Reading aloud in clear speech reduces sentence recognition memory and recall for native and non-native talkers

Sandie Keerstock and Rajka Smiljanic

Citation: The Journal of the Acoustical Society of America **150**, 3387 (2021); doi: 10.1121/10.0006732
View online: https://doi.org/10.1121/10.0006732
View Table of Contents: https://asa.scitation.org/toc/jas/150/5
Published by the Acoustical Society of America

JASA ARTICLE

Reading aloud in clear speech reduces sentence recognition memory and recall for native and non-native talkers[a)]

Sandie Keerstock[1,b)] and Rajka Smiljanic[2]

[1]*Department of Psychological Sciences, University of Missouri, 124 Psychology Building, 200 South 7th Street, Columbia, Missouri 65211, USA*

[2]*Department of Linguistics, University of Texas at Austin, 305 East 23rd Street STOP B5100, Austin, Texas 78712, USA*

ABSTRACT:

Speaking style variation plays a role in how listeners remember speech. Compared to conversational sentences, clearly spoken sentences were better recalled and identified as previously heard by native and non-native listeners. The present study investigated whether speaking style variation also plays a role in how talkers remember speech that they produce. Although distinctive forms of production (e.g., singing, speaking loudly) can enhance memory, the cognitive and articulatory efforts required to plan and produce listener-oriented hyper-articulated clear speech could detrimentally affect encoding and subsequent retrieval. Native and non-native English talkers' memories for sentences that they read aloud in clear and conversational speaking styles were assessed through a sentence recognition memory task (experiment 1; $N = 90$) and a recall task (experiment 2; $N = 75$). The results showed enhanced recognition memory and recall for sentences read aloud conversationally rather than clearly for both talker groups. In line with the "effortfulness" hypothesis, producing clear speech may increase the processing load diverting resources from memory encoding. Implications for the relationship between speech perception and production are discussed.

 https://doi.org/10.1121/10.0006732

(Received 29 December 2020; revised 12 August 2021; accepted 23 September 2021; published online 5 November 2021)

[Editor: Melissa Michaud Baese-Berk] Pages: 3387–3398

I. INTRODUCTION

Acoustic-phonetic enhancements in the form of listener-oriented intelligibility-enhancing clear speech can improve the listeners' retention of spoken information in memory. Clear speech enhanced the listeners' recognition memory (i.e., recognizing a previously heard item as old) for semantically meaningful and anomalous sentences heard in quiet (Van Engen *et al.*, 2012) and mixed with noise (Gilbert *et al.*, 2014). This benefit was shown to extend to both native and non-native English listeners of various first-language backgrounds (Keerstock and Smiljanic, 2018). In addition to sentence recognition memory, listening to clear speech improved the recall of words and entire sentences (Keerstock and Smiljanic, 2019). The clear speech benefit was also found for older adults with normal-to-moderately impaired hearing-listening abilities in the recall of medically relevant spoken information (DiDonato and Surprenant, 2015). The clear speech benefit on memory is in line with the processing models that invoke increased cognitive load and listening effort when speech comprehension is challenging because of signal degradation or listener characteristics [cf. framework for understanding effortful listening (FUEL), Pichora-Fuller *et al.*, 2016; effortfulness hypothesis (EH), McCoy *et al.*, 2005; and the ease of language understanding (ELU), Rönnberg *et al.*, 2013]. Within these models, the processing of degraded acoustic speech signals recruits additional cognitive resources, leaving fewer resources available for encoding speech in memory. Accordingly, facilitatory effects of clear speech during speech perception, due to the presence of optimal, unambiguous speech signals, allow greater cognitive resource allocation for encoding speech in memory.

The present study aims to expand this line of research, which primarily focused on speech perception, to speech production by examining the effect of speaking style on memory for the talkers themselves. To align the production effects to the perception-only effects from our previous work (Keerstock and Smiljanic, 2018, 2019), the memory of the native and non-native English talkers in the present study was also assessed via a recognition memory task (experiment 1) and a recall task (experiment 2). The crucial difference is that in the previous work, listeners heard sentences that were produced by a different talker. In the current work, talkers who produced conversational and clear speech sentences were themselves tested on the memory tasks. Both memory tasks consisted of two phases: an exposure phase followed by a test phase, which examined how well the material presented during the exposure phase was remembered. In a recognition memory task, the test phase involved presenting the material from the exposure phase

[a)]Portions of this research were presented at the 177th Meeting of the Acoustical Society of America, Louisville, KY, USA, May 2019; the Psychonomic Society's 61st Annual Meeting, November 2020; and the 179th Meeting of the Acoustical Society of America, December 2020.

[b)]Electronic mail: skeerstock@missouri.edu, ORCID: 0000-0001-5121-594X.

mixed with new/unfamiliar material and assessing participants' familiarity ("is this item familiar/is it 'old' or 'new?'") with a binary response (yes/no). In a recall task, the test phase involved asking participants to recall/write down as much as they can remember, which involved retrieving lexical items and entire units of connected meaning from memory and provided a quantitative assessment of the participants' memory. Combining the results from two different memory tasks, recognition memory and recall, helps assess the generalizability and robustness of the results to different yet ubiquitous memory processes. The goal was to provide a more comprehensive account of the speaking style production effect on memory for spoken information.

Although listening to clear speech (perception-only) has been linked to the subsequent enhancement of memory for speech, the effect of producing clear speech on subsequent memory is unknown. Two alternative conceptual frameworks (the "production effect" and "effortfulness models") predict opposite outcomes for reading aloud in clear compared to conversational speaking style on talkers' recognition memory and recall performance. The production effect (MacLeod *et al.*, 2010), derived from the "generation effect" (Slamecka and Graf, 1978), refers to the superior memory retention of material "produced" in some form relative to material read silently during an encoding phase. The benefit of producing items on recognition memory was found for a variety of production types: silent mouthing, saying nonwords aloud (MacLeod *et al.*, 2010), saying words loudly or singing words (Quinlan and Taylor, 2013), and for a wide range of materials, including word pairs, sentences and textbook passages (Ozubko *et al.*, 2012) and dialogues (Knutsen and Le Bigot, 2014). The production effect was observed in diverse modalities, such as writing, typing or spelling (Forrin *et al.*, 2012), although the benefit was lower compared to producing the words out loud. The distinctiveness account (Hunt, 2006, 2013) seems to be the primary explanation for the memory benefit. As described in Forrin and MacLeod (2018) and MacLeod *et al.* (2010), among others, reading aloud involves two distinctive processes, motor/articulatory processes and auditory processes, which help with discriminating produced from unproduced items during subsequent memory testing. Quinlan and Taylor (2013) tested the effects of additional distinctive elements in vocal productions and found that words produced loudly and words sung presented a memory advantage relative to words read aloud normally and words read silently. The authors proposed that higher intensity in loud speech and wider F0 modulation in sung productions provided an additional layer of salient acoustic information, which facilitated memory retrieval.

