Song and infant-directed speech facilitate word learning

Ma, Weiyi; Fiveash, Anna; Margulis, Elizabeth Hellmuth; Behrend, Douglas; Thompson, William Forde

Published in:
Quarterly Journal of Experimental Psychology

DOI:
10.1177/1747021819888982

In press at Quarterly Journal of Experimental Psychology

Song and Infant-Directed Speech Facilitate Word Learning

Weiyi Ma[1], Anna Fiveash[2], Elizabeth Margulis[3], Douglas Behrend[4], William Forde Thompson[5]

[1] School of Human Environmental Sciences, University of Arkansas, Fayetteville, AR, 72701, U.S.A.

[2] University of Lyon 1 and 2, Lyon Neuroscience Research Centre, Auditory Cognition and Psychoacoustics Team, Lyon, France

[3] Department of Music, Princeton University, Princeton, NJ 08544

[4] Department of Psychological Science, University of Arkansas, Fayetteville, AR, 72701, U.S.A.

[5] Department of Psychology, and Department of Cognitive Science, and ARC Centre of Excellence in Cognition and Its Disorders Macquarie University, Sydney, NSW, 2109, Australia

Date: October 25, 2019

Please address correspondence to:
Weiyi Ma
School of Human Environmental Sciences
University of Arkansas
Fayetteville AR, 72701
Weiyima@uark.edu

Acknowledgments: W.M. is supported by The University of Arkansas Startup Fund and the Provost's Collaborative Research Grant. A.F. is in the Auditory cognition and psychoacoustics team that is supported by ANR-10-LABX-60.

Abstract

Two separate lines of research have examined the influence of song and infant-directed speech (IDS – a speech register that includes some melodic features) on language learning, suggesting that the use of musical attributes in speech input can enhance language learning. However, the benefits of these two types of stimuli have never been directly compared. In this investigation, we compared the effects of song and IDS for immediate word learning and long-term memory of the learned words. This study examines whether the highly musical stimuli (i.e., song) would facilitate language learning more than the less musical stimuli (i.e., IDS). English-speaking adults were administered a word learning task, with Mandarin Chinese words presented in either adult-directed speech (ADS), IDS, or song. Participants' word learning performance was assessed immediately after the word learning task (immediate word learning) and then one day later (long-term memory). Results showed that both song and IDS facilitated immediate word learning and long-term memory of the words; however, this facilitative effect did not differ between IDS and song, suggesting that the relationship between the degree of musicality and language learning performance is not linear. In addition, song and IDS were found to facilitate the word association process (mapping a label to its referent) rather than the word recognition process. Finally, participants' confidence in their answers might not differ among ADS, IDS, and sung words.

Key words: music, infant-directed speech, song, language learning, memory.

Song and Infant-Directed Speech Facilitate Word Learning

Language learning is critical to human development, as language enables information communication, self-expression, and the formation and maintenance of social relationships. Like language, music serves as a major auditory channel of communication, and is widespread across human cultures (Merriam, 1964). An extensive body of research shows that sung words can enhance language learning and memory in children and adults (e.g., Good, Russo, & Sullivan, 2015; Ludke, Ferreira, & Overy, 2014; Rukholm, Helms-Park, Odgaard, & Smyth, 2018; Schön et al., 2008; Thaut, Peterson, McIntosh, & Hoemberg, 2014; Wallace, 1994). Another line of research suggests that adults use a special form of speech known as infant-directed speech (IDS) when they talk to infants and children that features exaggerated prosody (e.g., McRoberts & Best, 1997), slower transitions (e.g., Gleitman, Newport, & Gleitman, 1984), and increased repetition e.g., Fernald & Simon, 1984): all elements characteristic of music (Fernald, 1989). The use of IDS can facilitate language learning in children and adults (e.g., Estes & Hurley, 2013; Golinkoff & Alioto, 1995; Foursha-Stevenson, Schembri, Nicoladis, & Eriksen, 2017; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Thiessen, Hill, & Saffran, 2005; Singh, Nestor, Parikh, & Yull, 2009). Despite the findings showing that both singing and IDS can improve language learning, the effect of these two types of stimuli have never been directly compared. This study examines the impact of varying word presentation among ordinary adult-directed speech (ADS), IDS, and song on word learning in adults.

Speech and music share underlying cognitive and neural resources (e.g., Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Musso et al., 2015; Patel, 2003, 2008; Ross, Choi, & Purves, 2007) and draw on a common code of acoustic attributes to communicate emotional states (e.g., Heffner & Slevc, 2015; Ilie & Thompson, 2006, 2011; Juslin & Laukka, 2003; Quinto, Thompson, Keating, 2013); however, the existence of neural and acoustic commonalities between language and music does not necessarily imply that music exposure can facilitate language abilities. Yet a growing body of research suggests that music training may enhance language processing ability (Fujii & Wan, 2014; Kraus & Chandrasekaran, 2010; Maess et al., 2001; Patel, 2011; Tierney & Kraus, 2014), although there is ongoing discussion about how such data should be interpreted (Swaminathan, Schellenberg & Venkatesan, 2018). One common approach compares language processing in musicians versus non-musicians. Compared to non-musicians, musicians tend to have enhanced syllable discrimination (Zuk, 2014), speech in noise perception (Du & Zatorre, 2017), word segmentation (François, Jaillet, Takerkart, & Schön, 2014), and even pitch change tracking of speech at the brainstem level (Bidelman, Gandour, & Krishnan, 2011; Wong, Skoe, Russo, Dees, & Kraus, 2007). Furthermore, developmental research shows that music training is associated with enhanced speech segmentation skills in children (François, Chobert, Besson, & Schön, 2013), and that exposing infants to music leads to enhanced neural responses to violations of temporal structure in speech (Zhao & Kuhl, 2016). Such findings imply that music exposure has an impact on speech processing, but the processes that lead to such cross-domain influence are not fully understood (Mankel & Bidelman, 2018). Furthermore, research on music training does not speak to the influence of the use of musical attributes in speech input on language learning per se.

Two separate lines of research examined the influence of the use of musical attributes in speech – namely IDS and song – on language learning. Research on IDS suggests that the use of IDS towards infants and young children enhances language acquisition through attention directed to important elements of the speech stream (Karzon, 1985; Saint-Georges et al., 2013; Soderstrom, 2007), cues to word segmentation and grouping (Thiessen et al., 2005), and enhanced neural tracking of the speech envelope (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018). Compared to ADS, IDS can facilitate word learning and recognition in infants (Estes & Hurley, 2013; Foursha-Stevenson et al., 2017; Ma et al., 2011; Singh et al., 2009) and adults (Filippi, Gingras, & Fitch, 2014; Golinkoff & Alioto, 1995). At the neural level, compared to ADS, IDS enhances the processing of statistical regularities in speech (Bosseler, Teinonen, Tervaniemi, & Huotilainen, 2016), the brain response to familiarized words in infants (Zangl & Mills, 2007), and cerebral blood flow in infants (Saito et al., 2007). Notably, although the speech processing tasks used in these studies required memory of the speech input, none of these studies examined the influence of IDS on learners' *long-term* memory of the newly learned words after an extended period of time. Thus, it is still unclear whether IDS facilitates long-term word memory. Furthermore, although developmental research revealed a decline across age in learners' dependence on IDS in word learning (Ma et al., 2011), other evidence shows that IDS facilitated word learning even in adults (e.g., Golinkoff & Alioto, 1995). Thus, it is still unclear whether the benefits of IDS extend beyond child language acquisition, and beyond its capacity to direct attention.

Furthermore, different experimental paradigms point to different aspects of IDS that contribute to word learning and segmentation, making it difficult to distinguish the specific acoustic cues that facilitate language learning. Pitch (Filippi et al., 2014; Filippi, Laaha, & Fitch, 2017), hyper-articulation of vowels and slow speech (Song, Demuth, & Morgan, 2010), stronger rhythmic synchronization and acoustic temporal regularity (Leong, Kalashnikova, Burnham, & Goswami, 2017), and prosodic intonation (Räsänen, Kakouros, & Soderstrom, 2018; Thiessen et al., 2005) have all been implicated as important to word segmentation and learning. Further, not all studies have shown a facilitative effect of IDS for infants in word segmentation (Floccia et al.,

2016) or word learning (Schreiner, Altvater-Mackensen, & Mani, 2016), making it difficult to draw strong conclusions as to the role of IDS in word learning. However, a systematic review of IDS studies found that IDS enhanced infant's attention and language learning (Saint-Georges et al., 2013), and Nygaard, Herold, and Namy (2009) showed consistency across speakers producing IDS to communicate specific semantic concepts, suggesting a link between acoustic features and speech meaning. It therefore appears that on the whole, an important link exists between IDS and language acquisition, which could be exploited for the learning of new material in adults.

Another line of research examined the influence of song on word processing. Throughout history, songs have been used as a means of transmitting messages through generations (the oral tradition), suggesting that the combination of music and speech may also enhance learning and recall of the material the music is paired with (Milliron, 2017). Experimentally, results have been mixed. A seminal study by Wallace (1994) showed that sung text can both facilitate and interfere with language recall, which is necessary for word learning. Through a series of experiments, Wallace (1994) showed that listening to three verses of a sung ballad led to greater written recall (and delayed recall) than listening to three verses of a spoken ballad. Participants also showed greater recall in a sung condition compared to a rhythmically spoken condition, suggesting that the effect was based on more than regular rhythmic structure. However, when only one verse was presented, spoken text was recalled better than sung text, suggesting that repetition was important for learning sung material. Repeated pairing of notes and words was also shown to be important, with sung melodies repeated three times with the same melody resulting in greater recall than sung text with three different melodies or spoken text. From these experiments, Wallace (1994) suggested that both rhythmic and melodic information can have a facilitative effect on language recall if listeners are able to familiarize themselves with the melody.

Other research has also shown conflicting effects of song on word recall and learning. Racette and Peretz (2007) found that sung lyrics resulted in *poorer* word recall than spoken lyrics, suggesting that sung words made word learning more difficult. However, it should be noted that both the sung and spoken conditions included the melody in the background – this experimental paradigm could have unintentionally resulted in a facilitative effect of the melody on speech (i.e., paired-associate learning), but interference from a dual task when speech was sung to the background melody. Neither Rainey and Larsen (2002) nor Tamminen, Rastle, Darby, Lucas, and Williamson (2017) found an immediate influence of sung material on word recall or novel word learning, respectively. However, one week later, both studies found evidence that the sung material had been encoded better than the spoken material, suggesting a stronger representation in long-term memory. A similar result was found in older adults and adults with Alzheimer's Disease (AD; Moussard, Bigand, Belleville, & Peretz, 2014). Immediate recall of sung lyrics was comparable to spoken lyrics for controls, and worse for sung lyrics in AD. After a 10-minute delay, both groups showed enhanced memory for sung words compared to spoken words. The authors suggested that the immediate disadvantage for sung words in the AD patients could stem from the fact that the melodic information in the song represented a dual task. However, after a delay, a benefit from the melody could be observed, perhaps due to a stronger memory representation for the sung words (enhanced memory for word order after sung words compared to spoken words also occurs in individuals with multiple sclerosis; Thaut et al., 2014; Thaut, Peterson, Sena, & Mcintosh, 2008). Research has also shown that sung words enhance foreign language learning in both children (Good et al., 2015) and adults (Ludke et al., 2014; Rukholm et al., 2018), suggesting an enhanced memory representation for these words. Further, Schön et al. (2008) showed that novel, sung syllables were learned more effectively in a statistical learning paradigm compared to novel spoken syllables, and that this effect was enhanced when syllables were consistently matched with particular tones.

