EFFECTS OF VISUALLY PRESENTED LYRICS ON SONG RECALL:
A COGNITIVE LOAD PERSPECTIVE

A Dissertation in
Music Education
by
Yo-Jung Han

© 2018 Yo-Jung Han

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2018
The dissertation of Yo-Jung Han was reviewed and approved* by the following:

Joanne Rutkowski  
Professor Emeritus of Music Education  
Dissertation Advisor  
Chair of Committee

Linda Thornton  
Professor of Music Education

Peggy Van Meter  
Associate Professor of Educational Psychology

Ann Clements  
Associate Professor of Music Education  
Graduate Program Chair for Music Education

*Signatures are on file in the Graduate School
Abstract

When learning a song aurally, verbal information (i.e., lyrics) and musical information (e.g., pitches and rhythm) are processed through the aural processor (aural channel). If a song has much information, only using the aural channel might cause cognitive overload thus hindering one’s ability to learn the song. However, if the learner sees the lyrics, verbal information can be processed dually – aurally and visually. This dual processing of verbal information might reduce cognitive load in the aural channel, leading to more capacity to process musical information. This dissertation comprises two studies based on these assumptions.

In the first study, I investigated the effect of visually presented lyrics on song recall with members from auditioned choirs. While individually learning a difficult song, one group saw the lyrics, but the other group did not. When controlling for participants’ phonological working memory, the efficacy of instructional conditions depended on the participant’s level of musical expertise. Non-music majors benefited from seeing the lyrics whereas music majors did not.

In the second study, I examined whether showing the lyrics to non-music majors induced less cognitive load in the aural channel compared to not showing the lyrics, leading to better recall accuracy of the learned song. Cognitive load was measured through reaction times in a single sound monitoring task while learning songs. Non-music majors, but not auditioned choir members, learned two songs: For one song, they saw the lyrics, and for the other song, they did not see the lyrics. The presentation order of instructional conditions and songs was counterbalanced. When seeing the lyrics, participants reacted faster than when not seeing the
lyrics. A mediation analysis revealed instructional condition (seeing the lyrics or not) indirectly affected recall accuracy of lyrics and rhythm through cognitive load.

Instructional design should be based on many considerations such as instructional time, goals, and characteristics of students. Given limited instructional time, several strategies should be considered to prevent learners from experiencing cognitive overload while learning a difficult song aurally. Showing the lyrics of a new song could be one strategy for that purpose, at least for young adults with low levels of musical expertise.

*Keywords:* visually presented lyrics, song recall, song learning, cognitive load, modality effect, verbal redundancy effect
Table of Contents

List of Tables ......................................................................................................................... vii
List of Figures ......................................................................................................................... viii
Acknowledgement .................................................................................................................. ix

Chapter 1 Introduction ............................................................................................................. 1
  Background of the Problem .................................................................................................... 1
  Need for the Study ................................................................................................................ 3
  Purpose of the Studies .......................................................................................................... 6
  Assumptions, Limitations, and Scope .................................................................................. 6
  Definition of Key Terminology ............................................................................................ 8
  Organization of the Dissertation .......................................................................................... 9
  References for Chapter 1 ...................................................................................................... 11

Chapter 2 Review of Literature .............................................................................................. 17
  The Role of Lyrics in Singing and Song Acquisition .......................................................... 17
    Summary ............................................................................................................................. 21
  Relationship between Tunes and Words in Memory ............................................................ 22
    Summary ............................................................................................................................. 28
  Effects of Presentation Modalities on Music Instruction ...................................................... 29
    Summary ............................................................................................................................. 35
  Cognitive Load Effects .......................................................................................................... 36
    Modality Effect .................................................................................................................. 39
    Verbal Redundancy Effect ............................................................................................... 40
    Summary ............................................................................................................................. 42
  Chapter Summary ................................................................................................................ 43
  References for Chapter 2 ...................................................................................................... 45

Chapter 3 Study One: Effects of Visually Presented Lyrics on Song Recall ....................... 56
  Introduction ........................................................................................................................... 57
    Other Considerations ......................................................................................................... 61
    Research Questions ........................................................................................................... 63
  Method ................................................................................................................................ 63
    Design and Participants .................................................................................................... 63
    Instructional Materials ..................................................................................................... 64
    Data Collection Tools ....................................................................................................... 66
List of Tables

Table 3-1. Example of Contingency Tables Used to Calculate the Kappa Statistics .. 69
Table 3-2. Interrater Reliability based on Percentage Agreement, Kappa Agreement, and Pearson Correlation .............................................................. 70
Table 3-3. Recall Accuracy of Lyrics, Pitches, and Rhythm by Group and Major .... 72
Table 4-1. Interrater Reliability based on Percentage Agreement, Kappa Agreement, and Pearson Correlation .............................................................. 100
Table 4-2. Descriptive Statistics of Recall Accuracy (%) in terms of Instructional Condition, Song, and Presentation Order .............................................................. 102
Table 4-3. Descriptive Statistics of Recall Accuracy (%) by Song, Presentation Order, and Instructional Condition .............................................................. 103
Table 4-4. Reaction Times (RTs) for Instructional Condition, Song and Presentation Order ........................................................................................................ 104
Table 5-1. Means and SDs of the First (results form non-music majors) and Second (results from April presented in the first order) Studies ........................................ 126
Table 5-2. Nine Strategies to Alleviate Cognitive Load in Multimedia Learning (from Mayer & Moreno, 2003) .................................................................................. 134
List of Figures

Figure 3-1. The score of “April” ........................................................................................................... 65

Figure 3-2. Interactions between Group and Major. ................................................................. 71

Figure 4-1. Music scores of “April” and “Februay Twilight”................................................. 96

Figure 4-2. Bar graphs of reaction times for “April” and “Februay Twilight” by instructional condition and presentation order ................................................................. 104
Acknowledgment

I would like to express my deepest gratitude to my advisor and mentor, Dr. Joanne Rutkowski. Without her continuous support, close guidance, and unwavering patience, I would not have been able to complete this dissertation. She has demonstrated how to work with integrity as a scholar, teacher, and human being, and how to guide students with respect and care. It has been a blessing and privilege to work with her.

My sincere thanks also go out to Dr. Ann Clements and Dr. Linda Thornton, who supported me through the many phases of my doctoral program with personal care and kindness. I am grateful for their thoughtful guidance and constructive feedback on my work. I am particularly thankful to Dr. Peggy Van Meter, who guided me through the early stages of dissertating with her expertise in the area of instructional psychology. The course experiences with her shaped the direction and theoretical framework of my dissertation. Also, I appreciate her specific feedback on statistical analysis in this dissertation.

I extend my heartfelt gratitude to my colleagues and friends for their support throughout my journey at Penn State. Chianyi Ang, Dr. Mara Culp, Gregory Drane, Val Flamini, Dr. Lindsay Fulcher, Dr. Jason Gossett, Anne-Marie Hildebrandt, Mac Himes, Gretchen Lee, Dr. Dan Shevock, and Dr. Krissie Weimer. Thank you for the encouragement, support, and love you have shown me during the ups and downs of my doctoral life.

I am greatly indebted to my research participants in my two experimental studies of this dissertation. Your willingness to help someone out will be remembered and I hope to repay it by helping others. Particular thanks to Mary Ellen Pinzino, the composer of “April” and “February Twilight,” who generously allowed me to use her songs in the studies of this dissertation. My sincere gratitude to Anne-Marie Hildebrandt, Dr. Sangmi Kang, Seohyang
Kim, and Dr. Karen Thomas, who served as a rater of my participants’ singing in my studies. Special thanks to Anne-Marie Hildebrandt and Mitchell Auger, who audio-recorded the song instruction and test instruction for the studies.

Lastly, I would like to thank my family members who have always stayed by my side. To my devoted husband, Jaesung, I cannot thank you enough for your ongoing support, wise advice, and sincere love. Marrying you has been one of the best things I have done. To my adorable two sons, Daniel and Kai, thank you for your being. To my parents, mother-in-law, and brother, thank you for your prayers and support over many years. To my Heavenly Father, thank you for everything. You are the reason I live and love.
Chapter 1 Introduction

Background of the Problem

Singing a song is a unique form of human expression and communication. Recognizing the importance of singing in one’s personal life and the society, music educators in research and practice have long been interested in how to help students develop their singing ability and provide effective song instructions from various aspects (Goetze, Cooper, & Brown, 2010; Hedden, 2012). While various factors might affect singing accuracy and song acquisition (Goetze, Cooper, & Brown, 2010; Hedden, 2012; Nichols, 2016; Pfordresher et al., 2015), Hedden (2012) categorized the affecting variables related to children’s singing into two aspects: internal and external. Internal variables include vocal range, pitch matching ability, and sex difference whereas external variables include solo versus group singing, use of accompaniment, integration of text, and vocal modeling. Among various factors, the role of the lyrics in singing accuracy and song acquisition - whether singing with or without lyrics affects singing accuracy and song acquisition - has been examined by several researchers (Andress, 1986; Berkowska & Dalla Bella, 2009; Gault, 2002; Goetze, 1985, Goetze, Cooper, & Brown, 2010; Jacobi-Karna, 1996; Levinowitz, 1987; Levinowitz, 1989, Sims, Moore, & Kuhn, 2002; Smale, 1988; Welch, Sergeant, & White, 1996) because of the unique character of a song, which consists of verbal (i.e., lyrics) information as well as musical (e.g., pitches, rhythm) information.

When a song is sung, the verbal and musical information of the song are delivered as a single entity since they are temporally integrated. However, when it comes to perception and cognition, two distinct cognitive systems – language and music – are involved to process the information (Asano, 2016) although some functions and processing of two systems are shared.
(Peretz, Vuvan, Lagrois, & Armony, 2015; Sammler et al., 2015). Thus, the nature of the relationship between tune and lyrics has been of central interest for several researchers. The central questions on this issue are whether tune and lyrics of a song are integrated or separated in memory and if integrated, to what extent they are processed together (Ginsborg & Sloboda, 2007).

Serafine and her colleagues (1984, 1986, 1990), in their classical studies, demonstrated the integration effect in memory for songs, suggesting that words and tunes of a song are integrated in memory. Ginsborg and Sloboda (2007) argued the term “integration” might have a strong connotation of “all-or-nothing” (p. 434) thus suggesting the term “association” to indicate the probabilistic relationship that memory for one component (e.g., words) increases the likelihood of retrieving the other component (e.g., tune). On the other hand, neuroimaging studies found lyrics and tunes are integrated at prelexical, phonemic processing levels in the auditory analysis stage as well as in the preparation of motor output while they are processed independently at structural and semantic levels (Sammler et al., 2010).

In the area of music education, many researchers examined the effect of singing with or without the lyrics on singing accuracy and song acquisition. The underlying assumption of these studies is that lyrics may detract the learners’ attention from melodic information hindering song learning (Andress, 1986; Berkowska & Dalla Bella, 2009; Goetze, Cooper, & Brown, 2010; Levinowitz, 1987), thus singing without lyrics (singing on a syllable such as ‘la’) may help persons sing more accurately than singing with lyrics. However, results have been inconsistent: Some studies revealed participants, including young children and adults, sang more accurately on a neutral syllable (Berkowska & Dalla Bella, 2009; Goetze, 1985, Levinowitz, 1989, Welch, Sergeant, & White, 1996), while others found no difference between
singing with and without lyrics (Sims, Moore, & Kuhn, 2002; Smale, 1988). In Gault’s study (2002), the results regarding the text condition depended on the song taught. It is noteworthy the studies mentioned above included only aural instructions, not any visual materials.

