment 1) and influencing subsequent perception (rate normalization; cf. Experiment 3).

Thus, this study should be viewed as providing a neural implementation of cognitive principles proposed to underlie rate normalization, such as the principle of durational contrast (Diehl & Walsh, 1989; Oller, Eilers, Miskiel, Burns, & Urbano, 1991; Wade & Holt, 2005). This principle was formulated to have the psycholinguistic function of biasing the perception of ambiguous speech segments towards longer (shorter) percepts when they occur in the context of other shorter (longer) surrounding segments. The neural implementation proposed here refines the description of this cognitive principle by showing that rate normalization is governed by the syllable/packet rate of the carrier (in line with Bosker, 2017a). Furthermore, it provides important constraints, derived from current oscillations-based models of speech perception, revealing an upper limit to rate normalization effects (i.e., at the upper frequency of the theta range; 9 Hz).

Note, however, that we do not claim that this is the only mechanism driving rate normalization effects in speech perception. The perception of an ambiguous target word can also be biased by rate manipulations in the following, rather than preceding, context (although typically with smaller effect sizes; J. L. Miller & Liberman, 1979; Newman & Sawusch, 1996; Sawusch & Newman, 2000). Moreover, subjective impressions of fast speech (i.e., without an acoustic increase in syllable rate) can induce rate normalization, such as when listening to speech with segmental reductions (Reinisch, 2016a), habitually fast talkers (Maslowski, Meyer, & Bosker, in press; Reinisch, 2016b), an unfamiliar language (Bosker & Reinisch, 2017), or while dual-tasking (Bosker et al., 2017). We adopt the recently proposed two-stage model of normalization processes in speech perception (Bosker et al., 2017), in which, at an early stage, an automatic general-auditory mechanism is operating independent from attentional demands. At a later stage, higher-level influences, such as subjective impressions, come into play involving cognitive rather than perceptual adjustments. The neural mechanism proposed here would be a candidate mechanism for the first general-auditory stage of normalization.

In sum, based on behavioral findings from three psychoacoustic experiments, the present study (1) concludes that rate normalization is governed by the syllable/packet rate of the carrier; and (2) puts forward an oscillations-based neurobiological mechanism of rate normalization. This account is in line with other behavioral studies (Bosker, 2017a; Bosker & Kösem, 2017; Ghitza, 2012, 2014; Ghitza & Greenberg, 2009; Hickok et al., 2015), neuroimaging data (Doelling et al., 2014; Kösem et al., 2017), and previously formulated cognitive principles, such as durational contrast (Diehl & Walsh, 1989). Thus, it augments our understanding of the functional role of entrainment in speech comprehension by proposing that entrainment not only guides the processing of *concurrent speech* but also shapes the perception of *following speech content*.

ACKNOWLEDGEMENTS

The first author was supported by a Gravitation grant from the Dutch Government to the Language in Interaction Consortium. The second author was supported by a research grant from the Air Force Office of Scientific Research. We would like to thank two anonymous reviewers for helpful comments and suggestions, Phillip Alday for advice on statistical analyses, and Milou Huijsmans for her help in testing participants.

REFERENCES

Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America, 126*(5), 2649-2659.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences, 98*(23), 13367-13372.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

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

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [computer program].

Bosker, H. R. (2017a). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception & Psychophysics, 79*(1), 333-343. doi:10.3758/s13414-016-1206-4

Bosker, H. R. (2017b). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(8), 1225-1238. doi:10.1037/xlm0000381

Bosker, H. R., & Cooke, M. (in press). Talkers produce more pronounced amplitude modulations when speaking in noise. *Journal of the Acoustical Society of America*.

Bosker, H. R., & Kösem, A. (2017). *An entrained rhythm's frequency, not phase, influences temporal sampling of speech.* Paper presented at the Proceedings of Interspeech 2017, Stockholm.

Bosker, H. R., & Reinisch, E. (2017). Foreign languages sound fast: evidence from implicit rate normalization. *Frontiers in Psychology, 8*, 1063. doi:10.3389/fpsyg.2017.01063

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast but does not modulate acoustic context effects. *Journal of Memory and Language, 94*, 166-176. doi:10.1016/j.jml.2016.12.002

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (Melopsittacus undulatus). *The Journal of the Acoustical Society of America, 102*(3), 1891-1897.

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*(5), 2154-2164.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*(11), 1664-1670.

Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews, Online version*. doi:10.1016/j.neubiorev.2017.02.011

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85*, 761-768.

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech recognition,". *J. Acoust. Soc. Am, 95*(5), 2670-2680.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America, 95*(2), 1053-1064.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 914.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology, 5*(3), e1000302.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37*(4), 452-465.

Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology, 2*(130).

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology, 3*, 238.

Ghitza, O. (2014). Behavioral evidence for the role of cortical θ oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology, 5*, 652.

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience, 32*(5), 545-561. doi:10.1080/23273798.2016.1232419

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica, 66*(1-2), 113-126.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511-517.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43*(2), 137-146.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11*(12), e1001752.

Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science, 26*(7), 1006-1013. doi:doi:10.1177/0956797615576533

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 736-748.

Kösem, A., Bosker, H. R., Takashima, A., Jensen, O., Meyer, A., & Hagoort, P. (2017). Neural entrainment determines the words we hear. *bioRxiv*. doi:10.1101/175000

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America, 115*(1), 362-378.

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron, 77*(4), 750-761.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001-1010.

Maslowski, M., Meyer, A. S., & Bosker, H. R. (in press). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000579

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology, 59*(3), 203-243.

Miller, G. A., & Licklider, J. C. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America, 22*(2), 167-173.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*(6), 457-465.

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics, 58*(4), 540-560.

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*(1), 46-65.

Obleser, J., Herrmann, B., & Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Frontiers in Human Neuroscience, 6*.

Oller, D. K., Eilers, R. E., Miskiel, E., Burns, R., & Urbano, R. (1991). The stop/glide boundary shift: Modelling perceptual data. *Phonetica, 48*(1), 32-56.

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

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378-1387.

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A.-L. (2017). θ-band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience, 37*(33), 7930-7938.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech, 3*(1), 11-17.

Pitt, M. A., Szostak, C., & Dilley, L. (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*(1), 334-345. doi:10.3758/s13414-015-0981-7

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245-255.

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*(4), 413-425.

R Development Core Team. (2012). R: A Language and Environment for Statistical Computing [computer program].

Reinisch, E. (2016a). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics, 78*(4), 1203-1217. doi: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. doi:10.1017/S0142716415000612

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*(2), 101-116.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*. doi:10.1016/j.cub.2017.11.033

Saberi, K., & Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature, 398*(6730), 760-760.

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics, 62*(2), 285-300.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303.

Spaak, E., de Lange, F. P., & Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *The Journal of Neuroscience, 34*(10), 3536-3544. doi:10.1523/JNEUROSCI.4385-13.2014

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience, 30*(5), 529-543.

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports, 7*.

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America, 142*(4), 1976-1989.

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics, 67*(6), 939-950.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392-393.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. *Current Biology*. doi:10.1016/j.cub.2017.11.071

