**The Dissertation Committee for Sandie Keerstock Certifies that this is the approved version of the following Dissertation:**

# Memory for speech of varying intelligibility: effects of perception and production of clear speech on recall and recognition memory for native and non-native listeners and talkers

**Committee:**

Rajka Smiljanic, Supervisor

Scott Myers

Megan Crowhurst

David Quinto-Pozos

Valeriy Shafiro

# Memory for speech of varying intelligibility: effects of perception and production of clear speech on recall and recognition memory for native and non-native listeners and talkers

by

**Sandie Keerstock**

## Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**
**August 2020**

# Acknowledgements

First and foremost, I want to thank my advisor, Rajka Smiljanic. I am unbelievably lucky to have been mentored by Rajka and my gratitude to her is endless: Rajka gave me the most generous academic and emotional support throughout the years and I cannot imagine what this Ph.D. would have been like without her guidance. As a researcher and role model, I would give Rajka most of the credit for becoming the kind of scientist I am today. Next, I want to express my sincere gratitude to all my committee members. Thanks to Scott Myers, who has taught me to build stronger arguments and to clearly elaborate on my contributions, and thanks to Megan Crowhurst who, by initiating me to the art of phonology, has taught me how to be a precise and concise writer and thinker. I am also very grateful to David Quinto-Pozos and Valeriy Shafiro for immediately agreeing to join this committee and for sharing their interdisciplinary insights that have inspired many discussions of this project.

I want to thank the UT Sound Lab, my home for the past 5 years, which contributed to my growth as a researcher, experimentalist and colleague. I want to thank Rajka, the PI, and all the graduate and undergraduate lab members for their discussions, support, feedback and immense help with data collection throughout the years: Kirsten Meemann, Steven Alcorn, Sarah Ransom-Laud, Zhe-Chen Guo, Frida Ballard, Gabby Shaddock, Irene Smith, Luis De La Cruz, Alexandra Brown, Riley Pruden, Meryl Xiong, Jailyn Pena, Maria Gavino, Hanna Rey, Olivia Naworol, Kevin Lilley. A huge thank you to all the participants in my experiments.

I want to thank all the graduate students who made this past 5 years memorable: Ambrocio Gutierrez Lorenzo, Gladys Camacho Rios, Hammal AlBulushi, Laura Faircloth, Justin Power, Se Yeon Park, Lizzy Wood, Karina High, Qianping Gu, Eric Holgate and Jaime Perez Gonzalez. Thank you to the faculty and staff in the Department of Linguistics at UT, in particular our generous department chair Richard Meier for moral and financial support and Ben Rapstine for his incredible helpfulness. I want to thank all the professors that I TAed for, for a truly enriching experience. A special thank you to Harvey Sussman: I am honored to have TAed Language and the Brain; it not only broaden my horizon to neurolinguistics, I also received great advice and mentorship. I also want to thank everyone who supported me when I applied to graduate school in the US (Ioana Chitoran, Cecile Fougeron, Philippe Martin and Amandine Vuattoux). Thank you to the scholars who gave me precious feedback whenever I presented this work at the Acoustical Society of America and LabPhon meetings. I gratefully acknowledge the generous financial support from the Continuing Dissertation Fellowship, Carlotta Smith Fellowships and Professional Development Awards.

Finally, I want to thank my wonderful friends and family for their love and support: Diego and Audrey for 20+ years of friendship; Rozen Neupané; Joshua Bryant; Lina & Patrick; Chantal; Jasme Lee; Sarah & Kavi; Taylor & Jason; Lauren & Josh; Shannon & Adley; Alisa; Jesus & Jose; Ashley Watts; Hannah Volpert-Esmond. I am exceptionally grateful to my husband, Roberto Cofresí, for his unconditional love and for his positive and joyful attitude which made the completion of this Ph.D. with minimal stress possible (his communicable passion for research and ability to troubleshoot most things in R was a nice bonus!). Thanks to the Bonilla-Pesante and Cofresí for loving me and making their home mine. Thanks to my kindhearted mother Isabelle, who I miss dearly. A huge thank you to my role model and pillar in life, my father Erick Keerstock, for being so supportive, for always being on my side, and for his regular visits to the US despite the long distance with France—a much needed source of love and comfort.

# Abstract

## Memory for speech of varying intelligibility: effects of the perception and production of clear speech on recall and recognition memory for native and non-native listeners and talkers

Sandie Keerstock, Ph.D.
The University of Texas at Austin, 2020

Supervisor: Rajka Smiljanic

This dissertation examines the effects of signal-related articulatory-acoustic enhancements in the form of clear speech on signal-independent processes and integration of information in memory. In a series of five experimental studies, this dissertation investigates the effect of clear speech production and perception on recognition memory and recall for native and non-native listeners and talkers. Two perception studies in Chapter 2 examined the effect of clear speech on within-modal (i.e., audio-audio) or cross-modal (i.e., audio-text) sentence recognition memory for native and non-native listeners. A perception study in Chapter 3 tested the effect of clear speech on recall, a more complex memory task, for native and non-native listeners. Finally, two production studies in Chapter 4 investigated the effect of producing clear speech on recognition memory and recall for native and non-native *talkers*. Key findings from this dissertation were that clear speech improved within- and cross-modal recognition memory and recall for native and non-native *listeners* but impaired recognition memory and recall for native and non-native *talkers*.

These seemingly disparate findings in perception and production are discussed in the light of the models that appeal to ‘effort’ and cognitive load as detrimental to memory. This dissertation provides novel theoretical insights into how lower-level acoustic-phonetic enhancements interact with higher-level memory processes in first and second-language speech perception and production. The results from this dissertation have practical implications in a variety of environments where retention of spoken information is essential, such as classrooms and hospitals.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Introduction

## 1.1 Background

### 1.1.1. Clear speech

On an articulatory continuum that reflects the trade-off between minimizing articulatory effort (hypo-) and maximizing acoustic distinctiveness (hyper-) (H&H model, Lindblom, 1990), clear speech is the hyper-articulated speaking style that talkers spontaneously adopt to make themselves better understood when listeners are experiencing perceptual difficulties (e.g., hearing loss, non-native speaker of the language or noisy environment). Clear speech is one of many speaking style adaptations in which talkers adjust their output in response to communication challenges. As such, it shares characteristics with other listener- and environment-oriented speaking style adaptations including noise-adapted speech (NAS also referred to as Lombard speech; Lombard, 1911) infant-directed speech, foreigner-directed speech, and speech produced in response to vocoded speech (Cooke & Lu, 2010; Cristia, 2013; E. K. Johnson, Lahey, Ernestus, & Cutler, 2013; Uther, Knoll, & Burnham, 2007; Van Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988). Simply instructing talkers to "speak clearly" with no further instructions as to how to modify their speech leads to significant acoustic-articulatory modifications and perceptual benefit relative to the habitual conversational style (Lam & Tjaden, 2013; Lam, Tjaden, & Wilding, 2012; Smiljanic & Bradlow, 2009; and work in progress by Keerstock, Smiljanic and Chandrasekaran). The acoustic-phonetic characteristics of clear speech typically include slower speaking rate, greater dynamic pitch range and amplitude, expansion of the vowel space, and enhancement of language-specific vowel and consonant contrasts (Cooke et al., 2013; Ferguson, 2012; Pichora-Fuller, Goy, & Van Lieshout, 2010;

Smiljanic & Bradlow, 2009). These modifications improve speech perception in noise (syllables, words or sentences) for a variety of listener groups and degraded listening conditions: children with or without learning disabilities (Bradlow, Kraus, & Hayes, 2003), young adult and older adult listeners (Schum, 1996; Smiljanic & Gilbert, 2017) with normal and impaired hearing and cochlear implant users (Ferguson, 2012; Krause & Braida, 2002; Payton, Uchanski, & Braida, 1994; Smiljanic & Sladen, 2013). Although the magnitude of the clear speech intelligibility advantage varies across talkers, listener groups, and conditions of presentation, the clear speech benefit was found to be a robust one, increasing keyword recognition accuracy from 12 to 34 percentage points (see review by Smiljanic & Bradlow, 2009; and Smiljanic, to appear). Paradoxically, even though clear speech involves imagining a non-native speaker interlocutor, studies have shown that clear speech improves speech perception in noise for non-native speakers to a smaller extent than for native speakers (Bradlow & Alexander, 2007; Bradlow & Bent, 2002). The smaller 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, Rochette, & Charest, 2002; Smiljanic & Bradlow, 2005, 2008b; Uchanski, 1988). In fact, Smiljanic & Bradlow (2011) found that the intelligibility benefit for highly proficient non-native listeners could be similar to that of native listeners providing a more favorable signal-to-noise ratio (SNR).

### 1.1.2. Signal clarity and memory

While research has examined extensively the effect of clear speech on peripheral auditory speech processing, fewer studies have examined the effect of clarity in the speech signal on higher-level cognitive processes such as memory. Yet, memory is a crucial component of successful verbal communication. During successful communication, listeners must map varying acoustic input onto stored phonological and lexical representations and retain those representations in memory so they can access them during retrieval. Throughout the dissertation, I refer to the process of perceiving and mapping acoustic features of the speech signal onto stored phonological and lexical representations as “encoding”. During encoding, auditory information is held in working memory for a short period of time allowing for further processing of that information. The “phonological loop” is part of working memory involved with spoken and written material. It comprises both a memory store, which holds speech information, and a rehearsal process, which serves to maintain decaying speech representations in the store (Baddeley, 2000; Baddeley & Hitch, 1974). As noted recently by the authors (Baddeley & Hitch, 2019), the term “phonological” is intended to be relatively atheoretical since the nature of the storage code (acoustic vs. articulatory) is still not completely understood.

The contribution of signal clarity (e.g. clear speech) to the process of encoding and retention of speech in memory is not well understood. In the first study to address this gap, Van Engen et al. (2012) examined recognition memory (i.e., recognizing previously heard speech as old) for speech of varying intelligibility (conversational and clear speech) for young adult native listeners of English with no history of hearing loss. In addition to considering phonetic enhancements, they looked at the effect of semantic context (semantically-meaningful and semantically-anomalous sentences) on recognition memory (RM). They found that RM was enhanced for meaningful sentences compared to anomalous sentences and for sentences produced in clear speech compared to sentences produced in a more casual speech. Gilbert et al. (2014) found the same benefit of clear speech and NAS on RM even when listeners were exposed to sentences mixed with noise. This clear speech benefit on memory was found to extend to older adults with normal-to-moderately impaired hearing-listening abilities in recall of medically-relevant spoken information and to reduce the negative impact of the competing noise on learning and memory  (DiDonato & Surprenant, 2015).

The link between perceptual clarity and memory has been found in domains other than speech. In the visual domain, decreased visual acuity lead to impaired memory for older adults in high sensory demand visuospatial tasks (Glass, 2007). Perceptual clarity, however, does not always correlate with improved memory. In fact, the “perceptual-interference effect” shows that partially masked information or information in hard-to-read fonts was better remembered than easier to read information (Besken & Mulligan, 2013; Diemand-Yauman, Oppenheimer, & Vaughan, 2010). As argued in Yue, Castel, & Bjork (2013), this discrepancy in findings may be due to the cognitive effort expended during memory encoding that varies as a function of processing time, task difficulty and type of disfluency. The authors found that visually intact words presented for 0.5 s and 2 s were better recalled than blurred words but recall was unaffected by blurring manipulation when sufficient processing time was provided, i.e., when the words were presented for 5 s (Yue et al., 2013). While some types of disfluency may create desirable difficulty, such that presenting textual information in an unusual or distinctive font leads to better memory (Diemand-Yauman et al., 2010), visual distortions can create too high a demand on the cognitive processes necessary to encode words in memory and therefore lead to reduced performance.

In speech perception, the idea that perceptual ‘ease’ can improve encoding of speech in memory and promote memory retention is 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, Rudner, Foo, & Lunner, 2008). These theories posit that effortful speech processing recruits more cognitive resources leaving fewer resources available for speech encoding in memory. Since clear speech is easier to understand and alleviates demands on processing resources, it is hypothesized that more cognitive resources remain available for storing information in memory (Van Engen et al., 2012; Gilbert et al., 2014; and DiDonato & Surprenant, 2015). Conversely, listening to acoustically challenging speech such as speech masked with noise, foreign-accented speech, interrupted speech or temporally altered speech (fast 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; Peng & Wang, 2019; Pichora-Fuller et al., 2016; Shafiro, Sheft, & Risley, 2016; Van Engen & Peelle, 2014). Clear speech presumably facilitates perceptual fluency based on the robust clear speech intelligibility benefit (Smiljanic & Bradlow, 2009; Smiljanic, to appear). In contrast, conversational speech can be challenging to process even in the absence of signal degradation such as noise due to extreme reduction and even deletion of many speech segments or whole syllables (Johnson, 2004; Pluymaekers, Ernestus, & Baayen, 2005b, 2005a; Warner, Fountain, & Tucker, 2009). Processing reduced forms, which deviate from expected targets and lexical representations, may incur additional costs in terms of cognitive resources and therefore lead to reduced memory retention for conversational speech compared to clear speech.

### 1.1.3. Non-native speech processing and memory

Speech processing is more difficult and effortful for non-native speakers compared to native speakers. This difficulty is reflected at all levels of processing, from perceptual discrimination of sound contrasts to phonotactics and prosody  (Best & Tyler, 2007; Cutler, Garcia Lecumberri, & Cooke, 2008; Flege, 1995; Francis, Kaganovich, & Driscoll-Huber, 2008; Iverson et al., 2003; Kondaurova & Francis, 2008). Recent work using physiological measures has provided corroborating evidence of greater listening effort reflected, for example, in greater mean and peak pupil dilation when listening in a second language (L2) compared to in a first language (L1) (Borghini & Hazan, 2018; Francis, Tigchelaar, Zhang, & Zekveld, 2018).

When it comes to memory retention for their non-native language, non-native listeners seem to be at a disadvantage compared to native listeners. They tend to recall fewer words presented in noise than native listeners (Hygge, Kjellberg, & Nöstl, 2015; Molesworth, Burgess, Gunnell, Löffler, & Venjakob, 2014). Reducing signal degradation through noise-cancelling headphones was shown to improve recall of audio information played through external speaker at a level consistent with operational environment in aircrafts (i.e., 70dB) for non-native listeners  (Molesworth et al., 2014). As for the quality and precision of the memories formed, findings with L1 and L2 speakers are mixed. Schweppe, Barth, Ketzer-Nöltge, & Rummer (2015) found that sentence recall was significantly worse for L2 listeners than L1 listeners. They argued that verbatim recall in L2 may overload the attentional system, which is in line with the “effortfulness hypothesis” and "ease of language understanding" models. However, in Sampaio & Konopka (2013), L2 listeners outperformed L1 listeners in memory for verbatim sentences during a recall task. Based on the revised hierarchical model (Kroll & Stewart, 1994), they argued that engaging the L2-L1 lexical access route leads non-native participants to devote more resources to individual L2 lexical items and that this benefits verbatim memory (i.e., retention of L2 surface form).

## 1.2. RESEARCH QUESTIONS

### 1.2.1. Within- and cross-modal recognition memory

The goal of the first two studies (**Chapter 2**, Experiment 1 & 2) was to examine whether enhancing the clarity of the speech signal through conversational-to-clear speech modifications improves sentence RM for non-native as well as native listeners even when the stimuli in the test phase are presented in orthographic instead of auditory form (cross-modal RM). This investigation aimed to enhance our understanding of whether clear speech facilitates encoding (e.g., by alleviating cognitive demands), or retrieval (e.g., increased confidence when matching new items against old) of spoken information.

Van Engen et al. (2012) and Gilbert et al. (2014) found that the significant d' scores difference between clear and conversational speech was due to a lower false alarm rate in clear speech; that is, distractor sentences in clear speech were significantly better identified as new than distractor sentences produced in conversational style. These results are in line with RM studies on face recognition which also noted that d' scores were determined by differences in false alarm rates, while hit rates did not contribute as significantly (Lamont, Stewart-Williams, & Podd, 2005). In line with face recognition studies, Gilbert et al. (2014) and Van Engen et al. (2012) proposed that the greater number of salient acoustic cues in clear speech enabled listeners to compare distractor sentences in clear to the encoded sentences and reject them with more confidence. On the other hand, when hearing distractor sentences in a conversational style, there are fewer salient acoustic cues available to match against content stored in memory, which increases the false alarm rate. This suggests that the clear speech benefit on memory arises during the retrieval process (matching new items against old). Chapter 2 tests this hypothesis. If the clear speech benefit on memory arises during the retrieval process, presenting the material in the test phase orthographically (where the test words are written on the screen rather than heard) should deprive listeners from the exact acoustic match and undo the clear speech benefit. However, if the clear speech benefit arises during the encoding phase, by alleviating cognitive resources for instance, the benefit should persist. Furthermore, since utilizing information in one modality to recognize later events in another modality (cross-modal RM) may be more challenging and cognitively more demanding compared to within-modal RM (Björkman, 1967; Greene, Easton, & LaShell, 2001), it is compelling to test whether clear speech can alleviate some of the cross-modal processing difficulty and enhance sentence RM.

### 1.2.2. Beyond recognition memory: recall

Expanding on the RM results in Chapter 2, the robustness of the clear speech representations in native and non-native listeners' memory was examined by looking at recall, an ubiquitous and complex type of memory process (**Chapter 3**). In contrast with sentence RM, where listeners have to give a binary response (old/new), sentence recall is an open-ended task during which listeners search and retrieve from their memory, words, chunks of sentences, and up to entire sentences. Therefore, recall involves processing and encoding at phonological, lexical-semantic, morphosyntactic, and syntactic levels. Recall is typically more difficult and more prone to failure compared to RM (Gillund & Shiffrin, 1984; Ratcliff, 1978). The greater difficulty for recall compared to RM is especially exacerbated by cognitive difficulty or for individuals with depleted cognitive resources, as is the case for example for older adults (Craik & McDowd, 1987; Danckert & Craik, 2013; Erber, 1974; Shafiro & Sheft, 2017; White & Cunningham, 1982; Whiting & Smith, 1997) or individuals with depression (Brand, Jolles, & Gispen-de Wied, 1992).