However, music teachers often use visual and/or kinesthetic aids to promote students’ musical skills and understanding (Sheldon, 1991; Tarnowski, 1986; Zikmund & Nierman, 1992) since visual and kinesthetic representations may help learners understand and experience the transient and intangible characteristics of musical elements. Several researcher have examined the efficacy of multimodal approaches regarding singing achievement (Apfelstadt, 1984; Hughes, 1991; Pautz, 1988; Persellin, 1994; Tarnowski, 1986). Apfelstadt (1984) and Persellin (1994) found young children who received multimodal instruction better performed pitch-pattern echoing than the other groups of children who received aural instruction (Apfelstadt, 1984) and visual or kinesthetic instruction (Persellin, 1994). Yet, Hughes (1991) and Pautz (1988) did not find any difference among instructional conditions. In these studies, the musical components represented visually and kinesthetically were melodic contour and melodic rhythm. Although some researchers (Persellin, 1994; Tarnowski, 1986) provided visual and kinesthetic reinforcement of lyrics along with other musical components, the visual representation of lyrics was in the form of pictures, not written text. To my knowledge, the effect of visually presented lyrics on song learning has not been investigated.

**Need for the Study**

Since music is inherently an aural art, improving learners’ aural capacity to process musical information is an essential task for music educators. In that sense, when teaching a song, a rote teaching procedure - teaching songs by imitation without using visual material - has been recommended and utilized, particularly for younger students, in music teaching
practice (Gordon, 2012; Klinger, Campbell, & Goolsby, 1998). Also, this is how songs are learned in many cultures: Musical learning occurs through oral transmission (Anderson, & Campbell, 2011; Lehmann, Sloboda, & Woody, 2007).

However, in practice, we encounter occasions when teachers show the lyrics without the staff notation while teaching or singing a song. I often observed classroom teachers displaying the lyrics on an easel or projection screen while teaching a song to their students. In vocal music concerts, program notes sometimes provide lyrics for the audience. Karaoke screens always show the lyrics; many places of worship display song lyrics while the congregation sings together; and music shows on television present lyrics while singers sing. Since often we see the lyrics while learning or singing a song, it seems of interest to determine how seeing the lyrics may affect song learning and singing.

When learning a song aurally, verbal information and musical information are perceived through the aural channel. If the song has less information, processing it in the aural channel likely will not be a problem. However, if a song has much information, only using the aural channel might cause cognitive overload to the learners, thus hindering one’s ability to learn the song. Cognitive overload occurs when incoming information exceeds the learner’s working memory capacity to process information (Mayer & Moreno, 2003).

Based on the understanding of how human minds work, a considerable number of studies in instructional psychology have focused on how to optimize human cognitive capacity in presenting instructional materials (Mayer, 2008). Those studies have been guided by three basic assumptions: dual channel, limited capacity, and active processing (Mayer, 2014; Mayer & Moreno, 1998, 2003). Humans possess separate systems to process visual and auditory information, called dual channel. Human beings have limited capacity to process novel
information in each channel at any given time. And humans learn through active processing by attending to the incoming information, organizing attended information into a coherent mental representation, and integrating mental representations with previous knowledge (Moreno & Mayer, 2003).

Based on these assumptions, several instructional effects and principles in presenting material in ways that reduce cognitive load leading to better comprehension and memory have been studied and suggested (Mayer, 2014; Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011). Those effects, demonstrated and explained from a cognitive load perspective, are called cognitive load effects. Among several cognitive load effects, the modality effect suggests using two sensory modalities – auditory and visual – when presenting information can reduce cognitive load (Low & Sweller, 2014; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995; Tabbers, Martens, & van Merriënboer, 2004; Tindall-Ford, Chandler, & Sweller, 1997). In multimedia instructional settings, Mayer and Moreno (1998) found students performed better on transfer tests when instructional materials were presented as pictures and narration rather than as pictures and on-screen text. Presenting both pictures and words visually might cause cognitive overload in the visual channel while the aural channel is unused. On the contrary, if words are presented aurally as a form of narration instead of on-screen text, the visual channel might be offloaded with regard to the verbal information, having more capacity to process the pictures. It is assumed the reduction in cognitive load in the visual channel enhances learning.

By the same logic, when learning a song, if the lyrics are visually presented, the verbal information can be processed in both the aural and visual channels. By splitting the load of verbal information processing into two channels, the amount of load occupied by verbal information in the aural channel might be reduced, resulting in more capacity to process
musical information. Therefore, it is assumed that the visual presentation of lyrics facilitates information processing of a song leading to better recall of the learned song.

It is noteworthy that in the settings of the two investigations in this dissertation, the verbal information remained in the aural channel: Verbal information was presented aurally or aurally and visually. In Mayer and Moreno’s (1998) studies, when words were presented as narration instead of on-screen text, the verbal information was taken out from the visual channel. However, in my studies, participants basically learned the songs through aural instruction. When seeing the lyrics, the verbal information was not taken out of the aural channel and was presented dually – aurally and visually. When learning a song, it is critical that verbal and musical information is integrated. Keeping the verbal and musical information integrated in the aural channel seems important to avoid split-attention effect. Split-attention occurs when learners need to split their attention between multiple sources of information, which have not been spatially or temporally integrated but should be integrated to be understood (Ayres & Sweller, 2014; Sweller, 2011). Hence, the possible overload on the aural channel may be outweighed by the benefits of presenting already integrated materials.

**Purpose of the Studies**

Therefore, the purpose of my line of research is to investigate the effect of visually presented lyrics on song learning. This work may illuminate the process and relationship of verbal and musical information in song learning from a cognitive load perspective.

**Assumptions, Limitations, and Scope**

A piece of music consists of numerous pieces of information such as rhythm, tempo, meter, pitches, tonality, texture, and dynamics. In most cases, vocal music is layered with an additional piece of information, lyrics. Although lyrics are an important component of a song,
lyrics are a distinct type of information compared to other musical information: Lyrics are based in a different syntactic and semantic system, the linguistic system (Asano, 2016). For this reason, in this dissertation, lyrics are called verbal information compared to other musical information.

Song learning includes not only accurate singing of the lyrics and tune, and remembering those components, but also understanding the musical context of the song, understanding of the key, tonality, meter, and other musical characters of the song, as well as singing with expression. However, in this dissertation, recall accuracy of the learned song was used as an outcome of song learning under two different instructional conditions (with and without visually presented lyrics). Specifically, recall accuracy of lyrics, pitches, and rhythm was the focus; other aspects of song performance were not of interest and were not assessed.

These studies were designed and hypothesized from a cognitive load perspective. Research investigating cognitive load effects has usually employed words and pictures as task materials (Mayer & Moreno, 2003). Only Owens and Sweller (2008) explored the transferability of cognitive load effects to music instruction measuring participants’ comprehension level. However, the task used in my studies was a song recall, and the materials were lyrics and tunes instead of words and pictures. Hence, careful consideration needs to be taken when applying cognitive load effects to my work and when comparing results of my studies to other cognitive load studies.

In the two studies reported in this dissertation, hypotheses were tested after one-time instruction. In reality, learners more typically learn a song over time. However, findings gained during a short instructional period, such as employed in this study, might be advantageous in order to understand the process the learners experience at that moment of learning.
Participants in both studies of this dissertation were undergraduates. Learning a song requires a certain level of competence in music and language. Since the studies in this dissertation tested the effect of visually presented lyrics, participants’ reading ability of words mattered. Thus, undergraduates were chosen as participants assuming undergraduates are competent at reading the words and understanding the lyrics. The results gained from this population may be age-specific.

**Definition of Key Terminology**

The definitions of key terminology used in this dissertation came from cognitive load literature including Mayer (2014), Sweller, (2011), and Sweller, Ayres, & Kalyuga, (2011). They are used throughout this dissertation as defined below.

**Dual channel.** Dual channel refers to a theoretical assumption is that humans have separate processors for visual and auditory information. Auditory information is processed in the aural channel, and visual information is processed in the visual channel. (Baddeley, 1992; Mayer & Moreno, 2003; Penney, 1989)

**Limited capacity.** The characteristic of working memory in which only a limited amount of new information can be processed in each channel at any given time (Baddeley, 1992; Cowan, 2000; Miller, 1956)

**Cognitive load.** The total amount of information imposed on working memory (Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011)

**Cognitive overload.** The mental phenomenon humans experience when incoming information exceeds their working memory capacity to process information (Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011)
Difficult song. A song has much new information possibly leading to cognitive overload in the learner while learning the song was defined as a difficult song. The level of difficulty of a song could be different for learners depending on their levels of musical expertise and experience.

Organization of the Dissertation

This dissertation includes two research studies and is organized as described below. Following this introduction chapter, in Chapter 2, I present a review of literature in four areas: a) the role of lyrics in singing and song acquisition, b) the relationship between tunes and words in memory, c) effects of presentational modalities on music instruction, and d) cognitive load effects related to my dissertation topic.

Chapter 3 is a self-contained article of the first research project, in which I investigated effects of visually presented lyrics on song recall. Additionally, effects of perceived task difficulty, preferred learning styles, and college major (music/non-music) were examined. Twenty-six members of auditioned choirs at a major mid-Atlantic university learned a difficult song individually through prerecorded aural instruction via computer: One group saw the lyrics and the other group did not. Data were analyzed through multivariate analyses of covariance controlling for phonological working memory.

Chapter 4 is a self-contained article of the second research study. The main purpose of this study was to ascertain the mediating effect of cognitive load in song recall with visually presented lyrics. In other words, I examined whether seeing the lyrics while learning a difficult song aurally induces less cognitive load in learners compared to not seeing the lyrics, leading to better recall accuracy of the learned song. Cognitive load was measured through reaction time measures based on a dual-task paradigm. Recall accuracy of the learned song was assessed in
terms of lyrics, pitches, and rhythm. Thirty-six non-music majors individually learned two songs through prerecorded aural instruction. During the instructional phase, for one song, they saw the lyrics, and for the other song, they did not see the lyrics. The presentation order of instruction condition (with and without visually presented lyrics) and song were counterbalanced. Data were analyzed through repeated measures of multivariate analysis of variance and MEMORE (Montoya & Hayes, 2016), a mediation analysis.

Lastly, discussion of the results of the studies and implications for future research and practice were discussed in Chapter 5. References for each chapter are provided at the end of the chapter.
References for Chapter 1


http://dx.doi.org/10.1037/1076-898X.3.4.257


https://doi.org/10.1177/0305735692201005
Chapter 2 Review of Literature

A song is an artistic representation of words with musical components. Thus, learning a song requires processing verbal (i.e., lyrics) and musical (e.g., pitches and rhythm) information. From a cognitive load perspective, how showing the lyrics to the learners affects their song recall is the main topic of this dissertation. In this chapter, to better understand this issue and provide further theoretical foundation for the studies, I reviewed studies investigating the role of lyrics in singing and song acquisition. Then, studies examining the relationship between tune and lyrics in memory as well as studies examining the effects of presentation modalities (auditory, visual, and kinesthetic) on music learning were reviewed. Lastly, the studies regarding cognitive load effects, which directly relate to my dissertation topic, were reviewed.

The Role of Lyrics in Singing and Song Acquisition

Singing is a basic activity in general music classrooms and developing students’ singing skills by teaching songs in an effective and efficient way has been an important topic in music education (Nichols, 2016; Rutkowski & Miller, 2003). Music educators have explored various factors that may affect children’s singing accuracy and using of singing voice (Goetze, Cooper, & Brown, 2010; Pfordresher et al., 2015). Hedden (2012) categorized the affecting variables into two aspects: internal and external. Internal variables included vocal range, pitch matching ability, and sex difference. Solo versus group singing, use of accompaniment, integration of text, and vocal modeling were considered external variables.

Among various factors, how singing with or without lyrics (the words of a song) affects a person’s singing accuracy and song acquisition has been closely examined by several music educators. Singing accuracy usually has been measured by imitation tasks (i.e., matching a single pitch, interval, patterns, or unfamiliar melodies) and/or recall tasks using familiar songs
that participants already know and can sing from their long-term memory (Berkowska & Dalla Bella, 2013; Nichols, 2016). On the other hand, studies focusing on the effects of particular instructional approaches or conditions on song acquisition have utilized recognition or reproduction tasks with unfamiliar songs, which means participants need to learn the song in the study (e.g., Gault, 2002; Levinowitz, 1989).

