

# The possible role of brain rhythms in perceiving fast speech: Evidence from adult aging

Lana R. Penn, Nicole D. Ayasse, and Arthur Wingfield

*Volen National Center for Complex Systems, Brandeis University, Waltham, Massachusetts 02454, USA*

Oded Ghitza<sup>a)</sup>

*Department of Biomedical Engineering, Hearing Research Center, Boston University, Boston, Massachusetts 02215, USA*

(Received 11 May 2018; revised 28 August 2018; accepted 31 August 2018; published online 11 October 2018)

The rhythms of speech and the time scales of linguistic units (e.g., syllables) correspond remarkably to cortical oscillations. Previous research has demonstrated that in young adults, the intelligibility of time-compressed speech can be rescued by “repackaging” the speech signal through the regular insertion of silent gaps to restore correspondence to the theta oscillator. This experiment tested whether this same phenomenon can be demonstrated in older adults, who show age-related changes in cortical oscillations. The results demonstrated a similar phenomenon for older adults, but that the “rescue point” of repackaging is shifted, consistent with a slowing of theta oscillations.

© 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5054905>

[MC]

Pages: 2088–2094

## I. INTRODUCTION

Spoken language is an inherently rhythmic phenomenon in which the acoustic signal is transmitted in syllabic “packets,” temporally structured so that most of the energy fluctuations occur in the range between 3 and 20 Hz (e.g., [Ding et al., 2017](#)). This rhythmic variation has been postulated to reflect properties of higher-order cortical processing, not just biomechanical and articulatory constraints. In particular, the supra-segmental properties of speech, especially in view of their variability from language to language, are more likely to be the consequence of factors other than articulation, reflecting temporal constraints associated with neural circuits in the cerebral cortex, thalamus, hippocampus, and other regions of the brain. That is, certain neural oscillations could be the reflection of both local and longer-range, trans-cortical processing (e.g., [Buzsáki, 2006](#); [von Stein and Sarnthein, 2000](#)).

The frequency range over which such oscillations operate (0.5–80 Hz) might thus serve as the basis for hierarchical synchronization through which the central nervous system processes and integrates sensory information (e.g., [Lakatos et al., 2005](#); [Singer, 1999](#)). In particular, there is a remarkable correspondence between the time scales of phonemic, syllabic, and phrasal (psycho)-linguistic units, on the one hand, and the periods of the gamma, beta, theta, and delta oscillations, on the other (e.g., [Ghitza, 2011](#); [Poepfel, 2003](#)). That is, phonetic features (duration of 20–50 ms) correspond to gamma (>50 Hz) and beta (15–30 Hz) oscillations, syllables, and words (mean duration of 250 ms) with theta (3–8 Hz) oscillations, and sequences of syllables and words embedded within a prosodic phrase (500–2000 ms) with delta oscillations (<2 Hz). This correspondence has inspired

recent hypotheses on the potential role of neuronal oscillations in speech perception (e.g., [Ahissar and Ahissar, 2005](#); [Ghitza, 2011, 2016](#); [Ghitza and Greenberg, 2009](#); [Giraud and Poeppel, 2012](#); [Peelle and Davis, 2012](#); [Poepfel, 2003](#)).

It has long been known that using artificial time-compression to increase speech rates to a level far beyond what is normally encountered in everyday listening hinders intelligibility at the word level and comprehension at the level of sentences (e.g., [Dupoux and Green, 1997](#); [Foulke and Sticht, 1969](#); [Garvey, 1953](#); [Peelle and Wingfield, 2005](#); [Reed and Durlach, 1998](#); [Versfeld and Dreschler, 2002](#); [Wingfield et al., 2003](#)). This could be due to a number of reasons, but one explanation is that in order to comprehend fast speech the brain needs extra “decoding time” (e.g., [Vagharchakian et al., 2012](#)).

[Ghitza and Greenberg \(2009\)](#) tested this possibility by presenting participants with increasing rates of time-compressed speech and then with “repackaged” speech, with packets containing speech at the same rates but now followed by silence gaps that provide extra decoding time (see Sec. [IIB](#) for a detailed explanation of the repackaging process). They found that while there was a sharp decline in intelligibility for speech compressed by a factor of 3, intelligibility was considerably restored in the repackaged condition. Crucially, the packaging rate within which intelligibility was restored was inside the theta frequency band. This finding led to the model TEMPO, which hypothesizes that the speech decoding process is performed within a hierarchical window structure generated by a cascade of oscillations driven by theta, which is capable of tracking the input syllabic rhythm ([Ghitza, 2011](#); [Ghitza and Greenberg, 2009](#). See also [Giraud and Poeppel, 2012](#); [Peelle and Davis, 2012](#)).