To summarize, two separate lines of research have found that IDS and song may enhance word learning. However, the influence of IDS and song on language learning has not yet been directly compared under the same experimental conditions. Thus, it is still unclear whether sung speech (song) facilitates language learning more than a speech register that includes some melodic features (i.e., IDS). Although song and naturally-produced IDS differ in various factors besides musicality (e.g., acoustic, emotional, and linguistic factors – e.g., Trainor & Desjardins, 2002), song tends to be more musical than IDS. Thus, an investigation of this issue can improve our understanding of the relationship between the degree of musicality and language learning. Furthermore, it is still unclear whether song and IDS can facilitate both immediate word learning and long-term (after 24 hours) memory of the newly learned words. Comparing these effects can help identify plausible mechanisms by which music exposure benefits language learning. This study investigated whether adult monolingual English-speaking participants learn words in a new language (Chinese) better when the words were presented in song and IDS than when they were presented in ADS. Adult participants were recruited to examine whether the benefits of IDS extend beyond child language acquisition, and beyond its capacity to direct attention (Ma et al., 2011, Golinkoff & Alioto, 1995). No participants had received prior instruction in Mandarin Chinese.

This study differs from previous research in several ways. First, the participants learned new words presented in three vocalization types (ADS, IDS, and song), allowing us to determine whether the highly musical stimuli (i.e., song) facilitate word learning and long-term memory more than the less musical stimuli (i.e., IDS). Second, each participant was tested in a word learning task immediately after exposure and then one day later to assess both immediate word learning and long-term memory of the learned words. Third, this study examined the influence of song and IDS on two word learning processes – word recognition (recognition of an earlier learned word) and word association (mapping a label to a referent), thus allowing us to understand the potential mechanisms of the linguistic benefits of song and IDS.

Furthermore, this study presented the Chinese words using both familiar and unfamiliar melodies, allowing us to determine whether participants' language learning performance was influenced by their familiarity with the melody used. Finally, this study also recorded participants' confidence in their answers in the language learning task in an effort to determine whether the learners were more certain of their answers in IDS and song than in ADS.

Experiment 1

Experiment 1 examined whether immediate word learning and long-term memory for the newly learned words would be facilitated by IDS and song. In Experiment 1, English-speaking participants were exposed to Chinese target words, of which they had no prior knowledge, embedded in a Chinese sentence that was presented in ADS, IDS, or with a sung melody. Each participant completed two tasks: A word learning task (on the first day) and a long-term memory task (on the second day – approximately 24 hours later).

Participants

The participants were 42 undergraduates (*Mean* = 19.26 years, range = 17 – 22 years; 38 females) at the University of Arkansas. According to a questionnaire they completed prior to the study, all participants were native speakers of English and were non-musicians, and had received no formal education of Chinese language.

Stimuli

Since each experimental session consisted of three blocks (i.e., the ADS, IDS, and song blocks) and each block had eight words, 24 slides (eight for each block) featuring prototypical exemplars of common objects, each labeled with a disyllabic Chinese word were used (Table 1). Items were selected to appear in one block based on the following criteria: their names in Chinese were minimally confusable according to the native speaker of Chinese who designed the study, and there were no rhyming words or words that began with the same initial sounds. In the word learning task, the Chinese ADS, IDS, and sung words were embedded in Chinese sentence frames produced in either ADS, IDS, or with sung melodies.

-------------------- Insert Table 1 about here ----------------------

A female native speaker of Mandarin Chinese, who was not a professional musician but was comfortable singing, produced the auditory stimuli. Each sentence and word was produced in four versions: ADS, IDS, sung to a familiar melody (i.e., Happy Birthday), and sung to an unfamiliar melody (e.g., What a Beautiful Jasmine [Chinese folk music]), respectively (Figure 1). For each vocalization type, the speaker produced 48 sentences (2 sentences for each Chinese word x 8 words in each block x 3 blocks), in each of which the target word appeared at the sentential final position. These sentences were used to expose the participants to the association between a label and a referent. In addition, for each sound version, the speaker produced 24 words in insolation (8 words in each block x 3 blocks), which were used in the word recognition task in the test.

--------------------------- Insert Figure 1 about here ----------------------------

For the ADS stimuli, the speaker was instructed to "speak naturally as if talking to an adult". For the IDS stimuli, the speaker was instructed to speak "as if talking to an infant"; meanwhile, she was provided with a picture of an infant to talk to throughout her production of

IDS, which served a prop to assist her in imagining talking to an infant. The speaker was *not* instructed to produce exceptionally happy IDS. For the sung vocalizations, the speaker was instructed to “sing naturally as if singing to a friend”, and she was *not* instructed to 1) exaggerate her emotional expressiveness in singing or 2) produce infant-directed singing. Due to the sentence/word length, each sentence/word carried only a *section* of a melody. The speaker was instructed to first produce a sentence (the target word appeared at the sentential final position) to a section of a melody and then to reproduce the target word in isolation using the *sub-section* of the melody, to which the target word was produced in that sentence. This ensured that the target word carried the same melody in a sentence and in isolation. Furthermore, for each block, the speaker was instructed to produce the 16 sentences (2 sentences for each Chinese word x 8 words in each block) using the sequential sections of a melody, which maximized the maintenance the flow of a melody between sentences and helped the participants recognize the familiar melody.

For each vocalization type, this study used 1) 48 complete sentences (2 sentences for each Chinese word x 8 words in each block x 3 blocks) in the training, with the target words appearing at the sentential final position in each sentence, and 2) 24 words (8 words in each block x 3 blocks) that were presented in isolation in the test in Experiments 1 and 2 and throughout Experiment 3. Then, we compared the acoustic features of the 48 sentences and 24 words across vocalization types, respectively.

Duration. Since the speaker was instructed to produce the vocalizations naturally, word length was not controlled across vocalization types. For the 48 sentences, paired sample *t*-tests compared sentence duration across the vocalization types. Adjusted *p* values were used for multiple comparisons throughout the study. Results showed that sentence duration significantly increased from IDS (M = 2.290 sec, SD = .391) to the familiar melody (M = 3.64 sec, SD = .45;

t(47) = 18.31, *p* < .001, Cohen's d = 2.64) and then to the unfamiliar melody (*M* = 4.23 sec, *SD* = .23; *t*(47) = 8.94, *p* < .001, Cohen's d =1.29). To our surprise, sentence duration did not differ between ADS (*M* = 2.287 sec, *SD* = .387) and IDS (*t*(47) = .13, *p* = .90). Nevertheless, this is consistent with a recent study that shows that Chinese-speaking mothers may *not* slow down in IDS (Han, de Jong, & Kager, 2018). For the 24 words, separate paired sample *t*-tests found that word duration significantly increased from ADS (*M* = .93 sec, *SD* = .12) to IDS (*M* = .99 sec, *SD* = .10; *t*(23) = 4.51, *p* < .001, Cohen's d = .92) and then to the unfamiliar melody (*M* = 1.74 sec, *SD* = .13; *t*(23) = 22.96, *p* < .001, Cohen's d = 4.69). However, word duration did not differ between the unfamiliar and familiar melodies (*M* = 1.75 sec, *SD* = .16; *t*(23) = .12, *p* = .90).

Mean pitch level. For each sentence and word, we obtained the mean pitch level using Praat (Pitch – Get Pitch). Then, paired sample *t*-tests compared the mean pitch level across the vocalization types. For the 48 sentences, separate paired sample *t*-tests showed that the mean pitch level increased from ADS (*M* = 195.96 Hz, *SD* = 10.91) to the unfamiliar (*M* = 238.65 Hz, *SD* = 26.67; *t*(47) = 11.05, *p* < .001, Cohen's d = 1.59) and familiar (*M* = 242.20 Hz, *SD* = 26.71; *t*(47) = 11.06, *p* < .001, Cohen's d = 1.60) melodies; then it increased from the familiar melody to IDS (*M* = 289.68 Hz, *SD* = 28.50; *t*(47) = 8.74, *p* < .001, Cohen's d = 1.26). The findings are consistent with research showing that IDS tends to have a higher pitch level than ADS (e.g., Fernald & Kuhl, 1987; Kitamura & Burnham, 2003; Trainor & Desjardins, 2002). Furthermore, the mean pitch level did not differ between the unfamiliar and familiar melodies (*t*(47) = .49, *p* = .63). For the 24 words, separate paired sample *t*-tests showed that the mean pitch level did not differ between ADS (*M* = 201.97 Hz, *SD* = 19.33) and the unfamiliar melody (*M* = 216.21 Hz, *SD* = 38.66; *t*(23) = 1.69, *p* = .11); furthermore, the mean pitch level significantly increased from the unfamiliar melody (*M* = 216.21 Hz, *SD* = 38.66) to the familiar melody (*M* = 266.62 Hz, *SD* =

11.32; $t(23) = 6.59$, $p < .001$, Cohen's d = 1.35) and then marginally significantly increased to IDS ($M = 288.90$ Hz, SD = 46.11; $t(23) = 2.31$, $p = .03$, Cohen's d = .47; a significance cutoff level of .017 was used).

Pitch range ratio. For each sentence and word, we obtained the maximum and minimum pitch levels using Praat (Pitch – Get Maximum/Minimum Pitch), and calculated its pitch range ratio by dividing the maximum pitch level over the minimum pitch level. This ratio is a valid measure of pitch variability given that the fundamental frequency of a voice has a logarithmic relationship to the perceived pitch of that voice (Ma, Fiveash, & Thompson, 2019). For the 48 sentences, separate paired sample *t*-tests showed that the pitch range ratio did not differ between the unfamiliar ($M = 1.91$, $SD = .55$) and familiar ($M = 2.05$, $SD = .68$) melodies ($t(47) = 1.15$, $p = .26$); furthermore, the pitch range ratio increased from the familiar melody to ADS ($M = 2.97$, $SD = 1.15$; $t(47) = 5.32$, $p < .001$, Cohen's d = .77) and then to IDS ($M = 3.57$, $SD = 1.19$; $t(47) = 3.08$, $p = .003$, Cohen's d = .44), which is consistent with research showing that IDS tends to have a wider pitch variation range than ADS (e.g., Cooper & Aslin, 1990; Fernald & Kuhl, 1987; Van de Weijer, 1997). For the 24 words, separate paired sample *t*-tests found that the pitch range ratio did not differ between the familiar ($M = 1.22$, $SD = .15$) and unfamiliar ($M = 1.23$, $SD = .17$) melodies ($t(47) = .15$, $p = .89$); furthermore, the pitch range ratio increased from the unfamiliar melody to ADS ($M = 1.70$, $SD = .59$; $t(23) = 3.74$, $p = .001$, Cohen's d = .76) and then to IDS ($M = 2.44$, $SD = 1.33$; $t(23) = 3.07$, $p = .005$, Cohen's d = .63).