To the extent that easier-to-understand speaking style can alleviate some of the processing difficulties and that is encoded more robustly (beyond the acoustic-phonetic features), it was expected that clear speech will enhance recall for both native and non-native listeners (Gilbert, Chandrasekaran, & Smiljanic, 2014; Van Engen et al., 2012). The sentence recall analysis in Chapter 3 tested the depth and content of what is remembered in clear and conversational speech and whether clear speech helps listeners recall entire units of connected meaning (i.e., recall of full sentences), or whether the boost of memory retention is idiosyncratic and limited to certain words only. The sentence recall analysis sought to contribute to the ongoing debate in the literature as to whether non-native listeners are more prone to recalling sentences verbatim (Sampaio & Konopka, 2013) or whether non-native listeners experience difficulties in making use of top-down knowledge and reconstructing the gist of information (Bradlow & Bent, 2002; Schweppe et al., 2015).

### 1.2.3. Recognition memory and recall in the production of clear speech

The final two studies (**Chapter 4**, Experiment 1 & 2) aimed to investigate the effect of clearly produced speech on talkers' RM (Experiment 1) and recall (Experiment 2). It further examined the role of effort in production and perception of clear speech and its effect on  memory retention. While clear speech facilitates speech perception for listeners presumably by alleviating some processing effort, the production of clear hyper-articulated speech may be more effortful for talkers. When communication conditions are optimal, talkers tend to revert to hypo-articulated speech in an attempt to minimize the physical "cost" of making articulatory movements ('economy of effort', Guenther, 1995). Clear speech involves greater articulatory effort (peak speed, longer movement durations, greater distances) than casual speech (Perkell, Zandipour, Matthies, & Lane, 2002). To the extent that effort during encoding is costly in terms of resources and thus detrimental to memory ("effortfulness hypothesis"), the speaking style that requires more effort to produce (i.e., clear speech) should lead to decreased memory performance.

A competing hypothesis from the "production effect" literature suggests that the more 'exaggerated' productions might, in fact, improve memory. The "production effect" (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010) is the superior retention of material read aloud relative to material read silently during an encoding phase. Words produced out loud loudly and singing were better remembered than words produced out loud normally or silently (Quinlan & Taylor, 2013). The authors argued that items read aloud have additional information (articulatory and acoustic) relative to items not read aloud which is then used at test for discrimination ("distinctiveness account"). Since clear speech provides additional articulatory and acoustic information relative to conversational speech, it may in fact improve memory retention. The two studies in Chapter 4 tested whether the effort in producing clear speech improves (cf., distinctiveness account, production effect) or interferes with memory retention (cf., effortfulness hypothesis).

### 1.3. Significance

The general aim of this dissertation is to provide new insights into the link between perceptual clarity in the form of an intelligibility-enhancing speaking style and integration of information in memory and memory retention for native and non-native speakers. This dissertation offers novel contributions in L1 and L2 speech perception and production as well as auditory memory by testing the generalizability of the clear speech effect on memory to a variety of modalities (within- and cross-modal RM; perception and production), memory tasks (RM, recall) and populations (L1, L2 talkers and listeners). This dissertation is an important step towards better understanding of intelligibility variation and processing beyond word recognition. It stands to shed light on how speech acoustic clarity interacts with integration of information in memory, the link between lower-level acoustic enhancements and holistic speech processing and retention, the link between production and perception, and L2 speech processing and retention.

The results of this dissertation have practical applications to the educational field (e.g., how to optimize learning strategies in mixed native-non-native classrooms) at a time where the percentage of U.S. public school students who are second-language English learners approaches 9.5 percent, or 4.8 million students in 2015 (McFarland et al., 2018). It also has practical applications in clinical fields (e.g., how to optimize health care provider-patient interactions). Comparing memory-enhancing strategies for native and non-native speakers is critical given the growing linguistic diversity in the U.S.

### 1.4. SUMMARY

In this dissertation, I conducted five experimental studies (Table 1). In **Chapter 2**, I examined the effect of enhancing the clarity of the speech signal through conversational-to-clear speech modifications on within-modal (i.e., audio-audio) or cross-modal (i.e., audio-text) sentence RM for native and non-native *listeners*. In **Chapter 3**, I tested whether the clear speech benefit on memory extends from RM to recall, a more complex memory process. Finally in **Chapter 4**, I investigated the effect of clearly produced speech on talkers' (instead of listeners') RM and recall. In **Chapter 5**, I conclude by summarizing key findings across my experiments and discuss the relevant implications of my work for our understanding of how speech acoustic clarity interacts with integration of information in memory.

Table 1. Overview of dissertation chapters and experiments

| | Clear speech Perception | Clear speech Production |
|---|---|---|
| Recognition memory | **Chapter 2**<br>Experiment 1 (*n*=60) &<br>Experiment 2 (*n*=60) | **Chapter 4**<br>Experiment 1 (*n*=90) &<br>Experiment 2 (*n*=75) |
| Recall | **Chapter 3**<br>Experiment 1 (*n*=88) | |

# Chapter 2: Within- and cross-modal recognition memory for speech of varying intelligibility in native and non-native listeners[1]

## 2.1. Abstract

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 1). Sixty different individuals (30 native and 30 non-native English) participated in a cross-modal (i.e., audio-textual) sentence recognition memory task (Experiment 2). 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 1) or absent (Experiment 2). 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.

[1] This work was previously published: Keerstock, S., & Smiljanic, R. (2018). Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners. *The Journal of the Acoustical Society of America*, 144(5), 2871–2881. https://doi.org/10.1121/1.5078589. Dissertator contributed to the conception, design, data collection, analysis, interpretation, and writing of the study.

## 2.2. INTRODUCTION

Understanding speech is implicit and automatic in favorable listening conditions (Rönnberg, 2003; Rönnberg et al., 2013, 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; Ljung, Israelsson, & Hygge, 2013; Pichora-Fuller, Schneider, & Daneman, 1995; Pichora-Fuller & Souza, 2003; Souza, Arehart, Shen, Anderson, & Kates, 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, Rogers, Van Engen, & Peelle, 2018; Rabbitt, 1968, 1990; Rönnberg et al., 2013; Tun, McCoy, & Wingfield, 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 & Kewley-Port, 2002;

Pichora-Fuller et al., 2010; Smiljanic & Bradlow, 2009). These modifications improve intelligibility for a variety of listener groups and degraded listening conditions (Bradlow & Bent, 2002; Ferguson, 2012; Krause & Braida, 2002; Payton et al., 1994; Picheny, M.A., Durlach, N.I., 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, 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 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 (K. Johnson, 2004; Pluymaekers et al., 2005a, 2005b; 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 & Tyler, 2007; Cutler et al., 2008; Flege, 1995; Francis et al., 2008; Iverson et al., 2003; Kondaurova & Francis, 2008). Bradlow & Bent (2002) and Bradlow & Alexander (2007) found that non-native listeners benefited from clear speech although the intelligibility benefit was smaller compared to native listeners. Smiljanic & 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; Smiljanic & Bradlow, 2005, 2008a; 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, Burgess, Gunnell, Löffler, & Venjakob (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 1). 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 (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 2). 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, Easton, & LaShell (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 (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 & Cobb, 1978). The cost of cross-modal integration of information could result in reduced memory retention. Experiment 2 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.[2]

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, Rudner, Johnsrude, & Rönnberg, 2013). In the present study, all participants completed a forward

[2] It 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 & Hitch, 2019). 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.

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, Gathercole, & Papagno, 1998) and of auditory short-term memory (Engle, Tuholski, Laughlin, & Conway, 1999; Hale, Hoeppner, & Fiorello, 2002; Rosenthal, Riccio, Gsanger, & Jarratt, 2006), 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, 2007; Pichora-Fuller & Singh, 2006; Rudner, Foo, Rönnberg, & Lunner, 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 (Olsthoorn, Andringa, & Hulstijn, 2012).

### 2.3. Experiment overview

The current paper presents the results from two experiments. Experiment 1 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 2 tested cross-modal (audio-textual) sentence recognition memory for 30 native monolingual English listeners and 30 non-native English listeners (different individuals from Experiment 1). 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 2 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 the same day 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 minutes) followed by the sentence recognition memory task (lasting approximately 20 minutes).

## 2.4. EXPERIMENT 1: WITHIN-MODAL SENTENCE RECOGNITION MEMORY

### 2.4.1. Participants

Thirty native English listeners between the ages of 18 and 23 (mean: 19 years old; 21 F) and 30 non-native listeners between the ages of 18 and 31 (mean: 23 years old; 24 F) 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 2). 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 twenties 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, Blumenfeld, & Kaushanskaya, 2007) and passed a hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

Table 2. Language background information for non-native listeners.

| Factor | Experiment 1 (*n*=30) | Experiment 2 (*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)<br>English: 4.7 (mean); 3-5 (range) | L1: 4.6 (mean); 3-5 (range)<br>English: 4.6 (mean); 3-5 (range) |
| Contexts for daily exposure to English | Professional setting only: *n*=21<br>Extended and/or immediate relatives: *n*=9 | Professional setting only: *n*=25<br>Extended and/or immediate relatives: *n*=5 |
| Self-estimated proficiency[2] | L1: 4.9 (mean); 0.22 (sd)<br>English: 4.1 (mean); 0.62 (sd) | L1: 4.7 (mean); 0.49 (sd)<br>English: 4.1 (mean); 0.56 (sd) |
| 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). |

[1] (For each language, self-estimated amount of daily exposure on a scale from 1 (no current exposure) to 5(constant exposure) [2] (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))

### 2.4.2. 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 & Smiljanic 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.  44% of the lexical items in the new sentences appeared in the set of old sentences (overlapping items) while 56% of the 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 & Smiljanic, 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 Smiljanic & Bradlow, 2005). Individual sentences were segmented from the long recording and equalized for RMS amplitude using the software Praat (Boersma & 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 (Cooke et al., 2013; Pichora-Fuller et al., 2010; Smiljanic & Bradlow, 2009), such as significantly longer durations, higher mean F0s, larger F0 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 F; 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 signal-to-noise ratio (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, Florentine, & Buus, 1997; Meador, Flege, & Mackay, 2000; Rogers, Lister, Febo, Besing, & Abrams, 2006). The combined results of the two word-recognition-in-noise tasks showed a clear speech benefit for both native and non-native listeners.

### 2.4.3. Procedure

First, participants completed the forward digit span task (as designed by MacWhinney, E-Prime scripts). Participants were seated in a sound-attenuated booth facing a computer monitor. Instructions and stimuli were presented with E-Prime 2.0 Psychology Software Tool 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 (one through nine) 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 3 pictures of animals (a puppy, a bird, and a monkey). The test phase involved randomly presenting 2 old and 2 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.

### 2.4.4. 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 & 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 & Corwin, 1988). C scores, also calculated following Snodgrass & 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 & 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, Kronenberger, Roman, & Geers, 2011). The individual digit span scores were included as co-variates 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, Mächler, Bolker, & Walker, 2015).

### 2.4.5. Results

The mean and range of the digit span test for native and non-native listeners are shown in Table 3. 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 Figure 1 and Table 4. 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 (Figure **1a**). There was a main effect of Speaking Style (*p*<.001) on d' scores, but no effect of Listener Group (*p*=.69), no effect of Digit Span (*p*=.13), and no significant interaction between Speaking Style and Listener Group (*p*=.73). Thus, clear speaking style improved recognition memory and the clear speech benefit was similar for native and non-native listeners.[3]

The hit rates were higher for clear than conversational sentences for non-native listeners, while this was not the case for native listeners (Figure **1b**). The linear mixed-effects regression showed that there was a significant interaction between Speaking Style and Listener Group (*p*<.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*<.05), but not significant for native listeners (*p*=.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 (Figure **1c**). There was a significant interaction between Speaking Style and Listener Group (p<.01). Post-hoc analyses revealed that the effect of speaking style on false alarm rate was significant for native listeners (*p*<.001), but not for

[3] Even 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 (Olsthoorn et al., 2012). The results were similar: there was a main effect of Speaking Style (p<.01) on d' scores, but no effect of Listener Group (p=.6), and no interaction (p=.75)).

non-native listeners (*p*=.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 for differences in accuracy. We found a main effect of Speaking style (p<.001) and a main effect of Listener Group (p<.05), such that response times were significantly faster for clear than for conversational sentences and faster for native listeners than for non-native listeners (Figure **1d**). 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. Moreover, 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.

Table 3. Digit span scores.

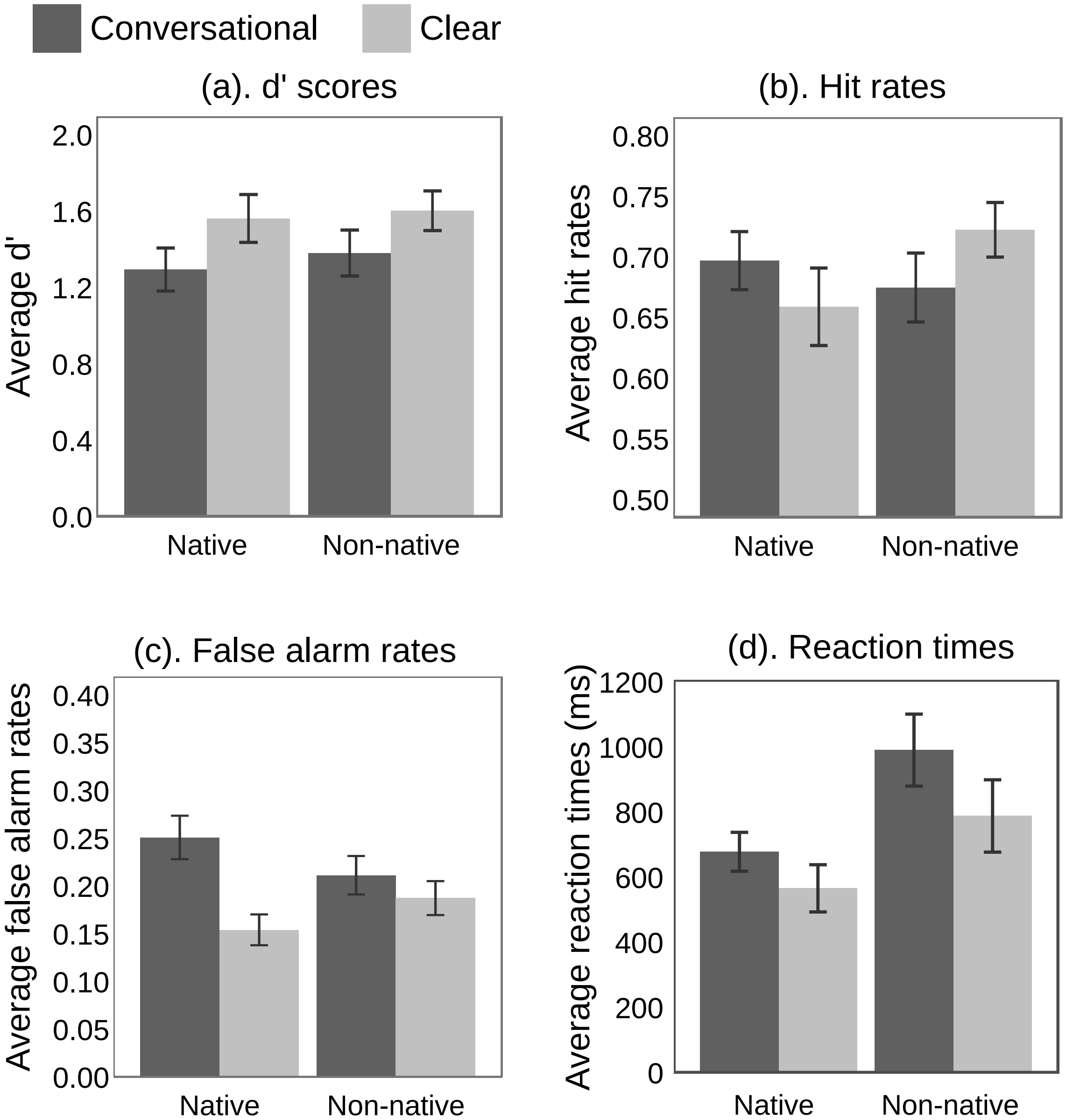
| **Experiment 1 (*N*=60 listeners)** | | | | |
|---|---|---|---|---|
| | mean | sd | min | max |
| native (*n*=30) | 8.00 | 1.29 | 6 | 10 |
| non-native (*n*=30) | 8.03 | 1.43 | 5 | 10 |

| **Experiment 2 (*N*=60 listeners)** | | | | |
|---|---|---|---|---|
| | mean | sd | min | max |
| native (*n*=30) | 8.17 | 1.23 | 6 | 10 |
| non-native (*n*=30) | 7.97 | 1.25 | 6 | 10 |

Table 4. 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 1 (*N*=60) | | Experiment 2 (*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) |

Figure 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 1 (within-modal). Error bars represent standard error.



## 2.5. EXPERIMENT 2: CROSS-MODAL SENTENCE RECOGNITION MEMORY

### 2.5.1. Participants

Thirty native English listeners between the age of 18 and 32 (mean: 20 years old; 17 F), and 30 non-native listeners between the age of 18 and 31 (mean: 22 years old; 18 F) participated in Experiment 2. They were different individuals from Experiment 1, but recruited from the same pool/community. As in Experiment 1, 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 2.

### 2.5.2. Stimuli

The stimuli were the same as in Experiment 1 and were also presented in quiet (i.e., without added noise).

### 2.5.3. Procedure

The procedure was identical to the one in Experiment 1. 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). 80 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 1). 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 1).

### 2.5.4. Analyses

Similar analyses of the digit span and sentence recognition memory tasks as in Experiment 1 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).

Linear mixed-effects regressions (LMER) 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).

