Published in final edited form as:

Phonetica. 2014 ; 71(3): 201–226. doi:10.1159/000370160.

Individual Differences in Learning Talker Categories: The Role of Working Memory

Susannah V. Levi

Department of Communicative Sciences and Disorders, New York University, New York, N.Y., USA

Abstract

The current study explores the question of how an auditory category is learned by having school-age listeners learn to categorize speech not in terms of linguistic categories, but instead in terms of talker categories (i.e., who is talking). Findings from visual-category learning indicate that working memory skills affect learning, but the literature is equivocal: sometimes better working memory is advantageous, and sometimes not. The current study examined the role of different components of working memory to test which component skills benefit, and which hinder, learning talker categories. Results revealed that the short-term storage component positively predicted learning, but that the Central Executive and Episodic Buffer negatively predicted learning. As with visual categories, better working memory is not always an advantage.

Introduction

Accurately perceiving speech is a nontrivial task that requires listeners to parse highly variable acoustic signals into stable categories. The speech signal simultaneously carries linguistic information (what is being said) and also talker information (who is saying it). Recovering the linguistic content in the face of talker variation has been examined extensively in the speech perception literature, where it has been shown that listeners process speech in a talker-contingent manner (Allen and Miller, 2004; Eisner and McQueen, 2005; Jesse et al., 2007; Kraljic and Samuel, 2007; Ladefoged and Broadbent, 1957; Levi et al., 2011; McQueen et al., 2006; Norris et al., 2003; Nygaard et al., 1994; Theodore and Miller, 2010; Yonan and Sommers, 2000).

Accurately perceiving information about who is talking has received less attention, although several factors that contribute to talker processing have been found. Aspects of the talkers (i.e., the signal) affect how well listeners can process talker-voice information; listeners are worse at processing talker information for foreign-accented speech than for speech produced by a native speaker (Thompson, 1987). In addition to the signal, a variety of factors related to the language skills or language competence of the listener affect how accurately listeners can process talker-voice information. First, age of the listener affects performance. While it is true that processing talker information occurs very early in development, in fetuses

(Kisilevsky et al., 2003), newborns (DeCasper and Fifer, 1980), and infants in their first year of life (DeCasper and Prescott, 1984; Johnson et al., 2011; Purhonen et al., 2004, 2005), it has been shown that this ability improves with age. When processing familiar voices, Bartholomeus (1973) found that preschool children ages 4–5 could identify their classmates' voices above chance (approximately 60% accuracy with 17–19 voices to identify), although this was highly variable across children. Spence et al. (2002) also found that children of this age could recognize the voices of familiar cartoon characters. Preschool children are also able to learn the voices of unfamiliar talkers above chance (Creel and Jimenez, 2012; Moher et al., 2010). Mann et al. (1979) found improvement in a voice-matching task of unfamiliar talkers across ages 6–10, a small drop during early teens, and then an increase in performance for 14-year-olds through adults. In a discrimination task also using unfamiliar voices, Levi and Schwartz (2013) found that discrimination accuracy in the native language improves with age, where 7- to 9-year-olds performed worse than 10- to 12-year-olds, who in turn performed worse than adults. Taken together, these results show that talker processing improves with age, for both familiar talkers and for unfamiliar talkers.

Second, exposure to a language (native vs. second language vs. unknown language) affects performance in talker processing tasks. Research on the perception of talker information has found that adult listeners are more accurate at perceiving talker information in their native language than in an unfamiliar language (Goggin et al., 1991; Goldstein et al., 1981; Hollien et al., 1982; Köster and Schiller, 1997; Levi and Schwartz, 2013; Perrachione and Wong, 2007; Thompson, 1987; Winters et al., 2008), suggesting that the lack of experience with the unfamiliar language – when compared to many years of experience in the native language – is a detriment to performance. In contrast, listeners learning a second language quickly exhibit native-like talker processing skills (Schiller and Köster, 1996; Sullivan and Kügler, 2001; Sullivan and Schlichting, 2000), showing that as experience with and exposure to the second language increases, talker processing improves. Interestingly, it is not clear that children show the same native language advantage as adults. Recently Levi and Schwartz (2013) reported no differences in talker discrimination accuracy for a known versus unknown language in children 7–12 years old, whereas adults with the same stimuli performed worse in the unknown language.

Third, dyslexic adults, who have poorer phonological awareness skills, perform differently than typical adults. Perrachione et al. (2011) found that Phonological Awareness skills for adults with a history of dyslexia positively correlated with learning to categorize talkers. Taken together, these studies show that language skills affect how well listeners are able to process information about the talker.

The current study contributes to the literature on the development of talker-voice processing by examining how a wide age range of children learn to identify talkers' voices from auditory-only stimuli. In addition to examining the effect of age, we also test the contribution of individual differences in children's memory and language skills. These subject variables will be examined in a talker categorization task in which children learn to identify the voices of three unknown talkers over 5 days. This study will also explore whether the impact of working memory that has been found to influence visual category learning in adults (e.g., Ashby and Maddox, 2005; DeCaro et al., 2008) also contributes to

learning talker categories in children. The remainder of the introduction provides brief overviews of category learning and of one prominent model of working memory.

Learning Novel Categories

In visual perception, there is evidence for two types of category structures, *rule-based* (also called reflective) and *information-integration* (also called reflexive) (Ashby and Maddox, 2005; Maddox and Ashby, 2004). Various types of evidence have been used for this distinction, including how working memory resources impact learning of different types of categories, whether learning is procedural or based on hypothesis testing, and what role feedback plays during learning. Although there is evidence that two category structures exist, there remains debate over whether these two types of category structures actually rely on different cognitive systems (Ashby et al., 1998; Ashby and O'Brien, 2005; Smith et al., 1998) or whether a single system can account for the different types of behaviors that are found in these two types of learning (Nosofsky, 1991; Nosofsky et al., 1989; Nosofsky and Johansen, 2000). Regardless of whether different systems underlie the processing of these two types of category structures, it is clear that the structure of different categories leads to different behaviors, as discussed below.

Rule-based category learning involves learning to apply a verbalizable rule to categorize (or sort) objects into different categories. These categories tend to have relatively simple rules, such as: 'If blue, put into category A, otherwise put into category B.' Rule-based category learning is assumed to involve explicit hypothesis testing and thus relies on working memory and other related cognitive skills (Ashby and Maddox, 2005; DeCaro et al., 2008, 2009; Grossman et al., 2002; Maddox and Ashby, 2004; Maddox et al., 2013; Miles and Minda, 2011; Waldron and Ashby, 2001; Zeithamova and Maddox, 2006), with better working memory skills leading to better learning. Information-integration category learning, in contrast, cannot be accomplished through a verbalizable rule, and instead relies on integrating information across a variety of dimensions prior to making a category decision (Ashby and Gott, 1988; Ashby and Maddox, 2005; Maddox and Ashby, 2004). This is typically considered to be a type of procedural learning (Ashby et al., 2003; Maddox et al., 2004). Whereas some studies have suggested that better working memory skills actually hinder information-integration category learning (DeCaro et al., 2008, 2009), others have disputed this conclusion (Craig and Lewandowsky, 2012; Lewandowsky et al., 2012; Tharp and Pickering, 2009). Unfortunately, these studies examining the role of working memory have used several measures of working memory combined into a single working memory score per participant, making it unclear which components of working memory affect category learning.

Most research on category learning and category structure has been in the visual domain. A few recent studies from the category learning literature have explored whether these two types of category structures exist in the auditory domain (Maddox et al., 2006, 2013) and which type of category structure speech stimuli belong to (Chandrasekaran et al., 2014; Maddox and Chandrasekaran, in press). Categories that are extracted from speech stimuli require information integration, as speech sound categories vary along multiple parameters, and listeners must integrate multiple aspects of the acoustic signal to correctly categorize the

signal. The famous lack of invariance problem (Klatt, 1979; Liberman et al., 1967) points to speech sound categories as requiring information-integration: cues to speech sound categories vary across phonetic environments (e.g., vowel length before voiced vs. voiceless consonants or mono- vs. polysyllabic words), within talkers (e.g., changes in speaking rate, register, health, emotional state), and across talkers (e.g., age, sex, dialect, idiolect). Thus, the acoustic cues for any particular speech sound category vary, requiring listeners to take many cues into account and be flexible enough to reweight the importance of different cues in different situations. The two studies that have examined optimal strategies for learning linguistic categories from the speech signal have found that speech categories are information-integration categories (Chandrasekaran et al., 2014; Maddox and Chandrasekaran, in press).

Similarly, categorizing the speech signal into talker categories also involves information-integration. The sounds of two words can be completely disjoint (e.g., *cat* and *mouse*), but listeners are still able to perceive similarities and differences across talkers (Levi and Schwartz, 2013; Wester, 2012; Winters et al., 2008). In order to correctly categorize a speech signal as having been produced by a particular talker, listeners must take information from the signal that pertains to anatomy (e.g., larger or smaller vocal tract, which determines resonance), to vocal fold vibration which is affected both by anatomy and by social factors (Yuasa, 2010), to accentedness (either regional or non-native), and to idiolectal differences. Additional evidence that talker categories involve information integration comes from studies of verbal overshadowing. In these studies (e.g., Mitchell and MacDonald, 2012; Vanags et al., 2005), listeners hear a voice which they later have to identify in a voice lineup. Some listeners are asked to describe the voice before identification, while others perform a filler task. Listeners who tried to verbalize the description of the voice were less accurate at identifying it, suggesting that generating verbalizable characteristics, or possible rules for identifying the voice, impaired talker-voice processing.