Results from studies measuring singing accuracy with and without lyrics have been inconsistent. When examining the text condition in Kindergarten, 1st grade and 3rd grade students, Goetze (1985) found singing with “loo” were more accurate than singing with lyrics, especially for Kindergarten and 1st grade students. In a replication study of Goetze’s work (1985), Smale (1987) did not find any significant difference between singing with and without lyrics in 4- and 5-year-old children. Jacobi-Karna (1996) reported the 4-year-old children sang more accurately with text while the other age groups, 3- and 5-year-old children did not. Whereas the study mentioned above investigated young children ranging from 3-year-old to 3rd grade, Berkowska and Dalla Bella (2009) studied adults’ singing proficiency. They found occasional singers who are not professional musicians sang more accurately when singing without lyrics compared to singing with lyrics. They claimed that singing on a syllable reduced the participants’ linguistic memory load and allowed them to focus on retrieving melodic information.

As mentioned above, unlike singing accuracy studies, research studies focusing on song acquisition required participants to learn an unfamiliar song and employed recognition or recall tasks to examine the effect of particular instructional conditions. Regarding the role of lyrics in song acquisition, two particular studies (Gault, 2002; Levinowitz, 1989) that employed recall task are discussed next.
Levinowitz (1989) examined the effect of learning songs with and without lyrics on singing accuracy of the learned songs in terms of tonal and rhythm accuracy. Two intact classes consisting of 4- and 5-year-old children received music instruction from the researcher once a week for five months including singing with and without text, moving, and chanting. During the last month of music class, they learned one song with lyrics and the other song with the syllable “bum”. Both songs were considered similar in that they have similar melodic, rhythmic, and harmonic components. Children’s singing of two songs with and without lyrics was rated based on two 5-point rating scales - one for tonal achievement and one for rhythmic achievement. The rhythmic performance scores did not differ, but the tonal scores differed between two test conditions, favoring singing on the neutral syllable. It is notable that in Levinowitz’s study (1989) while the order of songs when recording children’s singing was counterbalanced, the text condition for each song was not counterbalanced: one song was sung with text and the other song was sung on “bum.” However, the efficacy of learning songs with and without lyrics may depend on the song to be taught as discussed below.

Gault (2002) investigated effects of pedagogical approach (whole-song vs. phrase by phrase), text condition (singing with vs. without lyrics) along with developmental music aptitude on children’s song learning. Children in four Kindergarten and four 1st grade intact classes, randomly assigned to one of four treatment groups, learned two different folk songs. In the first four-weeks, they learned “Let Us Chase the Squirrel”, then learned “All Around the Buttercup” in the following four-weeks with their corresponding pedagogical approach/text condition. Their song performances were individually recorded during the fourth week of each four-week treatment period and evaluated in terms of tonal and rhythm based on a 5-point rating scale designed by the researcher. The results regarding pedagogical approach and text
condition depended on the song taught. For “Let Us Chase the Squirrel,” the presence of text condition and phrase by phrase approach were favored, but not for “All Around the Buttercup.” The author noted that participants sang “All Around the Buttercup” more accurately than “Let Us Chase the Squirrel” speculating that the experience gained while learning the first song “Let Us Chase the Squirrel” might have carried over to learning the second song “All Around the Buttercup,” reducing the effects of the text and pedagogical conditions since both songs have similar musical characteristics such as duple meter and Do tetratonic scale (Do, Re, Mi, Sol). Thus, when using two songs, counterbalancing the order of songs participants learn should be considered to ameliorate possible order effect. Students with high developmental music aptitude performed better than those with low music aptitude, but no significant interaction with other variables (text condition and pedagogical approach) was found.

Regarding singing with text, several music educators (e.g., Andress, 1986; Goetze et al, 2010; Levinowitz, 1987) expressed concern that the lyrics of a song might distract the children’s attention away from the musical aspects of the song. Levinowitz (1989) suggested whether the lyrics of the song distract children’s attention on the tonal/rhythmic aspects of the song might depend on the children’s language development.

[A] child whose language development is above average may not be distracted by the words and, therefore be able to attend to the music elements of the song. The young child who is facile with the language might learn the words first and then associate the melody with those words. The child whose language development is average or below average, however, may be so distracted by the words that his is unable to attend to the musical elements of that song. An alternative explanation is that with less language
facility a child first learns the melody and then associates the words with that particular melody (p. 15).

It seems an interesting point in that students’ language ability was considered in learning a song, but in Levinowitz’ (1989) study, the correlation between young children’s ability to sing the learned songs and their language development was low when the children’s ages in months were controlled for.

Summary

The role of lyrics in singing accuracy and song acquisition has been investigated with text condition (singing with lyrics or singing on a syllable). As Goetze, Cooper, and Brown (1990) stated, “Although there may be theoretical reasons to think text might distract young children’s attention from the singing process, there are no consistent findings to confirm that children learn to sing in tune most effectively without text” (p. 26). Further, children’s developmental music aptitude (Gault, 2002) and language development (Levinowitz, 1989) did not interact with text condition. Although in the case of adults who are occasional singers, singing without lyrics yielded more accurate performance than singing with lyrics (Berkowska & Dalla Bella, 2009). However, this does not guarantee singing without lyrics leads to better song acquisition.

The main focus of singing accuracy studies is different from the main interest of song acquisition studies, thus, the corresponding methods would be different. Song acquisition can be tested through singing accuracy, but it is tested with a newly learned song not through an imitation task or singing a familiar song stored in long-term memory. Hence, how singing with or without lyrics affects song acquisition could be another area to explore.
Relationship between Tunes and Words in Memory

The nature of relationship between words and tune of a song in memory has been investigated by several researchers in a variety of ways (see Ginsborg & Sloboda, 2007). The main questions posed in these studies were whether the words and tunes are independent or integrated in memory and if integrated, to what extent they are integrated in memory. This question was investigated largely through two types of memory retrieval tasks: recognition and recall. Recognition task in this area has been measured by asking participants to identify whether the tune or words is familiar (whether they heard it in the learning stage) or not (Crowder, Serafine, & Repp, 1990; Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). Recall task requires participants to reproduce the tune, words, or song from memory.

Serafine and her colleagues (1984, 1986), in their classical studies, demonstrated the integration effect in memory for songs, suggesting that words and tunes of a song are integrated in memory. The participants in all experiments in both studies (1984 & 1986) were undergraduate students; their level of musical expertise was not a variable of interest. Excerpts of American folksongs (considered unfamiliar to the participants) with interchangeable lyrics and melody were used for recognition tasks. After listening to several pairs of melody and words in a session, participants were asked to indicate whether they heard the melody or words in the previous session and if heard, which component (melody or words) they heard. The test items were slightly different in each experiment according to the specific research questions they posed, but the comparison of central interest was based on two types of test items - original pair and mismatched pair. In both types, the tune and words were previously presented, but in mismatched pair, the tune and words came from two different songs while original pair
was the exact same song presented previously. Melodies were better recognized when paired with the words originally heard with the melody than when being paired with the equally familiar words but from a different song. Similarly, words were better recognized in original pair items than mismatched pair items. This integration effect was found even when tunes were paired with nonsense texts to rule out the possible effect of semantic connotation of the words on melody (Serafine et al., 1986).

Later on, Crowder, Serafine, & Repp (1990) tested possible theories that may explain the integration effect. One was the physical interaction hypothesis, more specifically submelodic hypothesis, which asserts that different phonetic properties of different words change acoustic properties of notes such as articulation and this physical change might be subtle but memorable. The submelodic hypothesis was supported in that the melody was better recalled when being paired with nonsense words having similar phonetic properties with the original nonsense words than when being paired with nonsense words having different phonetic properties with the original nonsense words.

Another explanation was the association-by-contiguity hypothesis, which holds that any two events co-occurring (in close temporal proximity) may become connected in memory and serve as a retrieval cue for one another. To test this hypothesis, melody and words were presented at the same time, but separately (as a form of hummed melody and chanted words that retained the rhythm of the song). The co-occurrence of the hummed melody and chanted words was a sufficient condition to establish the integration effect in recognition tests. This finding was not in favor of the submelodic hypothesis in that the melody and words did not physically affect each other in the study condition. However, Crower et al. (1990) argued “Hearing a hummed melody at the same time as one hears a rhythmically compatible stream of
words might produce the experience of a song, whether the subject is deliberately trying to generate this song or not” (p. 475).

Although Serafine et al. (1984, 1986, 1990) did not examine participant’s musical experience or level of musical expertise in a series of their studies, they predicted stronger integration for musically inexperienced participants than for more musically experienced ones as well as stronger effect for younger than older participants (Crowder et al., 1990). In a replication study of the Serafine et al. study (1984) with preschool children and adults, Morrongiello and Roes (1990) found the integration effect was greater for adults than children unlike the prediction by Crowder et al. (1990). For the preschool children, the words played a critical role in the recognition task in that they indicated “exactly the same” if the words were the same regardless of the tune conditions (heard, not heard, or matched with a different text), but if the words were different, they indicated “not all the same” although the tune was heard in the learning session: Their judgment of the sameness of songs depended on the words of the song. However, as the authors (Morrongiello & Roes, 1990) pointed out, children considered the words as the song and thus their conception seemed to affect the test results instead of their cognition (Chen-Hafteck, 1999). This issue could be examined through a recognition task without words. Indeed, Feierabend, Saunders, Holahan, and Getnick (1998) found preschool children better recognized the songs without words than the songs with words. Also, as Morrongiello and Roes (1990) noted, if a different task such as recall instead of recognition is used in which children are required to produce the song, a stronger integration effect might be found for children. For example, Serafine et al. (1984) mentioned that a 2-year-old who was able to sing many songs accurately could not sing the songs on a syllable without text. Thus,
the recognition task might not be an optimal task to test the integration effect for young children.

Unlike recognition task, studies utilizing a recall task required participants to write down the words on paper (Calvert & Tart, 1993; Chazin & Neuschatz, 1990; Kilgour, Jakobson, & Cuddy, 2000; McElhinney & Annett, 1996; Purnell-Sebb & Speelman, 2008; Wallace, 1994), recite the words (Calvert & Billingsley, 1998; Racette & Peretz, 2007; Rainey & Larsen, 2002) or sing the song (Ginsborg & Sloboda, 2007; Racette & Peretz, 2007 in the sung-sung condition only). Some researchers used unfamiliar (not heard before) melodies and words asking participants to learn and remember through the study procedure (Ginsborg & Sloboda, 2007; Kilgour et al., 2000; McElhinney & Annett, 1996; Racette & Peretz, 2007; Wallace, 1994) whereas other researchers employed familiar melodies with new words/information set to it (Chazin & Neuschatz, 1990; Rainey & Larsen, 2002). Yet, Purnell-Webb and Speelman (2008) compared familiar and unfamiliar tunes on text recall. Participants in most studies mentioned above were undergraduates, otherwise preschool children (Calvert & Billingsley, 1998) or 8-year-old children along with young adults (Chazin & Neuschatz, 1990).

These different approaches might explain the inconsistent results in the studies directly and indirectly investigating the relationship between tunes and words in memory. Most studies mentioned above, except for Ginsborg and Sloboda (2007), investigated the efficacy of melody as a mnemonic tool in recalling words. Thus, their focus was on recall of text instead of recall of tune. Several studies demonstrated an advantage of sung over spoken materials in recalling text (McElhinney & Annett, 1996; Purnell-Webb & Speelman, 2008; Wallace, 1994), while some studies yielded contradictory findings. For example, two studies (Calvert & Tart, 1993; Rainey & Larsen, 2002) reported the benefit of sung words for long-term memory but not
immediate recall whereas Chazin and Neuschatz (1990) reported the opposite. Also, Kilgour et al. (2000) found when the presentation rate was the same, the spoken condition was superior in recalling the words.