Similar to singing and reading aloud loudly, listener-oriented clear speech involves increased f0 modulations and increased intensity relative to the conversational speech. Clear speech also typically includes numerous other articulatory/acoustic modifications: decrease in the speaking rate, more salient release of stop consonants, expansion of the vowel space, and enhancement of language-specific vowel and consonant contrasts (Smiljanic, 2021; Smiljanic and Bradlow, 2009). According to the production effect, the prediction is that producing clear speech should facilitate memory retention as it provides a number of additional salient articulatory and acoustic cues relative to conversational speech. However, clear speech production is complex and may be resource demanding. Compared to conversational speech and even to reading aloud loudly or singing, producing articulatory-acoustic modifications in clear speech requires accessing enhanced forms at multiple linguistic levels (phonemic, lexical, prosodic), which may impact speech planning and implementation. Articulating clear speech involves gestures of greater magnitude and peak speed, longer movement durations, greater distances than casual speech (Perkell *et al.*, 2002; Song, 2017; Tang *et al.*, 2015), and may involve increased time to inhale to a higher lung volume to generate greater vocal intensity or sound pressure level. At the conceptual level, addressing the listener's perceptual difficulty in a way that considers the nature of a specific communication barrier may be resource demanding. Furthermore, this requires continuous monitoring of the feedback from the listener as to whether the adaptations are successful in increasing intelligibility. The complex and dynamic nature of such adaptations suggest that clear speech production may be more effortful. To the extent that models like EF, FUEL, and ELU can be applied to predict memory performance following speech production, the more-effortful-to-produce speaking style (i.e., clear speech) should lead to reduced memory performance in contrast to the production effect.

The present study tested the two competing hypotheses. According to the production effect (MacLeod *et al.*, 2010), reading aloud in clear speech will enhance talkers' recognition memory (experiment 1) and recall (experiment 2) compared to reading aloud in conversational speaking style. Based on the effortfulness models (EF, FUEL, and ELU), we predicted that the additional cognitive and physiological effort expanded while reading aloud clearly will detrimentally affect encoding, leading to lower recognition memory (experiment 1) and recall (experiment 2) compared to reading aloud in conversational speaking style.

Adding insights from the effect of speech production on memory to the existing speech-perception literature presents theoretical interest. The relationship between speech perception and production is not well understood. Whereas a cooperative relationship is often assumed between the two modalities (Casserly and Pisoni, 2010; Denes and Pinson, 1963; Goldinger, 1996), mismatches between the two processes have also been reported in the literature (Baese-Berk, 2019; Dupoux *et al.*, 1999). Little is known about how cognitive resources are shared by the two modalities with regard to speaking style modifications and their impact on memory. In the present study, participants in the exposure phase read aloud sentences presented on the screen in either clear or conversational speech. They were asked to commit the sentences to memory and subsequently tested on how well they remembered the sentences that they produced in either

https://doi.org/10.1121/10.0006732

speaking style. If the processing resources are shared as the demands of the production task increase (producing clear vs conversational speech), processing resources will be diverted from memory encoding and allocated to the production task. This will result in lower recognition memory and recall for clear sentences compared to that for conversational sentences. If, on the other hand, memory encoding is not affected by the complexity of the production task, the memory outcomes should be the same when talkers produce conversational and clear speech sentences. This would suggest that the two modalities are dissociated to some extent.

Finally, the present study considered the effect of the native language background on the retention of different speaking styles in memory. In particular, we included native English talkers, those who acquired English from birth, and non-native talkers, those who acquired English later in life [second language (L2) learners] with various first-language backgrounds. Compared to native talkers, non-native talkers have less experience with the target language at all levels of linguistic structure, leading to processing difficulties in perception and production (Bradlow *et al.*, 2018; Reichle *et al.*, 2016). The cognitive demands of L2 speech processing typically result in a memory disadvantage relative to native listeners (Hygge *et al.*, 2015; Molesworth *et al.*, 2014). It is expected that the effort involved in L2 speech production, from lexical retrieval to articulation of L2 forms (Levelt, 1989), should detrimentally affect memory for non-native talkers. Producing clear speech is expected to further increase the cognitive demands for L2 speakers. Although they can typically implement global modifications in clear speech (increased F0 mean and energy between 1 and 3 kHz), non-native talkers may experience difficulties in producing language-specific enhancements (e.g., consonant and vowel phonemic contrasts, prosodic structure; Granlund *et al.*, 2011; Rogers *et al.*, 2010). This increased difficulty suggests that additional cognitive and physiological resources will be recruited, which can detrimentally affect memory. The current work aims to fill in the gap in our understanding of how increased production difficulty impacts memory for non-native talkers.

II. EXPERIMENT 1: SENTENCE RECOGNITION MEMORY

A. Participants

The participants included 60 native English talkers [37 female; $M_{age} = 19.6$, standard deviation $(SD)_{age} = 2$] and 30 non-native English talkers (22 female; $M_{age} = 23.2$, $SD_{age} = 4.1$), who partook in exchange for class credit or monetary compensation. The sample size was chosen to replicate Keerstock and Smiljanic (2018, 2019), and participants were pooled from the same community of students at the University of Texas (UT) at Austin. The native English talkers were all born and raised in the U.S., acquired English from birth, and reported being English dominant. The difference in sample size between native and non-native talkers resulted from merging into one group the native English talkers with exposure to another language from birth alongside English ($n = 27$) and the native English talkers without exposure to another language from birth ($n = 39$). As shown in the results, the effect of second language exposure did not significantly impact the results, thus, the two groups were combined. The non-native talkers acquired English after they had already acquired their first language, on average at age 7.7 years old (range 5–13 years old) and received no exposure to English from parents/caregivers while growing up. Information about the non-native talkers' language background is provided in the supplementary material.[1] They signed a written informed consent and filled out a detailed language background questionnaire adapted from the LEAP-Q (Marian *et al.*, 2007). All of them passed a hearing screening, administered bilaterally at 25 dB hearing level (HL) at 500, 1000, 2000, and 4000 Hz.

B. Material

The material consisted of 120 unique meaningful sentences taken from the Basic English Lexicon (BEL) sentence materials (Calandruccio and Smiljanic, 2012) and was used in previous work on clear speech memory (Keerstock and Smiljanic, 2018, 2019; Van Engen *et al.*, 2012; Gilbert *et al.*, 2014). The sentences contain high-frequency words that are familiar to non-native English speakers. The corpus was tested on a large cohort of non-native speakers of English and found to be appropriate for use in speech-perception testing with this population (Rimikis *et al.*, 2013). All sentences used here were composed of four content words and two function words organized in two syntactic structures: (1) an adjective and noun, followed by a verb and a single noun (e.g., "The hot sun warmed the ground."); or (2) a single noun followed by a verb and an adjective and a noun ("The mother baked the delicious cookies."). The sentences were, on average, 8 syllables long (range, 6–12 syllables; SD = 1.35). Of the 120 sentences, 60 were read aloud by the talkers in the exposure phase, and the remaining 60 sentences were used as a decoy in the recognition memory test phase during which sentences were read silently.

C. Procedure

Participants were seated facing a computer monitor in the sound-attenuated booth at the UT Phonetics Laboratory at the UT at Austin. The experiment consisted of a familiarization phase followed by a sentence recognition experiment and, finally, concluded with an additional recording phase.

The goal of the first phase was to familiarize the participants with the two speaking styles before the experiment. First, participants were instructed to read the practice sentence, "The dark house scared the baby," which was centered on the screen of a PowerPoint (Microsoft, Redmond, WA) slide once in each speaking style. The following instructions were written on the screen one at a time to elicit conversational and clear speaking styles: "Read this sentence in a normal, casual way, as if you were talking to a

family member or a close friend" and "Read this sentence clearly and carefully, as if talking to a non-native speaker of English or a person with hearing loss." Verbal feedback only consisted of reiterating word-for-word the instructions, and no other indication as to how to produce the speaking styles was provided. Then, participants listened to an audio recording example of "The dark house scared the baby," produced in each speaking style by the speaker who produced the stimuli in Keerstock and Smiljanic (2018). Each example was played only once to limit the imitation effects. Finally, participants were asked to read aloud one more time the practice sentence in each speaking style. This familiarization phase lasted approximately 5 min.