Using the term *categorization* to characterize talker identification from auditory information may seem odd, given that talkers are (usually) treated as distinct individuals rather than clusters of examples of a broader category (as penguins, cardinals, and parrots are all considered members of the category of birds). It is true that talkers can be categorized on this more coarse-grained level, such as female versus male talkers, children versus adults, or talkers with one regional accent versus another. In this way, these talker categories are truly examples of clusters of individual talkers. However, speech stimuli can be categorized on more fine-grained levels as well. The perception of a particular sound as one category versus another (e.g., /p/ vs. /b/) is more analogous to the perception of individual talkers. Within the speech perception literature, listeners are asked to categorize specific instances of speech sounds, such as categorizing a word as beginning with /p/ or /b/. The particular instances of /p/ will vary based on many factors such as speaker, speaking rate, phonetic environment, stress placement, and number of syllables in the target word, and yet listeners will categorize all the stimuli as beginning with /p/. In an analogous way, listeners hear many different words produced by a particular talker, but with sufficient experience with the talker's voice, listeners can accurately categorize the stimuli as having been produced by a particular talker. Thus, when a listener encounters a word and is asked to identify who produced it, he/she is categorizing the stimulus in terms of the talker.

Given that categories in speech (in the case of the current study, talker categories) involve information integration, it is likely that they will be sensitive to the same factors that affect categorization of visual information-integration categories. To promote optimal learning, learners received traditional feedback provided shortly after participants made their response (Ashby et al., 2002; Maddox et al., 2003) and also had a consistent mapping between a talker's voice and a response area on the computer screen (Ashby et al., 2003; Maddox et al., 2004). To test whether novel talker categories in speech are sensitive to the same factors affecting visual category learning, we collected a variety of measures of working memory to explore which components of working memory help or hinder learning of talker categories. The current study also explores whether these working memory components that affect adult category learning are also relevant for children.

Working Memory

Working memory involves the short-term storage or maintenance of incoming stimuli as well as the manipulation of information. One of the most influential models of working memory is the multi-component model proposed by Baddeley (1986, 2000, 2003, 2007, 2010), Baddeley and Hitch (1974), and Repovš and Baddeley (2006), which breaks up the construct of working memory into four components: Phonological Loop, Visuospatial Sketchpad, Central Executive, and Episodic Buffer (fig. 1). These components are distinct from long-term memory, but interface with it through the Episodic Buffer. An additional component, Phonological Awareness, has been found to affect children's academic success and will be discussed below.

The Phonological Loop is responsible for short-term storage of auditory information as well as verbal rehearsal. The short-term memory trace of auditory information decays within a few seconds unless refreshed by verbal or subvocal rehearsal. Studies of children with and without language impairments have found that the Phonological Loop strongly influences language learning (Baddeley et al., 1998). The Visuospatial Sketchpad forms the analogue for visually presented stimuli. The Central Executive is responsible for manipulation of information and attention and also controls the Phonological Loop and the Visuospatial Sketchpad. In newer versions of the model, the Central Executive does not include any storage capacity. Instead, the Episodic Buffer functions as a multimodal limited-capacity store, binding information from both visual and auditory domains into episodes and interfacing with long-term memory (Baddeley, 2000; Rudner and Rönnerberg, 2008). Thus, the Episodic Buffer is considered to have both a storage function and a processing function.

In a study examining the organization of these components of working memory and how they relate to children's academic success, Alloway et al. (2004) also found a unique contribution of Phonological Awareness. Phonological Awareness is a person's knowledge of the sound structure of language and is often tested by asking participants to distinguish, identify, or manipulate the sounds of words. Common types of tasks that are used to assess Phonological Awareness include rhyming tasks, deleting segments or parts of words (e.g., 'say [mit] *meat* without saying [m]'), and combining or *blending* sounds to produce a word (e.g., 'What word do these sounds make: [s]-[i]-[n]', *seen*) (Perfetti et al., 1987; Stahl and Murray, 1994; Stanovich et al., 1984). Alloway et al. (2004) found that the best-fitting model

for academic success included unique factors that were tied to the Central Executive, the Phonological Loop, and the Episodic Buffer, but also two additional factors, Phonological Awareness and nonverbal ability.

This particular model of working memory is used in the current study because it clearly delineates different components of working memory, because several tasks have been commonly used to assess the function of each of these components, and because it has been developed in part from evidence of children's data.

Current Study

The current study integrates research on talker processing, category learning, and working memory to explore what individual differences and other stimulus factors affect how listeners learn talker categories. The current study contributes to the talker processing literature by examining whether working memory plays a role and also by examining a wide age range of children. It also contributes to the category learning literature in general by testing whether the role of working memory that was found for visual category learning holds for auditory category learning and also which component(s) of working memory affect learning. It also contributes to the category learning literature by examining whether aspects found to affect performance in adults also affect performance in children. Here, we test the claim that information-integration category learning does not rely on high levels of an individual's working memory skills and in particular, we test whether better working memory skills actually hinder information-integration category learning, as was suggested by DeCaro et al. (2008). The previous studies that have tested the influence of a learner's working memory ability have used aggregate measures of working memory that do not differentiate which component of working memory is helpful (or harmful) when learning novel categories. To more specifically ask the question of which component(s) of working memory are important during category learning, the current study includes a measure of each of the relevant components of working memory (Phonological Loop, Visuospatial Sketchpad, Central Executive, and the Episodic Buffer) to test which components influence performance on an information-integration category learning task. In addition, the current study includes a measure of lexicon size and Phonological Awareness.

In the current study, children learned to identify the voices of three unfamiliar talkers ('Julia', 'Lisa', 'Anne') over 5 days of training. As previous studies have found that processing talker information in the speech signal is sensitive to age (Levi and Schwartz, 2013; Mann et al., 1979), we expect age to be a significant predictor of performance, with older children performing better than younger children. If categorizing talkers is similar to other information-integration categories, we also expect that the components of working memory that process auditory information or multimodal information will also predict performance, namely the Phonological Loop, the Central Executive, and the Episodic Buffer. In particular, it is expected that higher working memory function will negatively affect performance, as was found by DeCaro et al. (2008). In addition, a study with adult learners found that Phonological Awareness positively correlated with learning to categorize talkers (Perrachione et al., 2011), thus Phonological Awareness is expected to positively affect performance.

Methods

Participants

Forty-seven children ages 6;8–11;10 (mean: 9;3) participated in the study. Language and nonverbal cognition was assessed in all children using the Clinical Evaluation of Language Fundamentals-4 (CELF; Semel et al., 2003) and the Test of Nonverbal Intelligence-3 (TONI; Brown et al., 1997). Both tests are normed to 100 with a standard deviation (SD) of 15. To be included in the study, children had to score 85 or above on the Core Language composite of the CELF, which assesses both expressive and receptive language skills. For children ages 6–8, the Core Language composite consists of the subtests Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences. For children ages 9–11, the composite consists of the subtests Concepts and Following Directions, Recalling Sentences, Formulated Sentences, and Word Class. The mean CELF-Core Language score for the children was 103 (SD = 10.8), with a range of 85–126. The mean score on the TONI was 108 (SD = 16.5), with a range of 85–146. Thirty-two of the children passed a pure-tone hearing screening at 25 dB HL at 1,000, 2,000, and 4,000 Hz either with a portable Earscan3 Screening Audiometer (ES3S) in their school or home or with a GSI 61 Clinical Audiometer in a sound-attenuated IAC Booth in the Department of Communicative Sciences and Disorders at New York University. Of the remaining 15 children, 1 child only passed 4,000 Hz at 30 dB in the left ear and 14 children did not complete a hearing screening because the portable audiometer was not available at the time of testing. Of these 14, the parent questionnaire indicated that 5 passed a hearing screening at a doctor's office within 1 year prior to participating in the study and the remaining 9 had not experienced frequent ear infections and the parent report did not indicate any known hearing problems.

Stimuli

Three female bilingual German-L1/English-L2 talkers were recorded producing 360 monosyllabic CVC words in both English and German, but only the English words were used in the current study. Recordings were made in a sound-attenuated IAC booth at the Speech Research Laboratory at Indiana University using a SHURE SM98 head-mounted unidirectional (cardioid) condenser microphone with a flat frequency response from 40 to 20,000 Hz. Productions were digitized into 16-bit stereo recordings via Tucker-Davis Technologies System II hardware at 22,050 Hz and saved directly to a PC. Talkers read each word as it was presented to them on a computer monitor in random order, blocked by language. All sound files were normalized to have a uniform RMS amplitude. The 3 bilingual talkers used in the current study had similar intelligibility (table 1) and were selected from a larger group of bilinguals (Levi et al., 2011). Additionally, these 3 talkers were selected to have relatively different average fundamental frequency (F0) across productions. In addition to average F0, table 1 also provides information about the average word duration for these 3 speakers.