Wallace (1994) pointed out a critical condition is required for a tune to function as a mnemonic tool: The tune should be easily or sufficiently learned through repetition. When the structure and feature of a tune is well understood and matched to the words, the rich and interconnected information of tune could serve as an encoding device as well as a retrieval cue for recall of text; otherwise, the effect of music is eliminated. Thus, different study materials and conditions, a different scoring system (e.g., alteration allowed or not) and measurement (e.g., the number of words recalled vs. the number of trials to learn the words/song) might bring such different results.

It seems worthy to review Ginsborg and Sloboda’s (2007) study in detail since they investigated singers’ recall of a newly learned song unlike other studies that focused on recall of text. In Ginsborg and Sloboda’s (2007) study, participants were randomly assigned to three conditions in which they were asked to deliberately learn a song and memorize it by themselves from notated music including the words and audio recordings. Conditions 1 and 2 required participants to learn the words and tune separately with words first and then tune (condition 1) and vice versa (condition 2), and finally words and tune together. In condition 3, participants learned the words and melody together throughout the learning session. The session took 20 minutes in total and in conditions 1 and 2, participants took 5 minutes to learn and memorize the words and 7 minutes for melody and 8 minutes for both components. When recalling, no musical score was provided. Participants’ immediate and 10-minute delayed recall were measured in terms of recall accuracy, fluency, and the types of error based on the number of
substitutions and omissions in melody and words, and the place of hesitations. In this study, only the 10-minute delayed recall was analyzed to avoid ceiling effects of the immediate recall. Participants with high levels of musical expertise recalled more accurately and fluently than those with low levels of musical expertise but only when they learned the words and melody together. When examining the error types, there were more separate errors than conjoint errors, which indicates participants could preserve one component (e.g., words or melody) while the other components was recalled in error. Finally, more hesitations happened at the ends of phrases rather than at the beginning or middle of phrases, suggesting, “the formal structure of music provides a framework for recall” (Ginsborg & Sloboda, 2007, p. 421).

Ginsborg and Sloboda (2007) concluded their study as follows.

The relationship between the words and melodies of songs in memory, for singers who have deliberately memorized songs, is such that they are neither recalled entirely separately nor integrated to such an extent that if one component is recalled erroneously, recall or the other is inevitably affected. Rather, they are stored and retrieved in association with one another. Thus, the best way for an expert singer to learn and memorize a song, assuming that both the words and melody are novel, is to memorize the words and melody together. (p. 437)

Ginsborg and Sloboda (2007) also suggested using the term “association” instead of “integration” to explain the probabilistic relationship between the lyrics and tune of a song in memory. If one component is retrieved it will increase the likelihood of recall for the other component while the term “integration” has a strong connotation of a unity. On the other hand, Serafine et al. (1984, 1986, 1990) used the term “association” as “mere knowledge of co-occurrence” (1984, p. 287) and distinguished “integration” from a holistic conception.
The cognitive relationship between tune and lyrics in a song is a subject still debated. Numerous neuroimaging (e.g., functional magnetic resonance imaging (fMRI)) and electrophysiological studies (e.g., electroencephalography (EEG), and event-related brain potentials (ERPs)) investigating the relationship also yielded contradictory results (See Peretz, Vuvan, Lagrois, & Armony, 2015). Sammler et al. (2010) examined the degree of integration for song perception using a functional magnetic resonance (fMR)-adaptation paradigm. According to Peretz et al. (2015), fMRI adaptation paradigm is a decent technique that “can be exploited to distinguish domain-specific neural activation at a finer-grained level than the standard use of fMRI can offer” (p. 3). Sammler et al. (2010) found lyrics and tunes are integrated at prelexical, phonemic processing levels in the stage of auditory analysis as well as in the preparation of motor output while they are processed independently at structural and semantic levels, concluding “lyrics and tunes of unfamiliar songs are processed at different degrees of integration … at different stages of the processing” (p. 3576). Also, they pointed out, “beyond auditory perceptual processing, the degree of integration/ separation depends on the specific cognitive processes targeted by an experimental task (e.g., recognition vs recall or production of familiar vs unfamiliar songs), perhaps accounting for some of the conflicting results” (p. 3576).

Summary

Lyrics and tune of a song seem stored and retrieved in an integrated way to some degree (Serafine et al. 1984, 1986; Crowder et al. 1990). Several studies (Calvert & Tart, 1993; McElhinney & Annett, 1996; Purnell-Webb & Speelman, 2008; Wallace, 1994) investigating the effect of tune on recall of text also showed how musical structure and characteristics can function as a retrieval cue, suggesting some level of association between lyrics and tune in
memory. Also, the level of musical expertise matters on this issue: Individuals with musical training employ superior encoding strategies regarding musical materials (Kilgour et al., 2000) and singers with a high level of musical expertise seem to process the lyrics and melody in a more integrated way than singers with the lower level of musical expertise (Ginsborg & Sloboda, 2007).

While behavioral studies used a recognition or recall test to investigate how lyrics and tune are stored and retrieved in memory, brain studies employing electrophysiological and neuroimaging approaches allow examination of the relationship of lyrics and tune in the process of perception and retrieval. One current view is lyrics and tunes are integrated at prelexical, phonemic processing levels in the stage of auditory analysis as well as in the preparation of motor output while they are processed independently at structural and semantic levels (Sammler et al., 2010)

**Effects of Presentation Modalities on Music Instruction**

Music teachers often use visual and/or kinesthetic aids to promote students’ musical skills and understanding (Sheldon, 1991; Tarnowski, 1986; Zikmund & Nierman, 1992). Visual and kinesthetic representations may help learners understand and experience the transient and intangible characteristics of musical elements. While providing music instruction using several sense modalities have been advocated (Tarnowski, 1986), several researchers examined the efficacy of multimodal approaches regarding singing achievement (Apfelstadt, 1984; Hughes, 1991; Pautz, 1988; Persellin, 1994; Tarnowski, 1986), rhythm and/or melody retention (Persellin, 1992, 1994; Persellin & Pierce, 1988, Zikmund & Nierman, 1992), aural and instrumental performance skills (Kendall, 1988), and perception of musical form (Gomko & Poorman, 1998).
While various types of visual and kinesthetic aids have been used, the musical elements mostly used in research studies were melodic contour (Apfelstadt, 1984; Dunn, 2008; Gomko & Poorman, 1998; Hughes, 1991; Pautz, 1988; Persellin, 1993; 1994; Zikmund & Nierman, 1992; Tarnowski, 1986) and melodic rhythm (Dunn, 2008; Pautz, 1988; Persellin, 1988, 1992, 1994; Zikmund & Nierman, 1992; Tarnowski, 1986). Melodic contours are often visually represented through line drawings or kinesthetically experienced through finger tracing. Melodic rhythm, the duration of each note or intervals of two successive notes, is iconically represented with short and long lines or kinesthetically experienced by being tapped or clapping. In addition, Hughes (1991) provided traditional notation, stair-step melody bells and a vertical keyboard and had students feel the kinesthetic sensations of their larynx in order to teach melodic direction.

Besides melodic contour and rhythm, some researchers (Persellin, 1994; Tarnowski, 1986) provided visual and/or kinesthetic reinforcement of lyrics. In Persellin’s (1994) study, preschool children were assigned to one of four groups – aural, visual, kinesthetic, and multimodal. The visual group saw pictures representing lyrics as well as iconic notation (i.e., line drawings) representing melodic contour and rhythm, but other groups did not receive reinforcement of lyrics. Visual instruction was not effective for children in developing their ability to match pitch compared to aural and multimodal instruction. Persellin (1994) speculated the reason as below.

Children in the visual class were responsive to visual aids but would need to be reminded to sing while seeing the stimulus. Their attention was so fixed on the pictures that the music was of only secondary importance. This may be why their ability to match pitch did not improve over the ten-week period. (p(865,943),(891,975) 1233)
This speculation could be strengthened if the author ensured the children’s basic abilities to match pitch were equivalent across the groups by giving them the same test prior to the experimentation. If their pitch-matching abilities were different prior to instruction, it is hard to determine whether the treatment was effective or not.

Tarnowski (1986) tested three instructional modes - single (aural only), dual (aural-kinesthetic), and multimodal (aural-kinesthetic plus visual) - with 66 preschool children aged 24 months to 65 months. In dual and multimodal conditions, musical components such as steady beat, phrases, melodic contour, rhythm as well as lyrics were reinforced kinesthetically and/or visually. Lyrics were represented with gestures or pictures. Children’s singing of two learned songs were assessed in terms of phrase, lyrics, steady beat, melodic rhythm, melodic contour, tonality, and melodic interval based on 5-point scales in a phrase unit. No difference was found in seven outcome categories among instructional modes. The wide range of participants’ ages and repeated song instructions (they learn the same songs over 12 instructional sessions) might reduce the possible difference in learning outcomes.

Including studies reviewed above, results from studies investigating the effects of multimodal approaches on song instruction have been contradictory. Studies involving preschool and kindergarten children found group differences in pitch-pattern echoing (Apfelstadt, 1984; Pesellin, 1994). In Apfelstadt’s (1984) study, the group who learned the song with visual and kinesthetic reinforcements sang the pitch-pattern more accurately than the group who learned the song without any reinforcement. Persellin (1994) reported children receiving either the aural or multimodal treatment better matched a pitch pattern (i.e., G and E) than children receiving either the visual or kinesthetic treatment. On the other hand, studies
testing singing of 3rd and/or 4th grade students (Hughes, 1991; Pautz, 1988) did not find any difference among instructional conditions.

Whereas the effectiveness of adding iconic notation and gesture to music instruction was tested in general music class settings with children in preschool, kindergarten, 3rd or 4th grades in the aforementioned studies, the effects of using traditional notation were tested with an elementary instrumental ensemble (Kendall, 1988) and college students (Korenman & Peynircioglu, 2007). Kendall (1988) investigated whether adding music reading activities “during beginning instrumental music instruction impedes students’ development of aural musicianship and instrumental performance skills because of the division of their attention between aural and visual activities” (p. 215). Introduction of music reading activities did not impede students’ musical development in terms of ear-to-hand coordination and performance skills, rather it helped to develop students’ verbal association and sight-reading (Kendall, 1988).

Korenman and Peynircioglu (2007) investigated the effects of presentation modality and preferred learning style on learning efficiency and memory of unfamiliar melodies and sentences. Learning efficiency was measured by tallying the number of trials needed to reach the criterion set by the researchers and memory was assessed by recognition tests. No main effect of presentation modality was found, but a significant interaction between presentation modality and learning style was revealed. Participants learned faster and remembered better when the materials were presented in their preferred learning modality. It seems noteworthy that when participants were asked to recall visually presented items, they wrote the musical notes on a blank staff and the sentence on a blank sheet of paper. In most music education studies, visual and/or kinesthetic information was provided in addition to aural information.
whereas Korenman and Peynircioglu (2007) presented instructional materials either in the visual or aural form in the learning and testing stages.

As shown by Korenman and Peynircioglu (2007), the effect of presentation modality on music learning has been frequently examined in relation to students’ learning styles (Dunn, 2008). The definitions of learning styles are varied (Messick, 1984; Peterson, Rayner, & Armstrong, 2009; Willingham, Hughes, & Dobolyi, 2015), but perhaps the broad definition is “a combination of cognitive, affective, and physiological factors that merge to define each student’s unique approach to effective learning” (Wilson, 2012, p. 69). As the efficacy of certain instructional designs might depend on individual differences, learning styles theories hold the idea that “people learn in different ways and that learning can be optimized for an individual by tailoring instruction to his or her style” (Willingham et al., 2015, p. 266). This idea has been tested in the area of music education as well: Presentational modalities - presenting instructional materials in different perceptual modes - have been studied with preferred learning modalities (Zikmund & Nierman, 1992).

