

Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners^{a)}

Sandie Keerstock^{b)} and Rajka Smiljanić

Department of Linguistics, The University of Texas at Austin, Austin, Texas 78712, USA

(Received 25 March 2018; revised 10 October 2018; accepted 25 October 2018; published online 20 November 2018)

The goal of the study was to examine whether enhancing the clarity of the speech signal through conversational-to-clear speech modifications improves sentence recognition memory for native and non-native listeners, and if so, whether this effect would hold when the stimuli in the test phase are presented in orthographic instead of auditory form (cross-modal presentation). Sixty listeners (30 native and 30 non-native English) participated in a within-modal (i.e., audio-audio) sentence recognition memory task (Experiment I). Sixty different individuals (30 native and 30 non-native English) participated in a cross-modal (i.e., audio-textual) sentence recognition memory task (Experiment II). The results showed that listener-oriented clear speech enhanced sentence recognition memory for both listener groups regardless of whether the acoustic signal was present during the test phase (Experiment I) or absent (Experiment II). Compared to native listeners, non-native listeners had longer reaction times in the within-modal task and were overall less accurate in the cross-modal task. The results showed that more cognitive resources remained available for storing information in memory during processing of easier-to-understand clearly produced sentences. Furthermore, non-native listeners benefited from signal clarity in sentence recognition memory despite processing speech signals in a cognitively more demanding second language.

© 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5078589>

[SHF]

Pages: 2871–2881

I. INTRODUCTION

Understanding speech is implicit and automatic in favorable listening conditions (Rönnberg, 2003; Rönnberg *et al.*, 2013; Rönnberg *et al.*, 2008). However, daily communication often occurs in noise, in a foreign language, or with hearing loss. Under these circumstances, speech processing becomes more demanding, reducing recognition, understanding, and recall (Hygge *et al.*, 2015; Lecumberri *et al.*, 2010; Ljung *et al.*, 2013; Pichora-Fuller *et al.*, 1995; Pichora-Fuller and Souza, 2003; Souza *et al.*, 2015). Processing acoustically degraded or ambiguous signals requires listeners to engage more cognitive resources, leaving fewer of these resources for subsequent processing, such as storing linguistic information in memory (Koeritzer *et al.*, 2018; Rabbitt, 1990; Rönnberg *et al.*, 2013; Tun *et al.*, 2009). Here, we focus on how acoustic clarity of the speech signal (conversational and clear speech), listener characteristics (native and non-native speaker of English), and modality of presentation (within and cross modalities) affect these cognitive demands.

Two recent studies examined the effect of a listener-oriented clear speaking style on the robustness of memory representations in native English listeners (Gilbert *et al.*, 2014; Van Engen *et al.*, 2012). Talkers modify their spoken output when communicating with non-native speakers or

listeners with hearing impairments (Lindblom, 1990). Conversational-to-clear speech adjustments are typically characterized by: decreases in the speaking rate, increases in the dynamic pitch range and amplitude, more salient release of stop consonants, expansion of the vowel space, and enhancement of language-specific vowel and consonant contrasts (Cooke *et al.*, 2013; Ferguson and Kewley-Port, 2002; Pichora-Fuller *et al.*, 2010; Smiljanić and Bradlow, 2009). These modifications improve intelligibility for a variety of listener groups and degraded listening conditions (Bradlow and Bent, 2002; Ferguson, 2012; Krause and Braida, 2002; Payton *et al.*, 1994; Picheny *et al.*, 1985; Schum, 1996). In addition to improving intelligibility, Van Engen *et al.* (2012) found that meaningful sentences and sentences produced in clear speech were easier to recognize as previously heard than anomalous sentences or sentences produced in a more casual speech. The same benefit of clear speech and noise-adapted speech (another intelligibility-enhancing speaking style adaptation) on sentence recognition memory was found even when listeners were exposed to sentences mixed with noise (Gilbert *et al.*, 2014). These findings are in line with the “effortfulness hypothesis” (McCoy *et al.*, 2005; Rabbitt, 1968, 1990) and the “ease of language understanding” model (Rönnberg *et al.*, 2013; Rönnberg *et al.*, 2008) in that more cognitive resources remain available for storing information in memory during processing of easier-to-understand clear speech. Conversely, listening to acoustically challenging speech requires listeners to use more cognitive resources during speech processing, thereby depleting the cognitive resources available and needed to encode speech in memory (Peelle, 2018; Pichora-Fuller *et al.*, 2016). Even in the

^{a)}Portions of this work were presented at the 173rd Meeting of the Acoustical Society of America and the 8th Forum Acusticum, Boston, MA, USA, June 2017.

^{b)}Electronic mail: keerstock@utexas.edu

absence of signal degradation such as noise, casually produced conversational speech can be challenging to process due to the extreme reduction and even deletion of many speech segments or whole syllables (Johnson, 2004; Pluymaekers *et al.*, 2005a,b; Warner *et al.*, 2009). Processing reduced forms, which deviate from expected targets and lexical representations, may incur additional costs in terms of cognitive resources. Reduced memory retention for conversational speech compared to clear speech may reflect the use of more cognitive resources during perception of conversational speech.

In addition to variations in signal clarity, listeners may face linguistic challenges that require the use of additional cognitive resources during speech processing. Listening in a second language (L2) is difficult and effortful, and this is reflected at all levels of processing, from perceptual discrimination of sound contrasts to phonotactics and prosody (Best and Tyler, 2007; Cutler *et al.*, 2008; Flege, 1995; Francis *et al.*, 2008; Iverson *et al.*, 2003; Kondaurova and Francis, 2008). Bradlow and Bent (2002) and Bradlow and Alexander (2007) found that non-native listeners benefited from clear speech although the intelligibility benefit was smaller compared to native listeners. Smiljanić and Bradlow (2011) found that the intelligibility benefit for highly proficient non-native listeners could be similar to that of native listeners, but to achieve the same level of accuracy, non-native listeners needed a more favorable signal-to-noise ratio (SNR). The smaller clear speech intelligibility benefit for non-native listeners might in part arise from their lack of experience in attending to the relevant dimensions of vowel and consonant contrasts, which are enhanced in a language-specific way (Gagné *et al.*, 2002; Smiljanić and Bradlow, 2005, 2008; Uchanski, 1988). With regard to the effect of linguistic experience on recognition memory, Hygge *et al.* (2015) found that recall of words presented in noise was lower in L2 than in the first language (L1), but also that decreasing the SNR affected word recall equally in L1 and L2. Molesworth *et al.* (2014) also showed that recall in noise was lower for L2 words than L1 words, but recall in noise in L2 could be improved by noise cancelling headphones.

The present study builds on previous work in two ways. First, we examined whether the clear speech benefit on sentence recognition memory extends to non-native listeners (Experiment I). We predicted that speech clarity would enhance sentence recognition memory for both native and non-native listeners, but that the magnitude of the recognition memory benefit for sentences would be smaller in L2, as seen with word-recognition-in-noise and recognition memory for words (cf. Hygge *et al.*, 2015; Molesworth *et al.*, 2014). Second, we examined whether the clear speech benefit on sentence recognition memory would hold in a cross-modal presentation for native and non-native listeners (Experiment II). Specifically, we tested whether the clear speech benefit is facilitated by the presence of the same acoustic signal in the exposure and in the test phase (within-modal presentation) or whether the clear speech benefit persists when the stimuli in the test phase are presented in orthographic instead of auditory form (cross-modal presentation). Previous studies comparing within- and cross-modal

integration of information have suggested that utilizing information in one modality to recognize later events in another modality may be challenging, and therefore, cognitively more demanding. Greene *et al.* (2001) used visual-auditory events, such as a video of a baby crying and its corresponding audio clip, and showed that within-modal priming (audio-audio) and visual-to-audio cross-modal priming was superior to cross-modal (audio-visual) information integration (see also Björkman, 1967). In contrast, cross-modal integration of information was found to be as good as within-modal integration in a study using photographs and naturalistic sounds (Lawrence and Cobb, 1978). The cost of cross-modal integration of information could result in reduced memory retention. Experiment II tests whether clear speech can alleviate some of the cross-modal processing difficulty and enhance sentence recognition memory for native and non-native listeners.