The 360 English words were rated independently by 3 individuals (2 speech-language pathologists and the author) as likely to be known to children ages 6–9 ('high lexical familiarity'), likely to be unknown to these children ('low lexical familiarity'), or possibly known by these children. Out of these words, 133 were rated as high familiarity and 115

were rated as low familiarity by all 3 raters. These lexical familiarity ratings were compared with age of acquisition ratings from Cortese and Khanna (2008). For their ratings, lower numbers indicate that the words were acquired earlier. The mean age of acquisition for the high familiarity words was 3.01 (acquired between ages 4 and 6; range: 1.9–4.9; interquartile range: 2.65–3.35) and for the low familiarity words was 5.34 (acquired after ages 8–10; range: 3.5–6.6; interquartile range: 4.88–5.98). An independent-samples *t* test revealed a significant difference in age of acquisition ratings for the high versus the low familiarity sets [$t(226) = -29.13, p < 0.001$]. Only 3 of the high familiarity words were not found in the database of Cortese and Khanna (2008) (*house, mouse, pill*). Seventeen words from the low familiarity list were not found in the database; 15/17 have a frequency of 1 (out of one million) based on Kuera and Francis (1967) and 2 (*couth, goon*) are not included. This complete list of English words can be found in the Appendix.

Children were assigned to one of eight groups, representing a different random sampling of words. For each group, a distinct set of 52 high and 52 low familiarity words were selected for use in the talker learning task.

Procedure: Talker Learning

Stimuli were presented binaurally over Sennheiser HD-280 circumaural headphones using a Panasonic Toughbook CF-52 laptop running Windows XP and equipped with a touch screen. The experiments were created with E-Prime 2.0 Professional (Schneider et al., 2007). Children were tested in a quiet room either at school, in their home, or in the Department of Communicative Sciences and Disorders at New York University.

Children completed 5 days of talker training in which they learned to identify the voices of 3 unfamiliar talkers, presented as cartoon-like characters on a computer screen (fig. 2). Each day of training consisted of two training sessions with feedback and one test session without feedback. Children were instructed that they would hear a single word and have to decide which of three characters produced the word by tapping the screen. During the training sessions, children first had a familiarization phase in which they heard the same 4 words (2 high familiarity, 2 low familiarity) produced by each of the 3 talkers twice. Only the image of the actual character/talker appeared on the screen during the familiarization phase. After familiarization, children heard the same 5 high familiarity words and 5 low familiarity words produced by each of the 3 talkers for a total of 30 trials, which were split into three blocks to allow children to take breaks. On each trial, children heard a word and were asked to select which character had spoken the word. After their response, children received two forms of feedback: first they were shown a smiley or frowny face to indicate their accuracy and then they heard the word again while the image of the correct character/talker appeared on the screen. A sample of a feedback trial is presented in figure 3. The exact same training session was completed twice each day to provide children with additional exposure to the talkers (referred to as session 1 and session 2). Following the two training sessions with feedback, children then completed a test session with a distinct set of 5 high familiarity and 5 low familiarity words produced by all 3 talkers. Test sessions had the same format as the training sessions except no feedback was provided. Each child heard 52 high familiarity and 52 low familiarity words during the 5 days of training (e.g., 5 high familiarity words during

feedback trials + 5 high familiarity words on no-feedback trials $\times$ 5 days of training, plus 2 high familiarity words during familiarization). This is outlined in table 2. A critical component of the design of the current study is that a different set of lexical items was used on each day. Thus, these tasks represent true generalization.

Children were scheduled to complete the talker learning so that no more than 4 days intervened between consecutive sessions (mean: 3 days; SD: 1.65 days¹; median: 2 days; interquartile range: 1–3 days). Given the scheduling challenges with children, several children completed sessions that were further apart. Twelve children had 1–2 sessions that were up to 7 days apart. Three additional children had up to 11 days intervening between scheduled sessions. Four children had 2 days that were greater than 2 weeks apart. Despite this long delay, these 4 children showed improvement in talker identification accuracy across the delay. Thus, all children were included in the data analysis.

Procedure: Standardized Testing

Children completed several standardized tests designed to tap various components of working memory. These tests generate a raw score that is translated into a standard score that normalizes performance across age. Because age is expected to significantly predict performance on the task, statistical analyses were conducted using the standard scores to see if for a particular child's age, higher or lower ability in a particular domain (e.g., Phonological Loop or Central Executive) affects performance. Information about performance on these standardized tests for the participants is provided in table 3.

Forward Digit Span—Forward Digit Span was used to assess the Phonological Loop component of Baddeley's model of working memory, as it involves short-term storage of verbal information (Alloway et al., 2004; Pickering and Gathercole, 2001). In this task, children hear lists of numbers and have to repeat them in the same order as in the original list. Two lists are completed at each length (e.g., 2 digits long, 3 digits long, etc.) until children responded incorrectly on both lists at a particular length. The actual lists that were used were taken from the CELF-4, and thus each child received both a raw score and a standard score based on their age. Standard scores on this task are normed to 10 and have an SD of 3. The lists were recorded by a speaker of General American English and presented binaurally over Sennheiser HD-280 circumaural headphones. This ensured that each child was presented with the exact same stimuli.

Backward Digit Span—Backward Digit Span was used to assess the Central Executive component of Baddeley's working memory model, as it requires manipulating information (Alloway et al., 2004; Pickering and Gathercole, 2001). In this task, children hear lists of numbers and have to repeat the lists in the reverse order. As with the Forward Digit Span, children completed two lists at each length until they responded incorrectly to both lists at a particular length. Also like the Forward Digit Span, the actual lists used for the backward span were taken from the CELF-4, and were recorded by the same General American

¹This is the SD without the four outliers that were greater than 14 days (2 weeks) apart.

speaker to ensure a uniform presentation for all children. Standard scores on this task are normed to 10 with an SD of 3.

Block Span—The Block Span used in the current study was taken from the Children's Working Memory Test Battery (Pickering and Gathercole, 2001) and resembles the Corsi block span. Nine gray blocks (cubes) are attached to a board in a semirandom order so as to avoid any ability to verbally encode the location of the blocks. The experimenter points to the blocks on a board and the child is asked to point to the same blocks in the same order. Thus, the Block Span task is the nonverbal analogue to the Forward Digit Span and thus taps into the visuospatial sketchpad component of Baddeley's working memory model (Pickering and Gathercole, 2001). As the Block Span taps children's short-term memory for visual information, it was included as a control factor and was not expected to influence children's auditory category learning for the unfamiliar voices. Standard scores on this task are normed to 100 and have an SD of 15.

Recalling Sentences—The Recalling Sentences subtest of the CELF-4 resembles other sentence memory tasks that have been used to assess the role of the Episodic Buffer component of Baddeley's working memory model (Alloway and Gathercole, 2005; Alloway et al., 2004). In this subtest, the experimenter speaks a sentence out loud and the child repeats it. This is considered a test of the Episodic Buffer because it taps syntactic knowledge, which helps people recall sentences that are longer than the set of digits or words that people can recall. The CELF scoring procedure gives children 0–3 points based on the number of errors. Standard scores on this task are normed to 10 and have an SD of 3.

Comprehensive Test of Phonological Processing—The two Phonological Awareness subtests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 1999) were examined as Alloway et al. (2004) found that Phonological Awareness is an additional factor separate from the other components of working memory that contributes to academic success and also because Phonological Awareness was found to contribute to talker category learning in adults in Perrachione et al. (2011). The two subtests that make up the Phonological Awareness score of the CTOPP are Blending and Elision. In Blending, children hear prerecorded sounds which they have to concatenate to generate a word (e.g., [s]-[n] *sun*). In Elision, children have to produce a word after taking out a sound or sounds (e.g., 'say *tiger* without saying [g]'). The Phonological Awareness scores are normed to 100 and have an SD of 15.

Peabody Picture Vocabulary Test—The Peabody Picture Vocabulary Test (PPVT; Dunn and Dunn, 2007) tests children's receptive vocabulary and was used in the current study as a measure of children's lexicon size. In this test, children hear a word spoken by the experimenter and have to point to one of four pictures that matches the word. PPVT scores are normed to 100 and have an SD of 15.

Results

Training versus Test Condition

A repeated-measures ANOVA was conducted for the Training sessions and Test sessions with Day of Training (days 1–5) and Condition (Training, Test) as within-subjects factors (fig. 4). For this analysis, the average accuracy for each day was computed for each participant and served as the dependent variable. As expected, a main effect of Day of Training was found, indicating that children's accuracy improved across days of training [$F(4, 460) = 19.31, p < 0.001$]. In addition, there was a main effect of Condition, where children performed significantly worse on the Test conditions, where no feedback was provided [$F(1, 460) = 22.04, p < 0.001$]. The interaction was not significant ($p = 0.79$).

The histogram in figure 5 provides percent correct accuracy on day 5 for both the training condition (with feedback) and the test condition (no feedback). Forty-three out of 47 children reached above 60% accuracy on the test condition of the final day of learning, well above chance performance of 33.33%. Three children were below 60% for the test condition, but were above 60% (66%, 71%, 86%) for the training condition on the same day, suggesting that they did learn the talker categories. Only 1 child was below 60% on day 5 for both the training and test conditions. A binomial test confirmed that this child performed statistically above chance on 4 of the 5 days in the training condition and on 2 of the 5 days on the test condition (all $p < 0.05$), suggesting that this child also exhibited learning.

Standardized Test Measures

Correlations were conducted on the standardized test scores that were entered into the logit mixed-effects model (see next section) to confirm that these measures were not highly correlated. These correlations are provided in table 4.