The recognition memory experiment consisted of the exposure phase and test phase. All of the instructions and stimuli were presented in E-Prime 2.0 (Pittsburgh, PA). The E-Prime button box was used to navigate through the experiment and record participants' responses. Participants' productions were recorded in E-Prime using a Logitech head-mounted microphone (Fremont, CA). The experimental session started with four different practice sentences not used in the main experiment. The goal of the practice sentences was to ensure that the participants were comfortable using the button box and reading aloud the sentence in two speaking styles into the microphone as soon as the sentence appeared on the screen. Each sentence was written in the center of the screen against a uniform white background in black Arial size 25 pica font and remained on the screen for a duration of 6000 ms to give participants sufficient time to produce the sentence. Participants were not instructed to self-correct the errors that they produced and if they self-corrected, they were not encouraged to stop doing so. At the beginning of the exposure phase of the experiment, an instruction screen informed the participants that they had to commit to memory the sentences that they were reading aloud and there would be a memory test at the end. Participants produced 6 blocks of 10 randomized unique sentences for a total of 30 sentences in clear and 30 sentences in conversational speaking style. The speaking style presentation was counterbalanced such that half of the participants produced all of the sentences in blocks 1, 3, and 5 in conversational speech, and all of the sentences in blocks 2, 4, and 6 in clear speech, and half of the participants produced all of the sentences in blocks 1, 3, and 5 in clear speech and all of the sentences in blocks 2, 4, and 6 in conversational speech. This ensured that all of the sentences were produced in both conversational and clear styles across speakers. A screen instructing the participant as to which speaking style to adopt appeared before every block.

Immediately after producing all 60 sentences, the participants completed the recognition memory test. In the test phase, the participants were presented with all of the items from the exposure phase (60 old sentences) and 60 new items (sentences they did not produce in the exposure phase). The written sentences were randomly presented one at a time in the center of the screen against a uniform white background in black Arial size 25 pica font. Each sentence was presented only once. For each sentence, the participants used the button box to indicate whether the sentence was old (from the exposure) or new (distractor). The sentence remained on the screen until a response was recorded. The participants were instructed to respond as quickly and accurately as possible. The recognition memory experiment lasted approximately 25 min.

After the recognition memory experiment was completed, the participants recorded the same 60 sentences that they had produced during the exposure phase, but this time in the opposite speaking style (e.g., if they produced sentences 1–10 in the conversational style in the exposure, they now produced those same sentences in the clear speaking style). These additional recordings were used in acoustic analyses to assess whether the talkers had produced two distinct styles during exposure. These additional recordings took approximately 10 min to complete.

D. Analyses

1. Acoustic analyses

The acoustic analyses were conducted to verify that the talkers implemented two distinct (conversational and clear) speaking styles.[1]

Furthermore, we examined the speech onset latency as an indication of the cognitive and/or physiological need involved in speech planning and implementation of the conversational vs clear speaking style. The speech onset latency was measured on all of the 10 800 audio sentences as the duration (in ms) from the stimulus onset (i.e., when the target sentence first appeared on the screen) to the onset of the speech. The durations were entered in a LMER as the dependent variable, speaking style (conversational$_{[reference]}$vs clear), talker group (native$_{[reference]}$ vs non-native), and the speaking style by talker group interaction were entered in the model as independent variables. The subject and sentence were treated as random effects.

2. Recognition memory

The recognition memory data were analyzed within a signal detection framework (Snodgrass and Corwin, 1988) and followed previous analyses in Keerstock and Smiljanic (2018). The hit rates (recognizing an old item as old) and miss rates (recognizing an old item as new) were computed for each participant in each speaking style. One correct rejection rate (recognizing a new item as new) and one false alarm rate (recognizing a new item as old) per participant were computed as the new sentences were never produced in any speaking style because they were only orthographically presented and silently read during the test. To assess the discrimination sensitivity and accuracy independently of the response bias, the detection sensitivity (d') and response bias (C) were computed for each participant in each speaking style. d' scores were calculated for each participant by subtracting the normalized probability of the overall false alarm rate from the normalized probability of either the conversational or clear hit rate. These probabilities were

corrected to accommodate the hit and false alarm rates 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). d' measures the distance between the signal and the noise in SD units. A d' score value of 0 indicates an inability to distinguish signals from noise, whereas d' score values above zero indicate a greater detection sensitivity, an enhanced ability to distinguish signals from noise (Stanislaw and Todorov, 1999). C scores were calculated as in Snodgrass and Corwin (1988), wherein positive C values indicate a bias toward responding "new," and negative C values indicate a bias toward responding "old."

Finally, in the test phase, reaction times (RTs) were measured as the time between the onset of the on-screen-stimulus presentation to the time that the participant pressed a button on the button box to indicate their decision (old/new). The RTs of the recognition decisions for the sentences read aloud in conversational vs clear style in the exposure phase were compared. Only old trials were included in the analysis because new trials were made of distractor items that had never been read aloud. Both of the correct responses (i.e., "hit"; old item correctly recognized as old) and incorrect responses (i.e., "miss"; old item incorrectly recognized as new) were included in the analyses, and accuracy (1–0) was included as a covariate in the statistical models. No outliers were detected and, therefore, all of the data points were submitted for analysis.

The LMERs were conducted on d' scores with the RTs (s) as the dependent variables. The speaking style (conversational$_{[reference]}$ vs clear), talker group (native$_{[reference]}$ vs non-native), and speaking style by talker group interaction were included in the models. The fixed effects were contrast-coded zero vs one. The subject was treated as a random effect. The pairwise comparisons were performed with the emmeans package in *R* (Lenth *et al.*, 2018). The effect sizes were measured with Cohen's d (Cohen, 1988).

E. Results

1. Acoustics

Compared to the conversational sentences, the clear sentences were produced with a slower articulation rate, higher pause rate, increased energy in the 1–3 kHz range, and wider F0 range by both talker groups. The native talkers also produced a higher F0 mean in the clear compared to the conversational speech, but this was not significant for non-native talkers.[1] Overall, the results from the acoustic analyses confirmed that the talkers implemented conversational-to-clear acoustic-articulatory adaptations along the dimensions that are typically found in listener-oriented speaking style adaptations (Smiljanic and Bradlow, 2009).

Figure 1 (left panel) shows the speech onset latency (in ms) for native and non-native talkers in conversational and clear speech. Results from the LMER indicated a significant main effect of the speaking style ($\beta = 121.6$, $t = 34.8$, $p < 0.001$) such that the speech onset latency was significantly longer for the clear speech than it was for the conversational speech. The effect of the talker group ($\beta = 47$, $t = 1.7$, $p = 0.09$) and the speaking style by talker group interaction ($\beta = 12.1$, $t = 1.6$, $p = 0.1$) were not significant. Both native and non-native talkers required more time to initiate speech when reading sentences aloud in clear speech compared to conversational speech.

2. Recognition memory

Table I shows the mean (SD) of d', C, and the RTs (ms) for native and non-native listeners for conversational and clear sentences. Figure 2 (left panel) shows the distributions of the d' scores for individual native and non-native talkers in the two speaking styles. Average C scores for both talker groups were positive, indicating that the participants were generally biased to respond "new" more often than "old." This bias was stronger for the speech produced in a clear style for both talker groups. The results from the LMER that ran on d' scores as the dependent variable revealed a significant main effect of the speaking style ($\beta = -0.09$, $t = -2.6$, $p = 0.0102$, Cohen's $d = -0.56$) such that the d' scores, the ability to discriminate old from new sentences, were significantly lower in clear speech than in conversational speech. The effect of the talker group was not significant ($\beta = -0.09$, $t = -0.773$, $p = 0.4417$). The speaking style by talker group was also not significant ($\beta = 0.01$, $t = 0.135$, $p = 0.893$) and, therefore, was removed from the model before interpreting the main effects. Note that splitting the talker group into three levels [native English speakers with no other exposure to another language before age 6 years old ($n = 30$), native English speakers with exposure to both English and another language before age 6 years old ($n = 27$), and non-native English speakers ($n = 30$)] resulted in similar results and the groups were not significantly different from one another.