In testing cross-modal sentence recognition memory, we also aim to tease apart listener's reliance on linguistically encoded information from the reliance on surface features (i.e., acoustic cues). Since the test sentences are presented in orthographic form, memory traces can be activated via deeper linguistic processes at a level abstracted from the input speech. If the clear speech benefit persists in cross-modal recognition memory, it would suggest that the cognitive resources that remain available when listening to the easier-to-process clear sentences are used for deeper processing of the speech signal and storage in memory.¹

Finally, we also examined the role of working memory, defined as the ability to temporarily process and store information, in sentence recognition memory (Baddeley, 1992). During speech processing, listeners must map the acoustic information onto lexical and semantic representations. Working memory is then updated with new information from the auditory signal (Miyake *et al.*, 2000). When speech is degraded or differs from the expected form, as in casual reduced speech, it may be more difficult to match acoustic information to stored lexical information and working memory may be involved to a greater extent (Lunner, 2003; Rönnberg *et al.*, 2013; Souza *et al.*, 2015; Zekveld *et al.*, 2013). In the present study, all participants completed a forward digit span task (Wechsler, 1997). This task involves participants correctly recalling a sequence of digits they previously heard and testing increasingly longer sequences in each trial. This task was chosen to index working memory capacity because it is a widely used and accepted measure of the phonological loop (Baddeley, 2000; Baddeley *et al.*, 1998) and of auditory short-term memory (Hale *et al.*, 2002; Rosenthal *et al.*, 2006; Engle *et al.*, 1999), two processes that listeners engage in during sentence recognition memory. Since working memory is consumed by increased processing demands, we predicted that individuals with higher working memory capacity would be likely to cope better with the more-difficult-to-process speech signal than individuals with lower working memory capacity (Pichora-Fuller and Singh, 2006; Pichora-Fuller, 2007; Rudner *et al.*, 2009; Schneider, 2011; Zekveld *et al.*, 2013). We expected this to be true for individuals performing the task in their L1 or L2 even though listeners may be overall disadvantaged when doing a

digit span task in a non-native language (Olsthoom *et al.*, 2012).

II. EXPERIMENT OVERVIEW

The current paper presents the results from two experiments. Experiment I tested within-modal (audio-audio) sentence recognition memory for 30 native monolingual English listeners and 30 non-native English listeners. The goal of this experiment was to investigate whether the clear speech benefit on recognition memory observed for native listeners (Van Engen *et al.*, 2012, Gilbert *et al.*, 2014) extends to non-native listeners. Experiment II tested cross-modal (audio-textual) sentence recognition memory for 30 native monolingual English listeners and 30 non-native English listeners (different individuals from Experiment I). The goal was to examine whether the clear speech benefit persists even when the test stimuli are presented orthographically rather than auditorily. The cross-modal presentation increases cognitive demand at test and challenges listeners’ reliance on the specific acoustic-phonetic features (which may drive the benefit in the within-modal task) in sentence recognition memory. Thus, Experiment II may also speak to whether clear speech is better remembered due to its surface features or its potential to facilitate deeper linguistic encoding by freeing up cognitive resources. Experimental sessions took place in one-go and lasted less than one hour. First, each participant signed informed consent, completed a detailed language questionnaire, and passed a hearing screening. Participants then completed the forward digit span task (approximately 10 min) followed by the sentence recognition memory task (lasting approximately 20 min).

III. EXPERIMENT I: WITHIN-MODAL SENTENCE RECOGNITION MEMORY FOR NATIVE AND NON-NATIVE LISTENERS

A. Participants

Thirty native English listeners between the ages of 18 and 23 (mean: 19 years old; 21 Female) and 30 non-native listeners between the ages of 18 and 31 (mean: 23 years old; 24 Female) participated in the experiment. Native

monolingual speakers of American English were all born and raised in monolingual English households or communities in which English was the primary language and reported no current advanced proficiency in any other language. Non-native listeners reported having no exposure to English before the age of 6 (information about the non-native participants’ language background is provided in Table I). All the native monolingual English speakers and approximately half of the non-native English speakers were recruited via the Linguistics department subject pool. They were undergraduate students enrolled in a 12-week introductory course to Linguistics and received class credit for their participation. The other half of the non-native English speakers were recruited from the UT community (students and visiting scholars). They were paid \$10 for their participation. While the non-native participants’ background was somewhat more diverse, both groups were similar in age range and education levels (most participants were in their 20s and had some college education). Immediately before beginning the experiment, all participants signed written informed consent, filled out a detailed language background questionnaire adapted from the LEAP-Q questionnaire (Marian *et al.*, 2007), and passed a hearing screening, administered bilaterally at 25 dB hearing level (HL) at 500, 1000, 2000, and 4000 Hz.

B. Stimuli

The stimuli used in this study were the same 80 semantically-meaningful sentences used in Van Engen *et al.* (2012). The sentences (e.g., *The hot sun warmed the ground*) were produced in conversational and clear speaking styles by a 26-year-old female speaker of American English. The sentences contained high-frequency words familiar to non-native listeners (see Calandruccio and Smiljanić, 2012 for more details about the development of the materials). Forty sentences served as old/exposure sentences, and 40 as new/distractor sentences. Intelligibility of new and old sentences was equivalent as confirmed with a word-recognition-in-noise task (Van Engen *et al.*, 2012). There was some lexical overlap between old and new sentences. Forty-four percent of the lexical items in the new sentences appeared in the set of old sentences (overlapping items), while 56% of the

TABLE I. Language background information for non-native listeners. For each language, self-estimated amount of daily exposure on a scale from 1 (no current exposure) to 5(constant exposure). For each language, average of self-estimated proficiency for each skill, i.e., writing, speaking, reading, and listening on a scale from 1 (low) to 5 (high).

Factor	Experiment I (n = 30)	Experiment II (n = 30)
Age of first exposure to English (in years)	9 (mean); 6–17 (range)	8 (mean); 6–13 (range)
Age of arrival to USA (in years)	18.5 (mean); 6–30 (range)	16 (mean); 1–28 (range)
Daily exposure	L1: 4.6 (mean); 2–5 (range) English: 4.7 (mean); 3–5 (range)	L1: 4.6 (mean); 3–5 (range) English: 4.6 (mean); 3–5 (range)
Contexts for daily exposure to English	Professional setting only: n = 21 Extended and/or immediate relatives: n = 9	Professional setting only: n = 25 Extended and/or immediate relatives: n = 5
Self-estimated proficiency	L1: 4.9 (mean); 0.22 [standard deviation (s.d.)] English: 4.1 (mean); 0.62 (s.d.)	L1: 4.7 (mean); 0.49 (s.d.) English: 4.1 (mean); 0.56 (s.d.)
L1	Mandarin (n = 10), Korean (n = 7), Spanish (n = 5), French (n = 2), Farsi (n = 1), Turkish (n = 1), Cantonese (n = 1), Dutch (n = 1), Portuguese (n = 1), Amharic (n = 1).	Spanish (n = 11), Mandarin (n = 10), Korean (n = 4), French (n = 2), Turkish (n = 1), Hindi (n = 1), Indonesian (n = 1).

lexical items were unique to the new sentences. This amount of overlap made the task difficult enough, while still feasible. The overlapping items consisted mostly of the highly frequent words such as “old,” “girl,” “car,” “food,” etc. The unique words were also highly familiar and frequent words (as documented in [Calandruccio and Smiljanić, 2012](#)) such that the new sentences that contained these unique words could not be identified more accurately as new. Importantly, the unique and overlapping lexical items appeared equally in conversational and clear style and their distribution was not expected to affect sentence recognition pattern.