Category Learning

Given the difference in performance for Training versus Test noted above, separate analyses were conducted on children's performance on the Training conditions and on the Test conditions. To assess what components of working memory contribute to learning talker categories, a logit mixed-effects model (Baayen, 2008; Baayen et al., 2008; Jaeger, 2008) was fit to the raw data from the Training condition using the *glmer()* function with the binomial response family (Bates et al., 2010) in R (<http://www.r-project.org/>). The built-in p value calculations from the *glmer()* function were used. Because accuracy on this task is binary (correct or incorrect), a logit mixed-effect model can be used which directly generates z-scores and p values. A model was fit to the Training data with fixed effects for Day of Training (days 1–5), Age (in months), Lexical Familiarity (high, low), Session (1, 2), and Talker (F3, F8, F11) and standard scores for Forward Digit Span (Phonological Loop), Backward Digit Span (Central Executive), Block Span (Visuospatial Sketchpad), Sentence Recall (Episodic Buffer), CTOPP (Phonological Awareness), and PPVT (Lexicon Size), along with random intercepts by subjects with random slopes for Talker since children may differ in terms of which talkers are easier to identify and may also differ in their learning trajectories for the different talkers. The predictor variables for the working memory and language measures were mean-centered, which aids in interpretation and model fit (Gelman

and Hill, 2007). A similar logit mixed-effect model was fit to the Test data, but without Session, as only a single Test session was run each day.

As expected, both models revealed significant effects for Age and Day of Training, showing that older children performed better than younger children and that children's accuracy improved across the 5 days of training (table 5). The effect of Day is illustrated in figure 4 above. There was also a significant effect of Talker [average accuracy across days for Training: F3 (65.8%) < F11 (71.5%) < F8 (86.0%); average accuracy across days for Test: F3 (60.2%) < F11 (64.5%) < F8 (80.5%)]. In addition, the models revealed significant positive effects for the Phonological Loop (Forward Digit Span) and for Lexicon Size (PPVT). These positive effects indicate that children with a better Phonological Loop and with a larger vocabulary have an advantage when learning complex, novel auditory categories. The Episodic Buffer (Recalling Sentences), however, exhibited a negative influence on categorization, indicating that a better performing Episodic Buffer actually results in worse performance on this task. Despite the positive univariate correlations (table 4) between the Episodic Buffer (Recalling Sentences) and the Phonological Loop (Forward Digit Span) and the Lexicon (PPVT), and the fact that all univariate correlations with the outcome variable are positive, when all of these variables are included in the model the effect of the Episodic Buffer on the outcome is negative. This is consistent with a causal model where predictors are correlated (Cohen et al., 2003, fig. 12.1.2). Because all predictors are correlated, the effects of each predictor on the outcome are difficult to assess in the absence of the others because the negative effect of one variable (e.g., the Episodic Buffer) might be masked by the fact that a score for the Episodic Buffer is associated with a high score for the Phonological Loop. This can give rise to an apparent positive effect despite an underlying negative effect. To test the consistency of the data with such a causal model, additional models were created without the Phonological Loop and without the measure of lexicon size. The negative affect of the Episodic Buffer remained in the absence of some of these positively correlated predictors. Only when *all* other memory variables were ignored (i.e., removed), leaving a model with only Age, Day, Talker, frequency, and session) did the estimated effect of Episodic Buffer become positive. This is consistent with Simpson's paradox (or the reversal paradox) (Arah, 2008; Blyth, 1972; Pearl, 2014), where results should be interpreted in the context of other variables. In other words, the variables need to be included (i.e., controlled for) in the model rather than ignored (i.e., omitted).

Because the predictor variables show moderate correlations with each other (table 4), a set of checks was used to confirm that colinearity was not a problem for the estimation of effects in the models. Variance inflation factors were checked for both models. All variance inflation factors were less than 2, well under the conventional cutoff of 10, above which serious colinearity is involved (Cohen et al., 2003).² An additional check of colinearity was done using the condition number (kappa), where values over 30 represent serious issues of colinearity (Cohen et al., 2003). The kappa values for these two linear models were also well under this threshold (Training: 11.87; Test: 12.12).

²Even with a more conservative boundary at 2.5 (<http://hlplab.wordpress.com/2011/02/24/diagnosing-collinearity-in-lme4/>) the values here are still below this cutoff.

The remaining predictors were not significant. The lack of a significant effect of Lexical Familiarity indicates that children were similarly accurate when categorizing talkers for high and low familiarity words. A similar lack of effect for Lexical Familiarity was found in a talker discrimination task for both adults and school-age children (Levi and Schwartz, 2013). As expected, the Visuospatial Sketchpad (Block Span), which taps visuospatial short-term memory was not significant. The change in significance of Session between day 1 (significant) and day 5 (not significant) is similar to results found in a talker categorization task with adults (Winters et al., 2008). The lack of an effect of Session for the Training data is consistent with data from adults, where listeners improved across the first and second presentation of stimuli for day 1, but only showed consolidation of learning after sleep with improvement across days, but not within days. The full set of results can be found in table 5.

The results above show which predictors affect performance during the actual learning phase across the 5 days. In addition, we were interested in examining which factors affect performance on day 1 when children are first exposed to the talker categories and also on day 5 once the categories have been learned and the children have become experienced listeners. Logit mixed-effects models were fit to the data from only day 1 and from only day 5 for both Training and Test. The full results can be found in tables 6 and 7.

Similar to the previous set of results, the model for day 1-Training revealed significant positive effects for Age and Session, where children performed significantly better the second time through the training condition. The significant effect of Session indicates rapid learning on day 1, after only 30 training trials. There were also significant effects for the Talker comparisons [F_3 (51.8%) < F_{11} (58.1%) < F_8 (77.0%)]. The model also revealed a significant positive effect for the Phonological Loop (Forward Digit Span), along with a negative effect for the Episodic Buffer (Recalling Sentences). The effect of the Lexicon (PPVT) did not reach significance on day 1 Training.

The model for day 1-Test revealed similar results, including positive effects for Age, Lexicon (PPVT), and the Phonological Loop (Forward Digit Span), along with a negative effect for the Episodic Buffer (Recalling Sentences). The effect of Talker also reached significance for two of the three comparisons [F_3 (50.4%), F_{11} (56.6%) < F_8 (71.0%)]. As with the previous models, the tests of colinearity did not reveal serious problems.

Day 5 represents performance by experienced listeners. Both the Training and Test models for day 5 revealed significant positive effects for Age, Lexicon Size (PPVT), and the Phonological Loop (Forward Digit Span). In addition, they revealed negative effects of the Central Executive (Backward Digit Span). The Episodic Buffer negatively affected performance in the Training conditions, but this effect was reduced in the Test condition ($p = 0.08$). There were also significant effects for most of the talker comparisons [Training: F_3 (75.4%) < F_{11} (79.3%) < F_8 (87.7%); Test: F_3 (68.9%), F_{11} (71.7%) < F_8 (83.0%)]. The effect of Session for the training condition disappeared for day 5 where there was no difference in performance for the two Training sessions. As with the previous models, the tests of colinearity did not reveal serious problems.

As mentioned above, the change in significance of Session between day 1 (significant) and day 5 (not significant) for the Training data is similar to results found in a talker categorization task with adults (Winters et al., 2008) where adults only improve across sessions on day 1 and further improvements only occur across days of training.

Discussion

Taken together, the results reveal several consistent effects of the subject variables and also the experimental variables (e.g., Day). First, the results show a clear effect of age, where older children are better at learning novel talker categories than younger children. This is consistent with previous studies of the development of talker perception (Levi and Schwartz, 2013; Mann et al., 1979). There are many possible contributing factors that could explain the effect of age, such as an increase in language competence across development (as discussed in the Introduction). Another possible contributor is that older children are likely to have been exposed to more talkers and likely to have learned the voices of more talkers than younger children, which would have a facilitatory effect on learning an unfamiliar talker.

A second finding to emerge is that children who have larger lexicons *for their age* also have an advantage. The literature on word learning in children has found that children first process words holistically and then progressively change to lexical representations that are more phonetically specified (Jusczyk, 1986, 1992; Walley, 1993), and that this change in lexical representations extends into the school-age years (Metsala, 1997). Further, studies have found that toddlers are more sensitive to acoustic-phonetic detail in highly familiar words (Fennell and Werker, 2003; Stager and Werker, 1997; Swingley and Aslin, 2000; White and Morgan, 2008). Thus, children who have larger vocabularies for their age are likely to have more phonetically specified lexical representations and thus are better at pulling out that acoustic-phonetic information that is linked to an individual talker. This benefit of a larger lexicon suggests a rich-get-richer effect: a larger lexicon results in better categorization skills in terms of talker categories. That some children are better at learning the voices of previously unfamiliar talkers is also relevant for spoken language processing in general. Numerous studies have found that familiarity with a talker's voice improves spoken language processing in both children (Levi, 2014) and adults (Levi et al., 2011; Magnuson et al., 1995; Nygaard and Pisoni, 1998; Nygaard et al., 1994).

The results of the current study also shed light on the role of various components of working memory in category learning. Previous results from visual information-integration category learning have been equivocal: better working memory skills are sometimes advantageous (Craig and Lewandowsky, 2012; Lewandowsky et al., 2012) and sometimes not (DeCaro et al., 2008, 2009). Because these contradictory results may stem in part from the fact that a single working memory score was used per participant (Craig and Lewandowsky, 2012; Lewandowsky et al., 2012), we separated working memory into various components along the lines of Baddeley's model of working memory (Baddeley 1986, 2007; Baddeley & Hitch, 1974). The results of the current study suggest that the Phonological Loop enhances talker category learning, while the Episodic Buffer and the Central Executive seem to hinder learning of novel talker categories. The Phonological Loop is involved in short-term storage of auditory information and allows a learner to retain and rehearse a stimulus. Thus, it is

perhaps not surprising that it is advantageous to be able to retain information while comparing an incoming stimulus with other stored exemplars in order to arrive at a response.