Formant frequencies of vowels. We analyzed the formant of the monophthongs (i.e., /i/, /u/, /ɔ/) that are not adjacent to nasals, as nasalization interacts with vowel properties (e.g., Beddor, 1993). Notably, the target words used in the word learning task in this study did not contain the monophthong of /a/. Thus, this analysis included only eight words (fei1j**i**1, p**i**2xie2, t**u**2sh**u**1, qian1b**i**3, w**u**1gui1, dian4t**i**1, m**o**2g**u**1, b**o**1cai4), which contained four /i/s, four /u/s, and 2 /ɔ/s. F1 and F2 values of these vowels were obtained using Praat (Formant – Get F1/F2). IDS production has been described as mimicking a shorter vocal tract (Kent, 1976; 1992) that may result into enhanced formant frequencies (Peterson & Barney, 1952). It appears that the F1 and F2 values tend to bes smaller in ADS than in IDS and song (Table 2). However, caution should be taken in interpreting the generalizability of this finding because of the small sample size of vowels included in this analysis.

--------------------- Insert Table 2 about here ------------------------

Perceptual rating. To validate the auditory stimuli, 25 adult native English speakers completed an auditory perception task. On each trial, the participant heard a sentence or a word and was asked to rate the ADS-, IDS-, and song-likeness of the sound respectively, using a scale from 1 – 7 (1 = it does not sound like it at all, 7 = it sounds like it exactly). The presentation order of the auditory stimuli was randomized for each participant and counterbalanced across participants. Then, the average ADS-, IDS-, and song-likeness ratings were calculated for each vocalization type (ADS, IDS, the familiar melody, and the unfamiliar melody). The auditory stimuli were rated as more like what they were intended to be than the other vocalization types (Table 3, Figure 2). These findings verified the vocalization type of the auditory stimuli used in this study. Furthermore, paired sample *t*-tests showed that song-likeness ratings decreased from song vocalizations to IDS vocalizations ($t(24) = 7.87$, $p < .001$, Cohen's d = 1.61) then to ADS vocalizations ($t(24) = 6.31$, $p < .001$, Cohen's d = 1.29), suggesting that musicality decreased from songs to IDS then to ADS.

-------------------- Insert Table 3 about here ----------------------

-------------------- Insert Figure 2 about here ----------------------

Procedure

Participants completed a word learning task on the first day and a long-term memory task on the second day individually in a quiet room.

The word learning task consisted of three blocks (an ADS block, an IDS block, and a song block), each containing a familiarization phase and a test phase. Within each block, the experimental procedure was almost identical to that of Golinkoff and Alioto (1995). In the familiarization phase, participants first read the following instruction:

> *The purpose of this study is to examine the ways in which adults learn new words in a foreign language. There are three blocks. In each block, you will first see 8 slides of common, everyday objects. Along with these slides will be an audio of a speaker naming and talking about the objects in the slides in Chinese. Obviously, you will not be expected to understand the Chinese. We simply ask that you look at each slide and focus closely on what is being said. After the slide presentation, a test session will follow.*

The eight slides were presented as a synthesized video on a computer screen. Each slide was shown for a total of 12 seconds, while the accompanying audio, which was presented by the computer binaurally through headphones, started approximately two seconds after the onset of each slide. Thus, each slide was accompanied by 1) two seconds of silence at the beginning of the slide, 2) a pair of sentences/lyrics (see Table 1 for details) presented in ADS, IDS, or song depending on whether it was an ADS, IDS, or song block, and 3) approximately one to five seconds of silence towards the end of the slide depending on the length of the auditory stimuli. In addition, there was a 4-second silent, dark screen before the onset of the next slide. Thus, the total duration of the video was 124 seconds (eight slides and seven 4-second silent, dark screens) in each block. In order to examine the effect of familiarity of song melody on word learning, the eight slides were equally divided into two sets in the song block. Thus, four slides were accompanied by familiar melody vocalizations (the Familiar set) while the other four were accompanied by unfamiliar melody vocalizations (the Unfamiliar set). In the song block, we counterbalanced the melody assignment and the presentation order of the familiar and unfamiliar sets across participants; however, the presentation order of the four slides remained fixed within each set across participants in order to maintain the flow of melody across slides and therefore help the participants recognize the familiar melody. To keep the manipulation of presentation order of slides consistent across blocks, the eight slides in the ADS and IDS blocks were also equally divided into two sets. As in the song block, the presentation order of the four slides was fixed within each set and the presentation order of the two sets within the ADS and IDS blocks was counterbalanced across participants. Finally, the presentation order of the three blocks was also counterbalanced across participants (Table 4).

-------------------- Insert Table 4 about here --------------------

In each block, the *test phase* started immediately after the presentation of the eight slides and their accompanying Chinese audio. On each test trial, participants were provided with three different *written* English word choices. These choices, all from the pool of eight stimulus items, were randomly selected with the constraint that each alternative appear an equal number of times on the sheet and in each position (first, second, and third). Participants were then presented with an audio of the Chinese speaker reading a target word, which was presented only once, in isolation without the sentence frame, and in either ADS, IDS, or with a song melody corresponding to the experimental block. The presentation order of the target words was identical to their occurrence in the familiarization phase. On each trial, participants were instructed to 1) listen to the Chinese target word, and 2) select its English equivalent from the three choices given for that item to demonstrate word learning; if they were unsure, participants were told to make their best guess, and 3) rate their confidence in their answer, using a scale from 0 – 7 (0 = not confident at all, 7 = highly confident), by following the instruction, “Please indicate how confident you are in your answer.” When the test phase was completed, the next word learning block started.

The same group of participants were administered a long-term memory task approximately 20-28 hours after the completion of the word learning task. Like the word learning task, the long-term memory task consisted of an ADS block, an IDS block, and a song block. However, in the long-term memory task, the participants received *only* the test phase, which was identical to the one used in the word learning task. This test phase examined participants’ long-term memory of the words learned in the word learning task.

Results and Discussion

Each participant’s rate of correct responses and average confidence scores were calculated for each vocalization type (IDS, ADS, the familiar melody, and the unfamiliar melody) in both the word learning and long-term memory tasks, respectively. Since the aim of this study was to determine if IDS and song can facilitate word learning and memory, participants’ word learning and long-term memory performance were analyzed separately below.

Word learning. A paired sample t-test showed that the rate of correct responses did not differ between the familiar (M = .58, SD = .32) and unfamiliar (M = .61, SD = .29; $t(41)$ = .51, p = .61) melody sets in the song block. Thus, the sung words were analyzed together. Then, we asked whether the participants succeeded in learning the items above the level of chance (33%). Separate one-sample t-tests showed that the rate of correct responses was *marginally* significantly above chance in the ADS block (M = .41, SD = .22; $t(41)$ = 2.33, p = .025, Cohen’s d = .72; a significance cutoff level of .017 was used), and significantly above chance in both the IDS (*M* = .62, *SD* = .29; $t(41) = 6.52$, $p < .001$, Cohen's d = 2.01) and song blocks (*M* = .60, *SD* = .26; $t(41) = 6.65$, $p < .001$, Cohen's d = 2.05). Then, to determine if word learning performance differed across vocalization types, a one-way repeated measures ANOVA was conducted to analyze the rate of correct responses across the three vocalization types. A main effect of vocalization type emerged ($F(2, 82) = 13.80$, $p < .001$, $\eta_p^2 = .25$). Post-hoc analyses were conducted through separate paired sample *t*-tests. Results showed that the rate of correct responses was significantly lower in the ADS block than in the IDS ($t(41) = 4.78$, $p < .001$, Cohen's d = .74) and song ($t(41) = 4.14$, $p < .001$, Cohen's d = .64) blocks, but did not differ between the IDS and song blocks ($t(41) = .56$, $p = .58$). The findings suggest that IDS and song facilitated word learning, and that their facilitative strength did not differ from each other (Figure 3).

------------------------ Insert Figure 3 about here ----------------------

Long-term memory. Again, a paired sample *t*-test showed that the rate of correct responses did not differ between the familiar (*M* = .48, *SD* = .42) and unfamiliar (*M* = .44, *SD* = .31; $t(41) = .53$, $p = .60$) melody sets. Thus, the sung words were analyzed together. To determine whether the participants could reliably remember the items learned on the first day above the level of chance (33%), separate one-sample *t*-tests compared the rate of correct responses against chance. Results showed that the rate of correct responses did not differ from chance in the ADS block (*M* = .34, *SD* = .19; $t(41) = .29$, $p = .78$), but was significantly above chance in both the IDS (*M* = .46, *SD* = .30; $t(41) = 2.72$, $p = .009$, Cohen's d = .84) and song blocks (*M* = .47, *SD* = .28; $t(41) = 3.22$, $p = .003$, Cohen's d = .99), suggesting that participants reliably remembered the IDS and sung words but not the ADS words. To determine if memory performance differed across vocalization type, a one-way repeated measures ANOVA analyzed the rate of correct responses across the three vocalization types. A main effect of vocalization type emerged ($F(2, 82) = 4.23$, $p = .02$, $\eta_p^2 = .09$). Separate paired sample *t*-tests showed that the rate of correct responses in the ADS block was marginally significantly lower than that in the IDS block ($t(41) = 2.31$, $p = .026$, Cohen's d = .36; a significance cutoff level of .017 was used), and significantly lower than that in the song block ($t(41) = 3.04$, $p = .004$, Cohen's d = .47). However, the rate of correct responses did not differ between the IDS and song blocks ($t(41) = .27$, $p = .79$). These findings suggest that both IDS and song facilitated word long-term memory compared to ADS, and that their facilitative strength did not differ from each other.

Confidence scores. Separate paired sample *t*-tests showed that in the song block, confidence scores did not differ between the familiar and unfamiliar melody sets in either the word learning task ($t(41) = .54$, $p = .59$) or the long-term memory task ($t(41) = 0.80$, $p = .43$). Thus, the sung words were analyzed together in each task (Table 5). Separate one-way repeated measures ANOVAs were conducted to analyze the average confidence scores across the three vocalization types in the word learning and long-term memory tasks. The main effect of vocalization type did not approach significance in either the word learning task ($F(2, 82) = .11$, $p = .90$) or the long-term memory task ($F(2, 82) = .70$, $p = .50$), suggesting that confidence scores did not differ according to vocalization type.

------------------------ Insert Table 5 about here ------------------------

Before discussing the results, one concern about the auditory stimuli used in this study should be mentioned. This study did not control word length in order to ensure the naturalness of the vocalizations used. As Kilgour, Jakobson, and Cuddy (2000) found that the benefit of music on text recall could be due to duration, an important question then arises: Is it possible that the current findings were merely driven by sentence/word length that differed across vocalization types? This possibility is unlikely since neither the word learning nor the memory performance differed between the familiar and unfamiliar melody sets or between the IDS and song conditions, although word length was longer in the unfamiliar melody than in the familiar melody and was longer for songs than for IDS. Furthermore, although word length did not differ between ADS and IDS, word learning and memory performance were better for IDS than for ADS.

Experiment 1 showed that IDS and song facilitated immediate word learning and long-term memory of the newly learned words. Furthermore, the facilitative strength of IDS and song did not differ from each other. Thus, although additional musicality facilitated word learning and memory compared to ordinary speech, there was no increased benefit to highly musical stimuli (song) in comparison to less musical stimuli (IDS). In addition, confidence scores did not differ across vocalization types for either the word learning task or the long-term memory task. Finally, participants' word learning and long-term memory performance was *not* influenced by their familiarity with the melody used. Since the familiar and unfamiliar melodies sets did not differ in word learning performance, long-term memory performance, or confidence scores throughout this study, they will be reported as combined throughout the paper.