Recording took place in a sound-attenuated booth using a Shure SM10A head-mounted microphone and a Marantz solid-state recorder (PMD670). For the conversational speaking style, the speaker was asked to read sentences in a casual style, as if talking to someone who is familiar with their speech. For the clear speaking style, she was instructed to read the sentences as if talking to someone who is having difficulty understanding her, such as a non-native listener or a listener with hearing impairment (following [Smiljanić and Bradlow, 2005](#)). Individual sentences were segmented from the long recording and equalized for root-mean-square (rms) amplitude using the software Praat ([Boersma and Weenink, 2001](#)). The sentences were presented in quiet (i.e., without added noise) in the recognition memory task.

Acoustic analyses and word-recognition-in-noise intelligibility assessment for the sentences used in the present study were reported in [Van Engen et al. \(2012\)](#). The acoustic analyses showed that the sentences exhibited the acoustic-articulatory characteristics typically found in conversational-to-clear speech adaptations ([Smiljanic and Bradlow, 2009](#); [Pichora-Fuller et al., 2010](#); [Cooke et al., 2013](#)), such as - significantly longer durations, higher mean F_0 s, larger F_0 ranges, and greater energy in the 1–3 kHz range for clear than conversational speech. The intelligibility assessment showed a significant clear speech benefit for native English listeners. In the present study, we replicated the word-recognition-in-noise intelligibility assessment with 13 non-native English listeners [different individuals from the sentence recognition memory tests, but recruited from the same pool/community; 9 Female; mean age 22 years old, range: 18–31; first exposed to English at age 9, range: 3–17; first moved to the US at age 13, range: 1–30; first languages: Mandarin ($n=3$), Korean ($n=2$), Spanish ($n=4$), Nepali ($n=2$), Gujarati ($n=1$), Czech ($n=1$)]. The conversational and clear sentences were mixed with speech-shaped noise using the same SNR of 0 dB as in [Van Engen et al. \(2012\)](#). Participants were instructed to write down what they heard after the presentation of each sentence. Accuracy was higher for clear speech (68% keyword identification) than for conversational speech (27%), but non-native listeners were less accurate than native listeners in [Van Engen et al. \(2012\)](#) at the same SNR (95% and 79%, respectively). This is in keeping with previous work showing that the effect of the environmental signal distortion is greater even for highly proficient non-native listeners than for native listeners ([Mayo et al., 1997](#); [Meador et al., 2000](#); [Rogers et al., 2006](#)). The combined results of the two word-recognition-in-noise tasks

showed a clear speech benefit for both native and non-native listeners.

C. Procedure

First, participants completed the forward digit span task designed by [MacWhinney \(2016\)](#) (STEP E-Prime Scripts). Participants were seated in a sound-attenuated booth facing a computer monitor. Instructions and stimuli were presented with E-Prime ([E-Prime Psychology Software Tools](#), accessed 2017) and listener responses were collected using a computer keyboard. Participants were instructed to memorize numbers that were auditorily presented through Sennheiser HD570 headphones. All numbers (1–9) were digitized and presented randomly by a computer. Three sequences of a given length were presented per trial. Each sequence was presented alone. After each sequence, participants were instructed to type down on the keyboard the numbers in the correct order. The test started with a length of three-digits and increased in length by one digit following a successful recall (correct digits and serial order) of at least one of the three sequences of the same given length. Testing was discontinued after failure to identify three sequences of the same given length.

Participants then completed the recognition memory experiment. Instructions and stimuli were presented with E-Prime. Listener responses were collected using a button box. To familiarize participants with the button box and the task, a practice session was completed prior to the experiment. The instructions in the practice session were identical to the ones used in the experiment, but the stimuli were different. In the practice session, the exposure phase involved randomly presenting three pictures of animals (a puppy, a bird, and a monkey). The test phase involved randomly presenting two old and two new (a hat and a chair) pictures and asking the participant to categorize the picture as old or new. After each response, the participant was provided with feedback (correct/incorrect) on the computer screen. This feedback was only provided during the practice session. No feedback was ever provided as part of the experiment. In the exposure phase of the experiment, listeners heard 40 unique sentences (half in conversational, half in clear speech) in random order and were instructed to commit them to memory. Sentences were presented over headphones. Listeners heard each sentence only one time. Sentence presentations were separated by 1500 ms of silence. The display screen was always blank during the exposure phase. Immediately following the completion of the exposure phase, participants started the test phase. They were presented with 80 randomized sentences, 40 of which they heard during the exposure phase (old) and 40 sentences that they had not heard previously (new). Half of the sentences presented in the test phase were in conversational and half in clear speaking style. The test stimuli were presented in the same modality as in the exposure phase (audio). The old sentences were the same stimuli used in the exposure phase (i.e., same acoustic signal). In the test phase, participants were instructed to indicate for each sentence whether the sentence was old (i.e., heard during the exposure phase of the experiment) or new (i.e.,

never heard during the exposure phase of the experiment) by using the buttons labeled “old” and “new” on the button box. Participants were instructed to respond as quickly and as accurately as possible.

D. Analyses

In line with the previous studies (Gilbert *et al.*, 2014; Van Engen *et al.*, 2012), the recognition memory data was analyzed within a signal detection framework (Snodgrass and Corwin, 1988). Within this framework, when a stimulus from the exposure phase (old) is correctly identified by the listener, it is considered a hit; otherwise it is a miss. When a new stimulus is correctly identified, it is considered a correct rejection; otherwise it is a false alarm. In order to assess discrimination sensitivity and accuracy independently of response bias, detection sensitivity (d') and response bias (C) were computed for each participant in each speaking style. D' scores were calculated by subtracting the normalized probability of false alarms from the normalized probability of hits within each speaking style. Those probabilities were corrected to accommodate values of 0 and 1 in the d' calculation by adding 0.5 to each data point and dividing by $N + 1$, where N is the number of old or new trials within each speaking style (Snodgrass and Corwin, 1988). C scores, also calculated following Snodgrass and Corwin (1988), indicate whether participants are biased towards responding new (positive C values) or old (negative C values). Furthermore, we analyzed hit and false alarm rates separately in order to ascertain where the changes in d' occurred.

In addition to analyzing different type of responses (i.e., hit, false alarm), we analyzed reaction times (RTs) in order to evaluate participants' confidence in their responses. Faster responses indicate higher confidence than slower responses (Weidemann and Kahana, 2016). The RTs were calculated as the time elapsed from the onset of auditory stimulus presentation to the time the participant pressed the button on the button box to indicate their decision (old/new). The duration of each auditory stimulus was then subtracted from the RTs, thereby accounting for variability in the duration of the stimuli (i.e., different spoken sentences). This calculation yielded the true RT, that is, the time needed by the participant to make their decision (old/new) once they had finished hearing each auditory stimulus.

The digit span scores were calculated based on the longest digit list length correctly recalled, the “Longest Digit Span” (LDS), regardless of whether the subject passed one, two, or three trials at each length of digit span. The LDS was chosen as a measure because it provides a meaningful index of actual span length (Pisoni *et al.*, 2011). The individual digit span scores were included as co-variables in statistical analyses of the recognition memory results. This allowed us to control for individual differences in working memory when assessing sentence recognition memory.