In contrast to the storage component, the two components of working memory that involve some type of analysis of the stimulus have a negative influence on performance. Recall that the Central Executive is responsible for manipulating information and attentional control. The Episodic Buffer binds information for various sources and interfaces with long-term memory. That a better Central Executive and Episodic Buffer for a child's age actually hinder learning information-integration category structures may indicate that these individuals attempt to come up with clear analytical or verbalizable methods for categorizing the stimuli, thus favoring or biasing the learner towards a rule-based type of category structure as suggested by DeCaro et al. (2008). On the surface, this finding may seem at odds with the positive effect of age on talker category learning. While it is expected and normal for raw skills (scores) of the Central Executive and Episodic Buffer to improve with age, many other additional skills (e.g., perceptual, general language competence, vocabulary, phonological loop, additional exposure to different talkers) also improve as a child develops. The finding here for the negative effect of the Central Executive and Episodic Buffer suggests that for a child who has high Central Executive and Episodic Buffer skills *for their age* may be at a disadvantage for learning talker categories and possibly other types of information-integration categories.

Thus, our results show support for both the positive and negative roles of working memory during category learning. In line with DeCaro et al. (2008) we found that less is more when considering the analytical side of working memory and in line with Craig and Lewandowsky (2012) and Lewandowsky et al. (2012), we found that more is more when considering the auditory storage component of working memory. Critically, the current study shows that working memory is not neutral for learning auditory categories and thus is similar to learning visual categories.

In addition to uncovering the role of working memory in talker category learning, the working memory findings may provide new avenues of investigation for exploring how other auditory categories are learned in speech. What is the implication of these findings about the role of working memory during category learning for the acquisition of native phonological categories? If the interpretation of the negative impact of the Central Executive and Episodic Buffer is correct – namely that individuals with better performance of these components are biased towards trying to find a verbalizable rule and explicitly testing hypotheses about category membership – then it could explain in part why infants, who cannot generate and explicitly test verbalizable rules, are able to learn complex, highly variable phoneme categories so well and tune their perceptual systems within the first year of life (Best, 1993; Kuhl et al., 1992; Werker and Tees, 1984). That is, infants are unable to verbalize a rule for category membership and thus may actually have an advantage for learning information-integration category structures. We leave this issue to be sorted out by future research.

Research on how adults learn and perceive second language speech sounds has focused largely on the similarities and differences of the phoneme inventories and the language-specific phonetics of the L1 and L2 (Best, 1995; Best and Tyler, 2007; Flege, 1995; Kuhl

and Iverson, 1995; Strange, 2006). The results of the current study open an additional possible factor that could affect which individuals are better or worse at learning second language sound categories. If the interpretation of the Central Executive and Episodic Buffer components is correct, and if this means that individuals with highly analytical working memory components are at a disadvantage for learning information-integration categories (both visual and talker), then it may mean that those individuals with less advanced Central Executive and Episodic Buffer components are those people who are better at learning the phonology of a second language (i.e., those people with a better *ear* for language). Future research will need to explore this possibility.

These working memory findings may also have implications for children with language impairments who are considered to have poorly developed phonological categories (Bishop, 1992) and who also have a poorer Phonological Loop and Central Executive than children with typical language development (Archibald and Gathercole, 2006). While a poorer Central Executive may not be problematic for learning auditory categories, a poor Phonological Loop is problematic, at least based on the results of the current study. Future work will need to explore these various components in children with language impairments to help determine whether working memory impairments actually contribute to the development of poorer phonological categories in this population.

In addition to age, vocabulary, and working memory, several other findings emerged related to the experimental variables. First, learning to categorize speech into talker categories is not slow or effortful; children rapidly learn to categorize speech signals into talker categories, as evidenced by a significant increase in accuracy across the first and second training session on day 1. In addition, there was a clear effect of Day of Training, where children's performance improved over the 5 days. Furthermore, all children exhibited learning and most children performed well above chance on the last day. Critically, learning and accurate categorization were not specific to the particular materials that were used, as children were required to categorize different words on each day.

Second, there was a clear effect of talker, where children were more accurate at identifying F8, followed by F11, and with the lowest performance for F3. This order of identification accuracy lines up nicely with the talkers' intelligibility measures in the clear (table 1), which is how stimuli were presented to participants in the current study. As mentioned in the Introduction, listeners are worse at processing information about a talker in foreign-accented speech (Thompson, 1987). While intelligibility is not the same as degree of foreign accent, it is possible that reduced intelligibility is related to greater foreign accent; thus, one reason for the poorer performance in Thompson's (1987) study and also here may be tied to reduced intelligibility. In addition, F8 produced the lowest average F0 (170 Hz) of the 3 talkers and was the only talker with an average F0 less than 200, the commonly reported average F0 for women in this age group (Fitch, 1990; Linke, 1973; Stoicheff, 1981). This F0 value is also at the lowest end of F0 ranges for women in this age group. This low F0 likely aided in the identification of this talker. F11, who was the second best identified talker, had the highest average F0 (260 Hz). F3, whose F0 was between the other 2 talkers (222 Hz), may have been less well identified due to possible ambiguity or overlap with the other 2 talkers.

Third, the lack of an effect of lexical familiarity on category learning suggests that children were able to selectively attend to the talker dimension of the speech signal and were not hindered when presented with unfamiliar words. This is consistent with other studies that have examined the effect of lexical properties on talker processing. Similar results were found in a talker discrimination task for both school-age children and adults where lexical familiarity did not affect performance (Levi and Schwartz, 2013). Furthermore, in a similar categorization task with adults, lexical frequency did not significantly affect talker identification accuracy (Winters et al., 2008).

Fourth, Phonological Awareness skills did not significantly affect learning for children in the current study. Phonological Awareness was included as a possible predictor of category learning because a previous study with adult learners found it to be positively correlated with learning to categorize talkers (Perrachione et al., 2011). These different findings may stem from differences in the populations tested. The correlation between Phonological Awareness and talker category learning found in Perrachione et al. (2011) was for adults with *impaired* Phonological Awareness skills, whereas the current study examined children with typical language skills and thus mostly intact Phonological Awareness skills. While table 3 shows that there is a range of Phonological Awareness skills in the current set of participants (76–133), only 3 children fell below the normal range (below 85). Thus, Phonological Awareness skills may be predictive of talker learning only for individuals with impaired Phonological Awareness.

Conclusion

The current study drew on research on category learning, working memory, and talker processing to contribute to the research on how talker categories are learned. Not surprisingly, age was a significant predictor of performance, where older children learned the talker categories better than younger children. Beyond the contribution of age, results showed that working memory plays a role in talker category learning, but that some components benefit learning (Phonological Loop), while others hinder learning (Central Executive and Episodic Buffer). That working memory plays both a positive and negative role in learning novel talker categories may indicate that general cognitive mechanisms act on both auditory and visual category learning.

Acknowledgments

This work was supported by a grant from the NIH-NIDCD: 1R03DC009851-01A2 (Levi). We would like to thank Gabrielle Alfano, Josh Barocas, Jennifer Bruno, Stephanie Lee, Emma Mack, Alexandra Muratore, Sydney Robert, and Margo Waltz for help with data collection, Sean Martin for help with the statistics, and the children and families for their participation.

References

- Allen JS, Miller JL. Listener sensitivity to individual talker differences in voice-onset-time. *J Acoust Soc Am.* 2004; 115:3171–3183. [PubMed: 15237841]
- Alloway TP, Gathercole SE. The role of sentence recall in reading and language skills of children with learning difficulties. *Learn Individ Differences.* 2005; 15:271–282.