Participants’ learning styles have been assessed mainly through two types of instruments: a preference measure and a task-oriented measure (Dunn, 2008). The preference measure uses inventories such as the Learning Style Inventory (Dunn, Dunn, & Price, 1975) and asked people to respond to several questions or statements by checking one of the options provided or ranking responses (Guild & Garger, 1985). In the task-oriented measure such as the Swassing-Barbe Modality Index (SBMI, Swassing & Barbe, 1979), the learning style is identified based on the degree of success in completing tasks representing each perceptual modality (Pautz, 1988). People taking these instruments are identified as a visual learner, aural learner or kinesthetic learner, but the preference measure indicates learners’ preference for the
sense modality when learning while the task-oriented measure identifies learners’ perceptual modality strengths.

In the area of music education, studies investigating the effects of matching the instructional modes to children’s learning styles have yielded conflicting results: Some studies demonstrated the efficacy of tailoring instruction to students’ learning styles (Dunn, 2008; Persellin & Pierce, 1988; Korenman & Peynircioglu, 2007; Zikmund & Nierman, 1992) while others did not find any supporting evidence for customized instruction (Hughes, 1991; Pautz, 1988). Although some studies reported positive effects of style-based instruction, care should be taken when interpreting the findings of those studies because of the research design they employed. For example, Dunn (2008) investigated the effect of learners’ perceptual modality strengths on music listening with a qualitative approach, which is not designed to validate a theory or hypothesis. In the study by Zikmund and Nierman (1992), students identified as visual learners or kinesthetic learners received only the respective visual or kinesthetic reinforcement. In this design, it is hard to discern whether receiving visual or kinesthetic reinforcement was just beneficial regardless of students’ learning styles. An appropriate design should include all types of auditory, visual, and kinesthetic learners in every instructional condition to demonstrate cross-interaction between learning styles and presentational modalities (Rohrer & Pashler, 2012; Willingham et al., 2015).

In that sense, the study by Korenman and Peynircioglu (2007) seems valid in that all participants experienced two presentational modalities (auditory and visual) and interaction between presentational modalities and learners’ preferred learning modalities were revealed. However, as pointed out earlier, while most music education studies presented visual materials in addition to aural stimuli, Korenman and Peynircioglu (2007) presented instructional
materials either in the visual or aural form. For example, in the visual presentation mode, participants learned music (lyrics and melody) by seeing the music notation without any aural stimuli and they were asked to recall the musical information by writing it down on a blank sheet of paper. Although they found visual learners benefited most from the visual presentation mode, this would not be apply to learning music in music class because no aural experience was involved. Music is inherently an aural art and developing students’ aural capacity is one of goals of music education. In that sense, it is questionable whether the research findings are directly applicable to teaching practice.

**Summary**

Several researchers examined the efficacy of multimodal approaches regarding singing achievement (Apfelstadt, 1984; Hughes, 1991; Pautz, 1988; Persellin, 1994; Tarnowski, 1986), rhythm and/or melody retention (Persellin, 1992, 1994; Persellin & Pierce, 1988, Zikmund & Nierman, 1992), aural and instrumental performance skills (Kendall, 1988), and perception of musical form (Gomko & Poorman, 1998). In most studies, musical components represented visually and kinesthetically were melodic contour (Apfelstadt, 1984; Dunn, 2008; Gomko & Poorman, 1998; Hughes, 1991; Pautz, 1988; Persellin, 1993; 1994; Zikmund & Nierman, 1992; Tarnowski, 1986) and melodic rhythm (Dunn, 2008; Pautz, 1988; Persellin, 1988, 1992, 1994; Zikmund & Nierman, 1992; Tarnowski, 1986). Although some researchers (Persellin, 1994; Tarnowski, 1986) provided visual and kinesthetic reinforcements of lyrics along with other musical components, the visual representation of lyrics was in a form of pictures not written text. It is noteworthy that in most studies visual or kinesthetic information was presented in addition to aural information to reinforce certain musical concepts such as melodic contour and melodic rhythm.
Some studies examined the effects of presentation modalities on music learning in relation to students’ learning styles, but the results were inconsistent: Some studies demonstrated the efficacy of tailoring instruction to students’ learning styles (Dunn, 2008; Persellin & Pierce, 1988; Korenman & Peynircioglu, 2007; Zikmund & Nierman, 1992) while others did not find any supporting evidence for customized instruction (Hughes, 1991; Pautz, 1988). However, as noted earlier, many studies lacked a rigorous enough study design to validate the effectiveness of learning style-based instruction. Also, many scholars (Macdonald, Germine, Anderson, Christodoulou, & McGrath, 2017; Rohre & Pashler, 2012; Pashler et al., 2008; Willingham et al., 2015) pointed out the lack of sufficient evidence to tailor instruction to students’ preferred learning styles.

**Cognitive Load Effects**

In the area of cognitive psychology, mental processes such as attention, perception, memory, reasoning, and problem solving are explained based on the information processing model. In the information processing model, environmental energy is considered information, the processes of perception and comprehension is coding, and memory is storage and retrieval. This analogy provides a general framework to understand and describe human cognitive processes (Hunt & Ellis, 2004). The model posits three subsystems: sensory register, working (short-term) memory, and long-term memory (Atkinson & Shiffrin, 1969; Baddeley & Hitch, 1974). While all information enter through a sensory register, only attended information is processed in working memory. The encoded information is saved in and retrieved from long-term memory. The capacity of sensory register is vast, yet sensory memory decays quickly. Whereas long-term memory is assumed to be unlimited in terms of duration and capacity to
Cognitive load refers to the total amount of information imposed on working memory (Sweller, 1994). Given the limited capacity of working memory, it is assumed imposing a heavy cognitive load hinders learning (Mayer, 2014; Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011). A number of researchers have investigated how to optimize instructional design to reduce cognitive load in learners (e.g., Kalyuga, Chandler, & Sweller, 1999; Mayer & Moreno, 1998, 1999; Mousavi, Low, & Sweller, 1995; Tabbers, Marten, & van Merriënboer, 2004).

Cognitive load theory (Sweller, 1994) suggests three different types of cognitive load: intrinsic, extraneous, and germane. Intrinsic cognitive load relates to the “complexity of the knowledge” (Sweller, 2011, p. 57) and “inherent characteristics of the content” (de Jong, 2010, p. 106) to be learned. The more connections that are required among the elements of material, the more difficult it is to understand. The fewer connections required among the elements of material, the easier it is to understand (Sweller, van Merriënboer, & Paas, 1998). Extraneous cognitive load relates to how learning materials are presented. Extraneous cognitive load can be altered by different instructional designs. Lastly, germane cognitive load relates to the processes of construction of new schema and automation of schemas (Sweller et al., 1998). To construct a new schema, several processes such as differentiating, classifying, organizing, interpreting, and inferring (Mayer & Moreno, 2003) are involved. The load imposed on these processes is called germane load. Theoretically, intrinsic load cannot be changed, but the amount of extraneous and germane cognitive load can be altered by instruction. To reduce the extraneous cognitive load and provide an appropriate level of germane cognitive load is
considered one goal of instructional design (Sweller, 1994, 2011; Sweller, Ayres, & Kalyuga, 2011).

A large body of empirical research has investigated effects of particular instructional designs on learning outcomes. The effects demonstrated and explained from a cognitive load perspective have been called cognitive load effects (Mayer, 2014; Sweller, 2011). One theory regarding cognitive load is cognitive theory of multimedia learning (Mayer, 2008). In this theory, multimedia learning refers to “learning from words and pictures” (Mayer & Moreno, 2003, p. 43). Thus, studies based on this theory usually tested the instructional condition containing verbal (i.e., words) and visual (e.g., pictures) information. However, the principles of instructional design distilled from those studies have been applied to different type of information as well (e.g., Brünken, Plass, & Leutner, 2004; Owens & Sweller, 2008).

The cognitive theory of multimedia learning is based on three assumptions: dual channel, limited capacity, and active processing (Mayer, 2014; Mayer & Moreno, 1998, 2003). Humans possess separate systems to process visual and auditory information, called dual channel. Human beings have limited capacity to process novel information in each channel at any given time. And humans learn through active processing by attending to the incoming information, organizing attended information into a coherent mental representation, and integrating mental representations with previous knowledge (Mayer, 2003).

Based on these assumptions, a number of cognitive load effects were demonstrated (Mayer, 2014; Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011). Two of those effects - modality effect and verbal redundancy effect - are discussed next; both of which are closely related to this dissertation topic.
Modality Effect

In the cognitive load literature, the modality effect refers to the experimental demonstration that employing two sensory modalities rather than one is more effective when the instructional materials are unintelligible in isolation (Low & Sweller, 2014; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995; Tabbers, Martens, & van Merriënboer, 2004; Tindall-Ford, Chandler, & Sweller, 1997). When two sources of information should be mentally integrated for understanding, it is better to present one source visually and the other source aurally rather than to present both only visually. The modality effect is considered “one of the most reliable and valid instructional design effects in multimedia learning” (Brünken et al, 2004, p. 118).

Mayer and Moreno demonstrated this modality effect in multimedia instruction. Students performed better on transfer tests when instructional materials were presented as pictures and narration rather than as pictures and on-screen text (Mayer & Moreno, 1998, 1999). Presenting both pictures and words visually might cause cognitive overload in the visual channel while the aural channel is unused. On the contrary, if words are presented aurally as a form of narration instead of on-screen text, the visual channel might be offloaded with regard to the verbal information having more capacity to process the pictures. Mousavi, Low, and Sweller (1995) also reported that students performed better on geometry problem tests when learning with an audio-visual format rather than a visual only format. It is assumed the reduction in the cognitive load in the visual channel enhances learning.

Owens and Sweller (2008) examined the applicability of the modality effect as well as split-attention effect to music theory instruction. Split-attention occurs when learners need to split their attention between multiple sources of information, which hinders learning (Ayers &
Sweller, 2014; Sweller, 2011). By presenting instructional materials in audio-visual format, the split-attention effect can be avoided. When learning from spatially and temporally integrated materials, participants better performed in the knowledge and transfer tests than when learning from separated materials. The tests included filling a blank with a correct note in a bar based on time signature and identifying the type of meters.

The basic idea of the modality effect is reducing cognitive load by employing two channels so that the cognitive resources of both systems can be utilized. However, it is noteworthy that modality effect can be achieved under certain conditions. Under self-paced conditions, the visual-only format was better than the audio-visual format since learners could have sufficient time to transfer critical information to long-term memory, eliminating overload effects due to the time limitation of working memory (Tabbers, Marten, & van Merriënboer, 2004). If materials are easily understood without connecting to other elements of the material, it is said that element interactivity is low. If element interactivity is low, any cognitive load effects cannot be obtained including the modality effect (Sweller, 2014). The length of text does matter. If text is lengthy and complex, written format is better than auditory format or bi-sensory (written and aural) format because of the transient characteristics of auditory information (Low & Sweller, 2014).

**Verbal Redundancy Effect**

Verbal redundancy refers to “the simultaneous presentation of text and narration with identical words” (Moreno & Mayer, 2002, p. 156). In the literature of working memory, presenting identical words in two modalities helps learning, yielding better memory recall than presenting the words in only one modality. Attempting to maximize the working memory
capacity by employing two channels when presenting information is in line with the basic idea of modality effect.

Penney (1989) reviewed a large body of research indicating that presenting verbal material in two modalities yielded better memory recall than presenting verbal material in one modality. Lewandowski and Kobus (1993) also reported that participants recalled more words in the redundant verbal condition than the one modality condition. Verbal redundancy was helpful for less skilled readers to comprehend the text (Montali & Lewandowski, 1996). Furthermore, in Moreno and Mayer (2002)’s study, students better performed on retention, transfer and matching tests in the bi-sensory presentation condition compared to the one-sensory presentation condition.