How did IDS and song facilitate participants' performance? In order to succeed in Experiment 1, participants needed to accomplish 1) a word segmentation-association process *in the familiarization phase*, during which they needed to segment the target word from its sentence frame and then form an association between the target word and its accompanying image and 2) a word recognition process *in the test phase* for both the word learning and long-term memory tasks. Since the target words in the test phase of the IDS and song blocks were presented in IDS and song respectively, the benefit of IDS and song may be due to their facilitative effect on the word segmentation-association process or/and the word recognition process. Thus, it is still unclear whether IDS and song facilitate the word segmentation-association process. This issue was examined in Experiment 2.

Experiment 2

Experiment 2 examined whether IDS and song could facilitate the word segmentation-association process. Although Experiment 1 found that IDS and song facilitated participants' performance, the results did not allow us to determine whether IDS and song enhanced the word segmentation-association process in the familiarization phase or/and the word recognition process in the test, since the target words the IDS and song blocks were also presented in IDS and song at test. We predicted that if IDS and song facilitate the word segmentation-association process, better word learning and long-term memory should be observed in the IDS and song blocks compared to the ADS block, even if the target words are presented only in ADS at test.

Participants

The participants were 42 undergraduates at the University of Arkansas (Mean = 19.40 years, range = 18 – 23 years; 41 females) who had not participated in Experiment 1. Participant recruitment followed the same procedure and standard as Experiment 1.

Stimuli and procedure

The stimuli and procedure of Experiment 2 were almost identical to those of Experiment 1. Like Experiment 1, the auditory stimuli in the familiarization phase were presented in ADS, IDS, and song in the three blocks respectively (ADS, IDS, song). However, the target words in the *test* phase were presented *only* in ADS for both the word learning and long-term memory tasks throughout Experiment 2.

Results and discussion

Data analyses followed the procedure of Experiment 1.

Word learning. We examined whether the participants succeeded in learning the items above the level of chance (33%). Separate one-sample *t*-tests showed that the rate of correct responses did not differ from chance in the ADS block ($M = .38$, $SD = .24$; $t(41) = 1.46$, $p = .15$), but was significantly above chance in both the IDS ($M = .55$, $SD = .27$; $t(41) = 5.20$, $p < .001$, Cohen's d = 1.60) and song blocks ($M = .58$, $SD = .25$; $t(41) = 6.54$, $p < .001$, Cohen's d = 2.02). To determine if word learning performance differed across vocalization types, a one-way repeated measures ANOVA analyzed the rate of correct responses across the three vocalization types. A main effect of vocalization type emerged ($F(2, 82) = 9.30$, $p < .001$, $\eta_p^2 = .19$). Post-hoc analyses were conducted through separate paired sample *t*-tests. Results showed that the rate of correct responses was lower in the ADS block than in the IDS ($t(41) = 3.05$, $p = .004$, Cohen's d = .47) and song ($t(41) = 3.88$, $p < .001$, Cohen's d = .60) blocks, but did not differ between the IDS and song blocks ($t(41) = .67$, $p = .51$) (Figure 3). It is important to note that, although the stimulus sentences (and the target words contained therein) were presented in ADS, IDS, or song depending on the block in the familiarization phase, the target words were presented in isolation only in ADS in the test phase. Thus, the IDS and song blocks were presumably at a disadvantage at test, since they were no longer produced with the exaggerated musical properties that were available in the familiarization phase. Yet, participants still performed significantly better in the IDS and song blocks than in the ADS block.

Long-term memory. Separate one-sample *t*-tests compared the rate of correct responses against chance (33%). Results showed that the rate of correct responses did not differ from chance in the ADS block ($M = .36$, $SD = .27$; $t(41) = .79$, $p = .44$), but was significantly above chance in the IDS ($M = .46$, $SD = .23$; $t(41) = 3.47$, $p = .001$, Cohen's d = 1.07) and song blocks ($M = .47$, *SD* = .21; *t*(41) = 4.16, $p < .001$, Cohen's d = 1.28), suggesting that participants reliably remembered the IDS and sung words but not the ADS words. A one-way repeated measures ANOVA then analyzed the rate of correct responses across vocalization types. A main effect of vocalization type emerged ($F(2, 82) = 3.59$, $p = .03$, $\eta_p^2 = .08$), suggesting that memory performance differed across vocalization type. Then, separate paired sample *t*-tests revealed that the rate of correct responses in the ADS block was marginally significantly lower than that in the IDS block ($t(41) = 2.02$, $p = .05$, Cohen's d = .31; a significance cutoff level of .017 was used) and significantly lower than that in the song block ($t(41) = 2.65$, $p = .01$, Cohen's d = .41), but did not differ between the IDS and song blocks ($t(41) = .28$, $p = .78$) (Figure 3).

Word learning and long-term memory in Experiment 1 vs. Experiment 2. Experiments 1 and 2 used the identical experimental procedure except that in the test phase, the target words were presented only in ADS in Experiment 2. Did participants' performance differ between Experiments 1 and 2? Separate mixed ANOVAs for participants' word learning and long-term memory performance were run, with the within-subject factor of Vocalization Type (ADS, IDS, song), and the between-subjects factor of Experiment (1, 2). The analysis on word learning performance showed that the only significant result was the main effect of Vocalization Type ($F(2, 164) = 22.31$, $p < .001$, $\eta_p^2 = .21$). The main effect of Experiment ($F(1, 82) = .92$, $p = .34$) and the Experiment x Vocalization Type interaction ($F(2, 164) = .36$, $p = .70$) did not approach significance. Similarly, the analysis on participants' long-term memory performance showed that the only significant result was the main effect of Vocalization Type ($F(2, 164) = 7.80$, $p = .001$, $\eta_p^2 = .09$). Again, the main effect of Experiment ($F(1, 82) = .03$, $p = .86$) and the Experiment x Vocalization Type interaction ($F(2, 164) = .11$, $p = .90$) did not approach significance. Thus, participants' performance did not differ between Experiments 1 and 2, suggesting that presenting the target word only in ADS at test did not hinder participants' performance in the word learning or the long-term memory task.

Confidence scores. Separate one-way repeated measures ANOVAs were conducted to analyze the average confidence scores across the three vocalization types in the word learning and long-term memory tasks. The main effect of vocalization type did not approach significance in either the word learning task ($F(2, 82) = .45$, $p = .64$) or the long-term memory task ($F(2, 82) = .25$, $p = .78$), suggesting that confidence scores did not differ according to vocalization type.

Experiment 2 showed that IDS and song facilitated word learning and long-term memory even when the target words were presented only in ADS at test. Furthermore, participants' performance did not differ between Experiments 1 and 2. These findings suggest that presenting words in IDS and song did not facilitate the word recognition process at test. Thus, the benefit of IDS and song should be due to their facilitative effects on the segmentation-association process. However, since the word segmentation-association process consists of two sub-processes, word segmentation (segmenting the target word from the sentence frame) and word association (mapping a label to a referent), it is unclear whether IDS and song facilitated word association and/or word segmentation. Experiment 3 aimed to isolate the influence of IDS and song on *word association*.

Experiment 3

Experiment 3 examined whether the use of IDS and song can facilitate word association. We predicted that if IDS and song could facilitate the word association process, better word learning and long-term memory should be observed in the IDS and song blocks compared to the ADS block, even when the words are presented *in isolation* in the familiarization phase, which does not require word segmentation.

Participants

The participants were another group of 42 undergraduates at the University of Arkansas who did not participate in either of the previous studies (Mean = 19.71 years, range = 18 – 25 years; 40 females). Participant recruitment followed the same procedure and standard as Experiments 1 and 2.

Stimuli and procedure

The stimuli and procedure of Experiment 3 were almost identical to those of Experiment 2 except that the target words were presented twice *in isolation* without sentence frames in the familiarization phase. Thus, successful word learning only required word association in this experiment. Furthermore, since the auditory stimuli did not have sentence frames in the familiarization phase, the duration of each slide was only 8 seconds. The total duration of the movie in the familiarization phase was 92 seconds (eight slides and seven 4-second silent, dark screens) in each block. The test phases in the word learning phase and the long-term memory task were identical to Experiment 2. All words were presented in ADS in the test. We predicted that if IDS and song facilitate the word association process, better word learning and long-term memory should be observed in the IDS and song blocks than in the ADS block.

Results and discussion

Data analyses followed the procedure of Experiment 1 and 2.

Word learning. Separate one-sample *t*-tests showed that the rate of correct responses was significantly above chance (33%) in the ADS (M = .65, SD = .25; $t(41) = 8.29$, $p < .001$, Cohen's d = 2.56), IDS (M = .74, SD = .26; $t(41) = 10.24$, $p < .001$, Cohen's d = 3.16), and song blocks (M = .78, SD = .25; $t(41) = 11.53$, $p < .001$, Cohen's d = 3.56), suggesting that participants learned the target words regardless of the vocalization type. Then, a one-way repeated measures ANOVA was conducted to determine whether the rate of correct responses differed across vocalization types. A main effect of vocalization type emerged ($F(2, 82) = 6.14$, $p = .003$, $\eta_p^2 = .13$). Post-hoc analyses were conducted through separate paired sample *t*-tests. Results showed that the rate of correct responses in the ADS block was marginally significantly lower than that in the IDS block ($t(41) = 2.15$, $p = .038$, Cohen's d = .33; a significance cutoff level of .017 was used) and significantly lower than that in the song block ($t(41) = 3.53$, $p = .001$, Cohen's d = .54). However, the rate of correct responses did not differ between the IDS and song blocks ($t(41) = 1.20$, $p = .24$) (Figure 3).

Long-term memory. Separate one-sample *t*-tests showed that the rate of correct responses was significantly above chance (33%) in the ADS ($M = .52$, $SD = .23$; $t(41) = 5.28$, $p < .001$, Cohen's d = 1.63), IDS ($M = .62$, $SD = .27$; $t(41) = 6.91$, $p < .001$, Cohen's d = 2.13), and song ($M = .65$, $SD = .31$; $t(41) = 6.80$, $p < .001$, Cohen's d = 2.10) blocks, suggesting that the participants remembered the earlier learned Chinese words regardless of vocalization type. A one-way repeated measures ANOVA analyzed the rate of correct responses across vocalization types. A main effect of vocalization type emerged ($F(2, 82) = 3.54$, $p = .03$, $\eta_p^2 = .08$), suggesting that long-term memory performance differed across vocalization type. Separate paired sample *t*-tests showed that the rate of correct responses in the ADS block was marginally significantly lower than that in the IDS block ($t(41) = 2.04$, $p = .048$, Cohen's d = .31; a significance cutoff level of .017 was used) and significantly lower than that in the song block ($t(41) = 2.65$, $p = .011$, Cohen's d = .41). However, the rate of correct responses did not differ between the IDS and song blocks ($t(41) = .65$, $p = .52$) (Figure 3).

Word learning and long-term memory in Experiment 3 vs. Experiment 2. Experiments 2 and 3 used the identical procedure except that the fast mapping process in Experiment 2 required two sub-processes (word segmentation and word association) while Experiment 3 required only the word association sub-process. Separate mixed ANOVAs for participants' word learning and long-term memory performance were run, with the within-subject factor of Vocalization Type (ADS, IDS, song), and the between-subjects factor of Experiment (2, 3). The analysis on word learning performance revealed main effects of Vocalization Type ($F(2, 164) = 15.29$, $p < .001$, $\eta_p^2 = .16$) and Experiment ($F(1, 82) = 26.69$, $p < .001$, $\eta_p^2 = .25$), suggesting that word learning performance differed across vocalization types and between experiments. However, the Vocalization Type x Experiment interaction was not significant ($F(2, 164) = .85$, $p = .43$), suggesting that participants had a similar pattern of word learning performance between experiments. The analysis on long-term memory performance revealed a similar pattern of results. The main effects of Vocalization Type ($F(2, 164) = 7.02$, $p = .001$, $\eta_p^2 = .08$) and Experiment ($F(1, 82) = 16.86$, $p < .001$, $\eta_p^2 = .17$) were significant, but the Vocalization Type x Experiment interaction was not significant ($F(2, 164) = .10$, $p = .90$).