- Alloway TP, Gathercole SE, Willis C, Adams A-M. A structural analysis of working memory and related cognitive skills in young children. *J Exp Child Psychol.* 2004; 87:85–106. [PubMed: 14757066]
- Arah OA. The role of causal reasoning in understanding Simpson's paradox, Lord's paradox, and the suppression effect: covariate selection in the analysis of observational studies. *Emerg Themes Epidemiol.* 2008; 5:5. [PubMed: 18302750]
- Archibald LM, Gathercole SE. Short-term and working memory in specific language impairment. *Int J Lang Commun Disord.* 2006; 41:675–693. [PubMed: 17079222]
- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. *Psychol Rev.* 1998; 105:442. [PubMed: 9697427]
- Ashby FG, Ell SW, Waldron EM. Procedural learning in perceptual categorization. *Mem Cognit.* 2003; 31:1114–1125.
- Ashby FG, Gott RE. Decision rules in the perception and categorization of multidimensional stimuli. *J Exp Psychol Learn Mem Cogn.* 1988; 14:33. [PubMed: 2963894]
- Ashby FG, Maddox WT. Human category learning. *Annu Rev Psychol.* 2005; 56:149–178. [PubMed: 15709932]
- Ashby FG, Maddox WT, Bohil CJ. Observational versus feedback training in rule-based and information-integration category learning. *Mem Cognit.* 2002; 30:666–677.
- Ashby FG, O'Brien JB. Category learning and multiple memory systems. *Trends Cogn Sci.* 2005; 9:83–89. [PubMed: 15668101]
- Baayen, RH. *Analyzing Linguistic Data: A Practical Introduction to Statistics.* Cambridge: Cambridge University Press; 2008.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang.* 2008; 59:390–412.
- Baddeley, A. *Working Memory.* Oxford: Clarendon; 1986.
- Baddeley A. The episodic buffer: a new component of working memory? *Trends Cogn Sci.* 2000; 4:417–423. [PubMed: 11058819]
- Baddeley A. Working memory and language: an overview. *J Commun Disord.* 2003; 36:189–208. [PubMed: 12742667]
- Baddeley, A. *Working Memory, Thought, and Action.* Oxford: Oxford University Press; 2007.
- Baddeley A. Working memory. *Curr Biol.* 2010; 20:R136–R140. [PubMed: 20178752]
- Baddeley A, Gathercole S, Papagno C. The phonological loop as a language learning device. *Psychol Rev.* 1998; 105:158. [PubMed: 9450375]
- Baddeley A, Hitch G. Working memory. *Psychol Learn Motiv.* 1974; 8:47–89.
- Bartholomeus B. Voice identification by nursery school children. *Can J Psychol.* 1973; 27:464–472. [PubMed: 4766153]
- Bates, D.; Maechler, M.; Bolker, B.; Walker, S. *lme4: linear mixed-effects models using Eigen and S4.* 2010. <http://lme4.r-forge.r-project.org/>
- Best, CT. Emergence of language-specific constraints in perception of native and non-native speech: a window on early phonological development. In: de Boysson, B.; de Schonen, S.; Jusczyk, P.; McNeilage, P.; Morton, J., editors. *Developmental Neurocognition: Speech and Face Processing during the First Year of Life.* Dordrecht: Kluwer; 1993. p. 289-304.
- Best, CT. A direct realist view of cross-language speech perception. In: Strange, W., editor. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore: York Press; 1995. p. 171-204.
- Best, CT.; Tyler, MD. Honor of James Emil Flege. Amsterdam: Benjamins; 2007. Nonnative and second-language speech perception: commonalities and complementarities. *Language Experience in Second Language Speech Learning*; p. 13-34.
- Bishop DVM. The underlying nature of specific language impairment. *J Child Psychol Psychiatry.* 1992; 33:3–66. [PubMed: 1737831]
- Blyth CR. On Simpson's paradox and the sure-thing principle. *J Am Stat Assoc.* 1972; 67:364–366.
- Brown, L.; Sherbenou, RJ.; Johnsen, SK. *TONI-3: Test of Nonverbal Intelligence.* 3. Austin: Pro-Ed; 1997.

- Chandrasekaran B, Yi H-G, Maddox WT. Dual-learning systems during speech category learning. *Psychon Bull Rev.* 2014; 21:488–495. [PubMed: 24002965]
- Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 3. Mahwah: Erlbaum; 2003.
- Cortese MJ, Khanna MM. Age of acquisition ratings for 3,000 monosyllabic words. *Beh Res Methods.* 2008; 40:791–794.
- Craig S, Lewandowsky S. Whichever way you choose to categorize, working memory helps you learn. *Q J Exp Psychol.* 2012; 65:439–464.
- Creel SC, Jimenez SR. Differences in talker recognition by preschoolers and adults. *J Exp Child Psychol.* 2012; 113:487–509. [PubMed: 22958962]
- DeCaro MS, Carlson KD, Thomas RD, Beilock SL. When and how less is more: reply to Tharp and Pickering. *Cognition.* 2009; 111:415–421.
- DeCaro MS, Thomas RD, Beilock SL. Individual differences in category learning: sometimes less working memory capacity is better than more. *Cognition.* 2008; 107:284–294. [PubMed: 17707363]
- DeCasper AJ, Fifer WP. Of human bonding: newborns prefer their mothers' voices. *Science.* 1980; 208:1174–1176. [PubMed: 7375928]
- DeCasper AJ, Prescott PA. Human newborn's perception of male voices: preference, discrimination, and reinforcing value. *Dev Psychol.* 1984; 17:481–491.
- Dunn, LM.; Dunn, DM. *PPVT-4: Peabody Picture Vocabulary Test*. 4. Minneapolis: NCS Pearson; 2007.
- Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. *Percept Psychophys.* 2005; 67:224–238. [PubMed: 15971687]
- Fennell CT, Werker JF. Early word learners' ability to access phonetic detail in well-known words. *Lang Speech.* 2003; 46:245–264. [PubMed: 14748446]
- Fitch JL. Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *J Speech Hear Disord.* 1990; 55:360–363. [PubMed: 2329798]
- Fllege, JE. Second language speech learning: theory, findings, and problems. In: Strange, W., editor. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press; 1995. p. 233-277.
- Gelman, A.; Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press; 2007.
- Goggin JP, Thompson CP, Strube G, Simental LR. The role of language familiarity in voice identification. *Mem Cognit.* 1991; 19:448–458.
- Goldstein AG, Knight P, Bailis K, Conover J. Recognition memory for accented and unaccented voices. *Bull Psychon Soc.* 1981; 17:217–220.
- Grossman M, Smith EE, Koenig P, Glosser G, DeVita C, Moore P, McMillan C. The neural basis for categorization in semantic memory. *Neuroimage.* 2002; 17:1549–1561. [PubMed: 12414293]
- Hollien H, Majewski W, Doherty ET. Perceptual identification of voices under normal, stress, and disguise speaking conditions. *J Phon.* 1982; 10:139–148.
- Jaeger TF. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang.* 2008; 59:434–446. [PubMed: 19884961]
- Jesse, A.; McQueen, JM.; Page, M. The locus of talker-specific effects in spoken-word recognition. In: Trouvain, J.; Barry, WJ., editors. *International Congress of Phonetic Sciences (ICPhS 2007)*; Dudweiler, Pirrot. 2007. p. 1921-1924.
- Johnson EK, Westrek E, Nazzi T, Cutler A. Infant ability to tell voices apart rests on language experience. *Dev Sci.* 2011; 14:1002–1011. [PubMed: 21884316]
- Jusczyk, PW. Toward a model of the development of speech perception. In: Perkell, JS.; Klatt, DH., editors. *Invariance and Variability in Speech Processes*. Hillsdale: Erlbaum; 1986. p. 1-19.
- Jusczyk, PW. Developing phonological categories from the speech signal. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. *Phonological Development: Models, Research, Implications*. Parkton: York Press; 1992. p. 17-64.

- Kisilevsky BS, Hains SMJ, Lee K, Xie S, Huang H, Ye HH, Wang Z. Effects of experience on fetal voice recognition. *Psychol Sci.* 2003; 14:220–224. [PubMed: 12741744]
- Klatt DH. Speech perception: a model of acoustic-phonetic analysis and lexical access. *J Phon.* 1979; 7:1–26.
- Köster O, Schiller NO. Different influences of the native language of a listener on speaker recognition. *Forensic Linguist.* 1997; 4:18–28.
- Kraljic T, Samuel AG. Perceptual adjustments to multiple talkers. *J Mem Lang.* 2007; 56:1–15.
- Ku era, H.; Francis, WN. *Computational Analysis of Present-Day American English.* Providence: Brown University Press; 1967.
- Kuhl, PK.; Iverson, P. Linguistic experience and the ‘Perceptual Magnet Effect’. In: Strange, W., editor. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Baltimore: York Press; 1995. p. 121-154.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science.* 1992; 255:606–608. [PubMed: 1736364]
- Ladefoged P, Broadbent DE. Information conveyed by vowels. *J Acoust Soc Am.* 1957; 29:98–104.
- Levi SV. Talker familiarity and spoken word recognition in school-age children. *J Child Lang.* 2014 Epub ahead of print.
- Levi SV, Schwartz RG. The development of language-specific and language-independent talker processing. *J Speech Lang Hear Res.* 2013; 56:913–920. [PubMed: 23275425]
- Levi SV, Winters SJ, Pisoni DB. Effects of cross-language voice training on speech perception: whose familiar voices are more intelligible? *J Acoust Soc Am.* 2011; 130:4053–4062. [PubMed: 22225059]
- Lewandowsky S, Yang L-X, Newell BR, Kalish ML. Working memory does not dissociate between different perceptual categorization tasks. *J Exp Psychol Learn Mem Cogn.* 2012; 38:881. [PubMed: 22746954]
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev.* 1967; 74:431. [PubMed: 4170865]
- Linke CE. A study of pitch characteristics of female voices and their relationship to vocal effectiveness. *Folia Phoniatri Logop.* 1973; 25:173–185.
- Maddox WT, Ashby FG. Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behav Processes.* 2004; 66:309–332. [PubMed: 15157979]
- Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. *J Exp Psychol Learn Mem Cogn.* 2003; 29:650–662. [PubMed: 12924865]
- Maddox WT, Bohil CJ, Ing AD. Evidence for a procedural-learning-based system in perceptual category learning. *Psychon Bull Rev.* 2004; 11:945–952. [PubMed: 15732708]
- Maddox, WT.; Chandrasekaran, B. Tests of a dual-systems model of speech category learning. *Bilingualism: Language and Cognition*; in press
- Maddox WT, Chandrasekaran B, Smayda K, Yi H-G. Dual systems of speech category learning across the lifespan. *Psychol Aging.* 2013; 28:1042. [PubMed: 24364408]
- Maddox WT, Ing AD, Lauritzen JS. Stimulus modality interacts with category structure in perceptual category learning. *Percept Psychophys.* 2006; 68:1176–1190. [PubMed: 17355041]
- Magnuson, JS.; Yamada, RA.; Nusbaum, HC. The effects of familiarity with a voice on speech perception. *Proc of the 1995 Spring Meeting of the Acoustical Society of Japan*; 1995. p. 391-392.
- Mann VA, Diamond R, Carey S. Development of voice recognition: parallels with face recognition. *J Exp Child Psychol.* 1979; 27:153–165. [PubMed: 458368]
- McQueen JM, Cutler A, Norris D. Phonological abstraction in the mental lexicon. *Cogn Sci.* 2006; 30:1113–1126. [PubMed: 21702849]
- Metsala JL. An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Mem Cognit.* 1997; 25:47–56.
- Miles SJ, Minda JP. The effects of concurrent verbal and visual tasks on category learning. *J Exp Psychol Learn Mem Cogn.* 2011; 37:588. [PubMed: 21319921]
- Mitchell HF, MacDonald RAR. Recognition and description of singing voices: the impact of verbal overshadowing. *Musicae Scientiae.* 2012; 16:307–316.