According to this model, cortical theta cycles will align and synchronize with the syllabic pseudo-rhythm of the speech input such that intelligibility will remain high as long

<sup>a)</sup>Also at: Neuroscience Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany. Electronic mail: [oghitza@bu.edu](mailto:oghitza@bu.edu)

as theta is in sync with the input, but sharply decrease once the rhythm of the input exceeds the frequency range of theta (e.g., when speech is time compressed by a factor greater than 3). Ghitza and Greenberg posited that for fast speech, where theta is out of sync, the insertion of gaps is an act of providing extra decoding time, and the gradual change in gap duration should be viewed as tuning the packaging rate in a search for a better synchronization between the input information flow and the capacity of the auditory channel. And since the resulting optimal packaging rate was within the theta frequency range, they concluded that cortical decoding time is determined by theta. Subsequent studies with young adults (Ghitza, 2012, 2014) have been consistent with this hypothesis, finding that intelligibility decreases as speech rate increases, but is rescued by the insertion of silent gaps which, in terms of the TEMPO model, restores the input rhythm into the theta range. It is worth noting that an alternative, plausible interpretation to the intelligibility rescue phenomenon could be suggested. According to this alternative interpretation, when neurons are “bombarded” with information at a high rate they saturate, and inserting silent gaps allows them to recover from this “fatigue.” This process could explain the observed recovery without necessarily involving neural entrainment. Using this interpretation, however, it is hard to explain why the intelligibility recovery is in the form of a U-shape, with the bottom of the U coinciding with the theta-cycle duration range (see Fig. 3 and discussion in Ghitza, 2011).

Within the context of the TEMPO model, one might raise the question of whether this same “repackaging” effect can be extended to populations that have different neuronal properties from the typical young adult. It has been suggested, for example, that aging is associated with changes in slow wave neural oscillations (e.g., delta and theta waves). The exact nature of these changes has yet to be fully specified (e.g., Vlahou *et al.*, 2014), although several studies have reported decreases in theta power in older relative to young adults (cf. Hashemi *et al.*, 2016; Klimesch, 1999; Leirer *et al.*, 2011; Rondina *et al.*, 2016; Vlahou *et al.*, 2014), and slower or noisier oscillations in older adults (e.g., Voytek and Knight, 2015; Voytek *et al.*, 2015). It is certainly the case at the behavioral level that older adults show a consistent general slowing in their processing speed (e.g., Cerella, 1994; McCabe *et al.*, 2010; Salthouse, 1996) even when otherwise cognitively healthy, along with certain differences in temporal processing stemming from changes in the auditory brainstem (Pichora-Fuller and Souza, 2003; Peelle and Wingfield, 2016; Skoe *et al.*, 2015; Strouse *et al.*, 1998; Walton, 2010). In particular, intelligibility by older adults is affected much more by an increase in speech speed compared to that of young adults (e.g., Gordon-Salant and Fitzgibbons, 1993; Wingfield *et al.*, 1999).

According to the TEMPO model, if older adults exhibited an earlier and steeper decline in intelligibility, and an earlier “rescue” effect of repackaging, it would be because of slower oscillations, theta in particular. If such results were found, TEMPO would see this in terms of a shift in the oscillator’s frequency range, such that cortical theta oscillations are slower in older adults compared to younger adults.

These questions were addressed by conducting a two-part experiment using spoken random-digit strings. Digits strings were used in order to eliminate the effect of linguistic context and to focus exclusively on bottom-up processes. First, the accuracy of digit recognition as a function of speech speed by young and older adults was recorded and served as a baseline performance. The speech rates used ranged from 6 to 21 syllables per second. In agreement with other studies, we would expect accuracy to drop with increasing speech rates and with a steeper rate of decline for the older adults. Of special interest, however, is the extent to which repackaging of the speeded speech results in intelligibility recovery. To test this, time-compressed digit strings, as described above, were repackaged, with packaging rates ranging from 2 to 7 packets per second (corresponding to the theta frequency range). Following the TEMPO model, we postulated that repackaging would rescue intelligibility for both young and older adults, with the recovery for the older adults appearing at a lower packaging rate.

## II. METHODS

### A. Participants

Participants were 24 young adults aged 18 to 33 yrs ( $M = 21.6$ ) and 24 older adults aged 62–90 yrs ( $M = 74.5$ ). All participants spoke English as their first language and all reported good health, with no known history of stroke, Parkinson’s disease, or other neurologic disorders that might affect their ability to perform the experimental task.

Audiometric evaluation was carried out for each participant using a GSI 61 clinical audiometer (Grason-Stadler, Inc., Madison, WI) by way of standard audiometric techniques (Harrell, 2002). The young adults had a mean better-ear pure tone threshold average (PTA) of 7.1 dB hearing level (HL) averaged over 500, 1000 and 2000 Hz and a mean better-ear speech reception threshold (SRT) of 10.5 dB HL. The older adults had a mean better-ear PTA of 23.8 dB HL and a mean better-ear SRT of 26.7 dB HL. Although as is typical for their age ranges, the older adults as a group tended to have elevated thresholds relative to the young adults (Morrell *et al.*, 1996), both the young and older participants fell within a pure tone acuity range considered to be clinically normal hearing for speech (Katz, 2002). None of the participants were regular users of hearing aids and all testing was conducted unaided.

### B. Stimuli

The experimental stimuli consisted of 100 spoken digit strings of 4 digits recorded by a male speaker of American English at a rate of 3 syllables per second. The four items in each digit list were randomly drawn from the digits 0 to 9.