### 2.5.5. Results

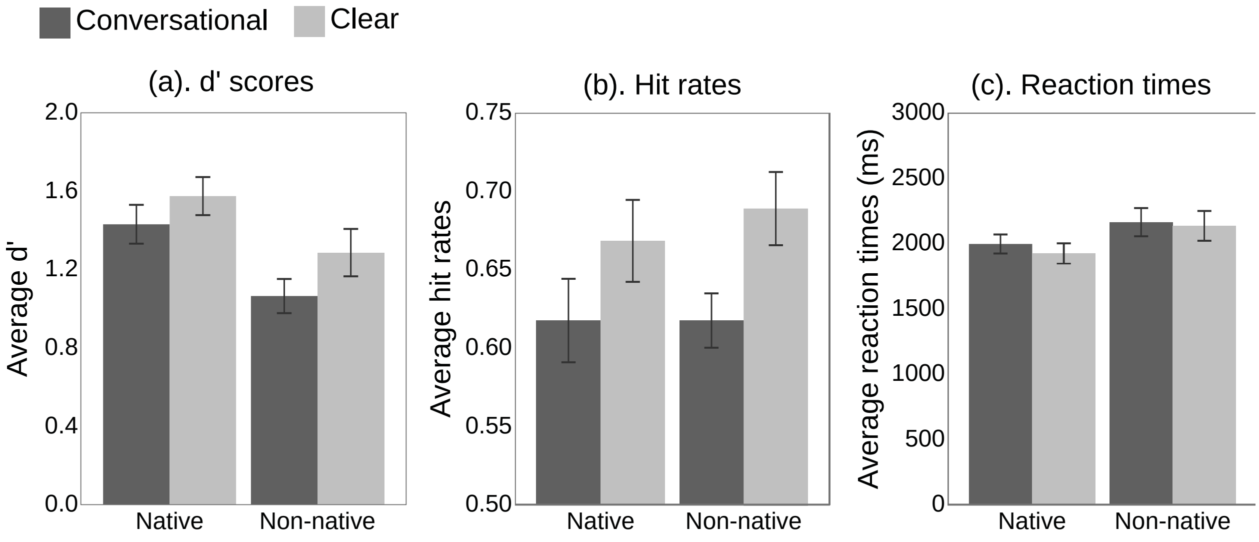
The mean and range of the digit span test for native and non-native listeners are shown in Table 3. As in Experiment 1, the distribution of the digit span scores revealed no differences between native and non-native listeners. The overall sentence recognition memory results are presented in Figure 2 and Table 4. 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 (Figure **2a**). There was a main effect of Listener Group (*p*<.05) on d’ scores and a main effect of Speaking Style (*p*<.01), but no effect of Digit Span (*p*=.93). No significant interaction between Speaking Style and Listener Group was found (*p*=.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.[4]

[4] As in Experiment 1, similar results were found when removing Digit Span from the model (main effect of Listener Group, p<.05; and main effect of Speaking Style, p<.01; no interaction, p=.48).

We also found that both native and non-native listeners had higher hit rate for sentences produced in clear than in conversational speech (Figure **2b**). The linear mixed-effects regression showed that there was a main effect of Speaking Style (*p*<.01), but no effect of Listener Group (*p*=.62), no effect of Digit Span (*p*=.15), and no interaction between Speaking Style and Listener Group (*p*=.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 (Figure **2c**). Linear mixed-effects regression analysis of RTs (including d' scores as a covariate) found that there was no effect of Speaking style (*p*=.26), no effect of Listener Group (*p*=.29), no effect of Digit Span (*p*=.11), no effect of d' scores (*p*=.9), and no interaction between Listener Group and Speaking style (*p*=.61).

Figure 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 2 (cross-modal). Error bars represent standard error.



## 2.6. Additional results across modalities

The above analyses assessed the effect of speaking style and language on sentence recognition memory within each modality. In order to compare the effect of modality (within- vs. across-) on sentence recognition memory, we conducted a linear mixed-effects regression 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*<.001) and a main effect of Modality (*p*<.05), but no effect of Digit Span (*p*=.47) and no interaction between Modality and Speaking style (*p*=.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 linear mixed-effects regression 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*<.01) and that d’ scores did not change as a function of stimulus presentation modality during the testing phase (no effect of Modality; *p*=.63; and no interaction with speaking style; *p*=.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.[5]

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 2). 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

[5] Due 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.

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, Smiljanic, & Calandruccio (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.

### 2.7. 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 1) and cross-modal (audio-textual; Experiment 2) 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. 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 (Lecumberri, Cooke, & Cutler, 2010; Mayo et al., 1997; Rogers et al., 2006). The lower non-native word recognition scores likely also have origin in the less efficient use of L2-specific clear speech enhancements (Bradlow & Alexander, 2007; Bradlow & 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 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 2) 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 & 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.[6]

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 & 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

[6] The longer RTs in Experiment 2 than in Experiment 1 could indicate cognitive effort, but could also reflect the time it takes to read printed information as opposed to process auditory information.

& 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 2). 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 2 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 & Merikle, 1996; Rönnberg et al., 2013; Unsworth & 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, this 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.

# Chapter 3: Recall of speech of varying intelligibility in native and non-native listeners[7]

## 3.1. Abstract

The present study examined the effect of intelligibility-enhancing clear speech on listeners' recall. Native (*n*=57) and non-native (*n*=31) English listeners heard meaningful sentences produced in clear and conversational speech, and then completed a cued-recall task. Results showed that listeners recalled more words from clearly produced sentences. Sentence-level analysis revealed that listening to clear speech increased the odds of recalling whole sentences and decreased the odds of erroneous and omitted responses. This study showed that the clear speech benefit extends beyond word- and sentence-level recognition memory to include deeper linguistic encoding at the level of syntactic and semantic information.

## 3.2. Introduction

Successful verbal communication involves mapping of variable acoustic input onto stored phonological and lexical representations and maintaining those representations in memory in order to extract sentence-level meaning. Processing degraded, masked or phonetically ambiguous acoustic signals requires additional cognitive resources leaving fewer resources available for encoding speech in memory (Rönnberg et al., 2013). The present study examined whether acoustic-phonetic enhancements in the form of listener-

[7] This work was previously published: Keerstock, S., & Smiljanic, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America, 146*(6), 4604–4610. https://doi.org/10.1121/1.5141372. Dissertator contributed to the conception, design, data collection, analysis, interpretation, and writing of the study.

oriented hyper-articulated clear speech facilitate recall of spoken information for native listeners and listeners who face additional difficulties associated with speech processing in second language (L2).

Adverse listening contexts (e.g., degraded signal quality, background noise, perceiving speech in L2) can raise speech processing demands, leaving fewer available cognitive resources for comprehension and recall of the message (cf. “effortfulness hypothesis” McCoy et al., 2005; Rabbitt, 1968, 1990; and cf. “ease of language understanding” model, Rönnberg et al., 2013). Foreign-accented speech, for instance, was shown to increase cognitive demands during speech perception (Van Engen & Peelle, 2014) and to be recalled less accurately compared to native-accented speech (K. Y. Chan, Chiu, Dailey, & Jalil, 2019). In this study, we examined the effect of intelligibility-varying speaking styles on subsequent recall. Unlike clearly spoken speech, casual, conversational speech produced by native speakers can be challenging to process due to pervasive reductions and deletions of speech segments or whole syllables such that it deviates from expected phonological and lexical representations (K. Johnson, 2004; Mattys, Davis, Bradlow, & Scott, 2012; Warner et al., 2009; Warner & Tucker, 2011).

The effect of listener-oriented clear speech on word recognition in noise is well documented (cf. reviews by Smiljanic & Bradlow, 2009 and Uchanski, 2005) but less attention has been given to how speech clarity affects memory for spoken language. Van Engen et al., (2012) showed enhanced sentence recognition memory for meaningful and clear sentences compared to anomalous and conversational sentences. Gilbert et al., (2014) extended these results to sentences presented in noise and to noise-adapted-speech, another intelligibility-enhancing speaking style. More recently, Keerstock & Smiljanic (2018) **(Chapter 2)** showed that the clear speech benefit on sentence recognition memory extended to non-native English listeners, and was evident when tested within (audio-audio) and across (audio-text) modalities, suggesting that acoustic-phonetic enhancements promote deeper linguistic encoding at a level abstracted from the input speech.

The current study tested the hypothesis that, by providing optimal and unambiguous speech signals (hyper-speech within Lindblom's 1990 H&H theory), clear speech may reduce cognitive effort during speech perception and thus improve memory for spoken language compared to conversational speech. We tested this hypothesis by examining memory for spoken language using a cued-recall task. To date, the effect of clear speech on memory has only been assessed via recognition memory, a familiarity decision task ('is this item familiar?') with a binary response (yes/no). In contrast, recall is a more complex task that requires that listeners process the incoming speech signals beyond the surface acoustic level at multiple levels of linguistic structure (phonological, lexical-semantic, morphosyntactic, and syntactic) in order to successfully search and retrieve lexical items and entire units of connected meaning from memory (Gillund & Shiffrin, 1984; Ratcliff, 1978). Limited cognitive resources (e.g., as the result of aging, depressive symptoms, or perceiving speech in noise) impair recall more than recognition (Brand et al., 1992; Ng, Rudner, Lunner, Pedersen, & Rönnberg, 2013; Rhodes, Greene, & Naveh-benjamin, 2019). To the extent that processing conversational speech demands more cognitive resources, we predicted that its recall might be further impaired in a recall task relative to clear speech.

Speech perception is additionally affected by fluency in the target language (Best & Tyler, 2007; Cutler et al., 2008; Flege, 1995; Iverson et al., 2003; Kondaurova & Francis, 2008). Non-native listeners recall fewer words in noise than native listeners, although the use of noise-cancelling headphones was shown to improve non-native listeners' performance (Hygge et al., 2015; Molesworth et al., 2014). The difficulty in remembering L2 speech may also arise from the increased recruitment of cognitive resources during speech perception at the expense of storing the information in memory (Best & Tyler, 2007; Flege, 1995; Iverson et al., 2003). Keerstock & Smiljanic (2018) showed that clear speech enhanced recognition memory for non-native listeners suggesting that some of the processing difficulty due to the lack of extensive familiarity with the target language was alleviated, and that sufficient cognitive resources remained available for memory encoding. Here, we extend that line of inquiry by examining whether the acoustic-phonetic clear speech modifications enhance native and non-native listeners' sentence recall. Similar to the improved sentence recognition memory, we expect that clear speech will enhance recall for non-native listeners. However, as some of the clear speech strategies are native-listener oriented (Smiljanic & Bradlow, 2009), the benefit for nonnative listeners may be smaller compared to the native listeners. The results will provide new insights into the link between the signal-related acoustic-phonetic enhancements and relatively signal-independent cognitive processes. Examining the retention of spoken information by L2 speakers also has practical implications as the number of L2 English students in U.S. public schools reached 9.5 percent, or 4.8 million students, in 2015 (McFarland et al., 2018).

Understanding whether the same memory enhancement strategies can apply to both L1 and L2 individuals can inform the use of these strategies in the classroom.

## 3.3. METHODS

### 3.3.1. Participants

Eighty-eight listeners participated in the study. They were recruited from the University of Texas community and received monetary compensation or class credit for their participation. The non-native English listener group consisted of 31 subjects (22 female; $M_{age}$ = 22.7, $SD_{age}$ = 3.8). They acquired English on average after age 7.6 (range 5-19) and received no exposure to English at home from parents/caregivers. Information about the non-native listeners' language background is provided in Table 5. The native English listeners group included 33 monolingual English listeners (18 female; $M_{age}$ = 19.6, $SD_{age}$ = 1.4) who reported no exposure to another language before age 6, and 24 native non-monolingual English listeners (15 female; $M_{age}$ = 18.8, $SD_{age}$ = 1.2) who were exposed to another language from birth alongside with English but reported being English dominant at the time of testing. The early exposure to another language in addition to English, however, did not have significant effect on recall (see the results below).

All participants signed a written informed consent and filled out a detailed language background questionnaire adapted from the LEAP-Q (Marian et al., 2007). All passed a hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

Table 5. Language background information for non-native listeners (n=31)

| | **Mean** | **SD** | **Range** |
|---|---|---|---|
| Age of first exposure to English (in years) | 7.6 | 2.9 | 5-19 |
| Age of arrival to USA (in years) | 16.1<br>(after 15yo: *n*=20) | 9.1 | 0-37 |
| Time spent in USA (in years) | 6.6 | 7.1 | 0-25 |
| Daily exposure[1] to English | 4.8 | 0.5 | 3-5 |
| Daily exposure to L1 | 4.5 | 0.7 | 3-5 |
| Self-estimated proficiency[2] in English | 4.3 | 0.6 | 3.25-5 |
| Self-estimated proficiency in L1 | 4.5 | 0.6 | 2.75-5 |
| L1 | Mandarin (*n*=8), Spanish (*n*=7), Hindi (*n*=3), Korean (*n*=3), Vietnamese (*n*=2), Cantonese, French, Gujarati, Indonesian, Malayalam, Marathi, Nepali, Serbian (*n*=1). | | |

[1] (For each language, self-estimated amount of daily exposure on a scale from 1 (no current exposure) to 5(constant exposure) [2] (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))

### 3.3.2. Stimuli

The stimuli consisted of 72 semantically-meaningful sentences from the same sentence pool as in Keerstock & Smiljanic (2018) . The sentences contained high-frequency words familiar to non-native listeners (see Calandruccio & Smiljanić, 2012 for details about the development of the materials). All sentences followed the same syntactic structure: they started with a determiner and a noun (e.g., the grandfather), followed by a verb, a determiner, an adjective and a noun (e.g., drank the dark coffee). The cue written on the page was always the first noun phrase (in italics) and the three keywords to be recalled were always the last three content words (underlined) e.g., "*The grandfather* drank the dark coffee" or "*The mother* baked the delicious cookies". The sentences were produced by a 26-year-old female speaker of American English in conversational and clear speaking style. For the conversational speaking style, the speaker was instructed to read sentences in a casual style, as if talking to someone who is familiar with her speech patterns. 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 (see Van Engen et al., 2012 for elicitation and recording details). The acoustic analyses showed significantly longer durations, higher mean F0s, larger F0 ranges, and greater energy in the 1–3 kHz range for clear compared to conversational speech (reported in Van Engen et al., 2012). For sentences used in the current study, we have previously found that word recognition in noise was higher for sentences produced in clear speech compared to conversational speech among native listeners (reported in Van Engen et al., 2012) and non-native listeners (reported in Keerstock & Smiljanic, 2018). For the current study, the sentences were equalized for RMS amplitude and presented to listeners in quiet (i.e., without added noise).

### 3.3.3. Procedure

Participants were seated in a sound-attenuated booth facing a computer monitor. Instructions and stimuli were presented with E-Prime 2.0 Psychology Software. The experimental session started with two practice sentences not used in the main experiment. Participants were asked to listen to the sentences and to try and memorize them. After hearing the sentences, they were instructed to write down what they remembered and to guess when uncertain. No feedback was provided. After the practice, listeners heard the 72 test sentences divided into six blocks of 12 sentences. The speaking style presentation was counterbalanced across listeners such that half of the participants heard all the sentences in Block 1, 3 and 5 produced in conversational speech and all the sentences in Block 2, 4 and 6 produced in clear speech, and half of the participants heard all the sentences in Block 1, 3 and 5 produced in clear speech and all the sentences in Block 2, 4 and 6 produced in conversational speech. This ensured that all sentences were heard in both conversational and clear style across listeners. No listener heard the same sentence twice. Sentences were presented through Sennheiser HD570 headphones while the screen display remained blank. Sentences were separated by 1500 milliseconds of silence.

After listening to each block of 12 sentences, participants wrote down their responses in a recall booklet. The participants were asked to recall and write down the rest of the sentence next to the cue on the recall booklet (e.g., "drank the dark coffee" or "baked the delicious cookies"). The 3 content words to be recalled were counted for keywords recall score. The sentences within each block contained no repetition of written cues or target words. The recall cues were provided in the booklet in the serial order of audio presentation; however, participants were not instructed to fill the recall booklet in any particular order. The recall test was self-paced. Once they were done with one block, participants pressed a button to initiate the audio presentation of the next block of 12 sentences. The whole experimental session lasted approximately 45 minutes.

### 3.3.4. Analysis

The effect of speaking style on recall was assessed in two ways. The percentage of keywords recalled per speaking style (*keyword recall*) provided a quantitative measurement of recall performance and an evaluation of how much of the speech content was recalled verbatim. There were 216 keywords (108 per speaking style) to be recalled. Each recalled keyword was scored as either correct (1) or incorrect (0). We adopted a strict scoring criterion whereby any morpho-phonological mismatch (e.g., “flowers” instead of “flower”) was scored as incorrect. Listeners were not penalized for obvious spelling errors. In the case of uncertainty due to handwriting, the first author consulted the second author and consensus was reached. Binomial logistic regressions were conducted using the generalized linear mixed-effects regressions (GLMER) function of the lme4 package in R (Bates et al., 2015) with keyword recall (0-1) as the dichotomous dependent variable. The model included Speaking Style (Conversational$_{\text{[reference]}}$ vs. Clear) and Listener Group (Native$_{\text{[reference]}}$ vs. Non-native) as the independent variables, Speaking Style X Listener Group as an interaction term, Word Position (1, 2, 3) as a covariate to account for the position of the word in the sentence, Block Position (1 - 6) as a covariate to account for practice effects, and Sentence Position (1 - 12) as a covariate to account for serial position effects within each block of 12 sentences (i.e., primacy and recency). Subject and Stimuli were modeled using a random intercept term.

The second measure addressed whether sentences were recalled as entire units of connected meaning (*sentence recall*). Responses were categorized as belonging to 1 of 5 categories (with no overlapping membership possible): verbatim, paraphrase, partial, error or omit (adapted from Brewer, Sampaio, & Barlow, 2005; Sampaio & Konopka, 2013; and to some extent, Chan et al., 2019). Scoring was done by the first author and the second author was consulted for ambiguous responses. To ensure consistency across multiple responses, a log with recurring paraphrase, partial or error responses was kept and referred to when scoring new sentences. Table 6 shows the scoring schema using the target example: “The grandfather drank the dark coffee.” Scoring beyond the individual target keywords correct allowed us to distinguish among varied responses where the intended message conveyed by the original sentence was recalled. It also allowed us to differentiate the missing responses from the responses where the recall deviated from the intended message (both scored as incorrect in the keyword recall analysis). A multinomial logistic regression (MLR) was conducted for the categorical dependent variable with multiple unordered recall response categories. MLR captures overall modulation of response probabilities while avoiding the statistical issues raised by non-independent tests such as repeated binary logistic regressions. Using the mlogit package in R (Croissant, 2015), we specified Category membership (Verbatim$_{[\text{reference}],}$ Paraphrase, Partial, Error or Omit) as the dependent variable and Speaking Style (Conversational$_{[\text{reference}]}$ vs. Clear) and Listener Group (Native$_{[\text{reference}]}$ vs. Non-native) and their interaction as the independent variables.