Linear mixed-effects regressions (LMER) were conducted on the following dependent variables: (1) d' scores, (2) normalized hit rates, (3) normalized false alarm rates, and (4) RTs. Speaking Style (conversational vs clear), Listener Group (native vs non-native) and the Speaking

Style by Listener Group interaction were included in the model. Digit Span Scores were included as a covariate. Subjects were modeled using a random intercept term. All regression models were fit using the lme4 package in R (Bates *et al.*, 2015).

E. Results

The mean and range of the digit span test for native and non-native listeners are shown in Table II. The distribution of digit span scores greatly overlapped between the two groups, such that native and non-native listeners performed equivalently on this task. The overall sentence recognition memory results are presented in Fig. 1 and Table III. Average C scores for both listener groups were positive, indicating that participants were generally biased to respond “new” more often than “old.” This bias was stronger for speech produced in a clear style for native listeners. D' scores were higher for clear than for conversational sentences for both native and non-native listeners [Fig. 1(a)]. There was a main effect of Speaking Style ($p < 0.001$) on d' scores, but no effect of Listener Group ($p = 0.69$), no effect of Digit Span ($p = 0.13$), and no significant interaction between Speaking Style and Listener Group ($p = 0.73$). Thus, clear speaking style improved recognition memory and the clear speech benefit was similar for native and non-native listeners.²

The hit rates were higher for clear than conversational sentences for non-native listeners, however, this was not the case for native listeners [Fig. 1(b)]. The linear mixed-effects regression showed that there was a significant interaction between Speaking Style and Listener Group ($p < 0.05$). *Post hoc* analyses to decompose the interaction revealed that the effect of speaking style on hit rates was significant for non-native listeners ($p < 0.05$), but not significant for native listeners ($p = 0.14$). The statistical results confirmed that hit rates, i.e., the ability to recognize previously heard sentences as old, significantly increased for non-native listeners as a result of speaking style enhancement, but the hit rate for conversational and clear sentences did not differ for native listeners.

False alarm rates were lower for clear sentences than conversational ones for native listeners, meaning that native listeners made fewer errors in identifying new clear sentences than new conversational sentences [Fig. 1(c)]. There was a significant interaction between Speaking Style and

TABLE II. Digit span scores.

Experiment I ($N = 60$)				
listeners	mean	s.d.	min	max
Native ($n = 30$)	8.00	1.29	6	10
Non-native ($n = 30$)	8.03	1.43	5	10
Experiment II ($N = 60$)				
listeners	mean	s.d.	min	max
Native ($n = 30$)	8.17	1.23	6	10
Non-native ($n = 30$)	7.97	1.25	6	10

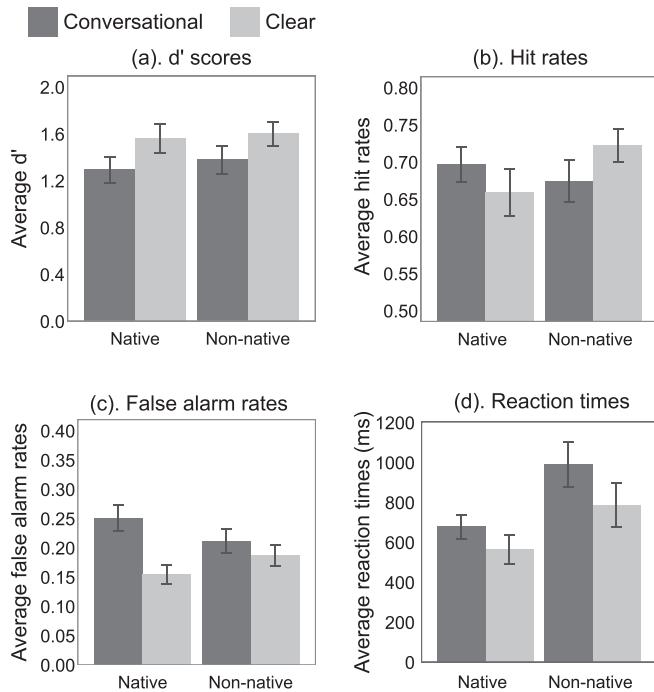


FIG. 1. Average of d' scores (a), normalized hit rates (b), normalized false alarm rates (c), and RTs (d) for native ($n = 30$) and non-native English listeners ($n = 30$) for sentences produced in clear (light grey) and conversational (dark grey) speaking styles in Experiment I (within-modal). Error bars represent standard error.

Listener Group ($p < 0.01$). *Post hoc* analyses revealed that the effect of speaking style on false alarm rate was significant for native listeners ($p < 0.001$), but not for non-native listeners ($p = 0.15$). In other words, the rate of false alarm significantly decreased for sentences in clear speaking style as opposed to sentences in conversational speaking style only for the native listener group.

Finally, we ran linear mixed-effects regression analysis of RTs with d' scores as an additional covariate to control

TABLE III. Mean and standard deviations (in parentheses) of hit rates, false alarm rates, d' , C , and reaction times (RTs) in milliseconds (ms) for native and non-native listeners for conversational and clear sentences in within- and cross-modal recognition memory tasks.

	Experiment I ($N = 60$)		Experiment II ($N = 60$)	
	Native ($n = 30$)		Native ($n = 30$)	
	Conversational	Clear	Conversational	Clear
Hit rate	0.7 (0.13)	0.66 (0.18)	0.62 (0.15)	0.67 (0.14)
False alarm rate	0.25 (0.12)	0.15 (0.09)	0.15 (0.1) (overall)	
d'	1.29 (0.62)	1.56 (0.69)	1.43 (0.56)	1.57 (0.51)
C	0.07 (0.3)	0.31 (0.33)	0.39 (0.33)	0.32 (0.33)
Mean RT	677 (326)	565 (397)	1994 (407)	1922 (386)
	Non-native ($n = 30$)		Non-native ($n = 30$)	
	Conversational	Clear	Conversational	Clear
Hit rate	0.67 (0.16)	0.72 (0.12)	0.62 (0.09)	0.69 (0.13)
False alarm rate	0.21 (0.11)	0.19 (0.1)	0.25 (0.12) (overall)	
d'	1.38 (0.66)	1.6 (0.57)	1.06 (0.48)	1.29 (0.66)
C	0.18 (0.32)	0.16 (0.3)	0.22 (0.27)	0.11 (0.26)
Mean RT	989 (605)	787 (608)	2162 (593)	2134 (626)

for differences in accuracy. We found a main effect of Speaking style ($p < 0.001$) and a main effect of Listener Group ($p < 0.05$), such that response times were significantly faster for clear than for conversational sentences and faster for native listeners than for non-native listeners [Fig. 1(d)]. The current RT analysis includes responses recorded before the stimuli offset (8.25% of the responses in total). The proportion of early responses was higher for clear sentences (13.5% for native and 10% for non-native listeners) than for conversational sentences (5.5% for native and 4% for non-native listeners). We decided to include these RTs in the analysis to not penalize participants following the instruction to respond as quickly and accurately as possible and because these early responses might be indicative of stronger participants' confidence rather than inattentive fast responses. If listeners were pressing the response button before the end of the stimuli in a random manner, we would expect this strategy to affect both conversational and clear sentences to the same degree. In contrast, this analysis revealed that listeners tended to respond more quickly when hearing clear sentences compared to conversational sentences. It remains to be determined in future work whether the faster RTs for clear speech truly reflect increased confidence about the accuracy of sentence recognition or are due to longer processing time afforded to the listeners by longer stimuli.