Each list was preceded by four, 400 Hz tones, at a rate designed to provide an opportunity for the presumed cortical theta oscillator to entrain to the input rhythm of the digit stimuli prior to their presentation. The structure of the tone packets is described below.

*Uniform compression.* The waveforms of the recorded digit strings were time-compressed using a pitch-synchronous,

overlap and add procedure (Moulines and Charpentier, 1990) incorporated into PRAAT, a speech analysis and modification package (Phonetic Sciences, Amsterdam, the Netherlands). With this method the formant patterns and other spectral properties of the time-compressed signal are preserved but altered in duration. The fundamental frequency (“pitch”) contour, however, remains the same.

The compression factor was gradually increased to generate time-compressed versions of the digit strings, labeled by speech speeds measured in terms of syllables per second. Six speech speeds were used in this study: 6, 9, 12, 15, 18, and 21

syllables per second. Six versions of core stimuli (sets of 4-digit lists preceded by four entrainment tones) were created, one for each speech rate, by concatenating the time-compressed digit strings with four tone packets at a packet frequency that is equal to the corresponding speech speed. For example, a digit string with a speed of 12 syllables per second would be preceded by four tone packets at 12 Hz packet frequency, recorded with a 50% duty cycle (about a 40 ms packet duration) and a 400 Hz tone inside the packet [see Fig. 1(A)].

*Compression with repackaging.* Repackaging refers to the process of dividing the time-compressed waveform into