The results from the LMER that ran on the RTs for old trials showed that there was a significant effect of the talker group ($\beta = 0.432003$, $t = 4.122$, $p < 0.001$) such that the non-native talkers were overall slower at responding than the native talkers, but there was no effect of the speaking style ($\beta = 0.001337$, $t = 0.052$, $p = 0.958$) and no interaction between the speaking style and talker group ($\beta = -0.004245$, $t = -0.078$, $p = 0.94$). The covariate, accuracy, was significant ($\beta = -0.189378$, $t = -6.831$, $p < 0.001$) such that the incorrect responses (i.e., Miss) were slower than the correct responses (i.e., hit). Taken together, the RT analyses pointed out a difference between the native and non-native talkers but showed that there were no statistically significant differences in the RTs between the sentences previously read aloud in conversational vs clear speech.

III. EXPERIMENT 2: SENTENCE RECALL

A. Participants

Forty-three native English talkers (24 female; $M_{age} = 19.3$, $SD_{age} = 1.7$) and 32 non-native English talkers (19 female; $M_{age} = 22.5$, $SD_{age} = 3.9$) participated in

experiment 2 in exchange for class credit or monetary compensation. They were different individuals than those is experiment 1 but were recruited from the same pool of students at the UT at Austin and had similar demographic and linguistic profiles. The native English talkers were born and raised in the U.S., acquired English from birth, and reported being English dominant. As in experiment 1, the native talker group was comprised of talkers with exposure to another language from birth alongside with English ($n = 20$) and talkers without exposure to another language from birth ($n = 23$), but this factor was not significant (cf. Sec. III E). The non-native talkers acquired English, on average, at age 7.9 years old (range 5–16 years old) and received no exposure to English from parents/caregivers while growing up. Information about the non-native talkers' language background is provided in the supplementary material.[1] They signed a written informed consent and filled out a detailed language background questionnaire adapted from the LEAP-Q (Marian *et al.*, 2007). All of them passed a hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

B. Material

The stimuli consisted of 72 unique sentences from the same BEL corpus that was used in experiment 1. The sentences used in experiment 2 included only one syntactic structure, consisting of a single noun phrase (determiner + noun) at the beginning of the sentence followed by a verb, an adjective, and a noun. The sentences were, on average, 8 syllables long (range, 6–11 syllables; SD = 1.42). The sentence-initial noun phrase was provided as a cue in the recall booklet (e.g., The grandfather), and the participants had to fill in the rest of the sentence (e.g., drank the dark coffee). This resulted in a considerable overlap between experiments 1 and 2: out of the 72 sentences used in experiment 2, 55 were also used in experiment 1, whereas the remaining 17 sentences were supplemented from the BEL corpus such that they would fit the syntactic requirement needed for the recall task.

C. Procedure

The participants were seated in a sound-attenuated booth facing a computer monitor. The instructions and stimuli were presented with E-Prime 2.0. As in experiment 1, experiment 2 consisted of a familiarization phase, followed by a recall task, and concluded with an additional recording phase. The familiarization phase was identical to the one described in experiment 1. The recall task was nearly identical to the one employed and described in Keerstock and Smiljanic (2019) except that participants were asked to read aloud instead of listen. The 72 test sentences were pseudo-randomized and divided into 6 blocks of 12 sentences to avoid repetitions of written cues or target words within the blocks. Each sentence was presented in the center of the screen against a uniform white background in black Arial size 25 pica font and remained on the screen for a duration of 6000 ms. The participants were asked to read aloud the sentence as soon as it appeared on the screen and commit it to memory. The speaking style presentation was counterbalanced across the blocks: half of the participants produced the sentences in blocks 1, 3, and 5 in conversational speech and blocks 2, 4, and 6 in clear speech, and half of the participants produced the sentences in blocks 1, 3, and 5 in clear speech and blocks 2, 4, and 6 in conversational speech. This ensured that all of the sentences were produced in both conversational and clear styles across the speakers. A screen instructing the participant which speaking style to produce appeared before every block. The productions were recorded in E-Prime using a Logitech head-mounted microphone.

After producing each block of 12 sentences, the participants were asked to write down what they remembered in the recall booklet. Each sentence was cued by the first noun phrase ("The grandfather," "The mother") written in the recall booklet. The participants were asked to recall and write down the rest of the sentence (e.g., "drank the dark coffee." or "baked the delicious cookies."). The recall cues written in the booklet were in the same presentation order as they were during the reading aloud phase; however, the participants were not instructed to fill the recall booklet in any particular order. The recall phase was self-paced. Overall, the recall experiment lasted approximately 30 min.

Finally, as in experiment 1, the participants were recorded again producing the same 72 sentences that they had produced during the memory task, but this time in the opposite speaking style (e.g., if they produced sentences 1–12 in the conversational style in the exposure, they now produced those same sentences in the clear speaking style) to assess whether the talkers had actually produced two distinct styles during exposure.

D. Analyses

1. Acoustic analyses

As in experiment 1, we also examined the speech onset latency (for the 10 800 audio sentences as the duration from the stimulus onset to the onset of the speech) as an indication of the cognitive and/or physiological need involved in speech planning and implementation of the conversational vs clear speaking style.[1]

2. Recall

Following Keerstock and Smiljanic (2019), each keyword to be recalled was scored as either correct (1) or incorrect (0). Because there were 36 sentences (with 3 keywords per sentence) in each speaking style, there were 108 keywords per speaking style to be recalled per subject. We adopted a strict scoring criterion whereby any morpho-phonological mismatch (e.g., "flowers" instead of "flower") was scored as incorrect. The participants were not penalized for obvious spelling errors. In the case of uncertainty due to the handwriting, the sentences were scored by another research assistant and a consensus was reached. To predict the recall outcome (0 or 1), we conducted binomial logistic

https://doi.org/10.1121/10.0006732

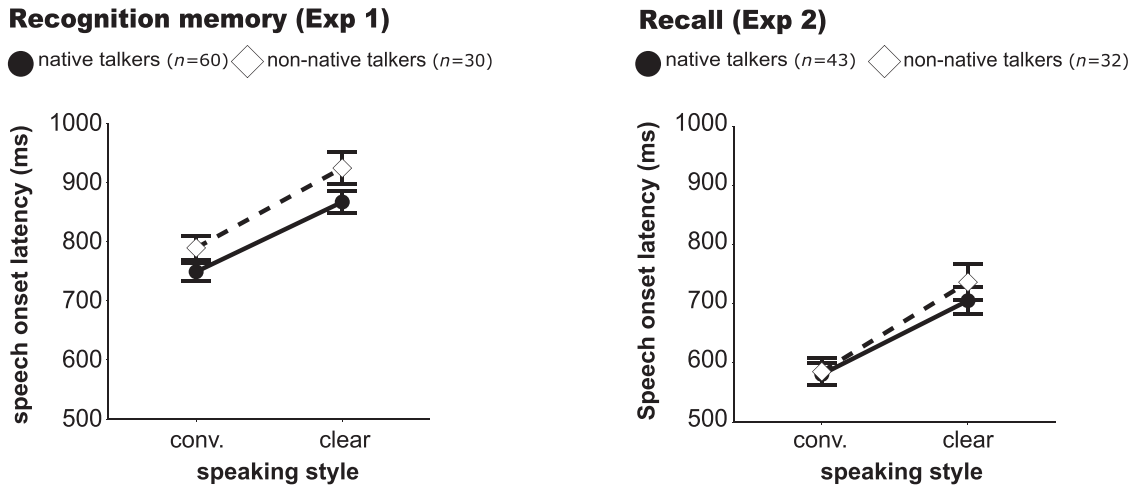

FIG. 1. The speech onset latency (in ms) in conversational ("conv.") and clear speech for native (solid) and non-native (dashed) talkers in the recognition memory experiment (left) and the recall experiment (right). The error bars represent the standard error.