IV. EXPERIMENT II: CROSS-MODAL SENTENCE RECOGNITION MEMORY FOR NATIVE AND NON-NATIVE LISTENERS

A. Participants

Thirty native English listeners between the age of 18 and 32 (mean: 20 years old; 17 Female), and 30 non-native listeners between the age of 18 and 31 (mean: 22 years old; 18 Female) participated in Experiment II. They were different individuals from Experiment I but recruited from the same pool/community. As in Experiment I, all participants signed written informed consent, filled out a detailed language background questionnaire, and passed a hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz before beginning the experiment. Information about the non-native participants' language background is provided in Table I.

B. Stimuli

The stimuli were the same as in Experiment I and were also presented in quiet (i.e., without added noise).

C. Procedure

The procedure was identical to the one in Experiment I. The only change was in the modality of presentation of the sentences in the test phase. Instead of hearing the sentences over headphones, participants saw the sentences orthographically displayed on the computer screen with no accompanying acoustic signal (and therefore, no speaking style associated to the sentences). Eighty sentences were presented (40 from the exposure phase and 40 new). Each sentence was presented in the center of the screen against a uniform white background

in black Arial size 25 font. Each sentence was displayed on the screen until participants recorded their response (old/new) via the button box. The decision to allow the sentence to remain visually available to the participant was based on pilot studies showing that listeners failed to process the written text when it was presented on the screen for only the duration of its spoken counterpart (which would have matched the time-limited availability of the auditory speech signal in Experiment I). This aspect of the design entails different demands on the listener's memory load across the two experiments. To ensure a timely response from participants, the instructions explicitly urged them to respond as quickly and as accurately as possible (as in Experiment I).

D. Analyses

Similar analyses of the digit span and sentence recognition memory tasks as in Experiment I were conducted here. A crucial difference was that since sentences were presented visually during the test phase, the distractor/new sentences had no speaking style associated with them. Consequently, although it remained possible to compute two hit rates per listener (one for each speaking style), only one false alarm rate per listener could be computed (over the entire set of new sentences). Thus, d' scores were calculated as the normalized probability of either clear or conversational hit rates minus the overall normalized probability of false alarms. Moreover, in order to compare the different speaking styles, we only analyzed RTs for the subset of stimuli that was presented to the participants in the exposure phase (i.e., 20 in conversational, 20 in clear).

LMEs were conducted on the following dependent variables: (1) d' scores, (2) normalized hit rates, and (3) RTs. Speaking Style (conversational vs clear), Listener Group (native vs non-native), and the Speaking Style by Listener Group interaction were included in the model. Digit Span scores were included as a covariate, and Subject was treated as a random effect. All regression models were fit using the lme4 package in R (Bates et al., 2015).

E. Results

The mean and range of the digit span test for native and non-native listeners are shown in Table II. As in Experiment I, the distribution of the digit span scores revealed no differences between native and non-native listeners. The overall

sentence recognition memory results are presented in Fig. 2 and Table III. The average C scores across listener groups were positive, indicating that participants were biased to respond "new" more often than "old." Contrary to the results of the within-modal task, the bias here was stronger for speech produced in a conversational style for both listener groups. D' scores were higher for native than non-native listeners and for clear than conversational sentences [Fig. 2(a)]. There was a main effect of Listener Group ($p < 0.05$) on d' scores and a main effect of Speaking Style ($p < 0.01$), but no effect of Digit Span ($p = 0.93$). No significant interaction between Speaking Style and Listener Group was found ($p = 0.49$). Thus, recognition memory was significantly better for clear sentences than for conversational sentences. The two listener groups also performed significantly differently: native listeners had overall higher d' scores than non-native listeners. Despite the absence of acoustic information in the test phase, both listener groups exhibited the clear speech benefit in sentence recognition memory. In other words, written information alone was enough to observe enhanced recognition memory for sentences in clear speech.³

We also found that both native and non-native listeners had higher hit rate for sentences produced in clear than in conversational speech [Fig. 2(b)]. The linear mixed-effects regression showed that there was a main effect of Speaking Style ($p < 0.01$), but no effect of Listener Group ($p = 0.62$), no effect of Digit Span ($p = 0.15$), and no interaction between Speaking Style and Listener Group ($p = 0.58$). Thus, when the sentences in the test phase were written on the screen, correct identification of old items was superior for clear sentences than conversational sentences in both listener groups.

Finally, there was no difference in RTs between speaking styles and listener groups (Fig. 2(c)). Linear mixed-effects regression analysis of RTs (including d' scores as a covariate) found that there was no effect of Speaking style ($p = 0.26$), no effect of Listener Group ($p = 0.29$), no effect of Digit Span ($p = 0.11$), no effect of d' scores ($p = 0.9$), and no interaction between Listener Group and Speaking style ($p = 0.61$).

V. ADDITIONAL RESULTS ACROSS MODALITIES

The above analyses assessed the effect of speaking style and language on sentence recognition memory within each

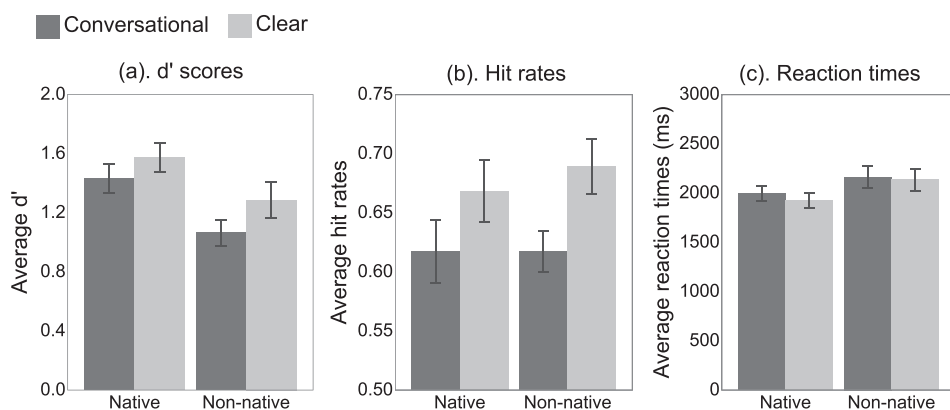


FIG. 2. Average of d' scores (a), normalized hit rates (b), and RTs (c) for native ($n = 30$) and non-native English listeners ($n = 30$) for sentences produced in clear (light grey) and conversational (dark grey) speaking styles in Experiment II (cross-modal). Error bars represent standard error.

modality. In order to compare the effect of modality (within- vs across-) on sentence recognition memory, we conducted a LMER analysis within each listener group. D' scores were the dependent variable, Speaking Style (conversational vs clear) and Modality (within vs cross) the independent variables. We added the interaction of Speaking Style by Modality in the model, Digit Span as a covariate, and Subject as a random effect. Results for the non-native listeners showed a main effect of Speaking style ($p < 0.001$) and a main effect of Modality ($p < 0.05$), but no effect of Digit Span ($p = 0.47$) and no interaction between Modality and Speaking style ($p = 0.99$). Thus, non-native listeners' d' scores were higher for clear than conversational sentences, and their d' scores were higher in the within-modal than in the cross-modal task. The LMER analysis of the native listener's data across the two modalities indicated that d' scores were higher for clear than conversational sentences (main effect of Speaking style; $p < 0.01$) and that d' scores did not change as a function of stimulus presentation modality during the testing phase (no effect of Modality; $p = 0.63$; and no interaction with speaking style; $p = 0.36$). Combined, the results of the two experiments showed that the cross-modal task was more difficult for non-native listeners than for native listeners. Importantly, both listener groups still showed higher accuracy for clear speech sentence compared to conversational sentences.⁴

Although this study was not designed to systematically investigate the effect of linguistic experience on recognition memory task performance, we conducted several analyses to explore its role. Our data set included a large number of non-native listeners with varied linguistic experiences and proficiency levels (although all had to be fully functional in the university setting, see Table I). We used linear mixed-effects regression to determine whether any of the following independent variables was predictive of d' scores in each experiment: L1 (e.g., Spanish, Mandarin), self-rated proficiency in L1 and L2, current daily exposure to each language, age of acquisition of English, and age of arrival in the US. Our analyses did not reveal a significant relationship between d' score and any of these linguistic experience factors. For instance, lower self-rated proficiency level in English did not predict lower d' scores. However, it is possible that none of our measures here were sensitive enough indicators of language proficiency. Rimikis *et al.* (2013) for instance, found that the best predictor of non-native listeners' performance in a speech-in-noise task was their spoken language proficiency as measured using an automated Versant test. It is also possible that we did not have enough variability in our non-native listeners' demographic characteristics to detect meaningful correlations. Ultimately, more research is needed to provide a more nuanced understanding of the effect of language experience on speech recognition memory.