FIGURES

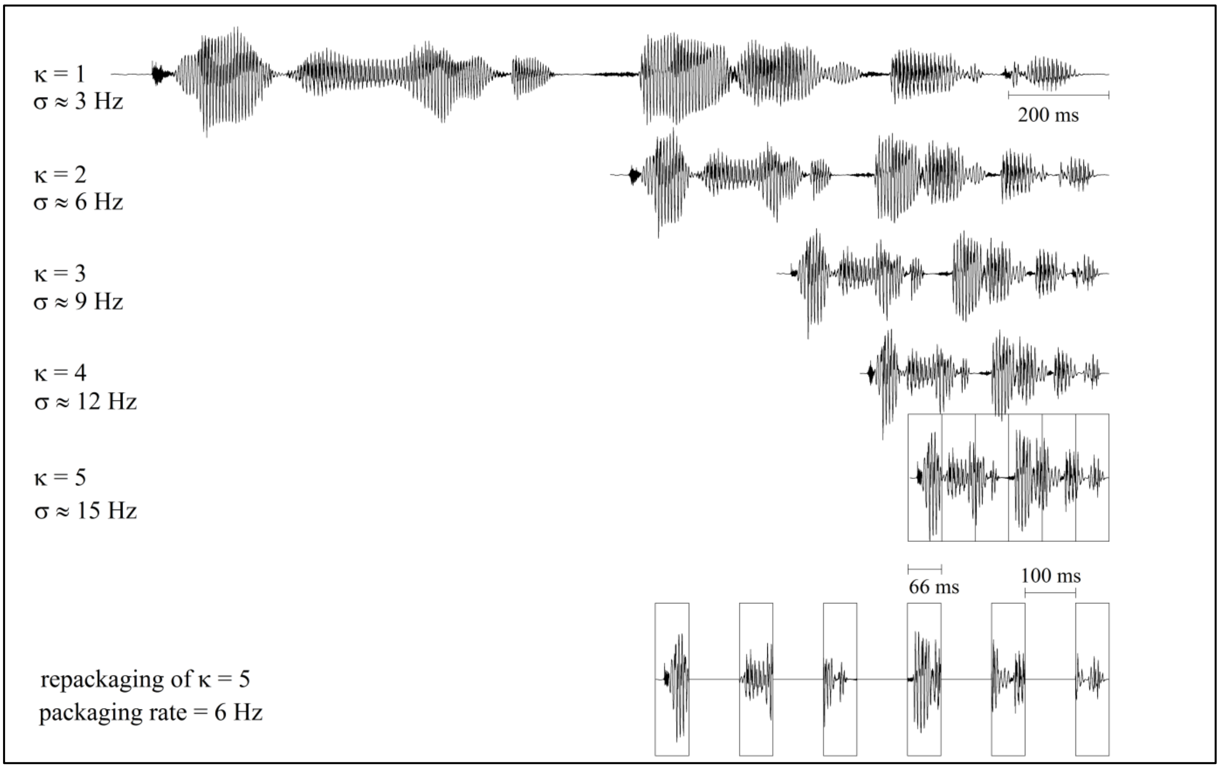

Figure 1. Waveforms for one example digit string, compressed at different time-compression factors *κ*, with syllable rate σ. The bottom waveform shows the repackaged condition, comprised of speech packets 66 ms long (taken from the *κ* = 5 condition with a syllable rate of about 15 syllables per second), spaced apart by 100 ms, resulting in a 6 Hz packet rate.