regressions with keyword recall as the dichotomous dependent variable using the generalized linear mixed-effects regressions (GLMER) function of the lme4 package in *R* (Bates *et al.*, 2015). The model included the speaking style (conversational$_{[reference]}$ vs clear) and talker group (native$_{[reference]}$ vs non-native) as the independent variables and the speaking style × talker group as an interaction term. In addition, we included the following covariates: word position (1, 2, 3; e.g., drank, dark, coffee, respectively) as a covariate to account for the position of the word in the sentence, sentence position (1–12) as a covariate to account for the serial position effects within each block of 12 sentences (i.e., primacy and recency), and block position (1–6) as a covariate to account for the practice and/or fatigue effects throughout the experiment. The subject and stimuli were modeled using a random intercept term.

E. Results

1. Acoustic analysis

The results from the acoustic analyses confirmed that both talker groups produced the typical acoustic–articulatory clear speech adaptations, namely, slower articulation rate, higher pause rate, increased energy in the 1–3 kHz range, and a wider F0 range. The F0 mean was not significantly different in the clear and conversational speech for either group.[1]

Figure 1 (right panel) shows the speech onset latency (in ms) for native and non-native talkers in conversational and clear speech. The results from the LMER indicated a significant speaking style by talker group interaction ($\beta = 37.8$, $t = 3.9$, $p < 0.001$), a main effect of the speaking style ($\beta = 122.4$, $t = 19.5$, $p < 0.001$) but no effect of the talker group ($\beta = -4$, $t = -0.13$, $p = 0.9$). Overall, both native and non-native talkers required more time to initiate speech when reading aloud clearly compared to conversationally. The simple effect analyses revealed that the speaking style by talker group interaction was driven by a larger effect of the speaking style for non-native talkers compared to native talkers. That is, non-native talkers required even more time to initiate speech when preparing to read clearly than when preparing to read clearly conversationally compared to native talkers.

2. Recall

Figure 2 (right panel) shows native and non-native talkers' keyword recall accuracy in the two speaking styles and Table I shows the mean (SD). The results from the logistic mixed-effect regression model with keyword recall (0–1) as the dependent variable are summarized in Table II. The speaking style × talker group interaction was not significant {odds ratio (OR) = 1.11 [95% confidence interval (CI), 0.96–1.28], $p = 0.17$} and, therefore, was removed from the model before interpreting the main effects. There was a

TABLE I. The mean (SD) of d', C, and RTs in seconds for native and non-native listeners for conversational and clear sentences in the recognition memory task (experiment 1) and the mean (SD) keyword recall in the recall task (experiment 2).

		Native talkers		Non-native talkers	
	Style	Conversational	Clear	Conversational	Clear
Recognition memory (experiment 1)	d'	1.47 (0.56)	1.37 (0.63)	1.37 (0.48)	1.29 (0.44)
	C	0.28 (0.32)	0.33 (0.35)	0.31 (0.32)	0.35 (0.34)
	RT	1.684 (0.961)	1.688 (0.925)	2.118 (1.218)	2.126 (1.221)
Recall (experiment 2)	% Keyword recall	48.02 (15.05)	45.11 (15.93)	48.61 (18.73)	47.57 (18.42)

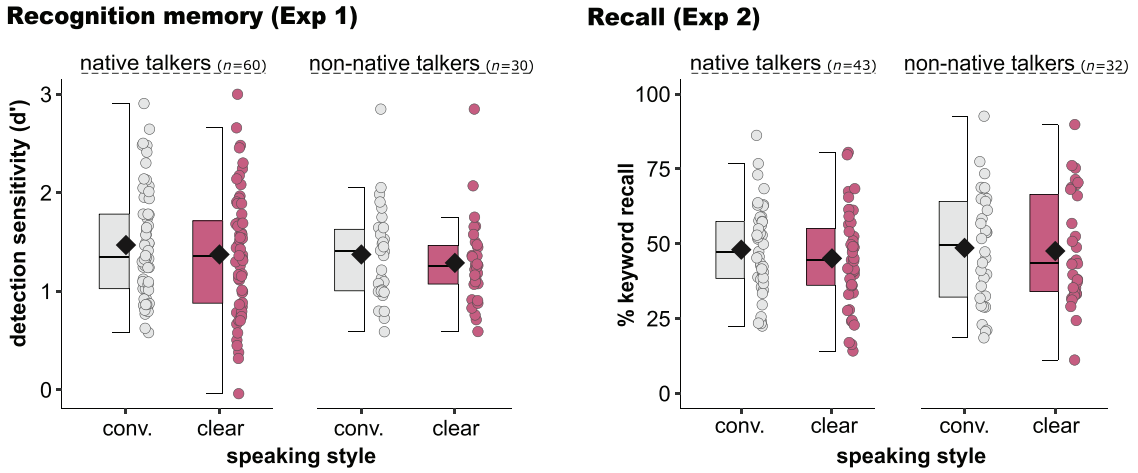

FIG. 2. (Color online) The d' distribution for native ($n = 60$) and non-native talkers ($n = 30$) in recognition memory (experiment 1; left) and correct keyword recall distribution for native ($n = 43$) and non-native talkers ($n = 32$) in recall (experiment 2; right) for sentences read aloud in conversational ("conv") and clear speech. The boxplots extend from the 25th to the 75th percentile with whiskers extending to points within 1.5 times the interquartile range. The central horizontal lines indicate the medians. The diamonds indicate the means. Individual circles adjacent to the boxplots represent the d' scores (left) and keyword recall percentage (right) for individual talkers (one data point per talker and per style).

significant main effect of the speaking style ($p < 0.01$) but no significant effect of the talker group ($p = 0.66$). All of the covariates (sentence position, block position, and word position) significantly affected the keyword recall. There was a large recency effect ($p < 0.001$) such that the 12th (and last) sentence within a block was recalled significantly better than sentences earlier in a block ($M_{\text{keyword}} = 84\%$ for sentence 12 vs $M_{\text{keyword}} = 43\%$ for sentences 1–11). There was also an increase in the keyword recall accuracy throughout the experiment, i.e., across all blocks ($p < 0.001$), consistent with the practice effects ($M_{\text{keyword}} = 39\%$ for the first block vs $M_{\text{keyword}} = 54\%$ for the sixth final block). Finally, a small increase in the keyword recall accuracy ($p < 0.01$) was found as a function of the word position within the sentence ($M_{\text{keyword}} = 46\%$ for the first position vs $M_{\text{keyword}} = 47\%$ for the middle position, $M_{\text{keyword}} = 50\%$ for the last position).

IV. DISCUSSION

Native and non-native talkers' memory for sentences read out loud in clear and conversational speaking styles was assessed either in an old/new recognition memory test (experiment 1) or in a recall task (experiment 2). Recognition memory and recall were significantly *reduced* for sentences produced in clear speech compared to sentences produced in conversational speech for both native and non-native *talkers*.