To decompose the main effects of Vocalization Type, separate paired sample *t*-tests compared participants' performance (with the data of Experiments 2 and 3 combined) across vocalization types. The analyses on word learning performance showed that the rate of correct responses was lower in the ADS ($M = .51$, $SD = .28$) than in the IDS ($M = .64$, $SD = .28$; $t(83) = 3.72$, $p < .001$, Cohen's d = .41) and song ($M = .68$, $SD = .27$; $t(83) = 5.23$, $p < .001$, Cohen's d = .57) blocks, but did not differ between the IDS and song blocks ($t(83) = 1.32$, $p = .19$). The analyses on long-term memory performance showed a similar pattern of results. The rate of correct responses was lower in the ADS ($M = .44$, $SD = .26$) than in the IDS ($M = .54$, $SD = .26$; $t(83) = 2.89$, $p = .005$, Cohen's d = .32) and song ($M = .56$, $SD = .28$; $t(83) = 3.74$, $p < .001$, Cohen's d = .41) blocks, but did not differ between the IDS and song blocks ($t(83) = .69$, $p = .49$).

Then, to decompose the main effects of Experiment, separate independent samples *t*-tests compared participants' rates of correct responses between Experiments 2 and 3. The analyses on word learning performance showed that the rate of correct responses was higher in Experiment 3 than in Experiment 2 for the ADS block ($t(82) = 4.93$, $p < .001$, Cohen's d = 1.10), the IDS block ($t(82) = 3.18$, $p = .002$, Cohen's d = .72), and the song block ($t(82) = 3.72$, $p < .001$, Cohen's d = .80). The analyses on long-term memory performance revealed a similar pattern of results. The rate of correct responses was higher in Experiment 3 than in Experiment 2 for the ADS block ($t(82) = 2.84$, $p = .006$, Cohen's d = .64), the IDS block ($t(82) = 2.93$, $p = .004$, Cohen's d = .64), and the song block ($t(82) = 3.21$, $p = .002$, Cohen's d = .68). These findings suggest that participants performed the worst in ADS compared to IDS and song in both Experiments 2 and 3, and that performance was overall higher in Experiment 3 than in Experiment 2 across the three vocalization types. Thus, it appears that excluding the word segmentation process enhanced word learning performance.

Confidence scores. Separate one-way repeated measures ANOVAs were conducted to analyze the average confidence scores across the three vocalization types in the word learning and long-term memory tasks respectively. The main effect of vocalization type was significant in both the word learning ($F(2, 82) = 7.33$, $p = .001$, $\eta_p^2 = .15$) and long-term memory tasks ($F(2, 82) = 8.36$, $p < .001$, $\eta_p^2 = .17$). Post-hoc analyses were run using separate paired sample *t*-tests. The analyses of word learning performance showed that the confidence scores were significantly lower in the ADS block than in the IDS ($t(41) = 2.51$, $p = .02$, Cohen's d = .39) and song ($t(41) = 3.82$, $p < .001$, Cohen's d = .59) blocks, but did not differ between the IDS and song blocks ($t(41) = 1.34$, $p = .19$). The analyses of long-term memory showed that the confidence scores in the ADS block were marginally significantly lower than those in the IDS block ($t(41) = 2.19$, $p = .03$, Cohen's d = .34; a significance cutoff level of .017 was used) and significantly lower than those in the song block ($t(41) = 4.30$, $p < .001$, Cohen's d = .66). Again, confidence scores did not differ between the IDS and song blocks ($t(41) = 1.86$, $p = .07$). This pattern of results suggests that in the easier task (word association only), participants were more certain of their answers in IDS and song than in ADS.

Experiment 3 showed that IDS and song facilitated word learning and long-term memory even when the target words were presented in isolation in the familiarization phase. The findings suggest that IDS and song enhanced word association.

General discussion

This study examined if word learning and long-term memory in adults differed depending on whether words were presented in ADS, IDS, or song. In Experiment 1, English-speaking participants were exposed to Chinese target words embedded in a Chinese sentence that was presented in ADS, IDS, or to a sung melody. After each block, the target words were again presented in ADS, IDS, or to a sung melody, depending on the block, and participants were asked to decide which English word the Chinese word referred to. This test phase was presented again the following day. In Experiment 2, the same procedure was used except that all target words in the test phase were presented only in ADS to distinguish whether IDS and song facilitated the initial word segmentation-association process or word recognition. In Experiment 3, target words were presented in isolation (without a sentence frame), to investigate whether IDS and song facilitated the word association process. The results suggested that IDS and song facilitated immediate word learning and long-term memory. In addition, the strength of the facilitative effect did not differ between IDS and song, suggesting that the highly musical stimuli (song) did not facilitate word learning more than the less musical stimuli (IDS). Furthermore, neither the word learning performance nor the long-term memory differed between the familiar and unfamiliar melody sets, suggesting that learners' familiarity with the melody did not influence their language learning performance. Finally, despite their superior performance in the IDS and song blocks, the participants' confidence scores did not differ across vocalization types in Experiments 1 and 2. However, in Experiment 3, participants' confidence scores were higher in the IDS and song blocks than in the ADS block, suggesting that learners were more certain of their answers in IDS and song than in ADS when the word learning task was simplified.

Why did song and IDS facilitate word learning and long-term memory?

To succeed in the word learning and long-term memory tasks, participants needed to complete 1) two sub-processes in the familiarization phase (word segmentation, word association) of Experiments 1 and 2, and only the word association sub-process in the familiarization phase of Experiment 3, and 2) a word recognition process in the test phase for both the word learning and long-term memory tasks in all three experiments. An important question then arises: How did song and IDS facilitate immediate word learning and long-term memory of the learned words? There are four possible reasons.

First, the prosodic cues in song and IDS, such as such as pausing, pre-boundary lengthening, and intonation, are important for speech segmentation. Developmental research has shown that IDS facilitates speech segmentation in English-learning infants (Johnson & Seidl, 2008; Seidl, 2007; Seidl & Cristiá, 2008; Thiessen et al., 2005). In addition, a series of studies on adult French speakers showed that presenting made-up words in songs rather than speech enhances speech segmentation (Francois & Schön; 2010; Schön et al., 2008). It is possible that the co-occurrence of syllable change and pitch change in song and IDS facilitates phonological unit identification and thereby speech segmentation (Thiessen & Saffran, 2009). However, since the participants were required to complete both word segmentation and word association to succeed in Experiments 1 and 2, this study does not allow us to directly evaluate whether the use of music attributes in speech input facilitated the speech segmentation in Experiments 1 and 2. Nevertheless, the comparison between Experiments 2 and 3 showed that participants in Experiment 3 outperformed those in Experiment 2, suggesting that excluding the word segmentation process enhanced participants' performance, and still resulted in a benefit of IDS and song for the word association process. More importantly, participants' word learning and memory performance with ADS words was above chance in Experiment 3. It appears that IDS and songs helped participants with mapping and retention, but they were exceptionally helpful when segmentation was required in the task.

Second, the sentence and word duration in IDS and song may facilitate word segmentation and word association. Perhaps participants could more readily encode phonological information from song and IDS than from ADS because song and IDS involve an expanded vowel space (Kuhl et al., 1997). This study found that the formant frequencies were greater in song and IDS than in ADS, although the generalizability of this finding should be interpreted with caution since the data analysis only included eight words. Furthermore, past research showed that the benefit of music on text recall could be due to duration (Kilgour et al., 2000). In this study, the ADS stimuli had shorter duration than the sung stimuli. However, the current findings cannot be solely driven by the stimulus duration because participants' performance was greater with IDS words than with ADS words although word/sentence duration did not differ between IDS and ADS stimuli. In addition, participants' performance did not differ between the familiar and unfamiliar melody sets although word/sentence duration was greater in the unfamiliar melody set than in the familiar melody set.

Third, song and IDS prosody may also facilitate novel word learning because they elicit more attention in participants than ADS. A defining feature of song and IDS is their deviation from a speaker's mean acoustic profile, functioning to attract attention (Ma & Thompson, 2015; Ma et al., 2019). Research has shown that IDS and song are more attention-getting than ADS (e.g., Cooper & Aslin, 1990; Fernald, 1985; Papousek, Bornstein, Nuzzo, Papousek, & Symmes, 1990; Shenfield, Trehub, & Nakata, 2003; Soderstrom, 2007). Attention serves as the foundation for active language learning in children (Moore, Angelopoulos, & Bennet, 1997; Shneidman, Buresh, Shimpi, Knight-Schwarz, & Woodward, 2009; Tomasello & Farrar, 1986; Yu & Smith, 2012). However, since this study did not collect attention data, we cannot evaluate the influence of the song and IDS on the participants' attention during the study. Future research should record participants' attention through eye movement and reaction time.

Fourth, musical cues may be a helpful mnemonic aid. Using typical adult listeners, research shows that hearing verbal material in songs improves memory for lyrics (Calvert & Tart, 1993) and proper names (Rainey & Larsen, 2002), recall of text (Wallace, 1994) and words, as measured by behavioral and/or EEG methods (Moussard, Bigand, Belleville & Peretz, 2012; Peterson & Thaut, 2007; Rainey & Larsen, 2002). Hearing words in songs also enhances word recall in patients with multiple sclerosis (Thaut et al., 2014) and word recognition (Simmons-Stern, Budson, & Ally 2010) and memory (Simmons-Stern et al., 2012) in patients with Alzheimer's disease. Research also shows that compared with their non-musician counterparts, musician adults (Chan, Ho, & Cheung, 1998) and school-aged musician children perform better in tasks of verbal memory (Ho, Cheung, & Chan, 2003) and recalling unfamiliar spoken lyrics (Kilgour et al., 2000). In addition, weekly instrument training lasting for 12- and 18-months enhanced verbal memory in 7-year-old children (Roden, Kreutz & Bongard, 2012). The current results are consistent with this literature and suggest that word learning and long-term memory for words is enhanced for words presented with a sung melody. They also extend the literature to show that both IDS and song enhance word learning and long-term memory at a similar level.

These four explanations may not be mutually exclusive. Furthermore, the comparisons between Experiments 1 and 2 showed no between-experiment differences in either the word-learning task or the long-term memory task, suggesting that the linguistic benefit of song and IDS was not related to the word recognition process in the test phase in this study. Thus, song and IDS were shown to facilitate the *encoding* process, either through enhanced word segmentation and/or enhanced word association. The results of Experiment 3 provide direct evidence that song and IDS facilitated word association in particular; however, it is still likely that song and IDS also facilitated word segmentation. Future research could directly compare the effect of song and IDS on word segmentation in the same paradigm, as it has been separately shown that both song (Schön et al., 2008) and IDS (Thiessen et al., 2005) influence word segmentation (read Schön & François, 2011 for a review).