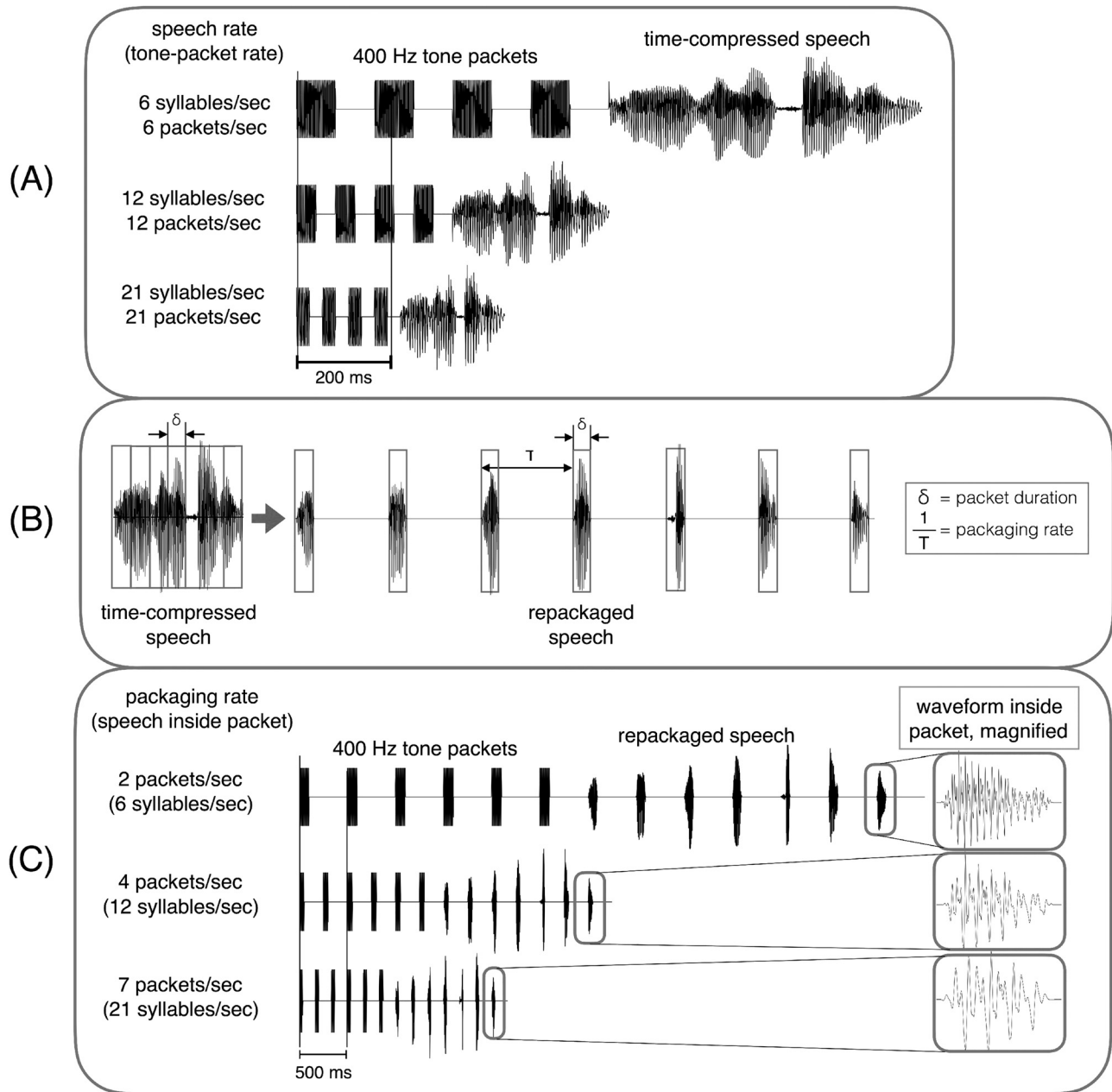


FIG. 1. Compression and repackaging process. (a) Example waveform of a 4-item stimulus list time-compressed to produce syllable rates of 6, 12, and 21 syllables per second. The speech waveform is shown preceded by a sequence of four 400 Hz tone-packets, delivered at the corresponding rate with a 50% duty cycle. (b) Illustration of the repackaging process. The left side shows the time-compressed waveform, blindly segmented into packets with equal duration of  $\delta$  (gray boxes). The right side shows the time-compressed waveform after repackaging, with a packaging rate of  $1/T$  packets per second. The acoustic signal inside a  $\delta$ -long packet is the time-compressed signal. (c) Repackaged condition for the digit string shown in (a) at packaging rates of 2, 4, and 7 packets per second, with the speech speed inside a packet being 6, 12, and 21 syllables per second, respectively, and a duty cycle of 20% across the six packaging rates. The speech waveform is shown preceded by a sequence of 400 Hz tone-packets, delivered at the corresponding packaging rate.

fragments, called packets, and delivering the packets at a prescribed rate (Ghitza and Greenberg, 2009). Figure 1(B) illustrates the repackaging process. The left panel shows the waveform of a sentence time-compressed by a factor of  $\kappa = 3$ . The compressed waveform is blindly segmented into packets with equal duration of  $\delta$  (gray boxes). The right panel shows the time-compressed waveform after repackaging, with a packaging rate of  $1/T$  packets per second (or Hz). The acoustic signal inside the  $\delta$ -long packet is the time-compressed signal. An exhaustive study on the interaction between the parameters of the repackaging process and its effect on the intelligibility and the quality of the speech output is yet to be conducted (see a few special cases in Ghitza and Greenberg, 2009; Ghitza, 2014). Although this would be an important area for future research, here we used a setting in which the duty cycle [ $\delta/T$  in Fig. 1(B)] is constant.

Six repackaging conditions were used in this study corresponding to the six speech speeds in the uniform time compression: 2, 3, 4, 5, 6, and 7 packets per second, with the speech rate inside a packet being 6, 9, 12, 15, 18, and 21 syllables per second, respectively. The packet duration was set to result in a constant duty cycle of 20% across the six packaging rates. Figure 1(C) shows example waveforms of repackaged digit-list stimuli at 2, 4, and 7 packets per second, with packets containing 6, 12, and 21 syllables per second uniformly time-compressed signals, respectively, each preceded by a sequence of tone packets recorded as described above, with a packet frequency corresponding to the packaging rate [see Fig. 1(C)].

### C. Procedure

Each participant received 120 experimental trials, each composed of a 4-digit string preceded by 4 tones. Half of the digit strings were time-compressed without repackaging while the other half were heard with the same compressed syllabic rate but with a repackaging rate at the respective  $\kappa$ . In order to keep the study at a manageable length and to avoid fatigue, half of the young and older adult participants received rates of 6, 12, and 18 syllables per second while the other half received rates of 9, 15, and 21 syllables per second for both the compressed and repackaged conditions. Each condition was preceded by three practice trials.

The repackaged and time-compressed conditions were presented in a blocked design, with the order of conditions counterbalanced across participants. The order of presentation of speech rates within conditions was also blocked, with the order of rates varied between participants.

Stimuli were presented binaurally over Eartone 3A (E-A-R Auditory Systems, Aero Company, Indianapolis, IN) insert earphones at 20 dB above each individual's better-ear SRT. The participant's task was to repeat back the four digits they heard. A trial was considered correct if all four digits were accurately repeated back and in the correct order.

An audibility check was conducted prior to the main experiment in which participants heard ten, 4-digit strings at a normal speech rate at the sound level that would be used for that participant in the main experiment. All participants

were able to correctly report accurately all 10 of the 4-item digit strings.

## III. RESULTS

Figure 2 shows the percentage of 4-item digit-lists reported correctly and in the correct order by the young and older adults as time-compression was used to systematically increase speech rate (indicated on the lower abscissa), and when the compressed signals were repackaged, with the repackaging rate, in packets per second (indicated in the upper-abscissa). Note that the syllable rate indicated in parentheses in the upper abscissa is the same as the rates indicated in the lower abscissa.

Visual inspection of the effects of the two conditions on speech intelligibility shows that for the uniform time-compression condition, while both age groups showed a performance decline with increasing speech rates, the older adults' accuracy was at floor level with speech rates as early as 12 syllables per second. Figure 2 also shows that both groups benefited from repackaging, with the vertical arrows indicating the young and older adults' "rescue points," defined as the earliest tested points at which repackaging yielded a significantly better performance relative to the same compression rates without repackaging. It can be seen that the rescue point occurred at an earlier for the older than that for the young adults; that is, the rescue point for the young adults occurred at 5 packets per second, at the center of the typical theta frequency range for young adults (3 to 8 Hz), while the rescue point for the older adults occurred at 4 packets per second.