VI. DISCUSSION

Understanding how acoustic and linguistic factors shape memory for speech sheds light into the cognitive processes involved in speech perception. This study examined the effect of speech clarity on sentence recognition memory in

within-modal (audio-audio; Experiment I) and cross-modal (audio-textual; Experiment II) tasks for native and non-native listeners. Accounting for individual differences in working memory, this study showed that native and non-native listeners performed similarly when sentence recognition memory was tested within modality, but non-native listeners performed worse than native listeners when memory was tested across modalities. Crucially, however, the study showed that in both modalities, both listener groups benefited significantly from clear speech enhancements in sentence recognition memory.

The within modality results showed that non-native listeners were able to utilize clear speech acoustic-phonetic enhancements to improve sentence recognition memory to the same extent as native listeners. Although both clear and conversation sentences were presented in quiet and were fully intelligible, casual reduced sentences required more cognitive effort to process and were thus remembered less accurately. This suggests that sentences produced in clear speech freed up cognitive resources and facilitated storage in memory for both native and non-native listeners, supporting the "effortfulness hypothesis" (McCoy *et al.*, 2005; Rabbitt, 1968, 1990) and the "ease of language understanding" model (Rönnberg *et al.*, 2013; Rönnberg *et al.*, 2008). While both listener groups demonstrated a clear speech benefit in discrimination sensitivity for clear sentences in the within-modality test, differences existed between native and non-native listeners. In line with Van Engen *et al.* (2012) and Gilbert *et al.* (2014), we found that native listeners were more accurate at identifying and rejecting distractor sentences produced in clear speech, that is, they had significantly lower false alarm rates. In the same within-modal task, non-native listeners were more accurate in recognizing clearly produced sentences as previously heard than casually produced sentences; that is, they had significantly higher hit rates. Enhanced capacity for identifying already heard sentences suggests that non-native listeners relied more heavily on episodic memory. The two listener groups also differed in the fluency of their performance. Despite similar discrimination accuracy, non-native listeners had significantly longer RTs than native listeners. This finding highlights the cost of L2 processing on cognitive resources. Future work should further examine the accuracy and speed trade-off in speech memory tasks for non-native listeners.

The within-modality results revealed an interesting discrepancy between the sentence recognition memory task and the word-recognition-in-noise task (intelligibility assessments discussed in Sec. III B of this paper). Even though clear speech improved sentence recognition memory for both native and non-native listeners equally, non-native listeners benefited less from the English-specific clear speech strategies in the word-recognition-in-noise task. The difference is in part due to the presence of noise during the word-recognition-in-noise task versus the absence of noise during the recognition memory task. Even when no differences between listener groups are found for word recognition in quiet, highly proficient non-native listeners were shown to be less accurate than native listeners when listening to speech mixed with noise (Mayo *et al.*, 1997; Rogers *et al.*, 2006; Lecumberri *et al.*, 2010). The lower non-native word

recognition scores likely also have origin in the less efficient use of L2-specific clear speech enhancements (Bradlow and Alexander, 2007; Bradlow and Bent, 2002). The difference between the word recognition and sentence recognition memory results could also be indicative of the different processes underlying the two tasks. In word-recognition-in-noise task, listeners need to map acoustic cues to the stored phoneme and lexical representations in order to write down what they heard, and this might decrease overall accuracy. In the sentence recognition memory task, on the other hand, listeners could store in memory only a few distinctive or salient acoustic cues without further mapping onto the lexicon or meaning, and this might increase overall accuracy. This, as was already argued above, could reflect greater reliance on signal-level information and episodic memory for non-native listeners.

The findings of the cross-modal task (Experiment II) allowed us to further probe what underlies the clear speech benefit on sentence recognition memory for both listener groups. Despite the cross-modal challenges reported in the literature (Björkman, 1967; Greene *et al.*, 2001; Lawrence and Cobb, 1978), our results showed that native listeners were successful in integrating information across modalities, demonstrating processing efficiency. The persistence of the clear speech benefit even when test sentences were presented in orthographic form suggests that the memory traces could be activated through deeper linguistic processes at a level abstracted from the input speech. Listening to the easier-to-process clear sentences may have freed up cognitive resources for deeper processing of the speech signal and storage in memory. Non-native listeners, however, were less successful in that task. When only written input was presented in the test phase, non-native listeners performed overall worse than native listeners, and worse than non-native listeners in the within-modal testing. This finding supports the idea that L2 language processing is costly for cognitive resources and that it may diminish resources needed for information integration across modalities. However, even in this overall more challenging task, the processing cost was offset by signal clarity. Additional cognitive resources remained available to the listeners for storing information in memory for clear speech sentences.⁵

Another possible account for poorer cross-modal recognition memory in non-native compared to native listeners is that non-native listeners might engage qualitatively different cognitive processes in L1 and L2. Sampaio and Konopka (2013) suggested that L2 speakers might rely to a greater extent on lower-level surface forms when recalling sentences than native speakers, who may instead rely more on “gist” memory (Fuzzy-Trace theory, Reyna and Brainerd, 2011). It is possible that in the present study, non-native listeners relied more heavily on the signal-level information and needed the specific acoustic signal to activate stored memory traces. This would account for the accuracy drop in the absence of acoustic input (Experiment II). A critical question of what is the precise nature of L1 and L2 memory traces for conversational and clear speech sentences that allows for cross-modal information integration merits further research.

One of the goals of the present study was to assess the role of working memory capacity on individual differences

in sentence recognition memory for native and non-native listeners. The finding that digit span did not predict performance in the recognition memory task contrasts with a number of studies that have found that individuals with higher working-memory capacity cope better with the more-difficult-to-process speech signal than individuals with lower working-memory capacity (Pichora-Fuller, 2007; Rudner *et al.*, 2009; Schneider, 2011; Zekveld *et al.*, 2013). Several reasons may account for the lack of a correlation between the working memory measure and the recognition memory performance in our study. First, it is possible that our sentence recognition task in quiet was not sufficiently difficult overall to reveal a correlation with working-memory capacity. Another possibility is that the task we chose to index working memory capacity, digit span, was not sensitive enough to use as a predictor of recognition memory performance. The digit span measure was chosen to account for the lower-level of speech processing that takes place during the recognition memory task (storage of acoustic cues, phonemes, salient words in short-term memory). However, as the results of Experiment II suggest, the fact that listeners’ recognition memory accuracy was well above chance even when provided with written text suggests that recognition memory may involve a more holistic approach to language comprehension beyond simply storing acoustic cues in the phonological loop. While digit span is an accepted measure of phonological loop and auditory short-term memory, it is not a sensitive enough predictor of language comprehension (Daneman and Merikle, 1996; Rönnerberg *et al.*, 2013; Unsworth and Engle, 2007), which may explain why it was not predictive of memory performance in our present study. Most importantly, for our findings of the clear speech benefit on sentence recognition memory, the distribution of digit span scores for our native and non-native listeners were similar (even though non-native speakers may be disadvantaged when completing an audio-based digit span task in a non-native language; Olsthoorn *et al.*, 2012). Further studies are needed to elucidate the relationship between the individual variation in working-memory capacity and sentence recognition memory task in L1 and L2. Future studies should consider using a more sensitive indicator of working-memory capacity, such as the visual digit-span task (Olsthoorn *et al.*, 2012).