However, cognitive load literature (Kalyuga, Chandler, & Sweller, 1999; Kalyuga & Sweller, 2014) yields conflicting results regarding verbal redundancy effects. As noted above, the modality effect is achieved only when the two sources cannot be understood in isolation, but verbal information provided as spoken words or written words can be understood by itself. Hence, Kalyuga and Sweller (2014) argued the additional presentation of verbal information in another format is redundant and causes unnecessary load for learners to coordinate two materials, which have the same information. Also, Kalyuga, Chandler, and Sweller (1999) reported eliminating redundant materials leads to better performance than including the redundant material. In computer-based instructional materials, simultaneous presentation of diagrams and auditory-verbal information was more effective than simultaneous presentation of diagrams with both auditory-verbal and visual-verbal information. Given the limited capacity of working memory, adding redundant information might impose unnecessary cognitive load that interferes with learning.
Moreno and Mayer (2002) explained the reason for the discrepancy found between working memory literature and cognitive load literature, distinguishing the two terms of mode and modality. According to Moreno and Mayer, mode refers to “the format used to represent the lesson, such as words versus pictures” while modality refers to “the information-processing channel used by the learner to process the information, such as auditory versus visual” (p. 156). They pointed out results may be contradictory because of the use of different instructional materials. In verbal redundancy studies in the working memory literature (Lewandowski & Kobus, 1993; Montali & Lewandowski, 1996), only verbal information was used. The materials were written words, spoken words, or both written and spoken words. However, in the cognitive load literature, when testing the effect of verbal redundancy, another mode of visual information such as diagrams or graphs was presented along with verbal materials. In this case, the visual channel might be overloaded by two modes of information. Since learners split their attention to two visual materials, the redundant visually presented verbal material might hinder learning. Mayer and Moreno argued if there is no competing visual information, even redundant information could be helpful for the learner. Therefore, when discussing the verbal redundancy effect, the number of information modes in each channel should be considered. If only one mode of information is presented in each channel, duplication of verbal materials in both channels may be helpful, maximizing the use of working memory capacity.

Summary

Cognitive load effect refers to effects found in studies investigating particular instructional design and explained from a cognitive load perspective (Mayer, 2014; Sweller, 2011). The studies acknowledged three basic assumptions: dual channel, limited capacity, and active processing (Mayer & Moreno, 1998, 2003; Mayer, 2014). Among several cognitive load
effects, modality effect (Mousavi, Low, & Sweller, 1995; Mayer & Moreno, 1998, 1999) and verbal redundancy effect (Lewandowski & Kobus, 1993; Moreno & Mayer, 2002; Montali & Lewandowski, 1996) suggest presenting instructional materials using two modalities (visual and aural) is more beneficial for learners than using only one modality. Given limited capacity of working memory, reducing unnecessary cognitive load to maximize cognitive capacity should be considered when designing and presenting instructional materials.

**Chapter Summary**

Since a song consists of verbal and musical information, the relationship of these two types of information has been closely examined from various aspects: Topics include the role of words in singing accuracy and song acquisition; the role of tune in text recall; and the integration effects of tune and lyrics of in recognition and recall of a song. Specifically, these studies examined whether singing with or without words increases singing accuracy and song acquisition; whether sung words is better recognized and recalled than spoken words; and whether the lyrics and melody are stored and retrieved in association with one another. The inconsistent findings from these studies may be due to different study methods including varying participants, design, materials, and tasks.

Interestingly, most studies concerning the aforementioned topics employed aurally presented materials. Although Ginsborg and Sloboda (2007) provided a musical score in addition to an audio recording, the musical score was not a variable of interest. Some studies investigated presentational modalities on music instruction including song learning, but the most tested components were visual and kinesthetic representations of melodic contour and melodic rhythm. Some researchers also represented the lyrics using pictures in addition to other musical components, but the unique effect of visually presented lyrics was not examined.
Instructional materials should be designed and presented based on understanding how human minds work. Theories concerning cognitive load provide a framework for testing and explaining how particular instructional conditions affect learning. A number of cognitive load effects have been demonstrated across various instructional domains, yet few research studies have examined whether those effects can be applied to musical materials. The Owens and Sweller (2008) study is one exception, wherein the modality effect and split-attention effect were demonstrated with music theory instruction. The basic idea of the modality effect is presenting two types of instructional information using two sensory modalities – one in visual modality and the other in aural modality – reduces learners’ cognitive load and thus leads to better learning from the presented instructional materials than using one modality. Even when only verbal information – one type of information – is used, presenting the identical information aurally and visually was beneficial for learners. Thus, whether this basic idea of the modality and verbal redundancy effects could be applied to song learning, specifically through visually presented lyrics while learning a song aurally, is a worthwhile topic to explore and the focus of the studies in this dissertation.
References for Chapter 2


https://doi.org/10.1207/s15327043hup0603_3


https://doi.org/10.3389/fpsyg.2017.01314

http://dx.doi.org/10.1037/0003-066X.63.8.760


https://doi.org/10.1207/S15326985EP3801_6

https://doi.org/10.2466/pms.1996.82.2.395

https://doi.org/10.1080/00461528409529283

http://dx.doi.org/10.1037/h0043158

https://doi.org/10.1177/002221949602900305

http://dx.doi.org/10.1037/0022-0663.91.2.358
http://dx.doi.org/10.1037/0022-0663.94.1.156


http://dx.doi.org/10.1037/0022-0663.87.2.319

https://doi.org/10.1177/8755123315576764


http://dx.doi.org/10.1037/1076-898X.3.4.257

http://dx.doi.org/10.1037/0278-7393.20.6.1471

https://doi.org/10.1177/0098628315589505

Doi:10.5296/ije.v4i3.1785

https://doi.org/10.1177/0305735692201005
Chapter 3 Study One

Effects of Visually Presented Lyrics on Song Recall

Abstract

The purpose of this study was to investigate effects of visually presented lyrics on song recall from a cognitive load perspective. Additionally, I examined effects of perceived task difficulty, preferred learning styles, and any differences between music and non-music majors. Twenty-six members of auditioned choirs at a major mid-Atlantic university learned a difficult song individually through prerecorded aural instruction via computer: One group saw the lyrics and the other group did not. Multivariate analyses of covariance controlling for phonological working memory revealed that music majors recalled the song better than non-music majors when not seeing the lyrics while no significant difference existed when seeing the lyrics. Further analysis revealed seeing the lyrics was beneficial for non-music majors in recalling pitches and rhythm, but not beneficial for music majors. Presenting verbal information dually might be redundant for music majors; they might process musical and verbal information in a more integrated way and have more aural capacity. However, for non-music majors, the dual processing of verbal information might reduce their cognitive load in the aural channel, providing more capacity to process musical information. Showing the lyrics can be one way to reduce learners’ cognitive load while teaching a difficult song, at least for young adults with lower levels of musical expertise.

Keywords: visually presented lyrics, song recall, cognitive load, dual processing, modality effect
Effects of Visually Presented Lyrics on Song Recall

Introduction

A song consists of two conceptually separable, but tightly bound types of information - verbal information (i.e., lyrics) and musical information (e.g., pitches and rhythm) (Sammler et al., 2010; Serafine, Crowder, & Repp, 1984; 1990). Since learning to sing a song requires a certain level of competence in music and language, a number of studies have been conducted to investigate the role of lyrics in singing and song learning. Research investigating singing accuracy with or without text has yielded contradictory results (Demorest et al., 2015). Some researchers claimed participants including young children and adults sang more accurately on a neutral syllable (Berkowska & Dalla Bella, 2009; Goetze, 1985, Levinowitz, 1989, Welch, Sergeant, & White, 1996), while other found no difference between singing with and without lyrics (Sims, Moore, & Kuhn, 2002; Smale, 1988). Some of these contradictory findings could be due to the lack of shared definition of accurate singing and a large variety of methodologies, measurements, and analysis techniques used in singing research (Demorest et al., 2015; Goetze, Cooper, & Brown, 1990).

Researchers in music psychology and cognitive neuroscience have investigated the relationships between lyrics and tunes in memory and neural processing (e.g., Ginsborg & Sloboda, 2007; Peretz, Vuvan, Lagrois, & Armony, 2015; Sammler et al., 2010; Serafine et al., 1984; 1990). The continuing questions in these areas are whether the processing of tunes and lyrics in perception and memory is separated or integrated and, if integrated, to what degree both components share the processing and representation systems (Sammler et al. 2010). Sammler et al. (2010) found lyrics and tunes are integrated at prelexical, phonemic processing levels in the stage of auditory analysis as well as in the preparation of motor output while they
are processed independently at structural and semantic levels, concluding “lyrics and tunes of unfamiliar songs are processed at different degrees of integration … at different stages of the processing” (p. 3576). Also, they pointed out, “beyond auditory perceptual processing, the degree of integration/ separation depends on the specific cognitive processes targeted by an experimental task (e.g., recognition versus recall or production of familiar versus unfamiliar songs), perhaps accounting for some of the conflicting results” (p. 3576).

Since music is inherently an aural art, improving learners’ aural capacity to process musical information is an essential task for music educators. To develop learner’s aural capacity, when teaching a song, a rote teaching procedure - teaching songs by imitation without using visual material - has been recommended and utilized, particularly for younger students, in music teaching practice (Gordon, 2012; Klinger, Campbell, & Goolsby, 1998). Also, in many cultures, musical learning occurs through oral transmission (Anderson, & Campbell, 2011; Lehmann, Sloboda, & Woody, 2007).

However, in practice, teachers sometimes show students the lyrics without the staff notation while teaching or singing a song. I often observe classroom teachers displaying the lyrics on an easel while teaching a song to their students. In vocal music concerts, program notes sometimes provide lyrics for the audience. Karaoke screens always show the lyrics. Music show on TV presents lyrics while singers sing. As we often see the lyrics while learning or singing a song, it seems of interest to determine how seeing the lyrics may affect song learning and singing.

When leaning a song aurally, verbal information (i.e., text) and musical information (e.g., pitches and rhythm) are perceived through the aural channel. Processing all information only using the aural channel might hinder one’s learning of the song. If the song has less
information, processing it in the aural channel likely will not be a problem. However, if a song has too much information, only using the aural channel might cause cognitive overload to the learners. From the cognitive load perspective, imposing a heavy cognitive load hinders learning (Mayer, 2008, 2014; Sweller, 1988, 2011; Sweller, Ayres, & Kalyuga, 2011). In this study, a song that has too much information is considered a difficult song from a cognitive load perspective. It is noteworthy that the amount of cognitive load induced from a certain song will be different among persons, depending their musical ability (e.g., musical understanding and memory capacity) and experiences: A song considered difficult to persons with lower levels of music expertise might be considered an easy song to persons with higher level of musical expertise.

A considerable number of studies in instructional psychology have focused on how to reduce cognitive load in presenting instructional materials based on understanding how human minds work (Mayer, 2014; Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011). One of the theories regarding cognitive load is the cognitive theory of multimedia learning (Mayer, 2008). This theory is based on three assumptions: dual channel, limited capacity, and active processing (Mayer & Moreno, 1998, 2003; Mayer, 2014). Humans possess separate systems for processing visual and auditory information, called dual channel. Human beings have limited capacity to process novel information in each channel at any given time. And humans learn through active processing by attending to the incoming information, organizing attended information into a coherent mental representation, and integrating mental representations with previous knowledge (Mayer & Moreno, 2003).

Based on these assumptions, numerous experiments have demonstrated effects of particular instructional conditions (Mayer, 2014; Sweller, 2011; Sweller, Ayres, & Kalyuga,
The effects demonstrated from a cognitive load perspective are called cognitive load effects. Among a number of cognitive load effects, modality effect and verbal redundancy effect are discussed next, both of which are closely related to this current study.

**Modality effect.** In the cognitive load literature, the modality effect refers to the experimental demonstration that employing two sensory modalities rather than one is more effective when the instructional materials are unintelligible in isolation (Low & Sweller, 2014; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995; Tabbers, Martens, & van Merriënboer, 2004; Tindall-Ford, Chandler, & Sweller, 1997). In multimedia instructional settings, Mayer and Moreno (1998) found students performed better on transfer tests when instructional materials were presented as pictures and narration rather than as pictures and on-screen text. Presenting both pictures and words visually might cause cognitive overload in the visual channel while the aural channel is unused. On the contrary, if words are presented aurally as a form of narration instead of on-screen text, the visual channel might be offloaded with regard to the verbal information, having more capacity to process the pictures. It is assumed the reduction in cognitive load in the visual channel enhances learning.