### A. Statistical analyses

The accuracy data for the syllabic rates of 6, 12, and 18 and for 9, 15, and 21 syllables were analyzed with separate 3 (Speech rate: 6, 12, 18 or 9, 15, 21) syllables per second  $\times$  2 (Time-compression condition: compressed, repackaged)  $\times$  2 (Age: Young, Older) mixed design analyses of variance

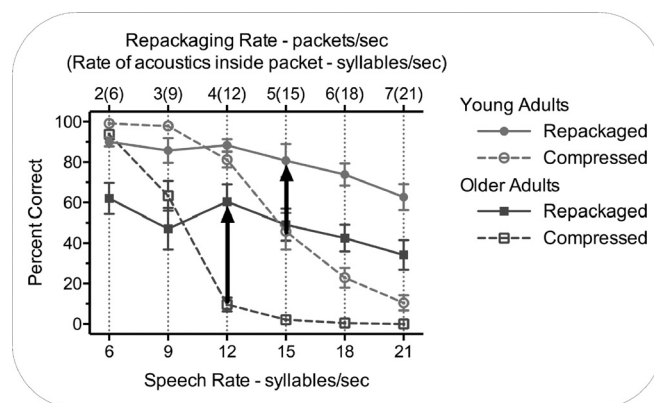


FIG. 2. Percentage of 4-item digit strings reported correctly and in the correct order for young and older adults as a function of increasing speech rate manipulated by time-compression of the speech signal indicated on the lower abscissa, versus the compressed signals repackaged, with the repackaging rate, in packets per second, indicated in the upper abscissa. The left and right arrows show the rescue points for the older and young adults, respectively. Error bars are one standard error.

(ANOVA), with syllabic rate and repackaging as within-participants' variables and age as a between-participants' variable. (Due to equipment failure the rate of 15 syllables per second in the repackaged condition was missing for one young adult participant).

The above-cited intelligibility decline with an increase in speech rate was reflected in a main effect of speech rate (Rates 6, 12, 18:  $F[2,44]=195.42$ ,  $p<0.001$ ,  $\eta_p^2=0.899$ ; Rates 9,15,21:  $F[2,42]=139.08$ ,  $p<0.001$ ,  $\eta_p^2=0.869$ ). The appearance of overall greater accuracy for the young adults relative to the older adults was confirmed by a significant main effect of age (Rates 6, 12, 18:  $F[1,22]=50.91$ ,  $p<0.001$ ,  $\eta_p^2=0.698$ ; Rates 9, 15, 21:  $F[1,21]=33.82$ ,  $p<0.001$ ,  $\eta_p^2=0.617$ ). Of special interest was the significant superiority with repackaging relative to time-compression without repackaging, which yielded a significant main effect of compression condition (Rates 6,12,18:  $F[1,22]=26.51$ ,  $p<0.001$ ,  $\eta_p^2=0.547$ ; Rates 9,15,21:  $F[1,21]=39.66$ ,  $p<0.001$ ,  $\eta_p^2=0.654$ ).

In addition to the main effect of compression condition, the ANOVA also confirmed that the intelligibility of the time-compressed speech was rescued by repackaging, especially at the faster speech rates. This was reflected in a significant Speech rate  $\times$  Compression condition interaction (Rates 6,12,18:  $F[2,44]=148.18$ ,  $p<0.001$ ,  $\eta_p^2=0.871$ ; Rates 9,15,21:  $F[2,42]=58.44$ ,  $p<0.001$ ,  $\eta_p^2=0.736$ ). There was also a significant Speech rate  $\times$  Age interaction (Rates 6,12,18:  $F[2,44]=21.33$ ,  $p<0.001$ ,  $\eta_p^2=0.492$ ; Rates 9,15,21:  $F[2,42]=7.78$ ,  $p=0.001$ ,  $\eta_p^2=0.270$ ). The benefit of repackaging was evident for both age groups, consistent with the absence of a Compression condition  $\times$  Age group interaction (Rates: 6,12,18:  $F[1,22]=0.33$ ,  $p=0.569$ ,  $\eta_p^2=0.015$ ; Rates 9,15,21:  $F[1,21]=0.85$ ,  $p=0.367$ ,  $\eta_p^2=0.039$ ). That the pattern was shifted for the two age groups, however, was reflected by a significant three-way Speech rate  $\times$  Compression condition  $\times$  Age group interaction (Rates 6,12,18:  $F[2,44]=37.64$ ,  $p<0.001$ ,  $\eta_p^2=0.631$ ; Rates 9,15,21:  $F[2,42]=3.73$ ,  $p=0.032$ ,  $\eta_p^2=0.0151$ ).

## B. Calculating the repackaging rescue point

In order to quantify the potential rescue effects of repackaging, and to clarify the source of the 3-way interaction obtained in the ANOVA, the rescue points between the compressed and repackaged conditions was examined. This was defined as the earliest tested point for each age group at which planned comparison testing showed that repackaging yielded significantly better performance than continuous time-compression.