Table 6. Scoring schema for sentence recall accuracy for the target sentence: "The grandfather (cue) drank the dark coffee."

| Score | Case | Example of response |
|---|---|---|
| Verbatim | Response included entire sentence with the original wording. | • drank the dark coffee |
| Paraphrase | Response included wording changes that did not alter the gist meaning of the original sentence (e.g., synonym shifts, or additions of implied information). | • drank the black coffee |
| Partial | Response contained some lexical information from the original sentence, but was deficient, lacking or deviated from the original meaning (e.g., loss of non-redundant information, non-synonymous word shifts, and additions of information not implied by the original sentence). | • drank the coffee<br>• drank the cold coffee |
| Error | Response contained no information from the original sentence. | • built the wooden table |
| Omit | No written response. | • |

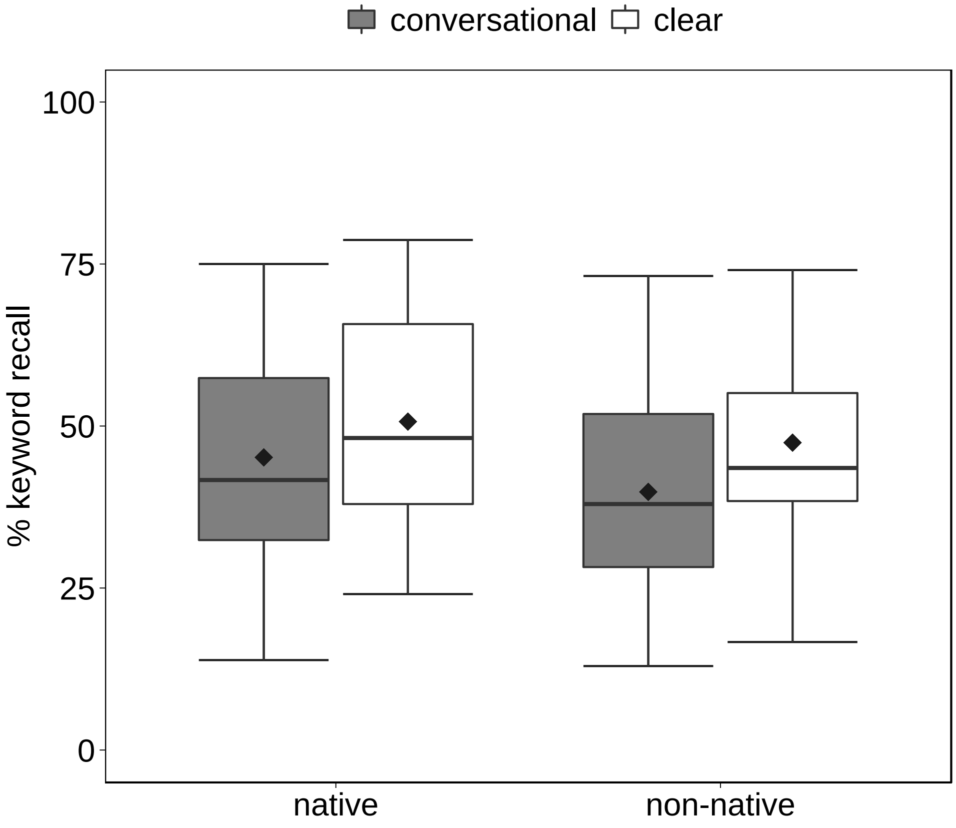
## 3.4. RESULTS

### 3.4.1. Keyword recall

Figure 3 shows the keyword accuracy results for native and non-native listeners in two speaking styles. Results from the logistic regressions on keyword recall showed a significant main effect of Speaking Style (*Odds Ratio* [*OR*] = 1.41 [95% *CI*: 1.32-1.50], p<.001) but no effect of Listener Group (*OR* = 0.8 [95% *CI*: 0.58-1.11], p=.18). The Speaking Style X Listener Group interaction was not significant (*OR* = 1.13 [95% *CI*: 0.99-1.30], p=.07) and therefore was removed from the model before interpreting the main effects. We tested an alternative model in which the Talker Group variable was split into 3 levels ('monolingual' (n=33), 'non-monolingual' (n=24) and non-native (n=31) English speakers) and found no significant effect of Talker Group on recall (p=.26). The parsimonious model with 2 levels (native vs. non-native) was elected as a better model in an ANOVA model comparison (Baayen, Davidson, & Bates, 2008) as the alternative model failed to improve the model fit ($\chi_2 = 0.82$, df = 1, p = 0.37). We concluded that a significant exposure to another language in addition to English did not differentially affect recall of our listeners in this study.

Furthermore, we tested whether recall accuracy could be accounted for by any of the following linguistic variables: age of exposure to English, age of arrival in the US, time spent in the US, self-rated proficiency in English and L1, current daily exposure to English and L1, number of languages reported before the age of 6, number of languages reported at any age, and type of L1. Model fit was only improved when both self-rated proficiency in English and current daily exposure to English were included the model ($\chi_2 = 7.16$, p=0.03). The main finding, however, remained unchanged (main effect of Speaking Style, p<.001; no effect of self-rated English proficiency, p=.18; no effect of current daily exposure to English, p=.23). The results thus showed that all listeners were able to recall more words from sentences produced in clear speech than in conversational speaking style regardless of their language experience.

Figure 3. Percent keyword accurately recalled in conversational (dark grey) and clear speech (white) and for native and non-native English listeners. Boxplots extend from the 25th to the 75th percentile, with whiskers extending to points within 1.5 times the inter-quartile range. The central horizontal lines indicate the medians and the diamonds indicate the means.



### 3.4.2 Sentence recall

Figure 4 shows the mean and standard error for each response category (in %) for the two speaking styles and the two listener groups. A summary of the direction of the effects of Speaking Style and Listener Group in the MLR is provided in Table 7. Results from the MLR showed an effect of Speaking Style on Verbatim, Error and Omit response rates in the direction expected. That is, the odds of Error responses (*OR* = 0.7 [95% *CI*:

0.59-0.82], p=.002) and Omit responses (*OR* = 0.74 [95% *CI*: 1.34-1.74], p<.001) were significantly *decreased* relative to the odds of Verbatim responses in clear speech compared to conversational speech. The odds of Paraphrase (p=.88) and Partial (p=.08) responses relative to the odds of Verbatim responses were not significantly affected by Speaking Style. As for the effect of Listener Group, results showed *increased* odds of Omit responses (*OR* = 1.52 [95% *CI*: 1.34-1.74], p<.001) and Partial responses (*OR* = 1.53 [95% *CI*: 1.31-1.80], p<.001) and *decreased* odds of Error responses (*OR* = 0.74 [95% *CI*: 0.62-0.90], p=.002) compared to Verbatim rates for non-native listeners relative to native listeners. The odds of Paraphrase relative to the odds of Verbatim responses were not significantly different for the two listener groups (p=.056). No significant interactions were found between Speaking Style and Listener Groups.

Figure 4. Proportion of response rate (mean ± SEM) in conversational (dark grey) and clear (white) speech for native and non-native English listeners.

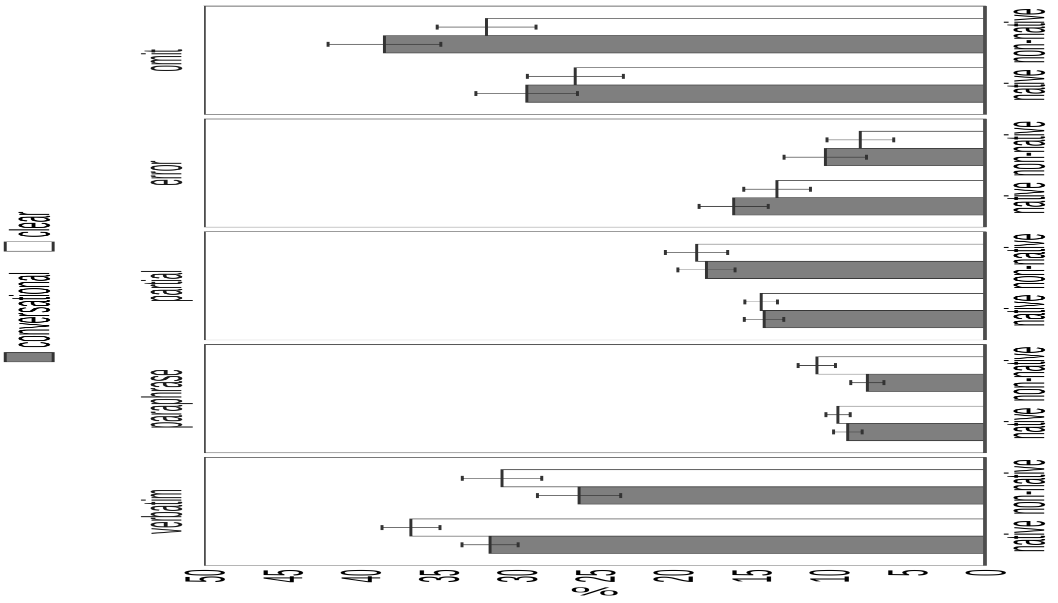


Table 7. Direction of the effects of Speaking Style and Listener Group on sentence recall response category in the MLR.

| | Effect of Clear Speech (relative to CO) | Effect of non-native listeners (relative to native listeners) |
|---|---|---|
| Verbatim | ↑ | ↑ |
| Error | ↓ | ↓ |
| Omit | ↓ | ↑ |
| Paraphrase | *n.s.* | *n.s.* |
| Partial | *n.s.* | ↑ |

*Note.* Arrows represent a significant increase or decrease in the odds of a particular response type (Error, Omit, Paraphrase, Partial) relative to a Verbatim response (set as the reference level in the MLR) in clear speech compared to conversational speech, and for non-native listeners compared to native listeners. In row 1, the odds of making a Verbatim response was evaluated by changing the reference level to Omit. *n.s.* indicate a nonsignificant effect.

### 3.5. Discussion

This study examined native and non-native listeners' recall of sentences produced in conversational and clear speech. We found that clear speech enhanced recall for both listener groups, regardless of linguistic experience. This benefit was evident when both keyword and sentence recall measures were considered. Listeners were able to recall more individual words as well as entire sentences verbatim in clear speech compared to conversational speech. The clear speech benefit was also manifested in lower error rates (where response contained no information from the original sentence) and in fewer omitted responses (where no response was provided at all). These results extend previous findings linking speech clarity to improved sentence recognition memory (Gilbert et al., 2014; Keerstock & Smiljanic, 2018; Van Engen et al., 2012) by providing evidence that clear speech enhances recall, a more complex and effortful form of memory.

The sentence results showed that the clear speech benefit goes beyond the recall of a 'list' of words to include deeper linguistic encoding at the level of syntactic and semantic information. This effect may not be attributed solely to enhanced clear speech intelligibility (Keerstock & Smiljanic, 2018; Van Engen et al., 2012) since both clear and conversational sentences in the current study were presented in quiet, at equal intensity levels, and were therefore presumably similarly intelligible. Listening to clearly produced sentences still led to better recall. This suggests that hearing conversational sentences, which are typically produced with reductions and even deletions of some speech segments, may be more effortful and require additional cognitive resources resulting in diminished recall. Conversely, clear speech may have freed up cognitive resources for deeper processing of the speech signal and storage in memory (cf. "effortfulness hypothesis" McCoy et al., 2005; Rabbitt, 1968, 1990 and "ease of language understanding" model, Rönnberg et al., 2013). It is further possible that the hyper-articulated clear speech provides listeners with higher certainty about what is being said so that they are less likely to record an erroneous response or omit a response altogether. Further work is needed to better understand the link between the varied recall responses and speaking style modifications.

One contributing factor to the memory benefit may be the duration of the clear speech sentences. Clear speech modifications involve a decrease in speaking rate and an increase in pausing. If the total time spent processing is correlated with subsequent memory performance (Total-Time hypothesis, Cooper & Pantle, 1967), it is possible that longer clear speech sentences provide listeners with more processing time compared to shorter conversational sentences which in turn benefits memory retention. However, the opposite could also hold in that cognitive performance is degraded when processing time is increased as the products of early processing may no longer be available by the time later processing is complete (cf., processing-speed theory, Salthouse, 1996). More work is needed to assess the contribution of duration and other conversational-to-clear speech modifications on linguistic processing and cognitive functioning.

Similar to the sentence recognition memory findings reported in Keerstock & Smiljanic (2018), the results here showed that non-native listeners were able to use conversational-to-clear speech modifications to significantly improve recall. This was true for both word and sentence recall measures; keyword and verbatim recall were higher while omit and error were lower in clear speech compared to conversational. Closer examination of the whole-sentence recall patterns, however, revealed some differences between the two listener groups. Non-native listeners overall recalled fewer entire sentences verbatim, recalled more incomplete (partial) sentences, and were more likely to omit a response than native listeners. These results are in line with findings showing reduced recall in L2 compared to L1 (Hygge et al., 2015; Molesworth et al., 2014; Schweppe et al., 2015). These differences likely reflect a difficulty in L2 processing found at all levels of linguistic structure, from sounds (Best & Tyler, 2007; Flege, 1995) to syntax (Ojima, Nakata, & Kakigi, 2005), even for highly proficient L2 speakers (Stepanov, Andreetta, Stateva, Zawiszewski, & Laka, 2019). Non-native listeners' higher omission rate and lower error rate compared to native listeners may suggest that non-native listeners were less likely to attempt responding or guessing when unsure. The higher rate of incomplete partial responses for non-native listeners may further highlight their difficulty in making use of top-down knowledge to fill in missing information (Bradlow & Alexander, 2007; Schweppe et al., 2015). Processing highly variable speech and committing information to memory is a difficult and effortful task for any listener, but is particularly challenging for non-native listeners. Despite these difficulties, the lack of an interaction between speaking style and listener group is notable because it shows that the clear speech benefit on recall is significant even among listeners who are not fully proficient in the target language.

This study represents the first examination of the effect of clear speech on memory using a cued-recall task. It is well established that speaking clearly enhances word recognition in noise for a variety of listener groups. The results presented here add further evidence that highly-intelligible clear speech enhances memory beyond recognition of spoken speech to recall of the message conveyed in the speech signal. The results support the idea that processing clear speech may reduce effort in memorizing spoken information in L1 and L2 processing. This research has implications for daily interactions in challenging environments, such as hospitals or classrooms, where successful information recall may impact health and learning outcomes (Bankoff & Sandberg, 2012; Latorre-Postigo et al., 2017; McGuire, 1996).

# Chapter 4: Recall and recognition memory of clearly produced speech by native and non-native talkers

## 4.1. ABSTRACT

Native and non-native listeners were more accurate in identifying sentences as previously heard (Keerstock & Smiljanic, 2018) **(Chapter 2)** and in recalling words and entire sentences (Keerstock & Smiljanic, 2019) **(Chapter 3)** if the sentences were heard in intelligibility-enhancing clear speech compared to conversational speech. This clear speech benefit on listeners' memory might in part be due to decreased listening effort in perceiving clear speech ("effortfulness hypothesis", Rabbit 1968, 1990). The effect of reading sentences aloud in clear speech on talkers' memory, however, is unknown. In the present study, native and non-native English speakers read sentences aloud in clear and conversational speaking styles. Their memory of the read sentences was assessed either via a sentence recognition memory task (Experiment 1; n=90) or a recall task (Experiment 2; n=75). Results from both experiments showed that reading aloud in a listener-oriented hyper-articulated clear speech lead to lower recognition memory (Experiment 1) and fewer keyword recalled (Experiment 2) compared to reading aloud in a casual conversational style. The results indicate that producing clear speech, unlike perceiving it, interferes with sentence recognition memory and recall. Production of listener-oriented hyper-articulated speech may recruit more cognitive resources, leaving fewer available for storing spoken information in memory. Implication for the relationship between speech perception and production are discussed.

## 4.2. INTRODUCTION

Acoustic-phonetic enhancements in the form of listener-oriented intelligibility-enhancing clear speech can improve listeners' retention of spoken information in memory. Clear speech enhanced listeners' recognition memory (i.e., recognizing 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). Clear speech benefit on memory extended to both native and non-native English listeners of various first-language backgrounds (Keerstock & Smiljanic, 2018, 2019). In addition to recognition memory, listening to clear speech improved recall of words and entire sentences (Keerstock & Smiljanic, 2018, 2019). The same benefit of clear speech was found for older adults with normal-to-moderately impaired hearing-listening abilities in recall of medically-relevant spoken information (DiDonato & Surprenant, 2015). The clear speech benefit on memory may be due to the decreased cognitive effort required to process easier-to-understand clear speech compared to conversational speech, which frees up more cognitive resources for encoding speech in memory (e.g., "effortfulness hypothesis", McCoy et al., 2005; Rabbitt, 1968, 1990).

The present study aims to expand this line of research—thus far focused on speech perception—to speech production by examining the effect of speaking style on memory for the talkers themselves instead of listeners. The 'production effect' (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010), or 'generation effect' (Jacoby, 1978; Slamecka & Graf, 1978) refers to the superior memory retention of material read aloud relative to material read silently during an encoding phase. The benefit on recognition memory was found for a variety of production types: mouthing and saying nonwords aloud (MacLeod et al. 2010) and saying words loudly or singing words (Quinlan & Taylor, 2013). The production effect was also observed for writing, typing or spelling (Forrin, MacLeod, & Ozubko, 2012), but the effect on memory was not as large as when producing the words out loud. The benefit was observed for word pairs, sentences and textbook passages (Ozubko et al., 2012), and dialogues (Knutsen & Le Bigot, 2014). The benefit was found to last over longer retention intervals (i.e., a week, Ozubko et al. 2012). The production effect is most commonly explained by the ‘distinctiveness account’: items read aloud have additional salient information (e.g., articulatory and acoustic) relative to items not read aloud which are used during the test for discriminating produced items from unproduced items. Thus, superior retention of words may arise through the addition of the salient acoustic information contained in loud speech (higher intensity) and sung production (wider f0 modulations) compared to read aloud normally and silent conditions (Quinlan & Taylor, 2013) .