This study also found that confidence scores did not differ across vocalization type in Experiments 1 and 2, despite the participants' superior performance with sung and IDS words. These findings suggest that the participants might not be more certain of their answer in song and IDS than in ADS. Furthermore, it is also possible that the participants might be unaware of the benefits of IDS and song on language learning when the new words were embedded in sentence frames. This is consistent with the research showing that background music in supermarkets could influence customers' purchasing behavior without the participants being aware of the influence of the music (Milliman, 1982). However, when the word learning task was simplified in Experiment 3, confidence scores were higher with the song and IDS words than with the ADS words.

What is the relationship between the musicality of speech input and language learning?

This study found that song did not facilitate language learning more than IDS, although both facilitated learning more than ADS. There are two possible reasons. First, the relationship between the degree of musicality of the listeners' speech input and their language learning performance may not be linear. Just like the classical Weber-Fechner Law on sensory perception threshold (Fechner, 1859; Weber, 1834), perhaps the linguistic benefit of musical stimuli also has its absolute threshold (the minimal amount of musicality to influence language learning) and difference threshold (the minimum difference in musicality between two stimuli to influence language learning differently). Perhaps, both the IDS and song stimuli used in this study reached the absolute threshold required to influence language learning, but the difference between IDS and song stimuli did not reach the difference threshold. Based on this explanation, we would predict that the IDS and sung sounds that are less musical than typical IDS and song may not facilitate language learning compared to ADS. In addition, when the musicality difference between IDS and song is maximized, one may observe the difference in language performance between IDS and song. These predictions should be explored by future research.

Second, IDS and song may facilitate language learning through different mechanisms. Thus, the finding that IDS and song have the same facilitative strength on word learning may be specific to this particular context used in this study. IDS generally does not feature a consistent pitch to word mapping. Rather, the more regular speech presentation, prolonged vowels, and enhanced pitch range and pitch height may instead influence segmentation and enhance attention to the speech signal (Saint-Georges et al., 2013; Soderstrom, 2007). On the other hand, song generally includes a consistent mapping between pitch and words, as well as a consistent rhythmic structure that underlies the melody. Wallace (1994) suggested that the combination of melodic, rhythmic, and text structures can enhance both encoding and retrieval of words. The addition of a melody may also enhance chunking and grouping processes necessary for encoding, and provide both melodic and rhythmic cues to retrieval, as well as a cognitive framework in which to store the verbal information. Nevertheless, one key attribute distinguishing speech and music is the greater preponderance of repetition in music in comparison to speech (Margulis, 2013). Perhaps repeated exposure to the same pitch mappings and syllables in music would enhance memory compared to different pitch mappings in IDS. Supporting this possibility is the finding that the repetition of melody is essential for speech memory in music (Wallace, 1994).

This study also found that participants' performance did not differ between the familiar and unfamiliar melodies. It is possible that this finding is related to the experimental procedure, where each slide was accompanied by two seconds of silence at the beginning of the slide, approximately one to five seconds of silence towards the end of the slide depending on the length of the auditory stimuli, and a 4-second silent, dark screen before the onset of the next slide. These silences may hinder the listeners' recognition of the familiar melody. Furthermore, in Experiment 3, the target words were presented in isolation in the familiarization phase. Thus, the duration of the words may not have been sufficiently long to carry the recognizable melody, although participants could still distinguish among vocalization types of these words.

Did the participants encode musical attributes in their word representation?

This study found no evidence that song and IDS facilitated word recognition in the test phase, as word learning performance did not differ between Experiments 1 and 2. Did the participants encode the music information – realized as tonal and pitch variation information – in their word representation? There are three possibilities. First, the participants might not encode the musical information in their word representation. Thus, the vocalization type of the target words used at the test did not influence the participants' word recognition. Supporting this possibility is the finding that children younger than two years of age tend to store highly specific acoustic features in their word representation, while older children do not (Hollich, 2006; Newman, 2006). Although these findings are not about adults, there appears to be an age-dependent decline in encoding non-phonemic information in ones' word representation. Second, the participants might have encoded the musical information in their word representation, but, as adults are presumably expert language learners and users, they could easily transfer the earlier learned sung and IDS words to another vocalization type (i.e., ADS). Finally, perhaps the participants encoded the musical information *at the prosodic level* rather than at the phonemic level. Thus, they recognized the newly learned IDS and sung words in ADS solely based on the segmental (vowel, consonant) information. This explanation is supported by the fact that English is a non-tone language, which does not encode tonal or pitch information in word identity. Furthermore, a recent study showed that the high priority assigned to segments in word recognition may be a universal feature of language (Ma, Zhou, Singh, & Gao, 2017).

Although this study does not allow us to tease apart these explanations, it should be noted that the latter two explanations are not mutually exclusive. Experienced language users tend to have a thorough phonological knowledge of the functions of segmental and supra-segmental (tone, pitch) information. This knowledge can help learners accept words that are produced with different acoustic attributes, which are not essential to word identity. Notably, the current findings are *inconsistent* with the past findings that words are recalled better if they are presented in the same medium as when they were encoded – the *encoding specificity principle* (Tulving & Thompson, 1973; see also context-dependent memory, Godden & Baddeley, 1975), and that detailed information about a talker's voice is retained in long-term episodic memory representation of spoken words (Palmeri, Goldiner, & Pisoni, 1993). Based on these past findings, perhaps the participants in this study might have encoded the musical attributes in memory, but the task (word recognition) and dependent variable (*accuracy* – rate of correct responses) used in this study were not sensitive enough to reveal the potential differences in word processing *efficiency* in the test phase between Experiments 1 and 2, especially for adult participants who are experienced language users. Thus, we would predict that the potential differences in word processing efficiency across vocalization types should be more evident when the participants are less experienced language users (e.g., children), and when more temporally sensitive methods (e.g., eye-tracking) and dependent variables (e.g., reaction time) are used.

Conclusion

The current study examined English-speaking participants' learning and long-term memory of Chinese words presented in ADS, IDS, and song. Results showed that song and IDS facilitated word learning and long-term memory, and that the facilitative strength of the added musical attributes did not differ between song and IDS. Furthermore, we showed that song and IDS facilitated language learning by enhancing *encoding* of the speech signal rather than facilitating word recognition. Encoding was enhanced specifically in relation to word association, but also potentially word segmentation. The current results show that using musical attributes in speech enhances both word learning and long-term memory for words.

References

Beddor, P. 1993. The Perception of Nasal Vowels. In M. Huffman, and R. Krakow. (eds.), *Nasals, nasalization and the velum*. (pp. 171-196). San Diego: Academic Press.

Bidelman, G. M., Gandour, J., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*, 425–434.

Bosseler, A. N., Teinonen, T., Tervaniemi, M., & Huotilainen, M. (2016). Infant directed speech enhances statistical learning in newborn infants: An ERP study. *Plos One*, *11*, e0162177. https://doi.org/10.1371/journal.pone.0162177

Calvert, S. L., & Tart, M. (1993). Song versus verbal forms for very-long-term, long-term, and short-term verbatim recall. *Journal of Applied Developmental Psychology*, *14*, 245–260. https://doi.org/10.1016/0193-3973(93)90035-T

Chan, A. S., Ho, Y., & Cheung, M. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*, 1584–1595.

Du, Y., & Zatorre, R. J. (2017). Musical training sharpens and bonds ears and tongue to hear speech better. *Proceedings of the National Academy of Sciences*, *114*, 13579-13584. https://doi.org/10.1073/pnas.1712223114

Estes, K. G., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*, 797–824. https://doi.org/10.1111/infa.12006

Fechner, G.T. (1859). *Elemente der Psychophysik.* Breitkopf and Hartel, Leipzig.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition*, *37*, 1–9. https://doi.org/10.3758/MC.37.1.1

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*, 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A, & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*, 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*, 104–113. http://dx.doi.org/10.1037/0012-1649.20.1.104

Filippi, P., Gingras, B., & Fitch, W. T. (2014). Pitch enhancement facilitates word learning across visual contexts. *Frontiers in Psychology*, *5*, 1468. https://doi.org/10.3389/fpsyg.2014.01468

Filippi, P., Laaha, S., & Fitch, A. T. (2017). Utterance-final position and pitch marking aid word learning in school-age children. *Royal Society Open Science*, *4*, 161035. https://doi.org/10.1098/rsos.161035

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., … Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Foursha-Stevenson, C., Schembri, T., Nicoladis, E., & Eriksen, C. (2017). The influence of child-directed speech on word learning and comprehension. *Journal of Psycholinguistic Research*, *46*, 329–343. https://doi.org/10.1007/s10936-016-9441-3

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*, 2038–2043. https://doi.org/10.1093/cercor/bhs180

François, C., Jaillet, F., Takerkart, S., & Schön, D. (2014). Faster sound stream segmentation in musicians than in nonmusicians. *Plos One*, *9*, e101340. https://doi.org/10.1371/journal.pone.0101340

Francois, C., & Schön, D. (2010). Learning of musical and linguistic structures: Comparing event-related potentials and behavior. *NeuroReport*, *21*, 928–932. https://doi.org/10.1097/WNR.0b013e32833ddd5e

Fujii, S., & Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: The SEP hypothesis. *Frontiers in Human Neuroscience*, *8*, 777. https://doi.org/10.3389/fnhum.2014.00777

Golinkoff, R. M., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language*, *22*, 703–726. https://doi.org/10.1017/S0305000900010011

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11*, 43- 79. https://doi.org/10.1017/S0305000900005584

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*, 325-33I.

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music*, *43*, 627–640. https://doi.org/10.1177/0305735614528833

Han, M., de Jong, N. H., & Kager, R. (2018). Infant-directed speech is not always slower: Cross-linguistic evidence from Dutch and Mandarin Chinese. In A. Bertolini and M. Kaplan (Ed.) *Proceedings of the 42nd annual Boston University Conference on Language Development* (pp. 331–344). Cascadilla Press.

Heffner, C. C., & Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Frontiers in Psychology, 6,* 1962. http://doi.org/10.3389/fpsyg.2015.01962

Ho, Y., Cheung, M., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*, 439–450. https://doi.org/10.1037/0894-4105.17.3.439

Hollich, G. (2006). Combining techniques to reveal emergent effects in infants' segmentation, word learning, and grammar. *Language and Speech*, *49*, 3–19. https://doi.org/10.1177/00238309060490010201

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*, 319–329.

Ilie, G., & Thompson, W. F. (2011). Experiential and cognitive changes following seven 32 minutes exposure to music and speech. *Music Perception, 28,* 247-264.

Johnson, E. K., & Seidl, A. (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, *13*, 440–455. https://doi.org/10.1080/15250000802329321

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*, 13745. https://doi.org/10.1038/s41598-018-32150-6

Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one- to four-month-old infants. *Journal of Experimental Child Psychology*, *39*, 326–342. https://doi.org/10.1016/0022-0965(85)90044-X

Kent, R. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research, 19*, 421–447.

Kent, R. (1992). The biology of phonological development. In C. Ferguson, L. Menn, & C. Stoel-Gammon, (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). York Press.

Kilgour, A. R., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory & Cognition*, *28*, 700–710.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in Mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*, 85–110. https://doi.org/10.1207/S15327078IN0401_5

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*, 599–605. https://doi.org/10.1038/nrn2882

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*, 684–686.

Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The Temporal Modulation Structure of Infant-Directed Speech. *Open Mind: Discoveries in Cognitive Science, 1*, 78–90. https://doi.org/10.1162/ opmi_a_00008

Levitin, D. J., & Menon, V. (2003). Musical structure processed in “language” areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage*, *20*, 2142–2152. https://doi.org/10.1016/S1053-8119(03)00482-8

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition*, *42*, 41–52. https://doi.org/10.3758/s13421-013-0342-5

Ma, W., Fiveash, A., & Thompson, W. F. (2019). Spontaneous emergence of language-like and music-like vocalizations from an artificial protolanguage. *Semiotica*, *229*, 1-23. https://doi.org/10.1515/sem-2018-0139

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*, 185–201. https://doi.org/10.1080/15475441.2011.579839

Ma, W., & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proceedings of the National Academy of Sciences*. *112,* 14563-14568. https://doi.org/10.1073/pnas.1515087112

Ma, W., Zhou, P., Singh, L., & Gao, L. (2017). Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. *Cognition*, *159*, 139–155. https://doi.org/10.1016/j.cognition.2016.11.011

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). Musical syntax is processed in Broca’s area: An MEG study. *Nature Neuroscience*, *4*, 540–545.

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences*, *115*, 13129–13134. https://doi.org/10.1073/pnas.1811793115

Margulis, E. H. (2013). *On repeat: How music plays the mind*. https://doi.org/10.1093/acprof:oso/9780199990825.001.0001

McRoberts, G. W., & Best, C. T. (1997). Accommodation in mean f0 during mother-infant and father-infant vocal interactions: A longitudinal case study. *Journal of Child Language, 24*, 719–736. https://doi.org/10.1017/S030500099700322X

Merriam, A.P. (1964). The Anthropology of Music. Northwestern University Press. Chicago.

Milliman, R. (1982). Using Background Music to Affect the Behavior of Supermarket Shoppers. *Journal of Marketing, 46*, 86-91.

Milliron, C. E. (2017). Next time won't you sing with me? The role of music rooted in oral tradition as a resource for literacy learning in the twenty-first century classroom. *Channels: Where Disciplines Meet*, *1*, 12.

Moore, C., Angelopoulos, M., & Bennett, P. (1997). The role of movement in the development of joint visual attention. *Infant Behavior & Development*, *20*, 83–92. https://doi.org/10.1016/S0163-6383(97)90063-1

Moussard, A., Bigand, E., Belleville, S., & Peretz, I. (2012). Music as an aid to learn new verbal information in Alzheimer's disease. *Music Perception*, *29*, 521–531. https://doi.org/10.1525/mp.2012.29.5.521

Moussard, A., Bigand, E., Belleville, S., & Peretz, I. (2014). Learning sung lyrics aids retention in normal ageing and Alzheimer's disease. *Neuropsychological Rehabilitation*, *24*, 894–917. https://doi.org/10.1080/09602011.2014.917982

Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., … Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *NeuroImage*, *117*, 267–283. https://doi.org/10.1016/j.neuroimage.2015.05.020

Newman, R. S. (2008). The level of detail in infants' word learning. *Current Directions in Psychological Science, 17,* 229-232. https://doi.org/10.1111/j.1467-8721.2008.00580.x

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*, 127–146. https://doi.org/10.1111/j.1551-6709.2008.01007.x

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 309–328.

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior & Development*, *13*, 539–545. https://doi.org/10.1016/0163-6383(90)90022-Z

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*, 142. https://doi.org/10.3389/fpsyg.2011.00142

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of The Acoustical Society of America, 24,* 175–184.

Peterson, D. A., & Thaut, M. H. (2007). Music increases frontal EEG coherence during verbal learning. *Neuroscience Letters*, *412*, 217–221. https://doi.org/10.1016/j.neulet.2006.10.057

Quinto, L. R., Thompson, W. F., & Keating, F. L. (2013). Emotional communication in speech and music: The role of melodic and rhythmic contrasts. *Frontiers in Psychology*, *4*. 184. https://doi.org/10.3389/fpsyg.2013.00184

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition*, *35*, 242–253.

Rainey, D., & Larsen, J. (2002). The effect of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception: An Interdisciplinary Journal*, *20*, 173–186.

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? – Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition*, *178*, 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: A longitudinal study. *Frontiers in Psychology*, *3*, 572. https://doi.org/10.3389/fpsyg.2012.00572

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences*, *104*, 9852–9857. https://doi.org/10.1073/pnas.0703140104

Rukholm, V. N., Helms-Park, R., Odgaard, E. C., & Smyth, R. (2018). Facilitating lexical acquisition in beginner learners of Italian through spoken or sung lyrics. *Canadian Modern Language Review*, 74, 153-175. https://doi.org/10.3138/cmlr.3789

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., … Cohen, D. (2013). Motherese in interaction: At the cross-road of emotion and cognition? (A systematic review). *Plos One*, *8*, e78103. https://doi.org/10.1371/journal.pone.0078103

Saito, Y., Aoyama, S., Kondo, T., Fukumoto, R., Konishi, N., Nakamura, K., … Toshima, T. (2007). Frontal cerebral blood flow change associated with infant-directed speech. *Archives* *of Disease in Childhood. Fetal and Neonatal Edition*, *92*, F113. https://doi.org/10.1136/adc.2006.097949

Schön, D., Boyer, M., Sylvain, M., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*, 975–983. https://doi.org/10.1016/j.cognition.2007.03.005

Schön, D., & Francois, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*, *2*, 167. https://doi.org/10.3389/fpsyg.2011.00167

Schreiner, M. S., Altvater-Mackensen, N., & Mani, N. (2016). Early word segmentation in naturalistic environments: Limited effects of speech register. *Infancy*, *21*, 625–647. https://doi.org/10.1111/infa.12133

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*, 24–48. https://doi.org/10.1016/j.jml.2006.10.004

Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, *11*, 596–606. https://doi.org/10.1111/j.1467-7687.2008.00704.x

Shenfield, T., Trehub, S. E., & Nakata, T. (2003). Maternal singing modulates infant arousal. *Psychology of Music*, *31*, 365–375. https://doi.org/10.1177/03057356030314002

Shneidman, L. A., Buresh, J. S., Shimpi, P. M., Knight-Schwarz, J., & Woodward, A. L. (2009). Social experience, social attention and word learning in an overhearing paradigm. *Language Learning and Development*, *5*, 266–281. https://doi.org/10.1080/15475440903001115

Simmons-Stern, N. R., Budson, A. E., & Ally, B. A. (2010). Music as a memory enhancer in patients with Alzheimer's disease. *Neuropsychologia*, *48*, 3164–3167. https://doi.org/10.1016/j.neuropsychologia.2010.04.033

Simmons-Stern, N. R., Deason, R. G., Brandler, B. J., Frustace, B. S., O'Connor, M. K., Ally, B. A., & Budson, A. E. (2012). Music-based memory enhancement in Alzheimer's Disease: Promise and limitations. *Neuropsychologia*, *50*, 3295–3303. https://doi.org/10.1016/j.neuropsychologia.2012.09.019

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy*, *14*, 654–666. https://doi.org/10.1080/15250000903263973

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*, 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America*, *128*, 389–400. https://doi.org/10.1121/1.3419786

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. 44, 992–999. https://doi.org/10.1037/xlm0000493

Tamminen, J., Rastle, K., Darby, J., Lucas, R., & Williamson, V. J. (2017). The impact of music on learning and consolidation of novel words. *Memory*, *25*, 107–121. https://doi.org/10.1080/09658211.2015.1130843

Thaut, M. H., Peterson, D. A., McIntosh, G. C., & Hoemberg, V. (2014). Music mnemonics aid verbal memory and induce learning - related brain plasticity in multiple sclerosis. *Frontiers in Human Neuroscience*, *8*, 395. https://doi.org/10.3389/fnhum.2014.00395

Thaut, M. H., Peterson, D. A., Sena, K. M., & Mcintosh, G. C. (2008). Musical structure facilitates verbal learning in multiple sclerosis. *Music Perception*, *25*, 325–330. https://doi.org/10.1525/MP.2008.25.4.325

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*, 53–71. https://doi.org/10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences*, *1169*, 225–233. https://doi.org/10.1111/j.1749-6632.2009.04547.x

Tierney, A., & Kraus, N. (2014). Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, *8*, 949. https://doi.org/10.3389/fnhum.2014.00949

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*, 1454–1463. https://doi.org/10.2307/1130423

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review*, *9*, 335–340. https://doi.org/10.3758/BF03196290

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. https://doi.org/10.1037/h0020071

Wallace, W. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1471–1485.

Weber, E. H. (1834) De pulsu, resorptione, auditu et tactu annotationes anatomic et physiologic,e, Leipsic (Author's summary: Ueber den Tastsinn, Arch. Anat. u. Physiol., 1835, 152.)

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422. https://doi.org/10.1038/nn1872

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*, 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Zangl, R., & Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6- and 13-month-old infants. *Infancy*, *11*, 31–62. https://doi.org/10.1207/s15327078in1101_2

Zhao, T. C., & Kuhl, P. K. (2016). Musical intervention enhances infants' neural processing of temporal structure in music and speech. *Proceedings of the National Academy of Sciences*, *113*, 5212-5217. https://doi.org/10.1073/pnas.1603984113

Zuk, J. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS One*, *9*, e99868. https://doi.org/10.1371/journal.pone.0099868

Table 1. Slides shown in one experimental block in the familiarization phase of Experiments 1 and 2 along with the accompanying Chinese sentences (presented in either ADS, IDS, or song) and their English translation

Block 1		
Slide	Sentence presented with slides	English translation
1. plane	Ni3 kan4! Zhe4 jiu4 shi4 **fei1ji1**. Kuai4 kan4! Zhe4li3 you3 yi2 ge4 **fei1ji1**.	Look! This is a plane. Look quickly! There is a plane.
2. ship	Na4 shi4 yi4 sou1 **lun2chuan2**. Ni3 kan4! Zhe4 jiu4 shi4 **lun2chuan2**.	That is a ship. Look! This is a ship.
3. apple	Zhe4li3 you3 ge4 **ping2guo3**. Zhe4 shi4 yi2 ge4 hao3chi1 de **ping2guo3**.	There is an apple. This is a delicious apple.
4. strawberry	Kan4 kan4! Zhe4 shi4 wo3 de **cao3mei2**. Zhe4 shi4 yi2 ge4 **cao3mei2**.	Look! This is my strawberry. This is a strawberry.
5. glass	Na4 shi4 yi4 zhi1 **shui3bei1**. Na2 zhe zi4ji3 de **shui3bei1**.	That is a glass. Hold your own glass.
6. leather shoe	Kan4 kan4! Wo zi4ji3 de **pi2xie2**. Kan! Zhe4 shuang1 **pi2xie2**.	Look! My own leather shoes. Look! This pair of leather shoes.
7. picture book	Wo3 kan4jian4 le **tu2shu1**. Wo3 zui4 xi3huan1 **tu2shu1**.	I saw picture books. My favorite things are picture books.
8. dolphin	Zhe4 shi4 yi4 zhi1 hen3 piao4liang4 de **hai3tun2**. Kan4! Zhe4li3 you3 yi4 zhi1 **hai3tun2**.	This is a very beautiful dolphin. Look! There is a dolphin.