*Young adults.* Although at or close to a performance ceiling at 6 and 9 syllables per second with continuous time compression, the young adults were nevertheless significantly more accurate in the time-compressed condition at syllable rates of 6 ( $t[11]=4.53$ ,  $p=0.001$ ) and 9 ( $t[11]=3.63$ ,  $p=0.004$ ) syllables per second (near or at performance ceiling). At a syllable rate of 12 syllables per second, the difference between conditions was not significant ( $t[11]=1.88$ ,  $p=0.087$ ), while by syllable rates of 15 ( $t[10]=3.33$ ,  $p=0.008$ ), 18 ( $t[11]=7.34$ ,  $p<0.001$ ), and 21

( $t[11]=10.49$ ,  $p<0.001$ ) syllables per second the young adults were significantly more accurate in the repackaged condition.

*Older adults.* At a syllable rate of 6 ( $t[11]=4.86$ ,  $p=0.001$ ) and 9 ( $t[11]=2.22$ ,  $p=0.049$ ) syllables per second the older adults were significantly more accurate in the time-compressed condition than in the repackaged condition. However, for syllable rates of 12 ( $t[11]=6.32$ ,  $p<0.001$ ), 15 ( $t[11]=6.00$ ,  $p<0.001$ ), 18 ( $t[11]=6.36$ ,  $p<0.001$ ), and 21 ( $t[11]=4.65$ ,  $p=0.001$ ) syllables per second, the older adults were significantly more accurate in the repackaged condition. Within the limits imposed by the number of rates tested, these data thus indicate a rescue point for the young adults by 5 packets per second, but that for the older adults this point has been reached by 4 packets per second, reflecting a shift in this rescue point in aging.

That the older adults show a significantly greater effect of compression condition at a syllable rate of 12 than do the younger adults was confirmed by a 2 (Time-compression condition: compressed, repackaged)  $\times$  (Age: Young, Older) mixed-design ANOVA for just the syllable rate of 12. This yielded a significant main effect of Compression condition ( $F[1,22]=42.52$ ,  $p<0.001$ ,  $\eta_p^2=0.659$ ) and a significant main effect of Age ( $F[1,22]=69.92$ ,  $p<0.001$ ,  $\eta_p^2=0.761$ ), as would be expected based on the previously reported results, as well as a significant interaction between Compression condition and Age ( $F[1,22]=24.26$ ,  $p<0.001$ ,  $\eta_p^2=0.524$ ). Bonferroni pairwise comparisons revealed that this interaction was due to the difference between Compression conditions for the older adults ( $p<0.001$ ) and the lack of a difference between Compression conditions for the young adults ( $p=0.272$ ). These analyses confirm the earlier rescue point in the older compared to the young adults.

## IV. DISCUSSION

The present study provides behavioral findings that are consistent with a postulated role of neuronal oscillations in speech perception. Oscillation-based models of speech perception (e.g., Ahissar and Ahissar, 2005; Ding and Simon, 2009; Ghitza and Greenberg, 2009; Ghitza, 2011; Giraud and Poeppel, 2012; Hyafil *et al.*, 2015; Lakatos *et al.*, 2005; Peelle and Davis, 2012; Poeppel, 2003; Shamir *et al.*, 2009) suggest a cortical computation principle by which speech decoding is performed within a time-varying window structure, synchronized with the input on multiple time scales. The window structure is generated by a cascade of flexible oscillations with theta as "master," capable of tracking the input pseudo-rhythm. From this it is argued that successful tracking can only be maintained if the input rhythm is within the theta frequency band (3 to 8 Hz).

These assertions were tested psychophysically in the present study, and the results were interpreted through the prism of TEMPO, a model that epitomizes oscillation-based models of speech perception. It had been shown that TEMPO is capable of explaining a variety of psychophysical and neuroimaging data difficult to explain by current models of speech perception, but emerging naturally from the

architecture of the model (e.g., Doelling *et al.*, 2014; Ghitza, 2012, 2014; Ghitza and Greenberg 2009). A key property that enables such accountability is the capability of the theta oscillator to track and stay locked to the input syllabic rhythm, hence providing a sufficient decoding time. (See also Peelle and Davis, 2012).

The current study aimed at examining whether the TEMPO model, which had previously only been tested in young adults, could be extended to older adults who are known to exhibit altered cortical oscillation patterns (cf. Hashemi *et al.*, 2016; Klimesch, 1999; Leirer *et al.*, 2011; Rondina *et al.*, 2016; Vlahou *et al.*, 2014; Voytek and Knight, 2015; Voytek *et al.*, 2015). One mechanism to explain these results, consistent with the TEMPO model, is that the decline in intelligibility that the older adults exhibit with time compression and the earlier rescue point are a consequence of altered neuronal oscillations (a shift downward) in the older adult compared to typical young adults, resulting in an early loss of synchronization with the input rhythm.