The production of conversational-to-clear speech adaptations typically consist of increases in the dynamic pitch range and increases in intensity (as in the loud and sing conditions in Quinlan & Taylor, 2013) but, they also include decreases in the speaking rate, more salient release of stop consonants, expansion of the vowel space, and enhancement of language-specific vowel and consonant contrasts. According to the ‘production effect’, clear speech production should lead to enhanced memory because it provides additional salient articulatory and acoustic cues relative to normal conversational speech which should facilitate memory retention. However, clear speech production is complex and may be more resource-demanding. For instance, producing articulatory-acoustic modifications in clear speech requires accessing representations at multiple linguistic levels (phonemic, lexical, prosodic, pragmatic) which may impact speech planning and memory retention. Furthermore, implementing clear speech involves greater articulatory magnitude (peak speed, longer movement durations, greater distances) than casual speech (Perkell et al., 2002; Song, 2017; Tang et al., 2015). This suggests that producing clear hyper-articulated speech is more effortful. To the extent that the 'effortfulness hypothesis' can predict memory performance in speech production as well as in speech perception, it would predict that the more-effortful-to-produce speaking style (i.e., clear speech) would lead to decreased memory performance. The present study therefore tested whether producing clear speech differentially affects memory retention compared to conversational speech and if so whether it improves ('production effect') or decreases memory retention ('effortfulness hypothesis').

Adding insights from speech production to the existing results from speech perception presents theoretical interest. Indeed, the relationship between speech perception and production is still elusive, and while a cooperative relationship is often assumed between the two modalities (Casserly & Pisoni, 2010; Denes & Pinson, 1963), mismatches between the two processes have also been reported in the literature (e.g., in perception and production of epenthetic vowels; (Baese-Berk, 2019; Dupoux, Hirose, Kakehi, Pallier, & Mehler, 1999). Crucially, little is known about how cognitive resources are shared by the two modalities and how this impacts auditory memory. By exploring whether producing clear speech can enhance talkers' memory for self-produced speech, this research also stands to shed light on potentially beneficial mnemonic strategies that have practical implications in the fields of education and psychology.

The present study also considers the effect of native language background on the retention of different speaking styles in memory. In particular, it examines how conversationally and clearly produced English sentences are retained by talkers for whom English is the first language (i.e., native English talkers), and talkers for whom English is the second language (i.e., non-native English talkers of various L1 background). Non-native talkers lack experience with all levels of linguistic structure in the target language resulting in systematic deviations from the target language norms (Bradlow, Blasingame, & Lee, 2018). With regard to the clear speech productions, non-native talkers can implement global modifications (increased F0 mean and energy between 1 and 3 kHz) but may find language-specific enhancements (consonant and vowel phonemic contrasts) challenging (Granlund, Baker, & Hazan, 2011; Rogers, DeMasi, & Krause, 2010; Smiljanic & Bradlow, 2005, 2011). The current work aims to fill in the gap in our understanding of how such increased difficulty impacts memory.

In this study, the production effect of hyper-articulated intelligibility-enhancing clear speaking style on native and non-native English talkers' memory was examined using a recognition memory task (Experiment 1) and a recall task (Experiment 2). The goal behind using two different memory tasks was to assess the generalizability and robustness of the results to different yet ubiquitous memory processes. Recognition memory assesses the familiarity process in recognizing previously read sentences ("is this item familiar?") with a binary response (yes/no), while recall assesses retrieval of lexical items and entire units of connected meaning from memory and provides a quantitative assessment of the subjects' auditory memory. Examining the two tasks can thus provide a more comprehensive account of the production effect of clear speech on memory for spoken information. This will also allow us to compare the production effects to the perception only effects on recognition memory and recall from our previous work (Keerstock & Smiljanic 2018, 2019).

## 4.3. EXPERIMENT 1: SENTENCE RECOGNITION MEMORY

### 4.3.1. Participants

Sixty native English listeners (37 female; $M_{age}$ = 19.6, $SD_{age}$ = 2) and 30 non-native English listeners (22 female; $M_{age}$ = 23.2, $SD_{age}$ = 4.1) participated in the experiment. The native English listeners were all born and raised in the U.S. and acquired English from birth. Approximately half of the native listeners (n=27) reported exposure to another language from birth alongside English; however, they all reported being English dominant at the time of testing. The difference in sample size between native and non-native listeners resulted from collapsing the native English listeners with and without exposure to another language from birth to one group, as the effect of second language exposure on recall was not significant (as discussed below). The non-native listeners acquired English on average after age 7.7 (range 5-13) and received no exposure to English at home from parents/caregivers. Information about the non-native listeners' language background is provided in Table 8. Participants were recruited from the University of Texas community and received monetary compensation or research credit for their participation. They signed a written informed consent and filled out a detailed language background questionnaire adapted from the LEAP-Q (Marian et al., 2007). All passed a hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

Table 8. Language background information for the non-native listeners in Experiment 1 RM (*n*=30) and Experiment 2 Recall (*n*=32)

| | Experiment 1 RM | | | Experiment 2 Recall | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Age of first exposure to English (in years) | 7.7 | 2.3 | 5-13 | 7.9 | 2.7 | 5-16 |
| Age of arrival to USA (in years) | 16.9 | 8.7 | 0-31 | 16.8 | 9.2 | 0-33 |
| Time spent in USA (in years) | 6.3 | 6 | 0-23 | 5.7 | 6.8 | 0-21 |
| Daily exposure to English[1] | 4.8 | 0.61 | 2-5 | 4.7 | 0.63 | 2-5 |
| Daily exposure to L1[1] | 4.2 | 1.1 | 1-5 | 4.3 | 0.78 | 2-5 |
| Self-estimated proficiency in English[2] | 4.3 | 0.7 | 2.25-5 | 4.1 | 0.66 | 3-5 |
| Self-estimated proficiency in L1[2] | 4.8 | 0.3 | 3.75-5 | 4.6 | 0.6 | 2.75-5 |
| First language | Spanish (n=12), Mandarin (n=4), Vietnamese (n=3), Cantonese (n=2), Korean (n=2), Arabic, French, German, Hebrew, Hindi, Hungarian, Turkish (n=1). | | | Spanish (n=11), Mandarin (n=6), Korean (n=3), Bahasa Indonesian (n=2), Vietnamese (n=2), Arabic, Greek, Hindi, Japanese, Persian, Portuguese, Taiwanese, Turkish (n=1). | | |

[1] (For each language, self-estimated amount of daily exposure on a scale from 1 (no current exposure) to 5(constant exposure) [2] (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))

### 4.3.2. Material

The material consisted of 120 unique meaningful sentences (e.g., “The hot sun warmed the ground”) from the same sentence pool used in Keerstock & Smiljanic (2018) and Keerstock & Smiljanic (2019). Sixty sentences were read aloud by the talkers. The remaining 60 sentences were used as decoy in the RM test phase where they were read silently only. The sentences contained high-frequency words familiar to non-native listeners (see Calandruccio & Smiljanić, 2012 for details about the development of the materials). All sentences were composed of 4 content and 2 function words and varied between 6 and 12 syllables.

### 4.3.3. Procedure

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

#### *4.3.3.1. Familiarization*

To familiarize the participants with the two speaking styles before the experiment, the practice sentence "The dark house scared the baby" appeared in the center of the screen of a PowerPoint slide. This sentence was not used in the main experiment. Participants were instructed to read the sentence 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. To further familiarize the participants with the speaking styles, they listened to an example of “The dark house scared the baby” produced in each speaking style by the speaker who produced the stimuli in Keerstock & Smiljanic (2018). Each example was played only once to limit imitation effects. Finally, participants were asked to read aloud the practice sentence in each speaking style one more time.

#### *4.3.3.2. RM experiment*

The RM experiment consisted of the exposure phase and the test phase. All instructions and stimuli were presented in E-Prime 2.0 Psychology Software. The E-Prime button box (SRbox) was used to navigate through the experiment and to record participants’ responses. Participants’ productions were recorded in E-Prime using a Logitech head-mounted microphone. The experimental session started with 4 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 the required speaking style into the microphone as soon as it appeared on the screen. Each sentence was presented in the center of the screen against a uniform white background in black Arial size 25 font and remained on the screen for a duration of 6000ms. Participants were not instructed to self-correct errors 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 that 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 the sentences in Block 1, 3 and 5 in conversational speech and all the sentences in Block 2, 4 and 6 in clear speech, and half of the participants produced all the sentences in Block 1, 3 and 5 in clear speech and all the sentences in Block 2, 4 and 6 in conversational speech. This ensured that all sentences were produced in both conversational and clear style across speakers. A screen instructing the participant which speaking style to adopt appeared before every block.

Immediately after producing all 60 sentences, participants completed the RM test. In the test phase, participants were presented with all the items from the exposure phase (60 old sentences) and 60 new items (sentences they did not produce in the exposure phase). The sentences were randomly presented one at a time in the center of the screen against a uniform white background in black Arial size 25 font. Each sentence was presented only once. For each sentence, 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. Participants were instructed to respond as quickly and accurately as possible.

#### *4.3.3.3. Additional recordings*

After the RM experiment was completed, participants recorded the same 60 sentences that they produced in the exposure phase, 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 actually produced two distinct styles during exposure.

### 4.3.4. Analyses

#### *4.3.4.1. Acoustic Analyses*

In order to verify that the talkers implemented two distinct (conversational and clear) speaking styles, we sampled recordings from every talkers to conduct global acoustic analyses[8], targeting acoustic metrics typically reported for conversational-to-clear speech modifications (see review by Smiljanic & Bradlow, 2009): articulation rate (syllables/second excluding pauses), pause rate (number of pauses/sentence), pause duration (in seconds), energy in the 1–3 kHz range (long-term average spectrum in 1-3k range) and F0 mean and range. Acoustic analyses were conducted using Praat (Boersma & Weenink, 2001). Articulation rate was calculated as the number of syllables produced per second after pauses were excluded. Pauses were defined as a period of silence exceeding 200ms excluding closures for stop-initial words. Sampling of sentences to analyze was performed at random, prior to conducting analyses. The sampled sentences were the same unique sentences for every talkers to maintain lexical consistency across talkers. In order to minimize fatigue effects across blocks (e.g. different speaking style production during the first vs. the middle vs. the last block), one sentence per block was selected to have a representative speech sample from each block in the subset. Since there were 6 blocks in

[8] Even though acoustic analyses are not typically conducted and reported in "production effect" studies (including studies that involve different production styles like singing or loud as in Quinlan & Taylor, 2013), we deemed this step necessary given that our predictions regarding RM depended on the production of two distinct speaking styles.

the exposure phase (3 in conversational, 3 in clear) and 6 blocks in the additional recording phase (3 in conversational, 3 in clear), 12 sentences were analyzed per talker (6 clear sentences and their conversational counterpart). Thus, the total analyzed sample consisted of 1080 sentences. To verify that the conversational productions were significantly different from the clear productions, we ran LMER models separately for native and non-native listeners on each of the 6 metrics as the dependent variables with Speaking Style (Conversational[reference]vs. Clear) as the fixed-effect and Subject and Sentence as random effects.

Furthermore, as a proxy of the effort involved in speech planning while reading aloud in conversational vs. clear speaking style, we examined speech onset latency. Speech onset latency was measured on the subset of sentences as the duration (in ms) from stimuli onset display on the screen to speech onset. The durations were entered in a LMER as the dependent variable, Speaking Style (Conversational$_{\text{[reference]}}$vs. Clear), Talker Group (Native$_{\text{[reference]}}$ vs. Non-native), and the Speaking Style by Talker Group interaction were entered in the model as independent variables. Subject and sentence were treated as random effects.

#### *4.3.4.2. RM*

The RM data was analyzed within a signal detection framework (Snodgrass & Corwin, 1988) and following previous analyses in Keerstock & Smiljanic (2018). 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 aloud in any speaking styles, they were only orthographically presented and silently read during the test. 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 for each participant by subtracting the normalized probability of the overall false alarm rate from the normalized probability of either conversational or clear hit rate. These 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 & Corwin, 1988). C scores were calculated as in Snodgrass and Corwin (1988) wherein positive C values indicate bias towards responding new and negative C values indicate bias towards responding old. Finally, we analyzed the reaction times (RTs) in responses to old items produced clearly and conversationally in the exposure phase to compare the processing time in recognition memory associated with the two speaking styles. The RTs were calculated as the time elapsed from the onset of written stimulus presentation on screen to the time the participant pressed the button on the button box to indicate their decision (old/new).

Linear mixed-effects regressions (LMER) were conducted on d' scores and RTs (in milliseconds) as the dependent variables. Speaking Style (Conversational$_{\text{[reference]}}$ vs. Clear), Talker Group (Native$_{\text{[reference]}}$ vs. Non-native), and the Speaking Style by Talker Group interaction were included in the model. Subject was treated as a random effect. All regression models throughout this paper were fit using the lme4 package in R (Bates et al., 2015). Significance values were computed using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Pairwise comparisons were performed with the emmeans package in R (Lenth, Singmann, Love, Buerkner, & Herve, 2018). Effect sizes were measured with Cohen's d (Cohen, 1988).

### 4.3.5. Results

#### *4.3.5.1. Speaking style acoustic differences*

Table 9 shows the mean (SD) articulation rate (syllables per second), pause duration (s), pause rate (number of pauses / sentence), 1–3 kHz energy (LTAS), F0 mean (Hz) and F0 range (Hz) for conversational (conv) and clear speech sentences produced by native and non-native talkers. Significance levels for the main effect of Speaking Style in lme4 models is reported in the table for each talker group. Overall, the acoustic analyses confirmed that conversational and clear sentences differed in their acoustic–articulatory characteristics along dimensions that are typically found in listener-oriented speaking style adaptations (Smiljanic & Bradlow, 2009). Compared to the conversational sentences, clear sentences had slower articulation rate, higher pause rate, longer pause duration, and increased energy in the 1–3kHz range for both talker groups. In addition, non-native talkers had a higher F0 mean and native talkers had a wider F0 range in CS than in conversational style.

Table 9. Mean (SD) articulation rate (syllables per second), pause duration (s), pause rate (number of pauses / sentence), 1–3 kHz energy (LTAS), F0 mean (Hz) and F0 range (Hz) for conversational (conv) and clear speech produced by native and non-native speakers in Experiment 1 (RM) and in Experiment 2 (Recall).

**Experiment 1 (RM)**

| | Native speakers (*n*=60) | | | Non-native speakers (*n*=30) | | |
|---|---|---|---|---|---|---|
| | conv | clear | Sig. | conv | clear | Sig. |
| Art. rate | 4.93 (0.99) | 3.37 (0.71) | *** | 4.4 (0.89) | 3.21 (0.56) | *** |
| Pause duration | 0.31 (0.18) | 0.41 (0.29) | ** | 0.29 (0.18) | 0.46 (0.28) | *** |
| Pause rate | 0.07 (0.27) | 1.15 (0.97) | *** | 0.2 (0.47) | 1.59 (0.94) | *** |
| LTAS | 19.03 (6.01) | 22.68 (6.39) | *** | 18.49 (6.02) | 22.18 (6.91) | *** |
| Pitch mean | 174.04 (47.3) | 174.7 (46.74) | n.s. | 181.42 (46.42) | 184.76 (46.48) | ** |
| Pitch range | 176.59 (121) | 192.88 (118) | * | 146.84 (90) | 162.63 (94) | n.s. |

**Experiment 2 (Recall)**

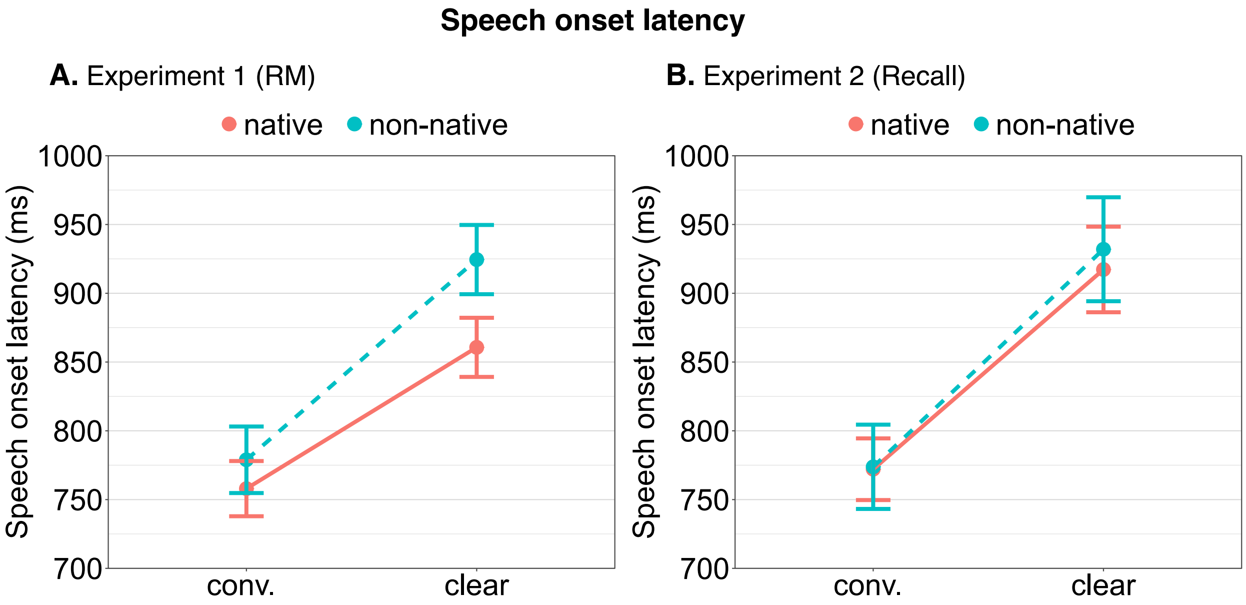
| | Native speakers (*n*=43) | | | Non-native speakers (*n*=32) | | |
|---|---|---|---|---|---|---|
| | conv | clear | Sig. | conv | clear | Sig. |
| Art. rate | 5.13 (0.92) | 3.32 (0.65) | *** | 4.58 (0.87) | 3.31 (0.64) | *** |
| Pause duration | 0.46 (0.22) | 0.46 (0.36) | n.s. | 0.44 (0.21) | 0.55 (0.33) | n.s. |
| Pause rate | 0.05 (0.24) | 0.72 (0.93) | *** | 0.07 (0.28) | 1.03 (1.03) | *** |
| LTAS | 18.48 (6.19) | 21.5 (6.45) | *** | 18.2 (6.45) | 21.04 (6.8) | *** |
| Pitch mean | 164.41 (46.08) | 166.48 (44.3) | n.s. | 170.11 (42.51) | 172.13 (43.73) | n.s. |
| Pitch range | 151.65 (106.13) | 193.53 (108.74) | *** | 154.75 (91.1) | 181.64 (102.43) | ** |

*Significance level reported for main effect of Speaking Style within talker groups in lme4 models*
*Signif. codes: '***' p<.001; '**' p<.01; '*' p<.05*

#### *4.3.5.2. Speech onset latency*

Figure 5A shows 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 Speaking Style (β = 114.3, t = 10.8, p<.001) such that speech onset latency was significantly longer for clear speech than conversational speech. The effect of Talker Group was not significant (β = 41.2, t = 1.3, p=.2). The Speaking Style by Talker Group was also not significant (β = 35.9, t = 1.6, p=.105). Results indicated that both native and non-native talkers required more time to initiate speech when reading sentences aloud in clear speech compared to conversational speech.