Taken as a whole, our study provides further evidence that acoustic clarity and language experience affect memory for spoken utterances. The results showed that clear speech improved not only speech perception, but also memorization and retention of information for both native and non-native listeners of the target language. These findings have implications for communication in challenging settings, such as noisy classrooms and doctor’s offices, where remembering spoken information is vital.

ACKNOWLEDGMENTS

We would like to thank the UT Sound Lab research assistants for their assistance with data collection.

¹It is possible that listeners rely on some acoustic cues even in the cross-modal recognition memory task through the phonological loop, which converts print to audio (Baddeley and Hitch, 2018). It is not clear though to

- what extent these acoustic cues would be exact matches to the specific acoustic cues heard in the exposure phase and thus, the extent to which these acoustic cues would facilitate sentence recognition memory. This possibility should be examined more closely in future work.
- ²Even though the digit span results were similar for the two listener groups, we compared the initial model to a statistical model without the digit span covariate to account for the fact that digit span could be an unreliable reflection of non-native listeners' working memory (Olsthoom *et al.*, 2012). The results were similar: there was a main effect of Speaking Style ($p < 0.01$) on d' scores, but no effect of Listener Group ($p = 0.6$), and no interaction ($p = 0.75$).
- ³As in Experiment I, similar results were found when removing Digit Span from the model (main effect of Listener Group, $p < 0.05$; and main effect of Speaking Style, $p < 0.01$; no interaction, $p = 0.48$).
- ⁴Due to logistical constraints on the number of stimuli that could reasonably be presented to the participants, we investigated the primary factor of interest, speaking style, as a within-subject factor and modality as a between-subject factor (i.e., in two separate experiments). While the participants in each experiment were different individuals, they were drawn from the same population (i.e., UT Austin community, similar education background, similar age range), and randomly assigned to different conditions. Moreover, we explicitly modeled idiosyncratic variation due to individual differences between participants by using a random intercept term in our mixed-effects regression models. For these reasons, we believe that conditions were met to allow for statistical inference.
- ⁵The longer RTs in Experiment II than in Experiment I could indicate cognitive effort, but could also reflect the time it takes to read printed information as opposed to process auditory information.
- Baddeley, A. (1992). "Working memory," *Science* (New York, N.Y.) **255**(5044), 556–559.
- Baddeley, A. (2000). "The episodic buffer: A new component of working memory?," *Trends Cogn. Sci.* **4**(11), 417–423.
- Baddeley, A., Gathercole, S., and Papagno, C. (1998). "The phonological loop as a language learning device," *Psychol. Rev.* **105**(1), 158–173.
- Baddeley, A., and Hitch, G. J. (2018). "The phonological loop as a buffer store: An update," *Cortex* (published online). DOI: 10.1016/j.cortex.2018.05.015.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* **67**, 1–48.
- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Learning & Language Teaching*, edited by O.-S. Bohn and M. J. Munro (John Benjamins Publishing Company, Amsterdam, the Netherlands), Vol. 17, pp. 13–34.
- Björkman, M. (1967). "Relations between intra-modal and cross-modal matching," *Scand. J. Psychol.* **8**(1), 65–76.
- Boersma, P., and Weenink, D. (2001). "Praat: Doing phonetics by computer [computer program]," <http://www.praat.org/> (Last viewed 8/10/2018).
- Bradlow, A. R., and Alexander, J. A. (2007). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," *J. Acoust. Soc. Am.* **121**(4), 2339–2349.
- Bradlow, A. R., and Bent, T. (2002). "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.* **112**, 272–284.
- Calandruccio, L., and Smiljanić, R. (2012). "New sentence recognition materials developed using a basic non-native English lexicon," *J. Speech Lang. Hear. Res.* **55**(5), 1342–1355.
- Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., and Tang, Y. (2013). "Evaluating the intelligibility benefit of speech modifications in known noise conditions," *Speech Commun.* **55**(4), 572–585.
- Cutler, A., Garcia Lecumberri, M. L., and Cooke, M. (2008). "Consonant identification in noise by native and non-native listeners: Effects of local context," *J. Acoust. Soc. Am.* **124**, 1264–1268.
- Daneman, M., and Merikle, P. M. (1996). "Working memory and language comprehension: A meta-analysis," *Psychon. Bull. Rev.* **3**(4), 422–433.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. A. (1999). "Working memory, short-term memory, and general fluid intelligence: A latent-variable approach," *J. Exp. Psychol.: General* **128**(3), 309–331.
- E-Prime. (Version 2.0). (Psychology Software Tools, Pittsburgh, PA) (accessed 2017).
- Ferguson, S. H. (2012). "Talker differences in clear and conversational speech: Vowel intelligibility for older adults with hearing loss," *J. Speech Lang. Hear. Res.* **55**(3), 779–790.
- Ferguson, S. H., and Kewley-Port, D. (2002). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **112**(1), 259–271.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (York Press, York, UK), pp. 233–277.
- Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. (2008). "Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English," *J. Acoust. Soc. Am.* **124**(2), 1234–1251.
- Gagné, J.-P., Rochette, A.-J., and Charest, M. (2002). "Auditory visual and audiovisual clear speech," *Speech Commun.* **37**(3), 213–230.
- Gilbert, R. C., Chandrasekaran, B., and Smiljanić, R. (2014). "Recognition memory in noise for speech of varying intelligibility," *J. Acoust. Soc. Am.* **135**(1), 389–399.
- Greene, A. J., Easton, R. D., and LaShell, L. S. (2001). "Visual-auditory events: Cross-modal perceptual priming and recognition memory," *Consciousness Cogn.* **10**(3), 425–435.
- Hale, J. B., Hoepfner, J.-A. B., and Fiorello, C. A. (2002). "Analyzing digit span components for assessment of attention processes," *J. Psychoed. Assess.* **20**(2), 128–143.
- Hygge, S., Kjellberg, A., and Nösl, A. (2015). "Speech intelligibility and recall of first and second language words heard at different signal-to-noise ratios," *Front. Psychol.* **6**, 1390.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition* **87**(1), B47–B57.
- Johnson, K. (2004). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa, Tokyo, Japan, pp. 29–54.
- Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., and Peelle, J. E. (2018). "The impact of age, background noise, semantic ambiguity, and hearing loss on recognition memory for spoken sentences," *J. Speech Lang. Hear. Res.* **61**, 740–751.
- Kondaurova, M. V., and Francis, A. L. (2008). "The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners," *J. Acoust. Soc. Am.* **124**(6), 3959–3971.
- Krause, J. C., and Braid, L. D. (2002). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," *J. Acoust. Soc. Am.* **112**, 2165–2172.
- Lawrence, D. M., and Cobb, N. J. (1978). "Cross-modal utilization of information: Recognition memory for environmental stimuli," *Percept. Motor Skills* **47**, 1203–1206.
- Lecumberri, M. L. G., Cooke, M., and Cutler, A. (2010). "Non-native speech perception in adverse conditions: A review," *Speech Commun.* **52**(11–12), 864–886.
- Lindblom, B. (1990). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling* (Kluwer Academic Publishers, Amsterdam, the Netherlands).
- Ljung, R., Israelsson, K., and Hygge, S. (2013). "Speech intelligibility and recall of spoken material heard at different signal-to-noise ratios and the role played by working memory capacity," *Appl. Cogn. Psychol.* **27**(2), 198–203.
- Lunner, T. (2003). "Cognitive function in relation to hearing aid use," *Int. J. Audiol.* **42**, S49–S58.
- MacWhinney, B. "Audio digit span," <http://step.talkbank.org/scripts-plus/> (Last viewed 10/2018).
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). "The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," *J. Speech Lang. Hear. Res.* **50**(4), 940–967.
- Mayo, L. H., Florentine, M., and Buus, S. (1997). "Age of second-language acquisition and perception of speech in noise," *J. Speech Lang. Hear. Res.* **40**(3), 686–693.
- McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (2005). "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," *Q. J. Exp. Psychol. A: Human Exp. Psychol.* **58**(1), 22–33.
- Meador, D., Flege, J. E., and Mackay, I. R. A. (2000). "Factors affecting the recognition of words in a second language," *Bilingual. Lang. Cogn.* **3**(1), 55–67.

- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). "The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis," *Cogn. Psychol.* **41**(1), 49–100.
- Molesworth, B. R. C., Burgess, M., Gunnell, B., Löffler, D., and Venjakob, A. (2014). "The effect on recognition memory of noise cancelling headphones in a noisy environment with native and non-native speakers," *Noise Health* **16**(71), 240–247.
- Olsthorn, N. M., Andringa, S., and Hulstijn, J. H. (2012). "Visual and auditory digit-span performance in native and non-native speakers," *Int. J. Bilingual.* **18**(6), 663–673.
- Payton, K. L., Uchanski, R. M., and Braida, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **95**(3), 1581–1592.
- Peelle, J. E. (2018). "Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior," *Ear Hear.* **39**(2), 204–214.
- Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* **28**(1), 96–103.
- Pichora-Fuller, M. K. (2007). "Audition and cognition: What audiologists need to know about listening," in *Hearing Care for Adults*, edited by C. Palmer and R. Seewald (Phonak, Stäfa, Switzerland), pp. 71–85.
- Pichora-Fuller, M. K., Goy, H., and Lieshout, P. Van. (2010). "Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments," *Seminars Hear.* **31**, 77–94.
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., and Wingfield, A. (2016). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)," *Ear Hear.* **37**, 5S–27S.
- Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). "How young and old adults listen to and remember speech in noise," *J. Acoust. Soc. Am.* **97**(1), 593–608.
- Pichora-Fuller, M. K., and Singh, G. (2006). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation," *Trends Amplif.* **10**(1), 29–59.
- Pichora-Fuller, M. K., and Souza, P. E. (2003). "Effects of aging on auditory processing of speech," *Int. J. Audiol.* **42**, 11–16.
- Pisoni, D., Kronenberger, W., Roman, A., and Geers, A. (2011). "Article 7: Measures of digit span and verbal rehearsal speed in deaf children following more than 10 years of cochlear implantation," *Ear Hear.* **32**(1), 60s–74s.
- Pluymackers, M., Ernestus, M., and Baayen, R. H. (2005a). "Articulatory planning is continuous and sensitive to informational redundancy," *Phonetica* **62**(2–4), 146–159.
- Pluymackers, M., Ernestus, M., and Baayen, R. H. (2005b). "Lexical frequency and acoustic reduction in spoken Dutch," *J. Acoust. Soc. Am.* **118**(4), 2561–2569.
- Rabbitt, P. M. (1968). "Channel-capacity, intelligibility and immediate memory," *Q. J. Exp. Psychol.* **20**(3), 241–248.
- Rabbitt, P. M. (1990). "Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ," *Acta Otolaryngol. Suppl.* **476**, 167–175, discussion 176.
- Reyna, V. F., and Brainerd, C. J. (2011). "Dual processes in decision making and developmental neuroscience: A fuzzy-trace model," *Dev. Rev.* **31**(2–3), 180–206.
- Rimikis, S., Smiljanić, R., and Calandruccio, L. (2013). "Nonnative English speaker performance on the Basic English Lexicon (BEL) sentences," *J. Speech Lang. Hear. Res.* **56**(3), 792–804.
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," *Appl. Psycholing.* **27**(3), 465–485.
- Rönnberg, J. (2003). "Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model," *Int. J. Audiol.* **42**, S68–S76.
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., and Rudner, M. (2013). "The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances," *Front. Syst. Neurosci.* **7**, 31.
- Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). "Cognition counts: A working memory system for ease of language understanding (ELU)," *Int. J. Audiol.* **47**, S99–S105.
- Rosenthal, E. N., Riccio, C. A., Gsanger, K. M., and Jarratt, K. P. (2006). "Digit span components as predictors of attention problems and executive functioning in children," *Arch. Clin. Neuropsychol.* **21**(2), 131–139.
- Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (2009). "Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids," *Scand. J. Psychol.* **50**(5), 405–418.
- Sampaio, C., and Konopka, A. E. (2013). "Memory for non-native language: The role of lexical processing in the retention of surface form," *Memory (Hove, England)* **21**(4), 537–544.
- Schneider, B. A. (2011). "How age affects auditory-cognitive interactions in speech comprehension," *Audiol. Res.* **1**(1), e10.
- Schum, D. (1996). "Intelligibility of clear and conversational speech of young and elderly talkers," *J. Am. Acad. Audiol.* **7**, 212–218.
- Smiljanić, R., and Bradlow, A. R. (2005). "Production and perception of clear speech in Croatian and English," *J. Acoust. Soc. Am.* **118**, 1677–1688.
- Smiljanić, R., and Bradlow, A. R. (2008). "Stability of temporal contrasts across speaking styles in English and Croatian," *J. Phon.* **36**(1), 91–113.
- Smiljanić, R., and Bradlow, A. R. (2009). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," *Lang. Ling. Compass* **3**(1), 236–264.
- Smiljanić, R., and Bradlow, A. R. (2011). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness," *J. Acoust. Soc. Am.* **130**(6), 4020–4031.
- Snodgrass, J. G., and Corwin, J. (1988). "Pragmatics of measuring recognition memory: Applications to dementia and amnesia," *J. Exp. Psychol. General* **117**(1), 34–50.
- Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., and Kates, J. M. (2015). "Working memory and intelligibility of hearing-aid processed speech," *Front. Psychol.* **6**, 526.
- Tun, P. A., McCoy, S., and Wingfield, A. (2009). "Aging, hearing acuity, and the attentional costs of effortful listening," *Psychol. Aging* **24**(3), 761–766.
- Uchanski, R. M. (1988). *Spectral and Temporal Contributions to Speech Clarity for Hearing Impaired Listeners* (Massachusetts Institute of Technology, Cambridge, MA).
- Unsworth, N., and Engle, R. W. (2007). "On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities," *Psychol. Bull.* **133**(6), 1038–1066.
- Van Engen, K. J., Chandrasekaran, B., and Smiljanić, R. (2012). "Effects of speech clarity on recognition memory for spoken sentences," *PloS One* **7**(9), e43753.
- Warner, N., Fountain, A., and Tucker, B. V. (2009). "Cues to perception of reduced flaps," *J. Acoust. Soc. Am.* **125**(5), 3317–3327.
- Wechsler, D. (1997). *WMS-III Administration and Scoring Manual* (The Psychological Corporation, Harcourt Brace Jovanovich, San Antonio, TX).
- Weidemann, C. T., and Kahana, M. J. (2016). "Assessing recognition memory using confidence ratings and response times," *Open Sci.* **3**(4), 150670.
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., and Rönnberg, J. (2013). "The effects of working memory capacity and semantic cues on the intelligibility of speech in noise," *J. Acoust. Soc. Am.* **134**(3), 2225–2234.