Note: The slides and words used in Block 2 are *skateboard* (hua2ban3), *bicycle* (dan1che1), *cherry* (ying1tao2), *pumpkin* (nan2gua1), *pencil* (qian1bi3), *desk lamp* (tai2deng1), *wristwatch* (shou3biao3), and *turtle* (wu1gui1). These used in Block 3 were *elevator* (dian4ti1), *rocket* (huo3jian4), *submarine* (qian2ting3), *mushroom* (mo2gu1), *spinach* (bo1cai4), *basketball* (lan2qiu2), *desk* (ke4zhuo1) and *whale* (jing1yu2). The same sentence frames were used across blocks except that different classers were used for grammatical purposes in sentences containing *skateboard* (**fu4** hua2ban3), *bicycle* (**liang4** dan1che1), *cherry* (**chuan4** ying1tao2), *desk lamp* (**ge4** tai2deng1), *elevator* (**bu4** dian4ti1), and *spinach* (**ke1** bo1cai4). The three blocks were produced in either ADS, IDS, or song, and the presentation of blocks was counterbalanced across participants.

Table 2. Mean formant frequencies for F1 and F2 (in Hertz) in ADS, IDS, and song for /i/, /u/, /ɔ/.

	ADS		IDS		Familiar melody		Unfamiliar melody	
	F1	F2	F1	F2	F1	F2	F1	F2
/i/	315.79	1984.15	347.50	2706.11	349.73	2565.60	437.84	2568.94
/u/	379.42	1018.43	448.06	1148.44	333.98	1023.19	414.74	1145.82
/ɔ/	502.02	1150.92	648.93	1195.01	708.08	1340.87	662.19	1193.79

Note: The data include only eight words.

Table 3. Means and standard deviations of the average ratings of the auditory stimuli used in this study

Vocalization type	ADS-likeness rating	IDS-likeness rating	Song-likeness rating
ADS audio	$M = 4.69$, $SD = 1.05$	$M = 2.96$, $SD = 1.06$	$M = 1.63$, $SD = .56$
IDS audio	$M = 3.26$, $SD = 1.15$	$M = 4.65$, $SD = 1.27$	$M = 3.17$, $SD = 1.27$
Familiar melody audio	$M = 1.71$, $SD = .64$	$M = 2.97$, $SD = 1.08$	$M = 5.18$, $SD = 1.14$
Unfamiliar melody audio	$M = 1.87$, $SD = .62$	$M = 3.10$, $SD = 1.18$	$M = 5.06$, $SD = 1.09$
Song (familiar and unfamiliar melodies averaged)	$M = 1.79$, $SD = .55$	$M = 3.03$, $SD = .91$	$M = 5.12$, $SD = .72$

Note: For the sung vocalizations, since paired sample *t*-tests showed that the familiar and unfamiliar melodies did not differ in their ADS-likeness ratings ($t(24) = 1.20$, $p = .24$), IDS-likeness ratings ($t(24) = 0.51$, $p = .61$), or song-likeness ratings ($t(24) = 0.36$, $p = .72$), the ratings were combined across the two song melodies.

Table 4. Examples of the experimental procedure (one trial in the familiarization phase and one trial in the test phase of the word learning task completed on Day 1 and one trial in the long-term memory task completed on Day 2)

		Visual	Audio
Day 1	**Familiarization phase** (Each familiarization phase contained eight trials, which were presented before the test phase)		"Ni3 kan4! Zhe4 jiu4 shi4 **fei1ji1**. Kuai4 kan4! Zhe4li3 you3 yi2 ge4 **fei1ji1**." [Look quickly! This is a plane. Look! There is a plane.]
	Test phase (The test phase started immediately after the familiarization phase. Each test phase contained eight trials, presented in the same order as the familiarization phase)	A multiple-choice question (shown on the computer monitor) A: plane B: apple C: strawberry	"**fei1ji1**." [plane]
		Approximately 24 hours	
Day 2	**Test phase** (the same as the test phase of Day 1)	A multiple-choice question (shown on the computer monitor) A: plane B: apple C: strawberry	"**fei1ji1**." [plane]

Note: On Day 1, the word learning task consisted of a familiarization phase and a test phase. Participants learned eight words – embedded in sentence frames (Experiments 1 and 2) or presented in isolation (Experiment 3) – in the familiarization phase. The test phase started immediately after the completion of the familiarization phase. On Day 2, the same test phase was administered to the participants to assess their long-term word memory.

Table 5. *Means and standard deviations of the average confidence scores*

	Word learning task			Long-term memory		
	ADS block	IDS block	Song block	ADS block	IDS block	Song block
Experiment 1	M = 2.96 SD = 1.65	M = 3.06 SD = 1.44	M = 2.98 SD = 1.31	M = 2.74 SD = 1.56	M = 2.49 SD = 1.24	M = 2.54 SD = 1.20
Experiment 2	M = 3.18 SD = 1.34	M = 3.00 SD = 1.55	M = 3.05 SD = 1.64	M = 2.31 SD = 1.19	M = 2.39 SD = 1.17	M = 2.32 SD = 1.16
Experiment 3	M = 3.38 SD = 1.54	M = 3.93 SD = 1.73	M = 4.25 SD = 1.65	M = 2.68 SD = 1.64	M = 3.11 SD = 1.78	M = 3.57 SD = 1.71

Note: Confidence scores did not differ across vocalization type in Experiments 1 and 2. In Experiment 3, confidence scores were lower with ADS words than with IDS and sung words for both the word learning and long-term memory tasks; confidence scores did not differ between IDS and sung words.

Figure 1. The melody and original lyrics of What a Beauty Jasmine.

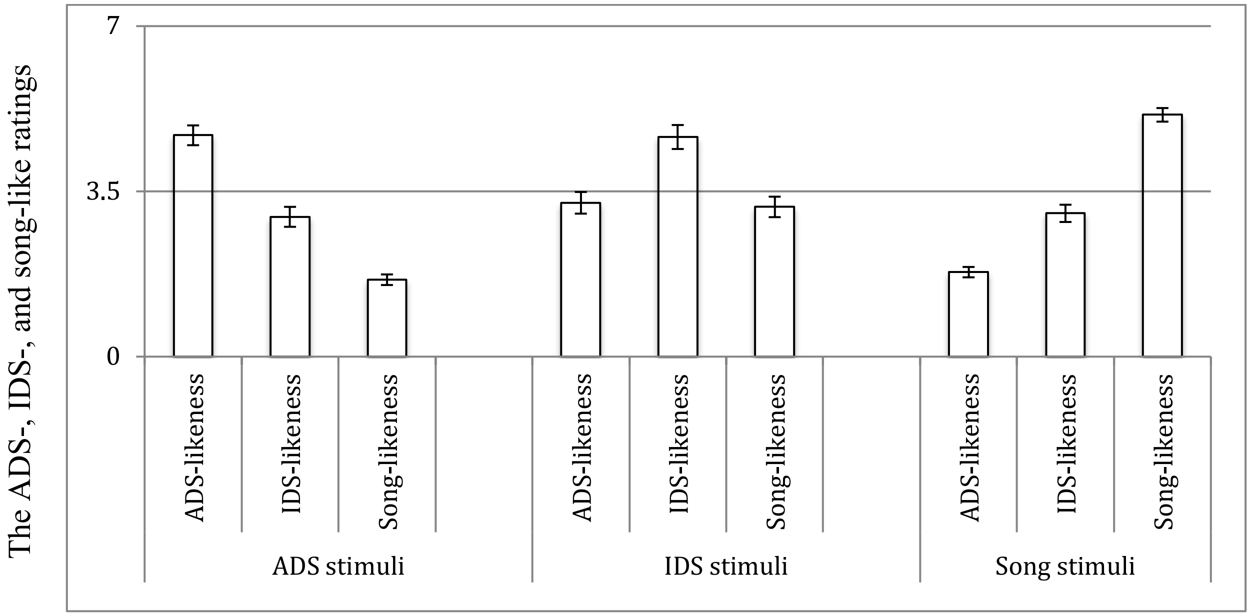

Figure 2. The ADS-, IDS-, and song-like ratings of the auditory stimuli used in this study. Paired sample *t*-tests compared the ratings within each vocalization type. For the ADS vocalizations, the ADS-likeness ratings were greater than the IDS-likeness ratings ($t(24) = 7.35$, $p < .001$), which in turn were greater than the song-likeness ratings ($t(24) = 6.94$, $p < .001$). For the IDS vocalizations, the IDS-likeness ratings were greater than the ADS-likeness ratings ($t(24) = 6.73$, $p < .001$) and song-likeness ratings ($t(24) = 7.31$, $p < .001$), but the ADS- and song-likeness ratings did not differ from each other ($t(24) = 0.66$, $p = .52$). For the song vocalizations, with the two melodies combined, the song-likeness ratings of the song vocalizations were greater than the IDS-likeness ratings ($t(24) = 12.98$, $p < .001$), which in turn were greater than the ADS-likeness ratings ($t(24) = 8.94$, $p < .001$).

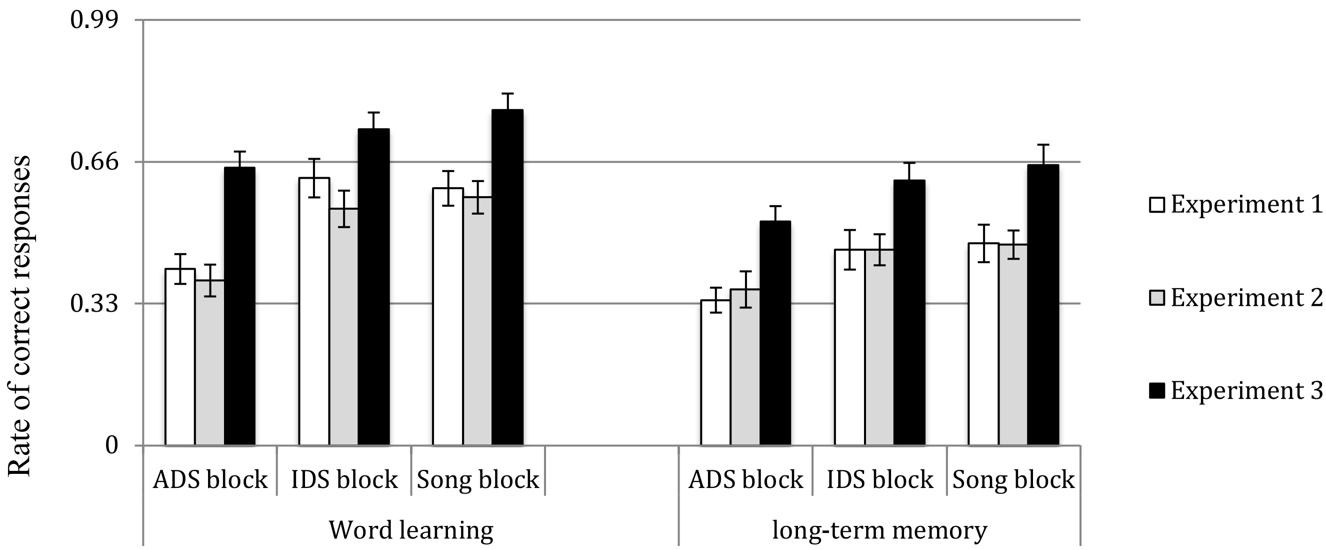

Figure 3. The rate of correct responses in the three experiments. In Experiments 1 and 2, the rate of correct responses in the word learning and long-term memory tasks were above chance (33%) for IDS and sung words, but did not differ from chance for ADS words. In Experiment 3, the rate of correct responses in the word learning and long-term memory tasks were above chance for ADS, IDS, and sung words. In the three experiments, word learning and long-term memory performance was better with IDS and sung words than with the ADS words, but did not differ between IDS and sung words. Error bars represent standard error means.