Our data show that, in the baseline time-compression condition, performance by the older adults was characterized by a steep decline in accuracy as a function of speech rate, a finding consistent with previous studies (e.g., Gordon-Salant and Fitzgibbons, 1993; Wingfield *et al.*, 1999, 2003). Indeed, the older adults performed at floor level at speeds as early as 12 syllables per second. Note that the knee-point in performance for young adults is a bit higher than the reported auditory channel capacity of 9 syllables/s (Ghitza, 2014). We believe that this is due to the low perplexity of our corpus (digits 0–9).

In the repackaging condition it was found that recovery occurs only at speeds higher than a “rescue point,” and performance at that region remains steady (i.e., the size of intelligibility recovery grows as the waveform speed increases). This behavior is in line with the notion that, when listening to time-compressed speech, additional decoding time is needed (e.g., Ghitza and Greenberg, 2009; Vagharchakian *et al.*, 2012), determined by neuronal entrainment to theta (as was elaborated in Sec. I). For the young adults, the rescue point occurs at 5 packets per second—at the center of the theta frequency range for typical young adults. This is in line with the assertion that the optimal range of packaging rate is inside the range of theta, and that the best synchronization between the speech stream and theta is achieved by tuning the packaging rate toward the mid-range of theta (Ghitza, 2011). Importantly, by contrast with the young adults, the rescue point for the older adults occurred by 4 packets per second, distinctively lower than that of the young adults. Interpreted through the prism of TEMPO, this result implies a decrease in the upper limit of the frequency range of the oscillations, or perhaps a potential shift downward in the frequency range of the theta oscillations with adult aging.

It is acknowledged that no hypothesis about internal physiological processes can be fully validated using only psychophysical methods. That the data are directly in line with the TEMPO model, however, motivate future psychophysical studies on the role neural oscillation in speech perception, and establish additional psychophysical context for

electrophysiological experiments that should use comparable tasks.

## ACKNOWLEDGMENTS

This research was supported by an AFOSR research Grant No. FA9550-11-1-0122 from the U.S. Air Force Office of Scientific Research (O.G.), NIH Grant No. AG019714 from the National Institute on Aging (A.W.), and NIH Grant No. T32 GM084907 from the National Institute of General Medical Sciences (N.D.A.).

- Ahissar, E., and Ahissar, M. (2005). “Processing of the temporal envelope of speech,” in *The Auditory Cortex. A Synthesis of Human and Animal Research*, edited by R. Konig, P. Heil, E. Buningner, and H. Scheich (Lawrence Erlbaum, Mahwah, NJ), pp. 295–312.
- Buzsaki, G. (2006). *Rhythms of the Brain* (Oxford University Press, New York).
- Cerella, J. (1994). “Generalized slowing and Brinley plots,” *J. Gerontol.* **49**, P65–P71.
- Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., and Poeppel, D. (2017). “Temporal modulations in speech and music,” *Neurosci. Biobehav. Rev.* **81**, 181–187.
- Ding, N., and Simon, J. Z. (2009). “Neural representations of complex temporal modulations in the human auditory cortex,” *J. Neurophysiol.* **102**(5), 2731–2743.
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). “Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual segmentation,” *NeuroImage* **85**, 761–768.
- Dupoux, E., and Green, K. (1997). “Perceptual adjustment to highly compressed speech: Effects of talker and rate changes,” *J. Exp. Psychol. Human Percept. Perform.* **23**(3), 914–927.
- Foulke, E., and Sticht, T. G. (1969). “Review of research on the intelligibility and comprehension of accelerated speech,” *Psychol. Bull.* **72**, 50–62.
- Garvey, W. D. (1953). “The intelligibility of speeded speech,” *J. Exp. Psychol.* **45**, 102–108.
- Ghitza, O. (2011). “Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm,” *Front. Psychol.* **2**, 130.
- Ghitza, O. (2012). “On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum,” *Front. Psychol.* **3**, 238.
- Ghitza, O. (2014). “Behavioral evidence for the role of cortical  $\theta$  oscillations in determining auditory channel capacity for speech,” *Front. Psychol.* **5**, 652.
- Ghitza, O. (2016). “Acoustic-driven delta rhythms as prosodic markers,” *Lang. Cogn. Neurosci.* **32**(5), 545–561.
- Ghitza, O., and 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**, 113–126.
- Giraud, A. L., and Poeppel, D. (2012). “Cortical oscillations and speech processing: Emerging computational principles and operations,” *Nat. Neurosci.* **15**(4), 511–517.
- Gordon-Salant, S., and Fitzgibbons, P. J. (1993). “Temporal factors and speech recognition performance in young and elderly listeners,” *J. Speech Hearing Res.* **36**, 1276–1285.
- Harrell, R. W. (2002). “Pure tone evaluation,” in *Handbook of Clinical Audiology*, 5th ed., edited by J. Katz (Lippincott, Williams, & Wilkins, Philadelphia, PA), pp. 71–87.
- Hashemi, A., Pino, L. J., Moffat, G., Mathewson, K. J., Aimone, C., Bennett, P. J., Schmidt, L. A., and Sekuler, A. B. (2016). “Characterizing population EEG dynamics throughout adulthood,” *eNeuro* **3**(6), ENEURO-0275.
- Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., and Giraud, A. L. (2015). “Speech encoding by coupled cortical theta and gamma oscillations,” *eLife* **4**, e06213.
- Katz, J. (2002). *Handbook of Clinical Audiology*, 5th ed. (Lippincott, Williams, & Wilkins, Philadelphia, PA).
- Klimesch, W. (1999). “EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis,” *Brain Res. Rev.* **29**, 169–195.

- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). "An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex," *J. Neurophysiol.* **94**(3), 1904–1911.
- Leirer, V. M., Wienbruch, C., Kolassa, S., Schlee, W., Elbert, T., and Kolassa, I. T. (2011). "Changes in cortical slow wave activity in healthy aging," *Brain Imaging Behav.* **5**(3), 222–228.
- McCabe, D. P., Roediger, H. L. III, McDaniel, M. A., Balota, D. A., and Hambrick, D. Z. (2010). "The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct," *Neuropsychology* **24**(2), 222–243.
- Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (1996). "Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level," *J. Acoust. Soc. Am.* **100**(4), 1949–1967.
- Moulines, E., and Charpentier, F. (1990). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Commun.* **9**, 453–467.
- Peelle, J. E., and Davis, M. H. (2012). "Neural oscillations carry speech rhythm through to comprehension," *Front. Psychol.* **3**, 320.
- Peelle, J. E., and Wingfield, A. (2005). "Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech," *J. Exp. Psychol. Human Percept. Perform.* **31**, 1315–1330.
- Peelle, J. E., and Wingfield, A. (2016). "The neural consequences of age-related hearing loss," *Trends Neurosci.* **39**(7), 486–497.
- Pichora-Fuller, M. K., and Souza, P. E. (2003). "Effects of aging on auditory processing of speech," *Intl. J. Audiol.* **42**(Suppl. 2), 11–16.
- Poeppel, D. (2003). "The analysis of speech in different temporal integration windows: Cerebral lateralization as asymmetric sampling in time," *Speech Commun.* **41**, 245–255.
- Reed, C. M., and Durlach, N. I. (1998). "Note on information transfer rates in human communication," *Presence* **7**(5), 509–518.
- Rondina, R., Olsen, R. K., McQuiggan, D. A., Fatima, Z., Li, L., Oziel, E., Meltzer, J. A., and Ryan, J. D. (2016). "Age-related changes to oscillatory dynamics in hippocampal and neocortical networks," *Neurobiol. Learn. Mem.* **134**, 15–30.
- Salthouse, T. A. (1996). "The processing-speed theory of adult age differences in cognition," *Psychol. Rev.* **103**, 403–428.
- Shamir, M., Ghitza, O., Epstein, S., and Kopell, N. (2009). "Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale," *PLoS Comput. Biol.* **5**(5), e1000370.
- Singer, W. (1999). "Neuronal synchrony: A versatile code for the definition of relations?," *Neuron* **24**(1), 49–65.
- Skoe, E., Krizman, J., Anderson, S., and Kraus, N. (2015). "Stability and plasticity of auditory brainstem function across the lifespan," *Cerebral Cortex* **25**(6), 1415–1426.
- Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). "Temporal processing in the aging auditory system," *J. Acoust. Soc. Am.* **104**(4), 2385–2399.
- Vagharchakian, L., Dehaene-Lambertz, G., Pallier, C., and Dehaene, S. (2012). "A temporal bottleneck in the language comprehension network," *J. Neurosci.* **32**(26), 9089–9102.
- Versfeld, N. J., and Dreschler, W. A. (2002). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," *J. Acoust. Soc. Am.* **111**(1), 401–408.
- Vlahou, E. L., Thurm, F., Kolassa, I. T., and Schlee, W. (2014). "Resting-state slow wave power, healthy aging and cognitive performance," *Sci. Reports* **4**, 5101.
- von Stein, A., and Sarnthein, J. (2000). "Different frequencies for different scales of cortical integration: From local gamma to long range alpha / theta synchronization," *Intl. J. Psychophysiol.* **38**, 301–313.
- Voytek, B., and Knight, R. T. (2015). "Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease," *Biol. Psychiatry* **77**, 1089–1097.
- Voytek, B., Kramer, M. A., Case, J., Lepage, K. Q., Tempesta, Z. R., Knight, R. T., and Gazzaley, A. (2015). "Age-related changes in 1/f neural electrophysiological noise," *J. Neurosci.* **35**(38), 13257–13265.
- Walton, J. P. (2010). "Timing is everything: Temporal processing deficits in the aged auditory brainstem," *Hear. Res.* **264**(1), 63–69.
- Wingfield, A., Peelle, J. E., and Grossman, M. (2003). "Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults," *Aging Neuropsychol. Cogn.* **10**(4), 310–322.
- Wingfield, A., Tun, P. A., Koh, C. K., and Rosen, M. J. (1999). "Regaining lost time: Adult aging and the effect of time restoration on recall of time-compressed speech," *Psychol. Aging* **14**, 380–389.