Figure 5. Speech onset latency (in ms) in conversational and clear speech for native (solid) and non-native (dashed) talkers in Experiment 1 (panel **5A**; right) and in Experiment 2 (panel **5B**; right).



### *4.3.5.3. RM*

Table 10 shows the mean (SD) of d', C, and reaction times (RTs) in milliseconds (ms) for native and non-native listeners for conversational and clear sentences. Figure 6A show the distributions of the d' scores by native and non-native listeners in the two speaking styles. 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 both talker groups. Results from the LMER ran on d’ as the dependent variable revealed a significant main effect of Speaking Style (β = -0.09, t = -2.6, p=.0102, Cohen’s *d* = -0.56) such that d’ scores in clear speech were significantly lower than in conversational speech. The effect of Talker Group was not significant (β = -0.09, t = -0.773, p=.4417). The Speaking Style by Talker Group was also not significant (β = 0.01, t = 0.135, p=.893).

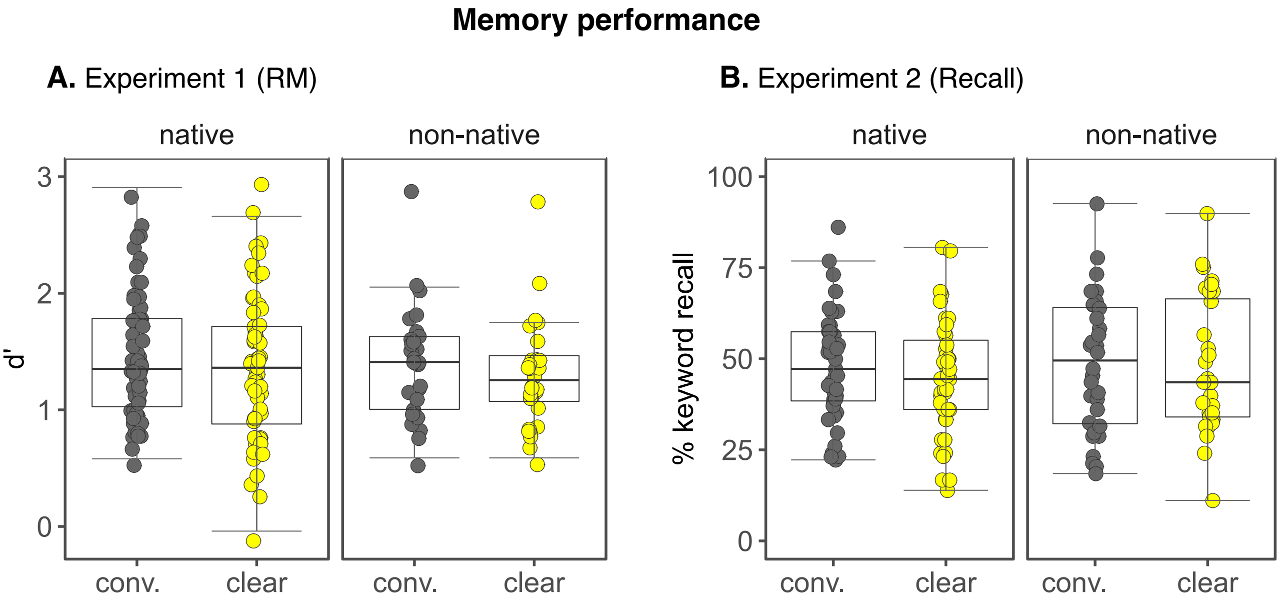
In order to explore the effect of early exposure to another language, we considered an alternative model in which the Talker Group variable was split into 3 levels: native English speakers with no other exposure to another language before age 6 (n=30), native English speakers with exposure to both English and another language before age 6 (n=27) and non-native English speakers (n=30)). The difference in Akaike Information Criterion (AIC) for the model with 2 levels (232) and for the alternative model with 3 levels (220) was not significant in an ANOVA model comparison (Baayen et al., 2008): $\chi_2$ = 2.1, df = 2, p = 0.35 and therefore we concluded that the parsimonious model with 2 levels (native vs. non-native) was a better model.

Results from the LMER ran on RTs (in milliseconds) as the dependent variable showed that there was no effect of Speaking Style (β = -24.44, t = -0.789, p=.432), no effect of Talker Group (β = 33.50, t = 0.492, p=.624), and no interaction between Speaking Style and Talker Group (β = -36.52, t = -0.553, p=.581).

Table 10. Mean (SD) of d’, C, and reaction times (RTs) in milliseconds (ms) for native and non-native listeners for conversational (conv) and clear sentences in the RM task.

| **Group** | **Style** | ***d’*** | ***C*** | **RT** |
|---|---|---|---|---|
| **Native** | **clear** | 1.37 (0.63) | 0.33 (0.35) | 1123 (323) |
| | **conv** | 1.47 (0.56) | 0.28 (0.32) | 1135 (362) |
| **Non-native** | **clear** | 1.29 (0.44) | 0.35 (0.34) | 1138 (290) |
| | **conv** | 1.37 (0.48) | 0.31 (0.32) | 1187 (365) |

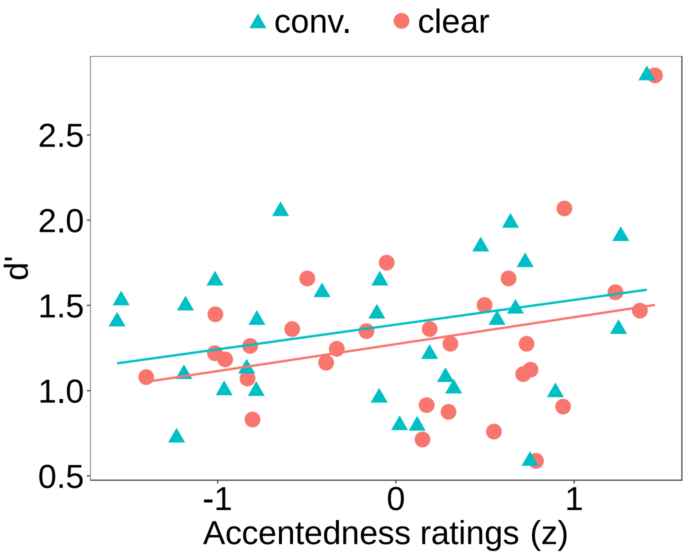
Figure 6. Memory performance for sentences read aloud in conversational (“conv” in grey) and clear (yellow) speech. Left panel **(6A)**: Experiment 1 RM shows the d’ distribution for native (*n*=60) and non-native talkers (*n*=30) in each speaking style. Right panel **(6B)**: Experiment 2 Recall shows the correct keyword recall distribution for native (*n*=43) and non-native talkers (*n*=32) in each speaking style. Boxplots extend from the 25th to the 75th percentile, with whiskers extending to points within 1.5 times the inter-quartile range. The central horizontal lines indicate the medians. Individual points represents data points for individual talkers.



### *4.3.5.4. Accentedness*

We conducted a number of exploratory analyses to investigate whether linguistic background variables (i.e., age English learned, age of arrival in an English speaking country) could predict behavioral differences in the recognition memory task, however, none of the variables were successful at predicting d’ scores. Therefore, in a separate experiment, we recruited 48 new native English listeners from the Linguistics subject pool at the University of Texas at Austin to provide accentedness ratings on the non-native talkers. To the extent that foreign-accentedness ratings can be used as an imperfect proxy for L2 (phonological) proficiency, we predicted that non-native talkers who were rated as more accented would score lower on the recognition memory task, as the less experience with producing L2 speech might recruit more cognitive resources resulting in decreased memory. Each listener heard all non-native talkers as well as 3 native talkers as controls (6 sentences*33 talkers). Each listener heard 6 sentences per talker (the same unique 3 sentences produced in clear and conversational style). Sentences were randomly presented one at a time to the listeners over headphones. Listeners were instructed to rate each sentence by clicking on a line representing a continuum from most native-like to most foreign-like on the computer screen. All instructions and stimuli were presented in E-Prime. The mouse x-axis coordinate response was recorded in E-Prime and normalized (z-scores) by listener for each speaking style and each talker. An LMER analysis was conducted with d' scores as dependent variable, Speaking Style and Averaged Normalized Accentedness Z-scores as independent variables and Talker as random intercept. Figure 7 shows the Normalized Accentedness Z-scores for mouse x-axis coordinate response. Negative values represent "native-like" ratings and positive values "foreign-sounding" ratings. Results from the LMER showed that Normalized Accentedness Z-scores did not significantly predict d' scores (p=.23), but the effect of Style was still significant (p= 0.0482), with higher d' for conversational sentences. Similar exploratory analyses were not conducted in Experiment 2.

Figure 7. Normalized z-scores for mouse x-axis coordinate response. Negative values represent “native-like” ratings and positive values “foreign-sounding” ratings. Clear (green) and conv. (red) *d’* scores represented across all 30 non-native speakers in Experiment 1 RM.



## 4.3. EXPERIMENT 2: SENTENCE RECALL

### 4.3.1. Participants

Forty-three native English listeners (24 female; $M_{age}$ = 19.3, $SD_{age}$ = 1.7) and 32 non-native English listeners (19 female; $M_{age}$ = 22.5, $SD_{age}$ = 3.9) participated in Experiment 2. They were all different individuals from Experiment 1 but had similar demographic and linguistic profiles. The native English listeners were born and raised in the U.S. and acquired English from birth. As in Experiment 1, approximately half of the native listeners (*n*=20) reported exposure to another language from birth alongside with English, but this factor had no significant effect on recall (as discussed below). The non-native listeners acquired English on average after age 5 (range 5-16) and received no exposure to English at home from parents/caregivers. Information about the non-native listeners' language background is provided in Table 7. Participants were recruited from the University of Texas community and received monetary compensation or research credit for their participation. They signed a written informed consent and filled out a detailed language background questionnaire adapted from the LEAP-Q (Marian et al., 2007). All passed a hearing screening, administered bilaterally at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

### 4.3.2. Material

The stimuli consisted of 72 unique sentences taken from the same sentence pool used in Experiment 1. The subset of sentences chosen for this experiment all had the same syntactic structure: a determiner and noun (e.g., the grandfather), followed by a verb, an adjective and a noun (e.g., drank the dark coffee). The cue in the recall booklet was always the first noun phrase (in italics) and the three keywords to be recalled were always the last three content words (underlined) e.g., "*The grandfather* drank the dark coffee" or "*The mother* baked the delicious cookies".

### 4.3.3. Procedure

Participants were seated in a sound-attenuated booth facing a computer monitor. Instructions and stimuli were presented with E-Prime 2.0 Psychology Software. As in Experiment 1, Experiment 2 consisted of the familiarization portion followed by the sentence recall test, and finally, concluded with an additional recording portion. In the familiarization phase, participants read one sentences in two speaking styles following the same instructions and details as in Experiment 1. The sentence used in the familiarization phase was not used in the subsequent cued-recall test. For the recall task, the 72 test sentences were divided into six blocks of 12 sentences. Each sentence was presented in the center of the screen against a uniform white background in black Arial size 25 font and remained on the screen for a duration of 6000ms. The participants were asked to read aloud the sentences as soon as it appeared on the screen and to try and memorize the sentences they were reading aloud. The speaking style presentation was counterbalanced such that half of the participants produced all the sentences in Block 1, 3 and 5 in conversational speech and all the sentences in Block 2, 4 and 6 in clear speech, and half of the participants produced all the sentences in Block 1, 3 and 5 in clear speech and all the sentences in Block 2, 4 and 6 in conversational speech. This ensured that all sentences were produced in both conversational and clear style across 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, participants were asked to write down their response on 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 provided in the booklet were in the same order as during the reading aloud (exposure) phase; however, participants were not instructed to fill the recall booklet in any particular order. The recall test was self-paced.

Finally, as in Experiment 1, after completion of the experiment participants recorded again the same 72 sentences that they had produced, 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) to assess whether the talkers had actually produced two distinct styles during exposure.

### 4.3.4. Analyses

#### *4.3.4.1. Acoustic analyses*

As in Experiment 1, a subset of 1080 sentences was analyzed to assess whether talkers implemented conversational-to-clear speech modifications. In addition to articulation rate (syllables/second excluding pauses), pause rate (number of pauses/sentence), pause duration (in seconds), energy in the 1–3 kHz range (long-term average spectrum in 1-3k range) and F0 mean and range, speech onset latency was measured as well to evaluate the impact of reading aloud in clear speech on speech planning.

#### *4.3.4.2. Recall*

Following Keerstock & Smiljanic (2019), each keyword to be recalled was scored as either correct (1) or incorrect (0). Since 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. Listeners were not penalized for obvious spelling errors. In the case of uncertainty due to handwriting, sentences were scored by another research assistant and consensus was reached. To predict the recall outcome (0 or 1), we conducted binomial logistic 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 Speaking Style (Conversational$_{\text{[reference]}}$ vs. Clear), and Listener Group (Native$_{\text{[reference]}}$ vs. Non-native) as the independent variables, Speaking Style X Listener Group as an interaction term, Word Position (1, 2, 3) as a covariate to account for the position of the word in the sentence, Block Position (1 - 6) as a covariate to account for practice effects throughout the experiment, and Sentence Position (1 - 12) as a covariate to account for serial position effects within each block of 12 sentences (i.e., primacy and recency). Subject and Stimuli were modeled using a random intercept term.

### 4.3.5. Results

#### *4.3.5.1. Speaking style acoustic differences*

Table 9 shows the mean (SD) articulation rate (syllables per second), pause duration (s), pause rate (number of pauses / sentence), 1–3 kHz energy (LTAS), F0 mean (Hz) and F0 range (Hz) for conversational (conv) and clear speech produced by native and non-native speakers. Significance levels for the main effect of Speaking Style in lme4 models is reported in the table for each talker group. The acoustic analyses confirmed the presence of the typical acoustic–articulatory adaptations found in listener-oriented clear speech (Smiljanic & Bradlow, 2009) for both native and non-native talkers: slower articulation rate, higher pause rate, increased energy in the 1–3kHz range, and wider F0 range for clearly produced sentences compared to conversationally produced sentences.

#### *4.3.5.2. Speech onset latency*

Figure 5B shows 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 Speaking Style ($\beta = 149.7$, $t = 9.7$, $p<.001$) such that speech onset latency in clear speech was significantly higher than in conversational speech. The effect of Talker Group was not significant ($\beta = 7.6$, $t = 0.2$, $p=.8$). The Speaking Style by Talker Group was also not significant ($\beta = 11.7$, $t = 0.37$, $p=.7$). Results indicated that both native and non-native talkers required more time to initiate speech when reading aloud clearly compared to conversationally.

#### *4.3.5.3. Recall*

Figure 6B shows native and non-native listeners' keyword recall accuracy in the two speaking styles. Results from the logistic regressions with keyword recall (0-1) as the dependent variable showed a significant main effect of Speaking Style (*Odds Ratio* [*OR*] = 0.9 [95% *CI*: 0.84-0.97], p=.004) but no significant effect of Listener Group (*OR* = 1.09 [95% *CI*: 0.74-1.59], p=.66). The Speaking Style X Listener Group interaction was not significant (*OR* = 1.11 [95% *CI*: 0.96-1.28], p=.17) and therefore was removed from the model before interpreting the main effects. As in Experiment 1, an alternative model in which the Talker Group variable was split into 3 levels was considered. The difference in Akaike Information Criterion (AIC) for the main model with 2 levels (18652) and for the alternative model with 3 levels (18656) was determined as not significant in an ANOVA model comparison (Baayen et al., 2008): $\chi_2$ = 0.2, df = 2, p = 0.9 and therefore we concluded that the parsimonious model with 2 levels (native vs. non-native) was a better model.

### 4.4. DISCUSSION

The overall purpose of this study was to assess if reading sentences out loud in clear speaking style confers memory benefit over reading sentences out loud in conversational style. Participants' memory for the sentences read aloud in conversational and clear speaking styles was assessed either in an old/new recognition memory test (Experiment 1) or in a recall task (Experiment 2). Consistent across the two tasks, the results showed that memory (indexed by *d'* in Experiment 1 and by keyword accuracy in Experiment 2) was significantly *reduced* for sentences produced in clear speech compared to sentences produced in conversational speech. The same decrease in d’ and keyword recall for clear compared to conversational speech was observed for native and non-native talkers. The lack of interactions between speaking style and listener group in the present study showed that clear speaking style adaptation was detrimental to memory retention for all talkers, including learners of the target language. Finer grained detail analyses exploring variation within the non-native talkers did not reveal different memory patterns. L2 proficiency proxies (i.e., age of learning English, age of arrival in English speaking country, foreign-accentedness) were not successful at predicting different memory patterns, regardless of the speaking style.

These results were exactly the opposite from the results obtained in speech perception for listeners hearing clear speech. Listening to clearly spoken sentences (without producing speech) resulted in higher recognition memory (Gilbert et al., 2014; Keerstock & Smiljanic, 2018; Van Engen et al., 2012) and in higher keyword recall (Keerstock and Smiljanic, 2019) compared to conversational sentences. Here, we found a negative effect of reading aloud in clear speech on talkers’ recognition memory and recall. The magnitude of the speaking style effect for the *talkers* in the recognition memory and recall tasks in the present study were compared to the magnitude of the results of within and cross-modal recognition memory for the *listeners* in Keerstock & Smiljanic (2018) and of recall for the *listeners* in Keerstock & Smiljanic (2019). For both RM and recall, the magnitude of the Speaking Style effect was somewhat smaller for the talkers compared to the listeners. The strength of the Speaking Style effect on talker’s recognition memory in the Experiment 1 was medium (Cohen’s *d*=0.56), whereas for the listeners in Keerstock &

Smiljanic (2018) it was strong (Cohen's *d*=0.89 for within-RM and *d*=0.85, for cross-modal RM). The magnitude of the Speaking Style effect on talkers' recall in Experiment 2 (*OR*=0.9; *OR-1*=-0.1) was smaller compared to the magnitude of the Speaking Style effect on listeners' recall in Keerstock & Smiljanic (2019) (*OR*=1.41; *OR-1*=+0.41). This difference in magnitude suggests caution in interpreting the effect of clear speech on talkers' memory. However, the fact that the pattern of results found for the talkers in sentence recognition memory (Experiment 1) replicated when testing sentence recall (Experiment 2) suggest that the negative memory effect of reading in clear speech is somewhat robust. A key question, then, is why memory was disrupted when sentences were read aloud in clear speaking style but was enhanced when only hearing clearly produced sentences (i.e., in the absence of generating hyper-articulated speech).