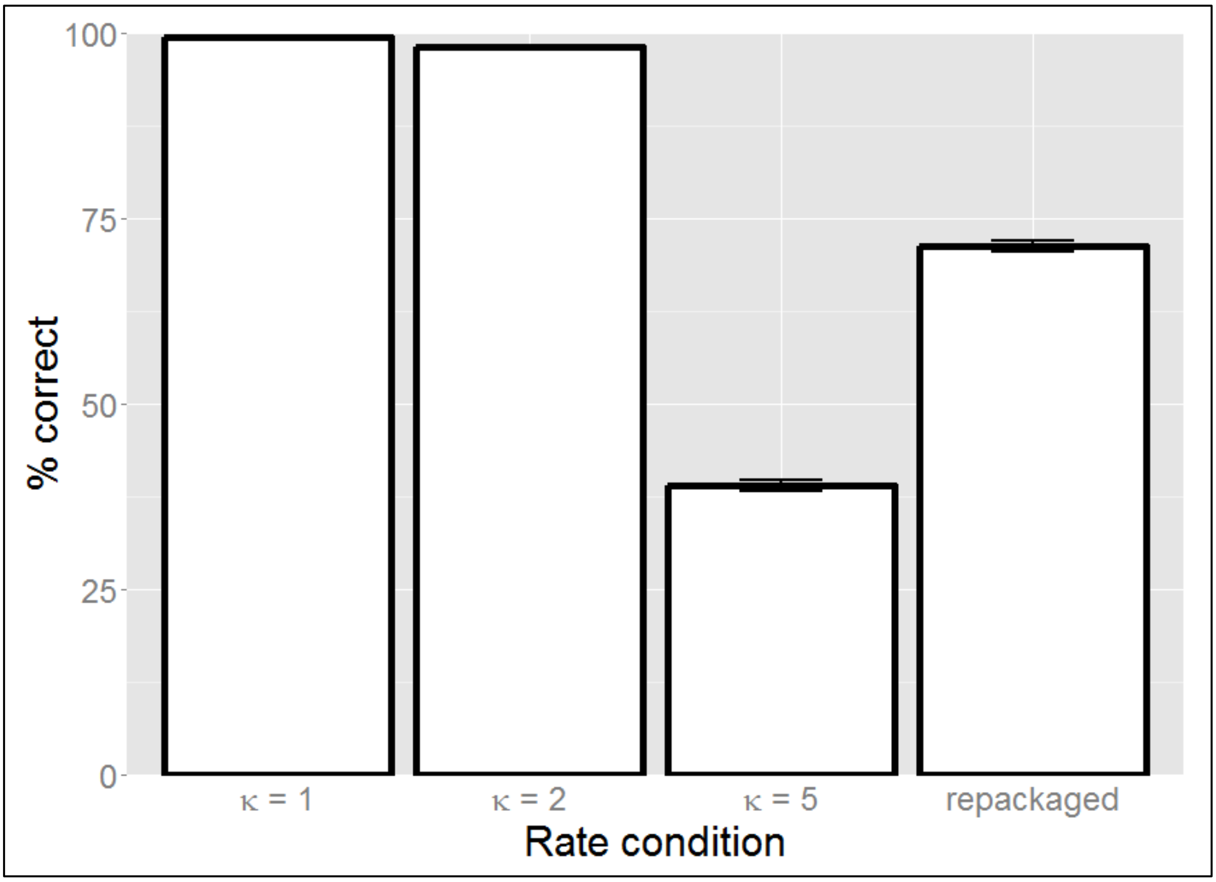

Figure 2. Percentage correct intelligibility scores for the four conditions with different time-compression factors *κ* from Experiment 1 (error bars show standard errors). Speech intelligibility is high for syllable rates within the theta range (i.e., *κ* = 1 and *κ* = 2 with syllable rates below 9 Hz) but deteriorates sharply for syllable rates outside the theta range (i.e., *κ* = 5 with syllable rates above 9 Hz). Moreover, when applying 'repackaging' such that the resulting packet delivery rate falls within theta range, speech intelligibility greatly improves.

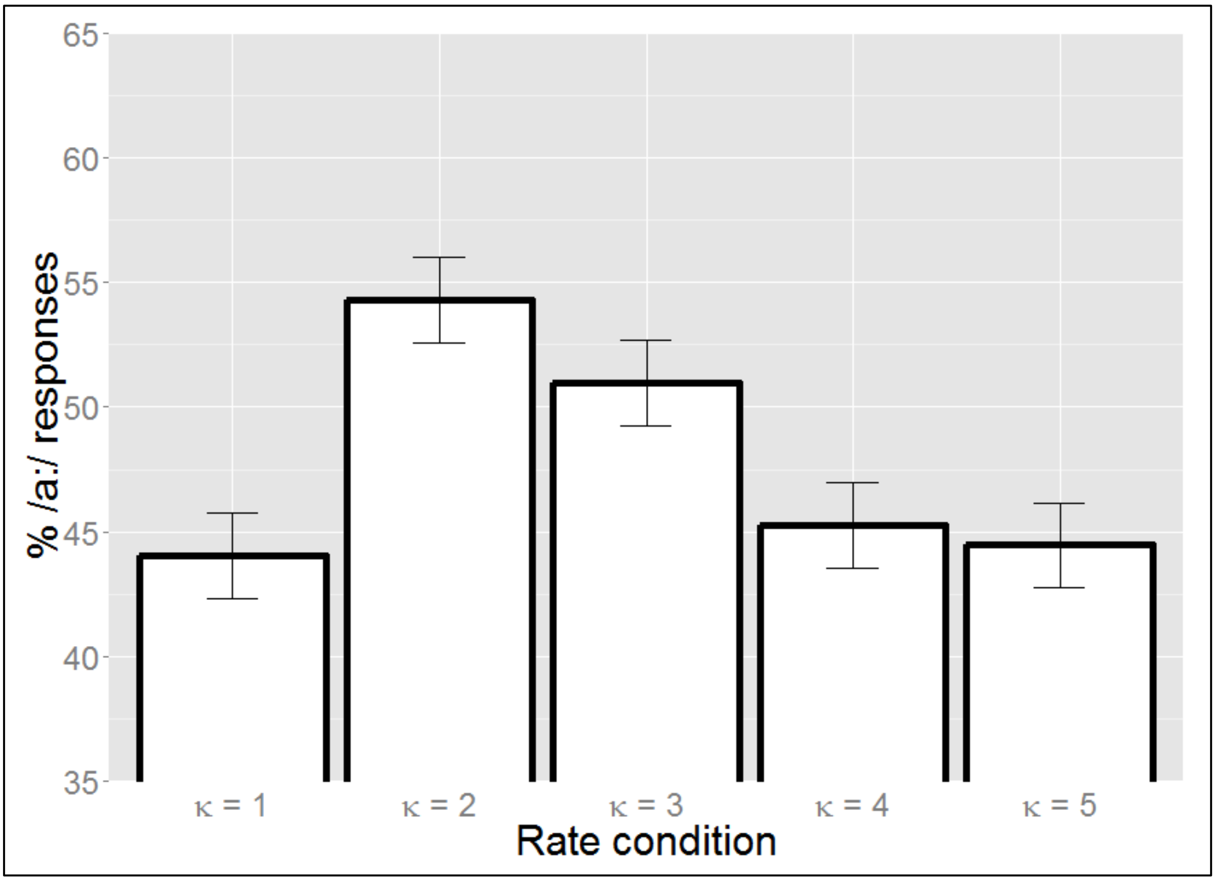

Figure 3. Average categorization data (in % long /a:/ responses) for the five conditions with different time-compression factors *κ* from Experiment 2 (error bars show standard errors). Compression of speech carriers by *κ* = 2, with syllable rates within the theta range, leads to an increase in % /a:/ responses. However, compression of carriers by *κ* = 4 and *κ* = 5, with syllable rates outside the theta range, does not lead to an increase in % /a:/ responses (comparable target categorization as in the baseline *κ* = 1 condition).