- Moher M, Feigenson L, Halberda J. A one-to-one bias and fast mapping support preschoolers' learning about faces and voices. *Cogn Sci*. 2010; 34:719–751. [PubMed: 21564234]
- Norris D, McQueen JM, Cutler A. Perceptual learning in speech. *Cognit Psychol*. 2003; 47:204–238. [PubMed: 12948518]
- Nosofsky RM. Typicality in logically defined categories: exemplar-similarity versus rule instantiation. *Mem Cognit*. 1991; 19:131–150.
- Nosofsky RM, Clark SE, Shin HJ. Rules and exemplars in categorization, identification, and recognition. *J Exp Psychol LearnMem Cogn*. 1989; 15:282.
- Nosofsky RM, Johansen MK. Exemplar-based accounts of 'multiple-system' phenomena in perceptual categorization. *Psychon Bull Rev*. 2000; 7:375–402. [PubMed: 11082849]
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Percept Psychophys*. 1998; 60:355–376. [PubMed: 9599989]
- Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychol Sci*. 1994; 5:42–46. [PubMed: 21526138]
- Pearl J. Comment: understanding Simpson's paradox. *Am Statist*. 2014; 68:8–13.
- Perfetti CA, Beck I, Bell LC, Hughes C. Phonemic knowledge and learning to read are reciprocal: a longitudinal study of first grade children. *Merrill-Palmer Q*. 1987:283–319.
- Perrachione TK, Del Tufo SN, Gabrieli JDE. Human voice recognition depends on language ability. *Science*. 2011; 333:595. [PubMed: 21798942]
- Perrachione TK, Wong PCM. Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. *Neuropsychologia*. 2007; 45:1899–1910. [PubMed: 17258240]
- Pickering, S.; Gathercole, SE. *Working Memory Test Battery for Children*. London: Psychological Corporation; 2001.
- Purhonen M, Kilpeläinen-Lees R, Valkonen-Korhonen M, Karhu J, Lehtonen J. Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *Int J Psychophysiol*. 2004; 52:257–266. [PubMed: 15094248]
- Purhonen M, Kilpeläinen-Lees R, Valkonen-Korhonen M, Karhu J, Lehtonen J. Four-month-old infants process own mother's voice faster than unfamiliar voices—electrical signs of sensitization in infant brain. *Brain Res Cogn Brain Res*. 2005; 24:627–633. [PubMed: 16099371]
- Reposš G, Baddeley A. The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*. 2006; 139:5–21. [PubMed: 16517088]
- Rudner M, Rönnerberg J. The role of the episodic buffer in working memory for language processing. *Cogn Process*. 2008; 9:19–28. [PubMed: 17917753]
- Schiller NO, Köster O. Evaluation of a foreign speaker in forensic phonetics: a report. *Forensic Linguist*. 1996; 3:176–185.
- Schneider, W.; Eschman, A.; Zuccolotto, A. *E-Prime 2.0 Professional*. Pittsburgh: Psychology Software Tools; 2007.
- Semel, E.; Wiig, EH.; Secord, WA. *Clinical Evaluation of Language Fundamentals*. 4. Toronto: Psychological Corporation/Harcourt Assessment Company; 2003. CELF-4
- Smith EE, Patalano AL, Jonides J. Alternative strategies of categorization. *Cognition*. 1998; 65:167–196. [PubMed: 9557382]
- Spence MJ, Rollins PR, Jerger S. Children's recognition of cartoon voices. *J Speech Lang Hear Res*. 2002; 45:214–222. [PubMed: 14748650]
- Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception tasks than in word-learning tasks. *Nature*. 1997; 388:381–382. [PubMed: 9237755]
- Stahl SA, Murray BA. Defining phonological awareness and its relationship to early reading. *J Educ Psychol*. 1994; 86:221.
- Stanovich KE, Cunningham AE, Cramer BB. Assessing phonological awareness in kindergarten children: issues of task comparability. *J Exp Child Psychol*. 1984; 38:175–190.
- Stoicheff ML. Speaking fundamental frequency characteristics of nonsmoking female adults. *J Speech Lang Hear Res*. 1981; 24:437–441.

- Strange W. Second-language speech perception: the modification of automatic selective perceptual routines. *J Acoust Soc Am*. 2006; 120:3137.
- Sullivan KPH, Kügler F. Was the knowledge of the second language or the age difference the determining factor? *Forensic Linguist*. 2001; 8:1–8.
- Sullivan KPH, Schlichting F. Speaker discrimination in a foreign language: first language environment, second language learners. *Forensic Linguist*. 2000; 17:95–111.
- Swingle D, Aslin RN. Spoken word recognition and lexical representation in very young children. *Cognition*. 2000; 76:147–166. [PubMed: 10856741]
- Tharp IJ, Pickering AD. A note on DeCaro, Thomas, and Beilock (2008) further data demonstrate complexities in the assessment of information-integration category learning. *Cognition*. 2009; 111:410–414.
- Theodore RM, Miller JL. Characteristics of listener sensitivity to talker-specific phonetic detail. *J Acoust Soc Am*. 2010; 128:2090–2099. [PubMed: 20968380]
- Thompson CP. A language effect in voice identification. *Appl Cogn Psychol*. 1987; 1:121–131.
- Vanags T, Carroll M, Perfect TJ. Verbal overshadowing: a sound theory of voice recognition? *Appl Cogn Psychol*. 2005; 19:1127–1144.
- Wagner, R.; Torgesen, JK.; Rashotte, C. *Comprehensive Test of Phonological Processing (CTOPP)*. Austin: Pro-Ed; 1999.
- Waldron EM, Ashby FG. The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychon Bull Rev*. 2001; 8:168–176. [PubMed: 11340863]
- Walley AC. The role of vocabulary development in children's spoken word recognition and segmentation ability. *Dev Rev*. 1993; 13:286–350.
- Werker JF, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav Dev*. 1984; 7:49–63.
- Wester M. Talker discrimination across languages. *Speech Commun*. 2012; 54:781–790.
- White KS, Morgan JL. Sub-segmental detail in early lexical representations. *J Mem Lang*. 2008; 59:114–132.
- Winters SJ, Levi SV, Pisoni DB. Identification and discrimination of bilingual talkers across languages. *J Acoust Soc Am*. 2008; 123:4524–4538. [PubMed: 18537401]
- Yonan CA, Sommers MS. The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychol Aging*. 2000; 15:88–99. [PubMed: 10755292]
- Yuasa IP. Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women? *Am Speech*. 2010; 85:315–337.
- Zeithamova D, Maddox WT. Dual-task interference in perceptual category learning. *Mem Cognit*. 2006; 34:387–398.

Appendix

Words with High Lexical Familiarity

bake	chop	hair	leash	peace	such
beach	coin	ham	leave	pet	tail
bean	comb	hang	lick	pill	then
bear	come	hatch	line	rash	these
beef	cop	hate	look	rat	thin
beer	couch	have	lose	ride	thumb
bib	date	head	mad	room	tight
big	dead	heel	map	rough	wag
bike	dish	here	match	run	web

bit	duck	him	meat	sale	whale
bake	chop	hair	leash	peace	such
both	fair	hit	miss	sauce	what
bug	fat	hoop	moose	seed	where
buzz	fill	hot	mop	shake	white
cab	fish	house	mouse	shape	win
cake	fit	hug	neck	share	wing
cash	foot	jail	need	shave	wish
cat	fun	jar	net	should	with
cave	gas	kill	nice	shout	write
cheek	goose	kite	night	shut	wrong
cheese	gun	knife	page	son	year
choke	gym	leaf	pan	soon	

Words with Low Lexical Familiarity

bane	fate	hick	mime	rife	vole
bid	faze	hitch	mock	rile	wade
bile	feign	hock	mode	rut	wane
cad	fib	hone	mope	sage	wean
chafe	foul	jeer	muss	sake	whim
char	gab	knack	neap	sate	whip
chum	gape	lass	nick	sheath	whiz
cog	gauge	laud	node	shuck	wick
cope	gauze	leach	noose	siege	wit
core	gawk	ledge	notch	souse	womb
couth	gig	loathe	peal	sup	wraith
cuff	gin	loom	peer	tithe	wrath
cull	gnash	lore	pith	tone	wreath
dame	gnat	luff	poach	tout	zeal
debt	goad	lush	puck	vague	
dock	goon	mace	raid	vain	
doff	gull	maim	reap	vat	
dose	hail	mauve	reek	veil	
dung	haze	mesh	retch	void

Fig. 1.
Model of working memory adapted from (Baddeley, 1986, 2000, 2007, 2010; Baddeley & Hitch, 1974; Repovš & Baddeley, 2006).