The results reported here contradict "production effect" predictions (MacLeod et al., 2010; Quinlan & Taylor, 2013). In Quinlan & Taylor (2013), the production of loud speech or singing resulted in superior retention of material compared to normal aloud production and silent readings. The authors argued that generating speech with additional salient cues (higher intensity, increased dynamic pitch range) led to enhanced memory for those items. Productions of conversational-to-clear speech adaptations also include higher intensity and increased pitch range yet they resulted in *inferior* memory retention compared to casual speech. One may wonder if our talkers did not produce distinct speaking styles when instructed to do so, however, acoustic analyses of the talkers' speech output demonstrated conversational-to-clear speech global modifications (slower articulation rate, higher pause rate, longer pause duration, increased F0 mean and range and increased LTAS within 1-3k energy range) consistent with previous findings (Smiljanic & Bradlow, 2009) thus confirming that they read target sentences in two distinct styles. It is possible that the talkers' memory in this study was taxed more because they were producing longer and more complex sentences compared to producing single words (as was done in a number of the studies examining the production effect on memory). While using different materials in these studies certainly affects memory processes, this however does not explain why producing sentences clearly is more detrimental for memory than producing sentences conversationally.

It seems more likely that the production of listener-oriented clear speech engages additional cognitive resources compared to following specific instructions to, for instance, read the words out loud. These recruited resources could be related to imagining an interlocutor and selecting the appropriate adjustments to address their perceptual difficulty (non-native interlocutor or listener with a hearing problem). Previous work has shown that talkers apply different acoustic-articulatory adjustments in response to different communicative challenges (Hazan & Baker, 2011; Cooke & Lu, 2010; Smiljanic & Gilbert, 2017). The results here suggest that cognitive resources may be taxed differently when producing clear speech than when implementing a more straightforward task such as reading out aloud or singing. The differences between the current results and those of Quinlan & Taylor (2013) could reflect the implementation of multiple modifications involved in clear speaking style, such as maximizing phonemic contrasts and prosodic information, in addition to slowing down, increasing intensity and F0 range (Smiljanic & Bradlow, 2009; Smiljanic, to appear). These acoustic-articulatory modifications are more complex than increasing loudness or pitch alone. This increased complexity requires accessing representations at multiple linguistic levels (phonemes, words, prosody), planning acoustic-articulatory movements for more exaggerated targets as well as increasing articulatory effort involved in these productions. These increased task, planning and implementation demands may shift resources away from encoding the produced information in memory.

These accounts are compatible with the processing models that invoke increased cognitive effort when speech comprehension is challenging due to signal degradation or listener characteristics . Within these models, memory encoding is negatively affected with increased processing effort. In previous perception work (Gilbert et al., 2014; Keerstock & Smiljanic, 2018, 2019; Van Engen et al., 2012), we argued that easier-to-understand clearly spoken sentences required fewer resources to process, compared to conversational speech, leaving more resources for encoding information in memory. To the extent that the articulatory and cognitive effort is increased when producing hyper-articulated and listener-oriented speaking style, we would expect the opposite effect than in perception alone which is exactly what we found in the current study. In the production-to-memory loop, fewer resources remain available for encoding information after producing clear speech compared to a less-effortful-to-produce casual speaking style.

The notion that producing clear speech is effortful aligns with the H&H theory (Lindblom, 1990; Perkell et al., 2002) which posits that talkers adjust their spoken output in a continuous manner varying from hypo- to hyper-articulated speech. Hypo-articulated speech arises from the talker-centric need for the economy of effort while listener-oriented hyper-articulated speech is on the opposite end of the spectrum involving increased effort aimed at maximizing intelligibility. The results showing delayed speech onset of clear speech compared to conversational speech support the idea that producing clear speech involves more effort. Both native and non-native talkers were consistently slower to initiate clear speech suggesting that planning and execution of clearly spoken sentences is more complex and may require additional resources. Evidence from the speech planning literature shows that increasing processing demands through speeded production of tongue twisters (Goldrick & Blumstein, 2006), when producing cognates in L2 (Jacobs, Fricke, & Kroll, 2016) or in reading paragraphs with increasing difficulty by older adults (Gollan & Goldrick, 2019), results in increased errors, reading times and errors in articulatory execution. To the extent that producing clear speech is similar to these tasks, we would expect the increased demands to affect processing from speech onset to memory.

Finally, selective attention likely plays a role in memory encoding for read clear speech sentences. When hearing clear speech sentences, listeners' attention may be drawn to the exaggerated acoustic-phonetic cues which then facilitates encoding of these features in memory. In reading sentences out loud, attention may be allocated differently leading to diminished memory due to, for instance, mind-wandering which is very common during reading and detrimental to comprehension (Mooneyham & Schooler, 2013). A recent study showed that reading aloud promotes mind-wandering even more relative to silent reading (Franklin, Mooneyham, Baird, & Schooler, 2014). It is possible that the reading aloud task in the current study led to greater mind-wandering, and lower memory, compared to when participants were only hearing sentences in the perception only study. The results further suggest that reading aloud in clear speech may result in more mind-wandering compared to conversational speech, which comes at the expense of encoding of the spoken information in 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 more comprehensive understanding of intelligibility variation and its impact on memory. This examination should include objective and subjective measures of the articulatory and cognitive effort involved in producing and perceiving clear and conversational speech as well as a close look at speech planning in different speaking styles. A pressing goal is to delineate how the production and perception modalities compete for various cognitive resources, such as selective attention and working memory.

# Chapter 5: General discussion and conclusions

In a series of 5 studies, I examined the effect of clear speech production and perception on RM and recall for native and non-native listeners and talkers. The goal of the dissertation was to shed light on how signal-related articulatory-acoustic enhancements in the form of clear speech affect signal-independent processes and integration of information in memory. Through this, I also examined the link between production and perception and L2 speech processing and encoding in memory. These questions were investigated using controlled experiments in which native and non-native English listeners and talkers completed RM or recall tasks. The dissertation produced a number of novel findings.

One of the main contributions of this dissertation is new empirical evidence for the *enhanced* RM and recall of clear speech among native and non-native *listeners*. In particular, my studies showed that the clear speech benefit on listeners' memory (Gilbert et al., 2014; Van Engen et al., 2012) extends from within-modal RM (Chapter 2, Exp 1) to cross-modal RM (Chapter 2, Exp 2) and to recall (Chapter 3), a more complex and effortful memory task (Gillund & Shiffrin, 1984; Rhodes et al., 2019). The finding that RM and recall were *enhanced* in clear speech did not align with the "perceptual-interference effect," which predicted that perceptually more difficult stimuli requiring some amount of effort (i.e., conversational speech) would have been better remembered (Besken & Mulligan, 2013; Diemand-Yauman et al., 2010). Instead the results aligned with the "effortfulness hypothesis" and the "ease of language understanding" model in that processing of easier-to-understand clear speech freed up some processing resources for encoding of speech in memory. The finding that listeners showed higher RM (d') for clear speech even in the cross-modal RM test (Chapter 2, Exp 2) suggests that the memory traces could be activated through deeper linguistic processes at a level abstracted from the input speech. Note that such an interpretation relies on the assumption that intelligibility-enhancing clear speech is less cognitively effortful to process. Future research is needed to provide direct evidence for the objective and subjective cognitive effort associated with perception of clear and conversational speech with physiological measures (e.g., pupillometry) or dual-task paradigms. An advantage of pupillometry as a physiological index of cognitive load is that it can identify processing differences even when behavioral results are similar (e.g., equivalent intelligibility levels for clear and conversational speech in quiet). For instance, pupillometry showed differences in cognitive effort for school-aged children listening to speech in "ideal" vs. "typical" listening environment despite them showing no difference in performance accuracy or reaction time (McGarrigle, Dawes, Stewart, Kuchinsky, & Munro, 2017).

Another novel finding concerns the enhanced recall of entire verbatim sentences in clear speech compared to conversational speech for native listeners (Chapter 3). This suggests that the clear speech benefit extended beyond recall of isolated words to entire units of connected meaning. By using meaningful sentences as stimuli instead of single words or lists of words, the dissertation revealed complex interactions between lower-level articulatory-acoustic modifications and integration and encoding of information at all levels of linguistic structure (e.g., syntax, semantics). Future work should examine whether the clear speech benefit on memory can generalize from sentences to an even larger unit of connected discourse, for instance entire paragraphs of connected meaning, as in DiDonato & Surprenant (2015) that used medically-relevant materials or in Ozubko, Hourihan, & MacLeod (2012) that used educationally-relevant paragraphs.

Another main contribution from this work regards the effect of clear speech *production* on RM and recall by the *talkers* themselves. The "generation effect" (Bertsch, Pesta, Wiscott, & McDaniel, 2007) and the “production memory effect” (MacLeod et al., 2010) predict that producing words out loud is a powerful mnemonic device, and that salient distinctive productions (i.e., loud, singing) enhance memory for those words compared to words read aloud normally or silently (Quinlan & Taylor, 2013). Since clear speech involves, among other articulatory-acoustic changes, speaking more loudly and variation in f0 range, increased memory for clear sentences was predicted. However, the results from Chapter 4 indicated that RM and recall were both *diminished* for native and non-native *talkers* for clear speech compared to conversational speech.

Taken together, the results from speech perception and production in the dissertation revealed that hearing and producing listener-oriented intelligibility-enhancing clear speech lead to different memory outcomes. Hearing clear speech enhanced listeners' auditory memory whereas producing clear speech articulatory-acoustic adaptations impaired talkers' verbal memory. These seemingly disparate findings in perception and production can be reconciled by models that appeal to ‘effort’ and cognitive load as detrimental to memory. It is possible that talkers spent more processing resources on addressing the perceptual difficulty on the part of the listener and on the planning and implementation of hyper-articulated clear speech compared to conversational speech. This may have led to fewer resources being available for encoding of the read information in memory. The reading task could have also resulted in more mind-wandering, which comes at the expense of encoding in memory for semantic content (Franklin et al., 2014). The results are compatible with Lindblom's H&H theory (1990), which posits a trade-off between economy of effort and intelligibility. The reason that talkers revert back to casual speech may be that engaging in the production of hyper-articulated speech for long durations is too costly for talkers and detracts talkers from the communicative message. The same principle of limited processing resources could thus account for interference with memory consolidation in the production domain (larger processing cost in producing hyper-articulated intelligibility-enhancing clear speech) and for facilitation of memory consolidation in the perception domain (diminished processing cost in hearing hyper-articulated intelligibility-enhancing clear speech). Crucially, this suggests that the two modalities compete for the same resources. Future research should collect objective and subjective measures of the articulatory and cognitive effort involved in producing and perceiving clear and conversational speech as well as examine speech planning in different speaking styles. A pressing goal is to delineate how the production and perception modalities compete for various cognitive resources, such as selective attention and working memory.

Finally, by including non-native participants in the experiments, the dissertation provided novel insights into L2 speech processing and integration of spoken information in memory. Overall, results for native and non-native listeners and talkers in the dissertation were similar. Even though the literature has largely demonstrated that speech perception and production are more difficult in L2, and furthermore, that non-native listeners seem to benefit to a smaller degree from clear speech compared to native speakers (Bradlow & Bent, 2002), my studies provided evidence that clear speech sentences are recognized and recalled more easily by both native speakers and non-native speakers. A few differences between the two groups were found, however. First, while native and non-native listeners performed similarly in within-modal RM, non-native listeners had lower d' in cross-modal RM compared to native listeners (Chapter 2). This indicated a cost for cross-modal integration and retrieval of information and greater reliance on acoustic input during test phase for non-native listeners. Second, compared to native listeners, non-native listeners recalled fewer entire sentences verbatim, recalled more incomplete (partial) sentences, and were more likely to omit a response than native listeners (Chapter 3). This was taken as further evidence of non-native's greater difficulty in using top-down knowledge to fill in missing information at the signal level (Bradlow & Alexander, 2007). Despite the few differences, the clear speech benefit on RM and recall for listeners was significant even among listeners who are not fully proficient in the target language. In production, the results for native and non-native talkers were again similar in that both groups showed lower RM and recall for clear compared to conversational speech. In terms of practical application and take-away message, the findings suggest that non-native as well as native listeners stand to benefit from listening to clear speech as it can enhance their memory for spoken information, and that reading aloud in clear speech might disrupt memory processes for both talker groups.

Building on findings form this dissertation, future research should examine the effect of clear speech on memory in more realistic environments by using educationally-relevant or medically-relevant paragraphs (Chan, McDermott, & Roediger, 2006; DiDonato & Surprenant, 2015) and more interactive tasks. It should also examine the clear speech benefit across the life span. As we age for instance, peripheral-auditory and cognitive-memory functions decline and lead to increased listening effort during speech processing (Anderson & Gagné, 2014; Committee Hearing and Bioacoustics and Biomechanics (CHABA), 1988; Salthouse, 1996). Examining whether clear speech can enhance verbal memory in older adults (or children) with and without hearing loss would provide novel insights into compensatory cognitive mechanisms that allow listeners to understand and remember speech under a wide range of communicative situations. Another important aspect that deserves further investigation is the temporal boundary of the clear speech benefit on listeners' memory. The existing literature (including this dissertation) has documented this clear speech benefit only for memory tested immediately following exposure. It remains to be seen whether the benefit persists after a delay, for instance after an intervening task or after overnight consolidation.

## References

Anderson, P., & Gagné, J. (2014). Older Adults Expend More Listening Effort Than Young Adults Recognizing Speech in Noise. *Journal of Speech, Language, and Hearing Research*, *54*(June 2011), 944–958. https://doi.org/10.1044/1092-4388(2010/10-0069)a

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D. (1992). Working Memory. *Science*, *255*(5044), 556–559. Retrieved from http://www.jstor.org.ezproxy.lib.utexas.edu/stable/2876819

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychological Review*. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation - Advances in Research and Theory*. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex*, *112*, 91–106. https://doi.org/10.1016/j.cortex.2018.05.015

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-019-01725-4

Bankoff, S. M., & Sandberg, E. H. (2012). Older Adults’ Memory for Verbally Presented

Medical Information. *Educational Gerontology*, *38*(8), 539–551. https://doi.org/10.1080/03601277.2011.595284

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

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory and Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory and Cognition*, *41*(6), 897–903. https://doi.org/10.3758/s13421-013-0307-8

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 1–47). Amsterdam, The Netherlands: John Benjamins.

Björkman, M. (1967). Relations between intra-modal and cross-modal matching. *Scandinavian Journal of Psychology*, *8*(1), 65–76. https://doi.org/10.1111/j.1467-9450.1967.tb01375.x

Boersma, P., & Weenink, D. (2001). Praat: doing phonetics by computer [Computer program]. *Glot International*.

Borghini, G., & Hazan, V. (2018). Listening Effort During Sentence Processing Is Increased for Non-native Listeners: A Pupillometry Study. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00152

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The*

*Journal of the Acoustical Society of America*, *112*(1), 272–284. https://doi.org/10.1121/1.1487837

Bradlow, A. R., Blasingame, M., & Lee, K. (2018). Language-independent talker-specificity in bilingual speech intelligibility: Individual traits persist across first-language and second-language speech. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 509–524. https://doi.org/10.5334/labphon.137

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*(1), 80–97. https://doi.org/10.1044/1092-4388(2003/007)

Brand, A. N., Jolles, J., & Gispen-de Wied, C. (1992). Recall and recognition memory deficits in depression. *Journal of Affective Disorders*, *25*(1), 77–86. https://doi.org/10.1016/0165-0327(92)90095-N

Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and nondeceptive sentences. *Journal of Memory and Language*, *52*(4), 618–627. https://doi.org/10.1016/j.jml.2005.01.017

Calandruccio, L., & Smiljanic, R. (2012). New Sentence Recognition Materials Developed Using a Basic Non-Native English Lexicon. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1342–1355. https://doi.org/10.1044/1092-4388(2012/11-0260)

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*. https://doi.org/10.1002/wcs.63

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006, November). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, Vol. 135, pp. 553–571. https://doi.org/10.1037/0096-3445.135.4.553

Chan, K. Y., Chiu, M. M., Dailey, B. A., & Jalil, D. M. (2019). Effect of Foreign Accent on Immediate Serial Recall. *Experimental Psychology*, *66*(1), 40–57. https://doi.org/10.1027/1618-3169/a000430

Cohen, J. (1988). Statistical power for the social sciences. *Hillsdale, NJ: Laurence Erlbaum and Associates*.

Committee Hearing and Bioacoustics and Biomechanics (CHABA). (1988). Speech understanding and aging. *Journal of the Acoustical Society of America*, *83*(3), 859–895. https://doi.org/10.1121/1.395965

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, *128*(4), 2059–2069. https://doi.org/10.1121/1.3478775

Cooke, M., Mayo, C., Valentini-Botinhao, C., Stylianou, Y., Sauert, B., & Tang, Y. (2013). Evaluating the intelligibility benefit of speech modifications in known noise conditions. *Speech Communication*, *55*(4), 572–585. https://doi.org/10.1016/j.specom.2013.01.001

Cooper, E. H., & Pantle, A. J. (1967). the Total-Time Hypothesis in Verbal Learning. *Psychological Bulletin*, *68*(4), 221–234. https://doi.org/10.1037/h0025052

Craik, F. I. M., & McDowd, J. M. (1987). Age Differences in Recall and Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 474–479. https://doi.org/10.1037/0278-7393.13.3.474

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech. *Linguistics and Language Compass*. https://doi.org/10.1111/lnc3.12015

Croissant, Y. (2015). *Estimation of multinomial logit models in R: The mlogit Packages*. Retrieved from https://cran.r-project.org/package=mlogit

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264–1268. https://doi.org/10.1121/1.2946707

Danckert, S. L., & Craik, F. I. M. (2013). Does aging affect recall more than recognition memory? *Psychology and Aging, 28*(4), 902–909. https://doi.org/10.1037/a0033263

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*.

https://doi.org/10.3758/BF03214546

Denes, P. B., & Pinson, E. N. (1963). *The Speech Chain: The Physics and Biology of Spoken Language*. Retrieved from https://books.google.com/books?id=eJocAAAAMAAJ

DiDonato, R. M., & Surprenant, A. M. (2015). Relatively effortless listening promotes understanding and recall of medical instructions in older adults. *Frontiers in Psychology*, *6*(JUN), 1–20. https://doi.org/10.3389/fpsyg.2015.00778

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2010). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Dupoux, E., Hirose, Y., Kakehi, K., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568–1578. https://doi.org/10.1037/0096-1523.25.6.1568

*E-Prime 2.0 Psychology Software Tool*. (n.d.).