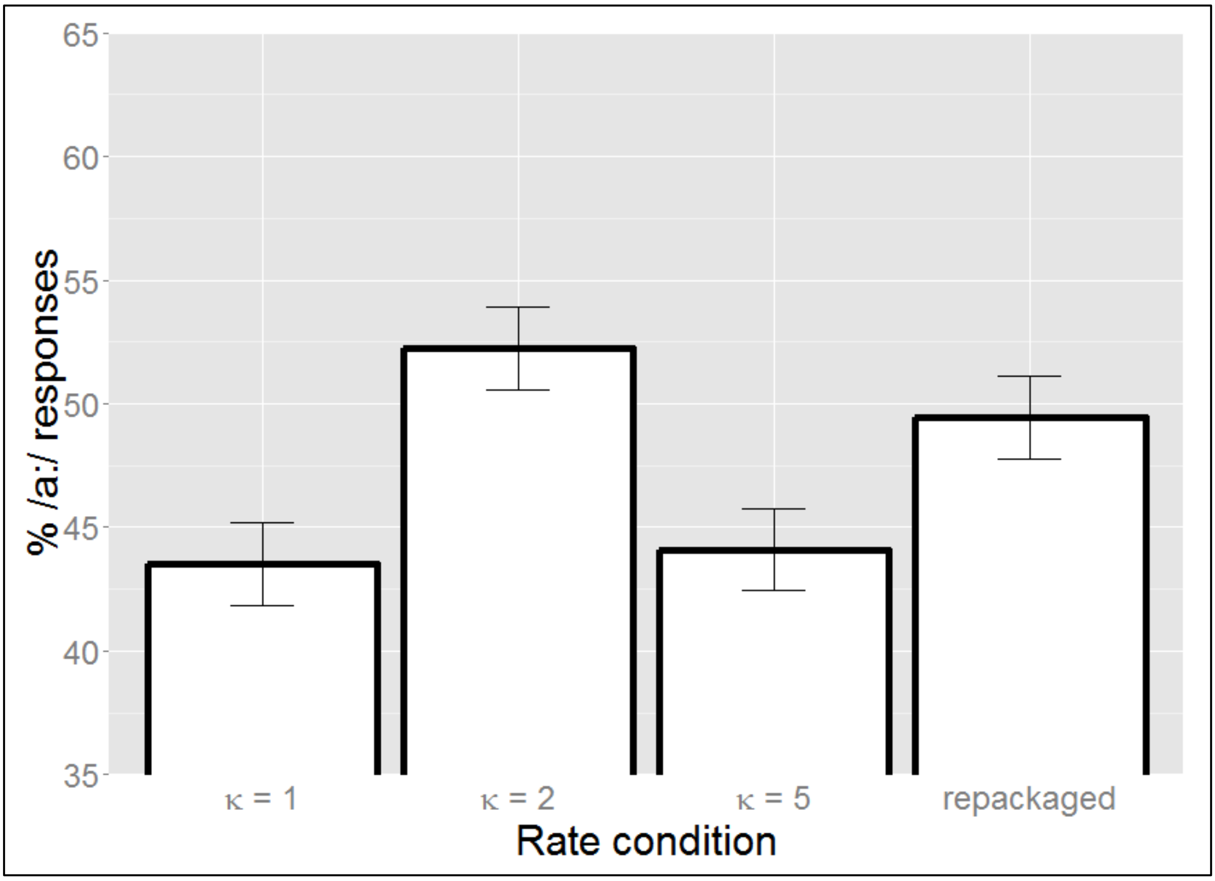

Figure 4. Average categorization data (in % long /a:/ responses) for the four conditions with different time-compression factors *κ* from Experiment 3 (error bars show standard errors). Similar to Experiment 2, compression of speech carriers by *κ* = 2, with syllable rates within the theta range, leads to an increase in % /a:/ responses. Also, compression of carriers by *κ* = 5, with syllable rates outside the theta range, does not lead to an increase in % /a:/ responses (comparable target categorization as in the baseline *κ* = 1 condition). However, when applying 'repackaging' to the *κ* = 5 condition, rate normalization is restored: the repackaged condition induces an increase in % /a:/ responses comparable to the *κ* = 2 condition.