**Verbal redundancy effect.** Verbal redundancy refers to “the simultaneous presentation of text and narration with identical words” (Moreno & Mayer, 2002, p. 156). In the working memory literature, presenting identical words in two modalities helps learning, yielding better memory recall than presenting the words in only one modality (Lewandowski & Kobus, 1993; Montali & Lewandowski, 1996; Penney, 1989). Attempting to maximize the working memory capacity by employing two channels when presenting information is in line with the basic idea of modality effect. This verbal redundancy effect is based upon a premise that only one mode of information is presented in each channel.
It is noteworthy that in this current (present) study, the verbal information still remains in the aural channel, which violates the condition that verbal redundancy is effective only when one mode of information is presented in each channel. However, it is critical that verbal and musical information should be integrated to learn and sing a song. In this sense, keeping the verbal and musical information integrated in the aural channel seems ultimately beneficial in song learning. The possible overload on the aural channel may be outweighed by the benefits of presenting already integrated materials.

When learning a song aurally, if learners see the lyrics, the verbal information can be processed in both aural and visual channels. The dual processing of verbal information might reduce the amount of load occupied by verbal information in the aural channel resulting in more capacity to process musical information. Therefore, I assumed that visual presentation of lyrics would facilitate processing of song information leading to better recall of the learned song.

Other Considerations

Learning style. Learning styles are “a combination of cognitive, affective, and physiological factors that merge to define each student's unique approach to effective learning” (Wilson, 2012, p. 69). Hall and Moseley (2005) viewed learning styles as personal preferences and orientations. Some learning style assessment instruments focus on identifying the individual’s preference for a learning modality such as visual, aural, and kinesthetic. Numerous studies have reported that students’ learning styles are different and their learning style influences their attitudes toward learning as well as their academic performance (see Minotti, 2005). In music education literature, studies investigating the effects of matching the instructional modes to children’s learning style yielded inconsistent results: Some studies
demonstrated the efficacy of tailoring instruction to students’ learning styles (Dunn, 2008; Persellin & Pierce, 1988; Korenman & Peynircioglu, 2007; Zikmund & Nierman, 1992) while others did not find any supporting evidence of customized instruction (Hughes, 1991; Pautz, 1988). This study investigated the possible influence of the preferred learning modality although it is acknowledged that, “There presently is no empirical justification for tailoring instruction to students’ supposedly different learning styles” (Rohre & Pashler, 2012, p. 636).

**Perceived task difficulty.** Self-reported task difficulty can be regarded as a direct and subjective measure of cognitive load (Brünken, Plass, & Leutner, 2003; Kalyuga et al., 1999). This also can be considered an indicator of the levels of participants’ musical competence in singing and learning a song. In this study, it was assumed if participants experienced less cognitive load by seeing the lyrics or participants are more capable of learning the song, they would report a lower level of task difficulty.

**Phonological working memory.** According to Baddeley (1992), working memory is a temporary storage and processor with limited capacity to process novel information at a given time. The phonological loop is one of subsystems of working memory, originally proposed as a processor of speech-based information. Baddeley (1992) extended the function of phonological loop including processing of musical information although some scholars questioned the model arguing the model does not provide adequate explanations for musical memory (e.g., Berz, 1995). In this study, I assumed if someone demonstrates high phonological working memory capacity, seeing lyrics or not seeing lyrics might not matter because they have enough capacity to process information without experiencing cognitive overload in the aural channel. The opposite may be true for someone with lower phonological working memory capacity. Hence, individual differences in participants’ phonological memory were controlled in this study.
Research Questions

To investigate the effect of visually presented lyrics (VPL) on song recall, the following questions guided the study: (a) Does the recall accuracy of the learned songs differ between groups who learn a difficult song with or without visually presented lyrics controlling for phonological working memory? (b) Do participants’ perceived task difficulty and preferred learning style play a role in their song recall accuracy? Specifically, do perceived task difficulty and preferred learning style interact with the instructional conditions?

Method

Design and Participants

This study utilized a two-group posttest-only randomized experimental design. The independent variable was the instructional condition in which two groups of participants learned a song aurally - one group saw the lyrics and the other group did not. I also investigated the effect of perceived task difficulty and preferred learning style. Dependent variables were the participants’ recall accuracy of lyrics, pitches, and rhythm of a learned song. Phonological working memory was utilized as a covariate.

Members from auditioned choirs at a Mid-Atlantic University were invited to participate in this study. I chose these particular groups to control some possible confounding factors such as singing and linguistic ability; I assumed that auditioned choir members could sing in tune and would have no hearing and text comprehension issues compared to other age groups, members of non-auditioned choirs, and non-choir members. Originally, I invited members of the auditioned choirs but excluded those in the most select choir that was mostly comprised of music majors. However, due to difficulty in securing a large enough sample size, I expanded the participants’ pool to include the most select choir as well as graduate students in
music education. Out of 26 total participants, 16 were female. From high school up to the time of data collection, the participants had participated in choirs for 1 to 20 semesters ($M = 9.73$, $SD = 5.127$) and had taken private voice lesson for 0 to 14 semesters ($M = 3.54$, $SD = 4.554$). Twelve participants were music majors. As a result, “major (music major or non-music major)” was added as an additional independent variable.

**Instructional Materials**

The song and song instruction used in this study are described below.

**Song.** This study hypothesized that visual presentation of lyrics may reduce cognitive load in the aural channel. However, if the amount of information is very small, resulting in low cognitive load in the aural channel, the instructional condition would not matter. Thus, I chose a song difficult enough to produce high cognitive load in the aural channel. Upon the recommendation of an expert in singing research, I used the song library of “Come Children Sing Institute” site (http://comechildrensing.com/), which includes more than 500 composed songs in various tonalities and meters. On this website, a song can be searched by the difficulty level as well as other categories such as meter and tonality. I selected a song called “April,” composed by Mary Ellen Pinzino with lyrics written by Sara Teasdale. As shown in Figure 3-1, “April” is in A Aeolian tonality and triple meter, consists of 16 measures (32 macrobeats), and is indicated as a difficult song. Each short phrase is similar to others in that each phrase always ends with the tonic, but there is no repetition of any phrase - the melody varies throughout the song. Such varying characteristics of the song seem to require learners to process much information. Thus, the song was considered a difficult song.
Song instruction. Teaching a song aurally can be approached in two ways: phrase-by-phrase and whole song. The phrase-by-phrase method is also called the echo-phrase, or repeated phrase approach and the whole song method is called the immersion or holistic approach (Gault, 2002; Klinger, Campbell, & Goolsby, 1998; Persellin & Bateman, 2009). In the phrase-by-phrase approach, the learners are invited to echo short phrases and then echo long phrases of the song. Music teachers have been trained to use this approach (Campbell & Scott-Kassner, 2006; Harrison, 1983; Herrold, 1991). In the whole song approach, the teacher sings the song in its entirety several times; the learners just sing along with the teacher when they are ready. This type of song teaching is “often practiced in more informal music-making
situations” (Persellin & Bateman, 2009, p. 800). Research comparing the effect of these two approaches has yielded inconsistent results (Gault, 2002; Klinger, et al., 1998; Persellin & Bateman, 2009). Persellin and Bateman (2009) argued teaching method might depend on “age of children, treatment duration and frequency, and songs” (p. 804). In this research, I used the phrase-by-phrase approach considering the short time of instruction and the age of the participants.

The song instruction for this study was audio-recorded by a native English speaker with the following procedure. First, the instructor sang the song (tune and lyrics) in its entirety. Second, the instructor sang the song by short phrases (two measures of the song at a time) with time gaps between short phrases in which participants could echo each short phrase. The instructor then sang long phrases of the entire song (four measures) with time gaps allowing participants to echo each long phrase. Third, the instructor sang the whole song again. Lastly, the instructor sang a “ready-sing,” a short singing cue: In this study the “ready-sing” included the beginning pitches of the song sung with appropriate tempo and meter. Then the participants sang the whole song by themselves. The script of song instruction is presented in Appendix A.

Data Collection Tools

Several data collection materials were used in this study as described below.

**Demographic questionnaire.** Using Qualtrics, an online survey system, participants were asked to indicate their sex, year in college, major, choral experiences, and private voice lesson experiences. This information was collected to describe the nature of participants.

**Perceived task difficulty report.** Participants were asked to rate their perceived task difficulty based on a continuous scale ranging from 1 (very easy) to 7 (very difficult) on the
computer. Using a sliding bar, participants indicated the difficulty levels of completing the song learning tasks.

**Phonological working memory test.** A sentence-span task, a subcomponent of a working memory test battery designed by Lewandowsky, Oberauer, Yang, & Ecker (2010) was used to measure phonological working memory. The test was run through MATLAB and the Psychophysics Toolbox. On each trial, the participants saw a sentence and to-be-remembered consonants. The participants had to judge the meaningfulness of each sentence and to memorize the following letter for later serial recall. In the program, the score was automatically calculated based on the proportion of the items correctly recalled in the correct list position. The task took approximately 10 minutes.

**Preferred learning style report.** An online version of the Barsch Learning Style Inventory (BLSI), available at Brett Bixler’s website (http://www.personal.psu.edu/bxb11/LSI/LSI.htm), was used to measure participants’ preferred learning style. This instrument focuses on identifying the individual preference of learning modalities: aural, visual, and kinesthetic. Krätzig and Arbuthnott (2006) reported the reliability coefficients of BLSI as .54 for visual, .56 for auditory, and .38 for kinesthetic items.

**Song recall test.** The lyrics were not presented for either group during the song recall test. The participants were aware of this recall condition before they learned the song. In the test, the entire song was played again before asking participants to sing the song from memory. The participant’s singing following a singing cue was audio-recorded through a microphone and recording program; the audio file was saved on the computer.
Procedures

Data were collected for each participant individually in a quiet room. After gaining consent from the participants, I provided an overview of the research procedure showing the outline on the computer screen. After completing the demographic questionnaire on the computer, the participant learned a song (April) through the prerecorded aural instruction on the computer using a headphone. While learning the song through aural instruction, participants in one group saw the whole lyrics of the song on the computer screen for the entire instructional time and participants in the other group did not see the lyrics but saw a blank screen. Participants then were asked to report the perceived task difficulty on the computer. They then took the phonological working memory test and completed the learning style inventory on the computer. Lastly, participants took the song recall test. The duration of the data collection process was about 30 minutes for each participant.

Ratings and Interrater Reliability

Two experienced music teachers rated the recall accuracy of the two songs based on the written instructions provided by the researcher. Recall accuracy was rated by a half bar unit (Ginsborg & Sloboda, 2007) for the components of lyrics, pitches, and rhythm. For each component, 32 points were possible.

Interrater reliability was calculated through three different measures: percentage agreement, Kappa agreement, and Pearson correlation coefficient. Percentage agreement was calculated based on the number of occurrences of the agreement on accuracy rating in each half bar unit for lyrics, pitches, and rhythm of each song for each participant. Then, the averages of percentage agreement for the all participants for each component of each song were used to represent interrater reliability. However, since the percentage of agreement does not account for
chance agreement, I also calculated Kappa agreement, which is commonly used as one of the chance-corrected agreement measures (McHugh, 2012). For each participant, contingency tables were created for each component of each song. Table 3-1 is an example of contingency tables I created for a particular participant.

Table 3-1.