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*. https://doi.org/10.1037//0096-3445.128.3.309

Erber, J. T. (1974). Age differences in recognition memory. *Journals of Gerontology*, *29*(2), 177–181. https://doi.org/10.1093/geronj/29.2.177

Ferguson, S. H. (2012). Talker Differences in Clear and Conversational Speech: Vowel Intelligibility for Older Adults With Hearing Loss. *Journal of Speech, Language, and Hearing Research*, *55*(3), 779–790. https://doi.org/10.1044/1092-4388(2011/10-0342)

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*(1), 259–271. https://doi.org/10.1121/1.1482078

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251. https://doi.org/10.1121/1.2945161

Francis, A. L., Tigchelaar, L. J., Zhang, R., & Zekveld, A. A. (2018). Effects of second language proficiency and linguistic uncertainty on recognition of speech in native and nonnative competing speech. *Journal of Speech, Language, and Hearing Research*, *61*(7), 1815–1830. https://doi.org/10.1044/2018_JSLHR-H-17-0254

Franklin, M. S., Mooneyham, B. W., Baird, B., & Schooler, J. W. (2014). Thinking one thing, saying another: The behavioral correlates of mind-wandering while reading aloud. *Psychonomic Bulletin and Review*, *21*(1), 205–210. https://doi.org/10.3758/s13423-013-0468-2

Gagné, J.-P., Rochette, A.-J., & Charest, M. (2002). Auditory, visual and audiovisual clear speech. *Speech Communication*, *37*(3–4), 213–230. https://doi.org/10.1016/S0167-6393(01)00012-7

Gilbert, R. C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *The Journal of the Acoustical Society of America*, *135*(1), 389–399. https://doi.org/10.1121/1.4838975

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Glass, J. M. (2007). Visual Function and Cognitive Aging: Differential Role of Contrast Sensitivity in Verbal Versus Spatial Tasks. *Psychology and Aging*, *22*(2), 233–238. https://doi.org/10.1037/0882-7974.22.2.233

Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683. https://doi.org/10.1080/01690960500181332

Gollan, T. H., & Goldrick, M. (2019). Aging deficits in naturalistic speech production and monitoring revealed through reading aloud. *Psychology and Aging*, *34*(1), 25–42. https://doi.org/10.1037/pag0000296

Granlund, S., Baker, R., & Hazan, V. (2011). Acoustic-phonetic characteristics of clear speech in bilinguals. *17th International Congress of Phonetic Sciences (ICPhS XVII)*, (August), 763–766.

Greene, A. J., Easton, R. D., & LaShell, L. S. R. (2001). Visual-auditory events: Cross-modal perceptual priming and recognition memory. *Consciousness and Cognition*, *10*(3), 425–435. https://doi.org/10.1006/ccog.2001.0502

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*. https://doi.org/10.1037/0033-295X.102.3.594

Hale, J. B., Hoeppner, J. A. B., & Fiorello, C. A. (2002). Analyzing digit span components for assessment of attention processes. *Journal of Psychoeducational Assessment*, *20*(2), 128–143. https://doi.org/10.1177/073428290202000202

Hygge, S., Kjellberg, A., & Nöstl, A. (2015). Speech intelligibility and recall of first and second language words heard at different signal-to-noise ratios. *Frontiers in Psychology*, *6*, 1–7. https://doi.org/10.3389/fpsyg.2015.01390

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. https://doi.org/10.1016/S0010-0277(02)00198-1

Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-Language Activation Begins During Speech Planning and Extends Into Second Language Speech. *Language Learning*, *66*(2), 324–353. https://doi.org/10.1111/lang.12148

Johnson, E. K., Lahey, M., Ernestus, M., & Cutler, A. (2013). A multimodal corpus of speech to infant and adult listeners. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4828977

Johnson, K. (2004). Massive reduction in conversational American English. *Workshop on Spontaneous Speech: Data and Analysis*. https://doi.org/10.13005/bbra/2050

Keerstock, S., & Smiljanic, R. (2018). Effects of intelligibility on within- and cross-modal sentence recognition memory for native and non-native listeners. *The Journal of the Acoustical Society of America*, *144*(5), 2871–2881. https://doi.org/10.1121/1.5078589

Keerstock, S., & Smiljanic, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America*, *146*(6), 4604–4610. https://doi.org/10.1121/1.5141372

Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2018). The Impact of Age, Background Noise, Semantic Ambiguity, and Hearing Loss on Recognition Memory for Spoken Sentences. *Journal of Speech, Language, and Hearing Research*, *61*(3), 740–751. https://doi.org/10.1044/2017_jslhr-h-17-0077

Kondaurova, M. V., & 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. *The Journal of the Acoustical Society of America*, *124*(6), 3959–3971. https://doi.org/10.1121/1.2999341

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America*, *112*(5), 2165–2172. https://doi.org/10.1121/1.1509432

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming. *Journal of Memory and Language*, Vol. 33, pp. 149–174.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software*. https://doi.org/10.18637/jss.v082.i13

Lam, J., & Tjaden, K. (2013). Intelligibility of Clear Speech: Effect of Instruction. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1429–1440. https://doi.org/10.1044/1092-4388(2013/12-0335)

Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of Clear Speech: Effect of Instruction. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1807–

1821. https://doi.org/10.1044/1092-4388(2012/11-0154)

Lamont, A., Stewart-Williams, S., & Podd, J. (2005). Face recognition and aging: effects of target age and memory load. *Memory & Cognition*, *33*(6), 1017–1024.

Latorre-Postigo, J. M., Ros-Segura, L., Navarro-Bravo, B., Ricarte-Trives, J. J., Serrano-Selva, J. P., & López-Torres-Hidalgo, J. (2017). Older adults' memory for medical information, effect of number and mode of presentation: An experimental study. *Patient Education and Counseling*, *100*(1), 160–166. https://doi.org/10.1016/j.pec.2016.08.001

Lawrence, D. M., & Cobb, N. J. (1978). Cross-modal utilization of infor- mation: Recognition memory for environmental stimuli. *Perceptual and Motor Skills*, 1203–1206.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*. https://doi.org/10.1016/j.specom.2010.08.014

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. https://doi.org/10.1080/00031305.1980.10483031>.License

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling, NATO ASI Series* (pp. 403–439). https://doi.org/10.1007/978-94-009-2037-8_16

Ljung, R., Israelsson, K., & 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. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.2896

Lombard, E. (1911). Le Signe de l'Élévation de la Voix. *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, *37*, 101–119.

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, *42*(sup1), 49–58. https://doi.org/10.3109/14992020309074624

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The Production Effect: Delineation of a Phenomenon. *Journal of Experimental*

*Psychology: Learning Memory and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

MacWhinney, B. (n.d.). *Audio Digit Span*. Retrieved from http://step.talkbank.org/scripts-plus/

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686–693. https://doi.org/10.1044/jslhr.4003.686

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., & Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(1), 22–33. https://doi.org/10.1080/02724980443000151

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., … Bullock Mann, F. (2018). *The Condition of Education 2018 (Compendium. Congressionally mandated annual report NCES 2018-144)*. Washington, DC.

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, *161*, 95–112. https://doi.org/10.1016/j.jecp.2017.04.006

McGuire, L. C. (1996). Remembering what the doctor said: Organization and adults' memory for medical information. *Experimental Aging Research*, *22*(4), 403–428. https://doi.org/10.1080/03610739608254020

Meador, D., Flege, J. E., & Mackay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, *3*(1), 55–67. https://doi.org/10.1017/S1366728900000134

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Molesworth, B. R. C., Burgess, M., Gunnell, B., Löffler, D., & Venjakob, A. (2014). The effect on recognition memory of noise cancelling headphones in a noisy environment with native and nonnative speakers. *Noise and Health*, *16*(71), 240. https://doi.org/10.4103/1463-1741.137062

Mooneyham, B. W., & Schooler, J. W. (2013, March). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology*, Vol. 67, pp. 11–18. https://doi.org/10.1037/a0031569

Ng, E. H. N., Rudner, M., Lunner, T., Pedersen, M. S., & Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *International Journal of Audiology*, *52*(7), 433–441. https://doi.org/10.3109/14992027.2013.776181

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212–1228. https://doi.org/10.1162/0898929055002436

Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2012). Visual and auditory digit-span performance in native and non-native speakers. *International Journal of Bilingualism*, *18*(6), 663–673. https://doi.org/10.1177/1367006912466314

Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*(7), 717–727. https://doi.org/10.1080/09658211.2012.699070

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*, *95*(3), 1581–1592. https://doi.org/10.1121/1.408545

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing*, *39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Peng, Z. E., & Wang, L. M. (2019). Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during english speech perception. *Journal of Speech, Language, and Hearing Research*, *62*(4), 1068–1081. https://doi.org/10.1044/2018_JSLHR-H-17-0423

Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America*, *112*(4), 1627–1641. https://doi.org/10.1121/1.1506369

Picheny, M.A., Durlach, N.I., and B. L. J. (1985). Speaking Clearly for the Hard of Hearing I: Intelligibility Differences between Clear and Conversational Speech. *Journal of Speech and Hearing Research*, *28*(March), 96–103.

Pichora-Fuller, M. K. (2007). Audition and Cognition: What Audiologists Need to Know about Listening. In C. Palmer & R. Seewald (Eds.), *Hearing Care for Adults*. Stäfa, Switzerland: Phonak.

Pichora-Fuller, M. K., Goy, H., & Van Lieshout, P. (2010). Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments. *Seminars in Hearing*. https://doi.org/10.1055/s-0030-1252100

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., … Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of*

*America*, *97*(1), 593–608. https://doi.org/10.1121/1.412282

Pichora-Fuller, M. K., & Singh, G. (2006). Effects of Age on Auditory and Cognitive Processing: Implications for Hearing Aid Fitting and Audiologic Rehabilitation. *Trends in Amplification*, *10*(1), 29–59. https://doi.org/10.1177/108471380601000103

Pichora-Fuller, M. K., & Souza, P. E. (2003). Effects of aging on auditory processing of speech. *International Journal of Audiology*. https://doi.org/10.3109/14992020309074638

Pisoni, D. B., Kronenberger, W. G., Roman, A. S., & Geers, A. E. (2011). Measures of digit span and verbal rehearsal speed in deaf children after more than 10 years of cochlear implantation. *Ear and Hearing*. https://doi.org/10.1097/aud.0b013e3181ffd58e

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*. https://doi.org/10.1159/000090095

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2011150

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*(8), 904–915. https://doi.org/10.1080/09658211.2013.766754

Rabbitt, P. (1968). Channel-capacity, intelligibility and immediate memory. *The Quarterly Journal of Experimental Psychology*, *20*(3), 241–248. https://doi.org/10.1080/14640746808400158

Rabbitt, P. (1990). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Oto-Laryngologica*, *111*(sup476), 167–176. https://doi.org/10.3109/00016489109127274

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*. https://doi.org/10.1037/0033-295X.85.2.59

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review*. https://doi.org/10.1016/j.dr.2011.07.004

Rhodes, S., Greene, N. R., & Naveh-benjamin, M. (2019). Age-Related Differences in Recall and Recognition : A Meta Analysis. *PsyArXiv*, *January*, 14.

Rimikis, S., Smiljanic, R., & Calandruccio, L. (2013). Nonnative English speaker performance on the Basic English Lexicon (BEL) sentences. *Journal of Speech, Language, and Hearing Research : JSLHR*, *56*(3), 792–804. https://doi.org/10.1044/1092-4388(2012/12-0178)

Rogers, C. L., DeMasi, T. M., & Krause, J. C. (2010). Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers. *The Journal of the Acoustical Society of America*, *128*(1), 410–423. https://doi.org/10.1121/1.3436523

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*. https://doi.org/10.1017.S014271640606036X

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., … Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, *7*(July), 1–17. https://doi.org/10.3389/fnsys.2013.00031

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *International Journal of Audiology*, *47*(SUPPL. 2). https://doi.org/10.1080/14992020802301167

Rosenthal, E., Riccio, C., Gsanger, K., & Jarratt, K. (2006). Digit Span components as predictors of attention problems and executive functioning in children. *Archives of Clinical Neuropsychology*, *21*(2), 131–139. https://doi.org/10.1016/j.acn.2005.08.004

Rudner, M., Foo, C., Rönnberg, J., & 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. *Scandinavian Journal of Psychology*, *50*(5), 405–418. https://doi.org/10.1111/j.1467-9450.2009.00745.x

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295X.103.3.403

Sampaio, C., & Konopka, A. E. (2013). Memory for non-native language: The role of lexical processing in the retention of surface form. *Memory*, *21*(4), 537–544. https://doi.org/10.1080/09658211.2012.746371

Schneider, B. A. (2011). How age affects auditory-cognitive interactions in speech comprehension. *Audiology Research*, *1*(1S). https://doi.org/10.4081/audiores.2011.e10

Schum, D. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *Journal of the American Academy of Audiology*, *7*(3), 212–218.

Schweppe, J., Barth, S., Ketzer-Nöltge, A., & Rummer, R. (2015). Does verbatim sentence recall underestimate the language competence of near-native speakers? *Frontiers in Psychology*, *6*(FEB). https://doi.org/10.3389/fpsyg.2015.00063

Shafiro, V., & Sheft, S. (2017). Hearing Loss and Ethnicity in Age-related Cognitive Decline. *Hearing Journal*, *71*(1), 8–9. https://doi.org/10.1097/01.HJ.0000529842.91837.39

Shafiro, V., Sheft, S., & Risley, R. (2016). The intelligibility of interrupted and temporally altered speech: Effects of context, age, and hearing loss. *The Journal of the Acoustical Society of America*, *139*(1), 455–465. https://doi.org/10.1121/1.4939891

Smiljanic, R. (n.d.). Clear speech perception. In L. C. Nygaard, J. Pardo, D. B. Pisoni, & R. Remez (Eds.), *Handbook of Speech Perception*. Wiley Publishing.

Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *116*(4), 2627–2627. https://doi.org/10.1121/1.4785477

Smiljanic, R., & Bradlow, A. R. (2008a). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36*(1), 91–113. https://doi.org/10.1016/j.wocn.2007.02.002

Smiljanic, R., & Bradlow, A. R. (2008b). Temporal organization of English clear and conversational speech. *The Journal of the Acoustical Society of America*, *124*(5), 3171–3182. https://doi.org/10.1121/1.2990712

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Linguistics and Language Compass*. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Smiljanic, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: intelligibility and accentedness. *The Journal of the Acoustical Society of America*, *130*(6), 4020–4031. https://doi.org/10.1121/1.3652882

Smiljanic, R., & Gilbert, R. C. (2017). Intelligibility of Noise-Adapted and Clear Speech in Child, Young Adult, and Older Adult Talkers. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3069–3080. https://doi.org/10.1044/2017_jslhr-s-16-0165

Smiljanic, R., & Sladen, D. (2013). Acoustic and Semantic Enhancements for Children With Cochlear Implants. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1085–1096. https://doi.org/10.1044/1092-4388(2012/12-0097)

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of Measuring Recognition Memory: Applications to Dementia and Amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34

Song, J. Y. (2017). The use of ultrasound in the study of articulatory properties of vowels in clear speech. *Clinical Linguistics and Phonetics*, *31*(5), 351–374. https://doi.org/10.1080/02699206.2016.1268207

Souza, P. E., Arehart, K. H., Shen, J., Anderson, M., & Kates, J. M. (2015). Working memory and intelligibility of hearing-aid processed speech. *Frontiers in Psychology*, *6*, 1–14. https://doi.org/10.3389/fpsyg.2015.00526

Stepanov, A., Andreetta, S., Stateva, P., Zawiszewski, A., & Laka, I. (2019). Anomaly detection in processing of complex syntax by early L2 learners. *Second Language Research*. https://doi.org/10.1177/0267658319827065

Tang, L. Y. W., Hannah, B., Jongman, A., Sereno, J., Wang, Y., & Hamarneh, G. (2015). Examining visible articulatory features in clear and plain speech. *Speech Communication*, *75*, 1–13. https://doi.org/10.1016/j.specom.2015.09.008

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, *24*(3), 761–766. https://doi.org/10.1037/a0014802

Uchanski, R. M. (1988). *Spectral and Temporal Contributions to Speech Clarity for Hearing Impaired Listeners*. Cambridge, MA: Massachusetts Institute of Technology.

Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of Speech Perception* (pp. 207–235). Malden, MA: Blackwell Publishers.

Unsworth, N., & 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. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2–7. https://doi.org/10.1016/j.specom.2006.10.003

Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of Speech Clarity on Recognition Memory for Spoken Sentences. *PLoS ONE*, *7*(9), e43753. https://doi.org/10.1371/journal.pone.0043753

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00577

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of* *the Acoustical Society of America*, *84*(3), 917–928. https://doi.org/10.1121/1.396660

Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.3097773

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, *130*(3), 1606–1617. https://doi.org/10.1121/1.3621306

Wechsler, D. (1997). WAIS-III administration and scoring manual. In *The Psychological Corporation, San Antonio, TX*.

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*. https://doi.org/10.1098/rsos.150670

White, N., & Cunningham, W. R. (1982). What is the evidence for retrieval problems in the elderly? *Experimental Aging Research*, *8*(3), 169–171. https://doi.org/10.1080/03610738208260276

Whiting, W. L., & Smith, A. D. (1997). Differential age-related processing limitations in recall and recognition tasks. *Psychology and Aging*, *12*(2), 216–224. https://doi.org/10.1037/0882-7974.12.2.216

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is-and is not-a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory and Cognition*, *41*(2), 229–241. https://doi.org/10.3758/s13421-012-0255-8

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225–2234. https://doi.org/10.1121/1.4817926