Fig. 2.
Response screen for talker training with and without feedback.

Fig. 3.
Procedure used for the talker training with feedback.

Fig. 4.

Average percent correct when categorizing the three talkers across the five days of learning along with Standard Errors. The solid lines represent Training and include the two sessions per day. The dashed circles represent Test and therefore only have a single session.

Fig. 5. Histograms of children's accuracy on the final day of learning for both training (with feedback) and test (without feedback).

Table 1

Intelligibility and F0 information for the talkers used during category learning

Talker	Intelligibility-average, %	Intelligibility-clear, %	F0, Hz	Duration, ms
F3	49.1	76.2	222	690
F8	54.7	83.3	170	628
F11	41.2	77.8	260	637

Intelligibility-average refers to the percent of words correctly identified averaged across three signal-to-noise ratios (0, +5, +10) and a clear listening condition (Levi et al., 2011). Intelligibility-clear refers to the percent of words correctly identified in the clear condition alone. F0 is the average F0 at the vowel midpoint for the 360 English words. Duration is the average whole-word duration across the full set of 360 English words.

Table 2

One day of the talker learning procedure

	Phase	Stimuli	Task
Training (session 1 and session 2)	Familiarization (×2)	2 high and 2 low familiarity words produced by all 3 talkers	Listen and attend to voice/character pair
	Recognition	5 high and 5 low familiarity words produced by all 3 talkers in random order	Identify talker of each word (with feedback)
Test	Familiarization (×2)	2 high and 2 low familiarity words produced by all 3 talkers	Listen and attend to voice/character pair
	Recognition	5 high and 5 low familiarity words produced by all 3 talkers in random order	Identify talker of each word (no feedback)

Listeners completed the training sessions with feedback twice (session 1 and session 2), followed by a single test session without feedback.

Table 3
Distribution of standardized scores for the children across the experimental standardized tasks

	Normed		Study		SD	range	interquartile
	mean	SD	mean	SD			
Inclusion criteria							
	100	15	103	10.8	85–126	95–110	
CELF-Core Language							
	100	15	108	16.5	85–146	96–119	
TONI							
Experimental measures							
	10	3	9.4	2.5	3–14	8–11	
Forward Digit Span (Phonological Loop)							
	10	3	9.8	2.6	5–15	8–12	
Backward Digit Span (Central Executive)							
	100	15	100.6	17	57–141	90–110	
Block Span (Visuospatial Sketchpad)							
	10	3	10.6	2.5	6–17	9–12	
Recalling Sentences (Episodic Buffer)							
	100	15	103.4	12.8	76–133	94–112	
CTOPP (Phonological Awareness)							
	100	15	106.4	16	78–140	95–118	
PPVT (Lexicon Size)							

The mean and SD from the normed tests are provided on the left with the mean, SD, range, and inter-quartile of the children in the current study on the right.

Table 4

Correlation matrix for the working memory and language measures

	Forward Digit	Backward Digit	Block Span	Recalling Sentences	CTOPP	PPVT
Forward Digit						
Backward Digit	0.39					
Block Span	0.29	0.32				
Recalling Sentences	0.47	0.30	0.19			
CTOPP	0.42	0.28	0.16	0.36		
PPVT	0.27	0.38	0.19	0.55	0.33	

Table 5

Results of the logit mixed-effects model across all days of learning for both the training portion (with feedback) and the test portion (no feedback)

	Training (feedback)				Test (no feedback)			
	estimate	SE	z	p value	estimate	SE	z	p value
Intercept	-1.60	0.490	-3.26	0.001	-2.16	0.548	-3.94	<0.001
Additional factors								
Day of Training	0.267	0.015	18.32	<0.001	0.194	0.019	9.84	<0.001
Age	0.016	0.004	3.78	<0.001	0.020	0.004	4.25	<0.001
Lexical Familiarity	-0.009	0.041	-0.23	0.817	0.000	0.055	0.00	0.999
Session	0.079	0.041	1.89	0.057	-	-	-	-
Talker: F11-F3	-0.362	0.090	-4.01	<0.001	-0.367	0.091	-2.92	0.009
Talker: F11-F8	1.07	0.128	8.41	<0.001	1.04	0.146	7.13	<0.001
Talker: F8-F3	1.44	0.111	12.86	<0.001	1.31	0.148	8.86	<0.001
Working memory								
Forward Digit Span (Phonological Loop)	0.102	0.037	2.73	0.006	0.111	0.040	2.70	0.006
Backward Digit Span (Central Executive)	-0.034	0.033	-1.02	0.305	-0.044	0.037	-1.16	0.245
Block Span (Visuospatial Sketchpad)	0.004	0.004	0.91	0.358	0.006	0.005	1.24	0.220
Recalling Sentences (Episodic Buffer)	-0.084	0.037	-2.42	0.024	-0.095	0.044	-2.17	0.029
Other language measures								
CTOPP (Phonological Awareness)	0.007	0.006	1.08	0.278	0.003	0.007	0.39	0.690
PPVT (Lexicon)	0.012	0.005	2.34	0.024	0.019	0.006	2.98	0.002
Collinearity checks								
variance inflation factors		all <2.0					all <2.0	
condition number (kappa)		11.87					12.12	

Positive z-scores indicate a positive influence on performance, and negative z-scores indicate a negative influence on performance.

Table 6

Results of the logit mixed-effects model for day 1 for both the training portion (with feedback) and the test portion (no feedback)

	Training (feedback)				Test (no feedback)			
	estimate	SE	z	p value	estimate	SE	z	p value
Intercept	-0.977	0.530	-1.84	0.065	-1.826	0.708	-2.58	0.009
Additional factors								
Age	0.010	0.004	2.18	0.028	0.019	0.006	3.09	0.002
Lexical Familiarity	0.013	0.082	0.16	0.868	-0.028	0.118	-0.24	0.810
Session	0.389	0.082	4.70	<0.001	-	-	-	-
Talker: F11-F3	-0.297	0.107	-2.75	0.015	-2.92	0.155	-1.88	0.141
Talker: F11-F8	1.03	0.148	7.01	<0.001	0.923	0.209	4.40	<0.001
Talker: F8-F3	1.33	0.140	9.50	<0.001	1.216	0.203	5.97	<0.001
Working memory								
Forward Digit Span (Phonological Loop)	0.095	0.038	2.45	0.014	0.100	0.050	1.97	0.048
Backward Digit Span (Central Executive)	-0.025	0.034	-0.59	0.550	-0.023	0.046	-0.49	0.619
Block Span (Visuospatial Sketchpad)	0.001	0.004	0.26	0.789	0.005	0.006	0.82	0.409
Recalling Sentences (Episodic Buffer)	-0.085	0.040	-2.12	0.033	-0.220	0.056	-3.87	<0.001
Other language measures								
CTOPP (Phonological Awareness)	0.001	0.006	-0.21	0.827	0.014	0.009	1.48	0.137
PPVT (Lexicon)	0.009	0.005	1.57	0.114	0.023	0.008	2.92	0.003
Colinearity checks								
variance inflation factors			all <2.0				all <2.0	
condition number (kappa)			12.08				12.47	

Positive z-scores indicate a positive influence on performance, and negative z-scores indicate a negative influence on performance.

Table 7

Results of the logit mixed-effects model for day 5 for both the training portion (with feedback) and the test portion (no feedback)

	Training (feedback)				Test (no feedback)			
	estimate	SE	z	p value	estimate	SE	z	p value
Intercept	-1.53	0.697	-2.20	0.027	-1.765	0.999	-1.76	0.077
Additional factors								
Age	0.028	0.006	4.72	<0.001	0.026	0.008	3.01	0.002
Lexical Familiarity	0.071	0.103	0.69	0.489	-0.156	0.133	-1.17	0.241
Session	0.149	0.104	1.43	0.151	-	-	-	-
Talker: F11-F3	-0.460	0.194	-2.36	0.047	-0.200	0.186	-1.07	0.525
Talker: F11-F8	0.64	2.40	2.69	0.019	0.950	0.232	4.08	<0.001
Talker: F8-F3	1.10	0.204	5.42	<0.001	1.15	0.220	5.22	<0.001
Working memory								
Forward Digit Span (Phonological Loop)	0.122	0.048	2.53	0.011	0.203	0.072	2.79	0.005
Backward Digit Span (Central Executive)	-0.106	0.043	-2.45	0.014	-0.141	0.066	-2.11	0.034
Block Span (Visuospatial Sketchpad)	0.003	0.005	0.61	0.542	0.007	0.008	0.83	0.402
Recalling Sentences (Episodic Buffer)	-0.113	0.051	-2.17	0.029	-0.132	0.077	-1.71	0.085
Other language measures								
CTOPP (Phonological Awareness)	0.015	0.008	1.79	0.072	-0.002	0.013	-0.14	0.882
PPVT (Lexicon)	0.024	0.008	3.09	<0.001	0.032	0.011	2.87	0.004
Colinearity checks								
variance inflation factors		all <2.0				all <2.0		
condition number (kappa)		11.23				12.47		

Positive z-scores indicate a positive influence on performance, and negative z-scores indicate a negative influence on performance.