The production effect would have predicted the opposite outcome to the one found here, namely, that clear speech production should enhance memory because of the availability of the auditory information and more distinctive acoustic-articulatory features (MacLeod *et al.*, 2010; Quinlan and Taylor, 2013). In Quinlan and Taylor (2013), for example, the production of loud speech or sung words resulted in superior retention of the material compared to normal aloud production and silent reading. As argued by the authors, generating speech with more distinct intensity and pitch features led to enhanced memory for those items. Presumably, the process of vocalizing words resulted in encoding and retention of these additional features. In explicit memory tests, the participants then retrieved the distinctive speech information and used it to recognize whether a word was studied or not. Instead, here, we found that clear speech production resulted in *inferior* memory retention compared to conversational speech.

The acoustic analyses of the talkers' speech output confirmed that the participants had produced distinct speaking styles consistent with typical clear speech modifications: slower articulation rate, higher pause rate, increased F0 mean, increased F0 range, and increased energy within the 1–3 kHz range (and likely some others we have not measured here; cf. Smiljanic and Bradlow, 2009). Although these analyses were conducted mainly to show that the talkers in the current study produced two distinct speaking styles as instructed, the results provide a large-scale (165 talkers who each recorded between 120 and 144 sentences for a total of 21 600 recordings) confirmation of the well-documented conversational-to-clear acoustic modifications,

TABLE II. The summary of the logistic mixed-effect regression model in recall (experiment 2). Bold: statistically significant p values ($p < 0.05$).

	Dependent variable: Accuracy binomial		
Predictors	Odds ratios	95% CI	p
(Intercept)	0.18	0.09–0.34	**<0.001**
Sentence position	1.14	1.08–1.21	**<0.001**
Block	1.18	1.05–1.32	**0.006**
Word position	1.11	1.07–1.16	**<0.001**
Speaking style (clear)	0.90	0.84–0.97	**0.004**
Talker group (non-native)	1.09	0.74–1.59	0.658

https://doi.org/10.1121/10.0006732

which had previously been demonstrated for fewer talkers (Lam *et al.*, 2012; Mefferd and Dietrich, 2020; Smiljanic and Gilbert, 2017). Furthermore, there is evidence that the production of clear speech includes distinct acoustic features that benefit memory during speech perception. Van Engen *et al.* (2012) and Keerstock and Smiljanic (2018) showed that the enhanced recognition memory for clear speech was largely due to a lower rate of false alarm responses. The lower false alarm rate suggests that a greater number of clear speech distinctive features are available to listeners in memory compared to conversational speech, allowing them to reject the new items that did not match those features (also, see similar reasoning for face recognition in Lamont *et al.*, 2005). In the present study, talkers, despite producing a number of distinct cues, were not able to encode them or use them to retrieve the spoken information.

We considered the ways in which our study differed from the production effect studies, which may have led to diverging results. In the present study, the talkers were instructed to imagine an interlocutor with a perceptual problem, specifically, a non-native interlocutor or a listener with a hearing problem. This task seems far more complex than the elicitation of loud items, for instance (Quinlan and Taylor, 2013). Extensive work demonstrated that the talkers apply different acoustic-articulatory "tweaks" in response to different communicative challenges (Cooke and Lu, 2010; Hazan and Baker, 2011; Smiljanic and Gilbert, 2017). The clear speech modifications are, thus, influenced by the non-linguistic context in an adaptive manner, requiring the talkers to modify their production online in response to their audience and listening conditions (Buz *et al.*, 2016; Lee and Baese-Berk, 2020). Although our talkers produced read speech in a laboratory setting, the task still required complex considerations and modifications along all levels of the linguistic structure (Smiljanic and Bradlow, 2009; Smiljanic, 2021). For instance, clear speech production requires retrieving hyper-articulated forms of phonemes, restructuring prosodic cues, planning acoustic-articulatory movements for more exaggerated targets, and increasing the articulatory effort. Furthermore, producing a significant unit of connected meaning (i.e., sentences) involves more complex encoding, planning, and implementation compared to producing single words. Whereas the production effect, originally shown with single words, has been replicated with sentences and paragraphs (Ozubko *et al.*, 2012), the effect of sustaining a distinct, listener-oriented speaking style may still be more complex, although a closer comparison of the two tasks warrants further examination.

The *reduced* memory, despite the presence of distinctive acoustic-phonetic features in clear speech production, contrasts notably with findings from the clear speech perception. In previous studies, listening to clear speech was found to enhance recognition memory (Gilbert *et al.*, 2014; Keerstock and Smiljanic, 2018; Van Engen *et al.*, 2012) and recall (DiDonato and Surprenant, 2015; Keerstock and Smiljanic, 2019) compared to fast and often reduced forms of conversational speech. This perceptual benefit is in line with the ELU, FUEL, and EH models, which invoke an increased cognitive load and listening effort when access to the speech signal is impeded or the speech signal itself is degraded (McCoy *et al.*, 2005; Peelle, 2018; Pichora-Fuller *et al.*, 2016; Rabbitt, 1968; Rönnberg, 2003; Rönnberg *et al.*, 2013; Van Engen and Peelle, 2014; Winn *et al.*, 2015; Winn and Teece, 2021; Zekveld *et al.*, 2010, 2011). Easier-to-understand clear speech, as shown in enhanced intelligibility in some of the previous work, required fewer processing resources during speech perception, leaving more resources for encoding the information in memory. The lack of a memory benefit in clear speech production seems, in fact, compatible with ELU, FUEL, and EH models. Implementing and sustaining a listener-oriented speaking style may require additional cognitive resources, which may shift the resources away from encoding the produced information in memory. Although we have not tested the intelligibility of the sentences elicited in the current study, the talkers implemented the typical conversational-to-clear acoustic-articulatory modifications, suggesting the perceptual benefit. The production of the distinct acoustic-articulatory features in clear speech compared to conversational speech is likely more resource demanding even in the absence of the intelligibility benefit. In the production-to-memory loop, then, fewer resources remain available for encoding the distinctive information during clear speech production compared to the casual speaking style, which may be relatively less demanding.

The notion that producing clear speech is effortful aligns with the H and H theory (Lindblom, 1990; Perkell *et al.*, 2002), which posits that talkers adjust their spoken output in a continuous manner, varying from hypo-articulated speech (i.e., talker-centric economy of effort) to hyper-articulated speech (listener-oriented maximizing intelligibility). From an articulatory standpoint, clear speech involves larger articulator movement and increased peak velocity compared to conversational speech (Song, 2017; Tang *et al.*, 2015; Tasko and Greilick, 2010). The current speech onset latency results, showing that the talkers were consistently slower to initiate clear speech than conversational speech, align with the notion that planning and execution of clear speech requires additional resources, whether these resources are cognitive in nature (e.g., speech planning) or physiological (e.g., higher lung volume). The evidence from the speech planning literature shows that increasing the processing demands through, for example, speeded production of tongue twisters (Goldrick and Blumstein, 2006), producing cognates in L2 (Jacobs *et al.*, 2016), or reading paragraphs with increasing lexical difficulty by older adults (Gollan and Goldrick, 2019) results in increased reading times and errors in articulatory execution. To the extent that producing clear speech is resource demanding, it is likely that upstream processes, such as encoding in memory, would be accordingly disrupted. Whereas suggestive of the increased effort, our results do not directly implicate speech planning or physiological resources that may underlie the clear speech production

effect on memory. This remains an important avenue for future work.

The current results suggest that the production and perception modalities may compete for central processing resources during speaking style modifications. A similar antagonistic relationship has been found in speech sound learning studies (Baese-Berk, 2019; Baese-Berk and Samuel, 2016). Using a distributional learning paradigm to teach participants novel phoneme categories, Baese-Berk (2019) showed that individuals who were trained in perception alone improved in the production of these contrasts. Individuals who were trained in both perception and production, however, did not show substantial learning in perception. That is, producing novel categories during training disrupted formation of the new sound categories. Although previous work tested the link between the modalities with two separate tasks, in the present study, the participant produced sentences and simultaneously encoded information in memory. Reduced memory for clear speech suggests that additional processing resources were diverted to the sentence production, disrupting memory encoding. More resources remained available for memory encoding when talkers produced casual speech. Future work should directly compare the effect of auditory-only, silent reading, and reading aloud in conversational and clear speaking styles on memory encoding to fully understand how the cognitive resources are allocated in these tasks.

It also remains to be determined where exactly the locus of the increased effort during clear speech production lies. We have hinted at various linguistic, nonlinguistic, and domain general possibilities. Selective attention likely plays a role in memory encoding for clear speech sentences. In perception, the listeners' attention may be drawn to the exaggerated clear speech acoustic-phonetic cues, which then facilitates encoding of these features in memory. In reading sentences out loud clearly, attention may be diverted to producing the needed acoustic-articulatory enhancements and away from encoding the information in memory. It is also possible that diminished memory arises from mind-wandering, which is common during reading and detrimental to comprehension (Franklin *et al.*, 2014; Mooneyham and Schooler, 2013). The reading aloud task in the present study could have led to greater mind-wandering and lower memory compared to when the participants were only hearing sentences in the perception-only study. Note also that we found sizeable individual variability in d' scores (0.56–0.63 in experiment 1) and percent keyword accuracy (15%–18% in experiment 2). This is in keeping with previous perception-only studies (Keerstock and Smiljanic, 2018, 2019) and can be attributed, in part, to the differences in working memory capacities across participants.[2] A pressing goal is to delineate how the production and perception of conversational and clear speaking styles compete for various cognitive resources such as selective attention and working memory.

Finally, the lack of a significant effect of the talker group on recognition memory and keyword recall accuracy and the lack of interactions between the speaking style and talker group is telling. The effect of reading aloud in clear speech was detrimental to the talkers' memory regardless of whether they acquired English as their first or second language. The non-native talkers who participated in the present study were highly proficient and used English in their daily lives, which could account for the lack of differences between the talker groups. However, the cost of performing the task in L2 did manifest itself in the overall slower RTs in responding to old/new items in experiment 1. In earlier perception studies (Keerstock and Smiljanic, 2018, 2019), the non-native listeners were shown to benefit as much as the native listeners from clear speech. This was true for the within-modal sentence recognition and word recall tasks. The L2 processing cost was found only when other responses were examined more closely. For instance, the non-native listeners had longer RTs in the within-modal recognition task and overall lower accuracy in the cross-modal recognition task. Thus, the differences between the native and non-native participants across the production and perception tasks emerge only in more challenging conditions and with more sensitive measures (e.g., RTs; cf. Van Wijngaarden, 2001; van Wijngaarden *et al.*, 2002). Further work should delineate how the clear speech memory benefit is modulated by variation in L2 proficiency and, in particular, for L2 learners who are in the early stages of L2 acquisition. Unlike the present study, which examined non-native speakers with heterogenous L1s, future studies could focus on specific L1-L2 pairs and, for instance, examine the effect of the speaking style on the memory for difficult-to-acquire phonemes given the speakers' L1. It is important to keep in mind that a correlation between intelligibility, listening effort, and memory does not provide the full picture of the processing cost be it in L1 or L2. A closer look at the errors (e.g., types of errors, position within a sentence, sources of ambiguity) could shed light on the mechanism underlying cognitive effort in clear speech production, perception, and memory.

This discussion examined some possible mechanisms underlying the differences between clear speech production and perception effects on memory. Rather than providing answers, it outlined much needed venues for future work with the goal of a more comprehensive understanding of the intelligibility variation and its impact on memory. One such direction is to conduct similar experiments in more naturalistic settings and with more interactive tasks to verify the different effects of clear speech production and perception on memory outcomes. Much needed examination should include objective and subjective measures of the physiological, articulatory, and cognitive efforts involved in producing and perceiving clear and conversational speech, as well as a close look at speech planning processes in different speaking styles.

ACKNOWLEDGMENTS

We would like to thank the UT Sound Laboratory research assistants for their assistance with the data

collection. This research was supported, in part, by the Carlotta Smith Fellowship, Linguistics, UT, to S.K.

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.11121/10.0006732 for detailed language background information about the non-native talkers (Supplement Table I) and for detailed methods and results of acoustic analyses (Supplement Table II).
[2]Although we did not directly measure the individual working memory capacities, the mixed-effect models and random intercept by subject allowed us to take into account the between-person variability when considering the speaking style and group effects.

Baese-Berk, M. M. (**2019**). "Interactions between speech perception and production during learning of novel phonemic categories," Atten., Percep., Psychophys. **81**, 981–1005.
Baese-Berk, M. M., and Samuel, A. G. (**2016**). "Listeners beware: Speech production may be bad for learning speech sounds," J. Mem. Lang. **89**, 23–36.
Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.
Bradlow, A. R., Blasingame, M., and Lee, K. (**2018**). "Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech," Lab. Phonol. **9**(1), 509–524.
Buz, E., Tanenhaus, M. K., and Jaeger, T. F. (**2016**). "Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations," J. Mem. Lang. **89**, 68–86.
Calandruccio, L., and Smiljanic, R. (**2012**). "New sentence recognition materials developed using a basic non-native English lexicon," J. Speech, Lang., Hear. Res. **55**(5), 1342–1355.
Casserly, E. D., and Pisoni, D. B. (**2010**). "Speech perception and production," Wiley Interdiscip. Rev. Cogn. Sci. **1**(5), 629–647.
Cohen, J. (**1988**). *Statistical Power for the Social Sciences* (Laurence Erlbaum and Associates, Hillsdale, NJ).
Cooke, M., and Lu, Y. (**2010**). "Spectral and temporal changes to speech produced in the presence of energetic and informational maskers," J. Acoust. Soc. Am. **128**(4), 2059–2069.
Denes, P. B., and Pinson, E. N. (**1963**). *The Speech Chain: The Physics and Biology of Spoken Language* (Waveland Press, Long Grove, IL).
DiDonato, R. M., and Surprenant, A. M. (**2015**). "Relatively effortless listening promotes understanding and recall of medical instructions in older adults," Front. Psychol. **6**(JUN), 1–20.
Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., and Mehler, J. (**1999**). "Epenthetic vowels in Japanese: A perceptual illusion?," J. Exp. Psychol.: Human Percept. Perform. **25**(6), 1568–1578.
Forrin, N. D., and MacLeod, C. M. (**2018**). "This time it's personal: The memory benefit of hearing oneself," Memory **26**(4), 574–579.
Forrin, N. D., MacLeod, C. M., and Ozubko, J. D. (**2012**). "Widening the boundaries of the production effect," Mem. Cognit. **40**(7), 1046–1055.
Franklin, M. S., Mooneyham, B. W., Baird, B., and Schooler, J. W. (**2014**). "Thinking one thing, saying another: The behavioral correlates of mind-wandering while reading aloud," Psychon. Bull. Rev. **21**(1), 205–210.
Gilbert, R. C., Chandrasekaran, B., and Smiljanic, R. (**2014**). "Recognition memory in noise for speech of varying intelligibility," J. Acoust. Soc. Am. **135**(1), 389–399.
Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol.: Learn., Mem., Cognit. **22**(5), 1166–1183.
Goldrick, M., and Blumstein, S. (**2006**). "Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters," Lang. Cognit. Process. **21**(6), 649–683.
Gollan, T. H., and Goldrick, M. (**2019**). "Aging deficits in naturalistic speech production and monitoring revealed through reading aloud," Psychol. Aging **34**(1), 25–42.
Granlund, S., Baker, R., and Hazan, V. (**2011**). "Acoustic-phonetic characteristics of clear speech in bilinguals," in *17th International Congress of Phonetic Sciences (ICPhS XVII)*, August, pp. 763–766.
Hazan, V., and Baker, R. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**(4), 2139–2152.
Hunt, R. R. (**2006**). "The concept of distinctiveness in memory research," in *Distinctiveness and Memory*, edited by R. R. Hunt and J. B. Worthen (Oxford University Press, Oxford, UK), pp. 2–25.
Hunt, R. R. (**2013**). "Precision in memory through distinctive processing," Curr. Directions Psychol. Sci. **22**, 10–15.
Hygge, S., Kjellberg, A., and Nöstl, A. (**2015**). "Speech intelligibility and recall of first and second language words heard at different signal-to-noise ratios," Front. Psychol. **6**, 1–7.
Jacobs, A., Fricke, M., and Kroll, J. F. (**2016**). "Cross-language activation begins during speech planning and extends into second language speech," Lang. Learn. **66**(2), 324–353.
Keerstock, S., and Smiljanic, R. (**2018**). "Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners," J. Acoust. Soc. Am. **144**(5), 2871–2881.
Keerstock, S., and Smiljanic, R. (**2019**). "Clear speech improves listeners' recall," J. Acoust. Soc. Am. **146**(6), 4604–4610.
Knutsen, D., and Le Bigot, L. (**2014**). "Capturing egocentric biases in reference reuse during collaborative dialogue," Psychon. Bull. Rev. **21**(6), 1590–1599.
Lam, J., Tjaden, K., and Wilding, G. (**2012**). "Acoustics of clear speech: Effect of instruction," J. Speech, Lang., Hear. Res. **55**(6), 1807–1821.
Lamont, A., Stewart-Williams, S., and Podd, J. (**2005**). "Face recognition and aging: Effects of target age and memory load," Mem. Cognit. **33**(6), 1017–1024.
Lee, D.-Y., and Baese-Berk, M. M. (**2020**). "The maintenance of clear speech in naturalistic conversations," J. Acoust. Soc. Am. **147**(5), 3702–3711.
Lenth, R., Singmann, H., Love, J., Buerkner, P., and Herve, M. (**2018**). "emmeans: Estimated marginal means, aka least-squares means.," *R* Package version 1.2.3, available at https://doi.org/10.1080/00031305.1980.10483031 (Last viewed August 1, 2021).
Levelt, W. J. M. (**1989**). "Speaking: From intention to articulation," in *Speaking: From Intention to Articulation* (The MIT Press, Cambridge, MA).
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling, NATO ASI Series*, edited by W. J. Hardcastle and A. Marchal (Springer, The Netherlands), pp. 403–439.
MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., and Ozubko, J. D. (**2010**). "The production effect: Delineation of a phenomenon," J. Exp. Psychol.: Learn. Mem. Cognit. **36**(3), 671–685.
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.
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. Sect. A: Human Exp. Psychol. **58**(1), 22–33.
Mefferd, A., and Dietrich, M. (**2020**). "Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers With dysarthria due to amyotrophic lateral sclerosis: Effects of loud, clear, and slow speech," J. Speech, Lang., Hear. Res. **63**(8), 2625–2636.
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 nonnative speakers," Noise Health **16**(71), 240–247.
Mooneyham, B. W., and Schooler, J. W. (**2013**). "The costs and benefits of mind-wandering: A review," Can. J. Exp. Psychol. **67**(1), 11–18.
Ozubko, J. D., Hourihan, K. L., and MacLeod, C. M. (**2012**). "Production benefits learning: The production effect endures and improves memory for text," Memory **20**(7), 717–727.
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.
Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," J. Acoust. Soc. Am. **112**(4), 1627–1641.
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.
Quinlan, C. K., and Taylor, T. L. (**2013**). "Enhancing the production effect in memory," Memory **21**(8), 904–915.
Rabbitt, P. (**1968**). "Channel-capacity, intelligibility and immediate memory," Q. J. Exp. Psychol. **20**(3), 241–248.
Reichle, R. V., Tremblay, A., and Coughlin, C. (**2016**). "Working memory capacity in L2 processing," Probus **28**(1), 29–55.
Rimikis, S., Smiljanic, 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., DeMasi, T. M., and Krause, J. C. (**2010**). "Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers," J. Acoust. Soc. Am. **128**(1), 410–423.
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**, 68–76.
Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., 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**(July), 1–17.
Slamecka, N. J., and Graf, P. (**1978**). "The generation effect: Delineation of a phenomenon," J. Exp. Psychol.: Human Learn. Mem. **4**(6), 592–604.
Smiljanic, R. (**2021**). "Clear speech perception," in *The Handbook of Speech Perception*, edited by L. C. Nygaard, J. Pardo, D. B. Pisoni, and R. Remez (Wiley, New York), pp. 177–205.
Smiljanic, R., and Bradlow, A. R. (**2009**). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Linguist. Lang. Compass. **3**, 236–264.
Smiljanic, R., and Gilbert, R. C. (**2017**). "Acoustics of clear and noise-adapted speech in children, young, and older adults," J. Speech, Lang., Hear. Res. **60**(11), 3081–3096.
Snodgrass, J. G., and Corwin, J. (**1988**). "Pragmatics of measuring recognition memory: Applications to dementia and amnesia," J. Exp. Psychol. **117**(1), 34–50.
Song, J. Y. (**2017**). "The use of ultrasound in the study of articulatory properties of vowels in clear speech," Clin. Linguist. Phonet. **31**(5), 351–374.
Stanislaw, H., and Todorov, N. (**1999**). "Calculation of signal detection theory measures," Behav. Res. Methods, Instrum., Comput. **31**(1), 137–149.
Tang, L. Y. W., Hannah, B., Jongman, A., Sereno, J., Wang, Y., and Hamarneh, G. (**2015**). "Examining visible articulatory features in clear and plain speech," Speech Commun. **75**, 1–13.
Tasko, S. M., and Greilick, K. (**2010**). "Acoustic and articulatory features of diphthong production: A speech clarity study," J. Speech, Lang., Hear. Res. **53**(1), 84–99.
Van Engen, K. J., Chandrasekaran, B., and Smiljanic, R. (**2012**). "Effects of speech clarity on recognition memory for spoken sentences," PLoS One **7**(9), e43753.
Van Engen, K. J., and Peelle, J. E. (**2014**). "Listening effort and accented speech," Front. Human Neurosci. **8**, 577.
van Wijngaarden, S. J. (**2001**). "Intelligibility of native and non-native Dutch speech," Speech Commun. **35**(1–2), 103–113.
van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native talkers," J. Acoust. Soc. Am. **112**(6), 3004–3013.
Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (**2015**). "The impact of auditory spectral resolution on listening effort revealed by pupil dilation," Ear Hear. **36**(4), e153–e165.
Winn, M. B., and Teece, K. H. (**2021**). "Listening effort is not the same as speech intelligibility score," Trends Hear. **25**, 1–26.
Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2010**). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," Ear Hear. **31**(4), 480–490.
Zekveld, A. A., Kramer, S. E., and Festen, J. M. (**2011**). "Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response," Ear Hear. **32**(4), 498–510.