*Example of Contingency Tables Used to Calculate the Kappa Statistics*

<table>
<thead>
<tr>
<th></th>
<th>Lyrics</th>
<th></th>
<th></th>
<th>Pitches</th>
<th></th>
<th></th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rater 1</td>
<td>Rater 2</td>
<td></td>
<td>Total</td>
<td>Rater 1</td>
<td>Rater 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Correct</td>
<td></td>
<td></td>
<td>Correct</td>
<td>Correct</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incorrect</td>
<td>Incorrect</td>
<td></td>
<td></td>
<td>Incorrect</td>
<td>Incorrect</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Total</td>
<td></td>
<td></td>
<td>Total</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Rater 2 Correct</td>
<td>28</td>
<td>12</td>
<td>31</td>
<td>Incorrect</td>
<td>0</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Incorrect</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>Total</td>
<td>31</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

In this example, Kappa values were 1 for lyrics, .868 for pitches, 0 for rhythm, although the percentage agreement were 100, 93.59 and 90.32, respectively for this sample. This result indicates the limitation of Kappa agreement in that the Kappa statistics are highly dependent of trait prevalence and marginal probabilities (see Feng, 2013), which makes the Kappa statistics often challenging to interpret and unstable (Gwet, 2002). For this reason, I calculated both percentage and Kappa agreement as suggested by McHugh (2012). Lastly, Pearson correlation coefficient was calculated to estimate the interrater reliability since the accuracy scores were ratio data, which has zero and the score interval is interpretable. Correlation coefficient represents the ordinal nature of ratings. However, correlation coefficient has a limitation in that it ignores systematic differences in ratings. For example, one rater might give continuously higher scores than the other rater, but the correlation coefficient can be high if their rank orders are the same. According to Liao, Hunt, and Chen (2010), the correlation coefficient is useful to ascertain the consistency of scores across different participants from different rater, whereas interrater agreement is useful to determine the stability of scores a participant receives from
different raters. For these reasons, interrater reliability was calculated through those three measures and presented in Table 3-2.

Table 3-1.

**Interrater Reliability based on Percentage Agreement, Kappa Agreement, and Pearson Correlation**

<table>
<thead>
<tr>
<th></th>
<th>Lyrics</th>
<th>Pitches</th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Agreement</td>
<td>91.55</td>
<td>92.94</td>
<td>89.81</td>
</tr>
<tr>
<td>Kappa Agreement</td>
<td>.73</td>
<td>.81</td>
<td>.75</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.98</td>
<td>.97</td>
<td>.95</td>
</tr>
</tbody>
</table>

Percentage agreement seemed high considering the total possible scores were only 32 for each component. The Kappa statistics indicated very good agreement (ranged from .80 to 1) for pitches and good agreement (ranged from .60 to .80) for lyrics and rhythm. The Pearson correlation coefficients also yielded high coefficients. Hence, the averages of the two raters’ scores for each component were used for data analysis.

**Results**

The differences by instructional condition (group) for recall accuracy of lyrics, pitches, and rhythm were analyzed using Multivariate Analysis of Covariance (MANCOVA). Phonological working memory was used as a covariate. The test assumptions such as normality, homogeneity of variance/covariance matrices, absence of univariate outliers/multivariate outliers, linearity, and absence of multicollinearity (Mayers, 2013) were assessed and satisfied. The MANCOVA revealed no group difference in recall accuracy for any of the items, $F(3, 21) = 1.274, p = .309, \eta^2_p = .154$. This analysis indicated participants performed similarly on the recall tasks regardless of whether they saw the lyrics during instruction or not.

The interaction between groups and preferred learning styles as well as between groups and perceived task difficulty were analyzed by MANCOVA. No interactions existed ($p = .314$)
71

& $p = .777$, respectively). Therefore, perceived task difficulty and learning styles do not appear to be playing a role in recall accuracy related to the instructional condition.

As stated earlier, due to the expanded sample that included more music majors, I additionally ran a two by two factorial (instructional conditions and majors) MANCOVA. When phonological working memory was controlled for, a main effect for major [$F (3, 19) = 2.276, p = .045, \eta^2_p = .338$] and interaction between group and major [$F (3, 19) = 4.558, p = .014, \eta^2_p = .418$] were revealed, although a main effect for group was still not found [$F (3, 19) = 2.276, p = .113, \eta^2_p = .264$]. Univariate analyses showed music majors recalled lyrics [$M_D = 6.633, F (1, 21) = 4.772, p = .040, \eta^2_p = .185$] and pitches [$M_D = 4.846, F (1, 21) = 5.491, p = .029, \eta^2_p = .207$] better than non-music majors. As shown in Figure 3-2, there were statistically significant interactions between group and major - regarding pitches [$F (1, 21) = 12.573, p = .002, \eta^2_p = .375$] and rhythm [$F (1, 21) = 9.983, p = .005, \eta^2_p = .322$]. Table 3-3 shows the adjusted and unadjusted means and standard deviations of recall accuracy in terms of lyrics, pitches, and rhythm by group and major.

![Figure 3-2](image)

**Figure 3-2.** Interactions between Group and Major. VPL indicates Visually Presented Lyrics
Table 3-3

Recall Accuracy of Lyrics, Pitches, and Rhythm by Group and Major. VPL indicates Visually Presented Lyrics

<table>
<thead>
<tr>
<th></th>
<th>Lyrics</th>
<th>Pitches</th>
<th>Rhythm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Adjusted M (SD)</td>
<td>Unadjusted M (SD)</td>
</tr>
<tr>
<td>Without VPL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-music majors</td>
<td>9</td>
<td>10.99 (2.58)</td>
<td>12.44 (8.97)</td>
</tr>
<tr>
<td>Music major</td>
<td>4</td>
<td>20.95 (3.68)</td>
<td>21.63 (5.65)</td>
</tr>
<tr>
<td>With VPL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-music majors</td>
<td>5</td>
<td>10.26 (3.46)</td>
<td>8.30 (4.66)</td>
</tr>
<tr>
<td>Music major</td>
<td>8</td>
<td>13.57 (2.63)</td>
<td>12.81 (8.14)</td>
</tr>
</tbody>
</table>

The nature of the interaction between group and major controlling for phonological working memory were explored by simple effects tests. In the group without visually presented lyrics (VPL), music majors better recalled the learned song than non-music majors in terms of pitches ($M_D = 12.30, SD = 3.01, p = .001$) and rhythm ($M_D = 7.05, SD = 2.44, p = .009$).

However, in the group with VPL, there was no difference in the recall accuracy between music majors and non-majors. On the other hand, non-music majors better performed pitches ($M_D = 8.07, SD = 3.09, p = .016$) and rhythm ($M_D = 7.21, SD = 2.5, p = .009$) when they saw the lyrics compared to when they did not see the lyrics. Conversely, music majors better recalled the pitches when they did not see the lyrics ($M_D = 6.84, SD = 3.11, p = .04$).

Discussion

In this study, the effects of visually presented lyrics while learning a difficult song aurally on recall accuracy in terms of lyrics, pitches, and rhythm was investigated. Also, I explored whether perceived task difficulty, preferred learning style as well as major (music/non-music) interacted with the instructional condition. In all analyses, participants’ phonological working memory was controlled. No main effect for instructional condition was found, but an interaction between instructional condition and major was revealed. Non-music majors benefited from seeing the lyrics when recalling pitches and rhythm, while music majors
benefited from not seeing the lyrics when recalling pitches. Music majors outperformed non-music majors in terms of pitches and rhythm only when the participants did not see the lyrics while learning the song. When they saw the lyrics, there was no statistical difference between music majors and non-music majors.

Several studies investigating the effect of musical training and expertise found musical training enhances auditory temporal processing (Besson, Schön, Moreno, Santos, & Magne, 2007; Jakobson, Cuddy, & Kilgour, 2003) and verbal memory (Chan, Ho, & Cheung, 1998; Jellison & Miller, 1982; Kilgour, Jakobson, & Cuddy, 2000). Also, their superior understanding of musical structure and characteristics facilitates encoding and retrieving of musical information (Racette & Peretz, 2007) in that verbal information and musical information are stored and retrieved in association each other (Crowder et al., 1990; Ginsborg & Sloboda, 2007, Serafine et al., 1984, 1986). In this study, music majors’ enhanced encoding strategy and auditory temporal processing may allow them to have more capacity in the aural channel to process the verbal and musical information, which was supported by the results that music majors better recalled the lyrics and pitches than non-music majors over instructional conditions.

However, when instructional condition was taken into consideration, music majors outperformed non-music majors regarding pitches and rhythm only when not seeing the lyrics. According to Ginsborg and Sloboda (2007), “[Musicians] had been taught to memorize the lyrics and melodies of songs as composites, in such a way that memory for one component enhances memory for the other” (p. 435). They found singers with higher levels of musical expertise recalled the learned song more fluently and accurately than singers with lower levels of musical expertise, but only when they memorized the lyrics and tune together. Thus, music
majors may process verbal and musical information in a more integrated way than non-music majors.

These findings also seem congruent with the expertise reversal effect (Kalyuga & Sweller, 2014; Kalyuga, Chandler, & Sweller, 1998; Sweller, Ayres, Kalyuga, & Chandler, 2003). Expertise reversal effect means that the modality effect may not be effective or may even impede learning for high-knowledge learners, since information beneficial for less knowledgeable learners may be redundant for more knowledgeable learners. The dual presentation of verbal information might be redundant for music majors, whereas it seems beneficial for non-music majors.

On the other hand, if non-music majors have less capacity in the aural channel compared to music-majors, they could benefit by processing verbal information dually in reducing their cognitive load in the aural channel. As hypothesized in this study, the reduction of cognitive load in the aural channel might provide more capacity in the aural channel to process musical information. Thus, the levels of musical expertise do matter for which instructional condition was more effective. It is noteworthy that although non-music majors were considered to have low levels of musical expertise compared to music majors, they were still members of an auditioned choir. Those with less musical expertise may find seeing the lyrics even more helpful. However, care should be taken to consider this explanation since this hypothesis was not directly assessed, although an attempt was made to examine participants’ cognitive load through self-reported task difficulty.

Participants’ perceived task difficulty had been considered an indicator of the cognitive load they experienced during song learning, but perceived task difficulty did not play a role in recall accuracy related to the instructional conditions. This self-report might not function well
as an indicator of cognitive load. Since it is a subjective measure, individual differences in rating styles could affect the results, and the levels of perception might not be related to the levels of performance. Originally, Kalyuga et al. (1999) asked participants to rate the difficulty of the materials and reported the ratings were highly sensible in finding a difference in training strategies. However, as Brünken et al. (2003) pointed out, task difficulty, the levels of competence of the learners, and different attentional processes could affect the difference. Brünken et al. (2003) suggested other approaches such as the use of neuroimaging techniques and the dual-task approach for measuring participants’ cognitive load in a direct and objective manner.

Like perceived task difficulty, preferred learning styles did not affect song recall. Research concerning the effects of preferred learning style in relation to instructional modality on musical tasks has yielded contradictory results (Mishra, 2007). It is noteworthy that previous studies employed either iconic representations of music (Apfelstadt, 1984; Dunn, 2008; Gomko & Poorman, 1998; Hughes, 1991; Pautz, 1988; Persellin, 1988, 1992, 1993; 1994; Zikmund & Nierman, 1992; Tarnowski, 1986) or notation (Korenman & Peynircioglu, 2007) as visual materials whereas this study used printed lyrics. Perhaps, the instructional condition with visually presented lyrics is not enough to represent the way that visual learners prefer to learn: Expecting a link between preferred learning modality and instructional condition in this study might not be reasonable.

Despite the significant findings in this study, the small and imbalanced sample size was not negligible. Although the test assumptions for MANCOVA such as normality, homogeneity of variance/ covariance matrices, absence of univariate outliers/ multivariate outliers, linearity, and absence of multicollinearity (Mayers, 2013) were satisfied, the numbers of music majors
and non-music majors across groups was small (ranged from four to nine) and not balanced since the music majors variable was added later. A larger and more balanced sample would provide more valid results allowing for greater generalizability.

In this study, the effect of visually presented lyrics was tested with auditioned choir members at a University. I chose this particular group in order to control for any confounding factors such as individual differences in reading comprehension and singing ability. I assumed auditioned choir members could sing in tune and would have no hearing and text comprehension issues compared to other age groups, members of non-auditioned choirs, and non-choir members. However, in typical practice, choir members learn songs from musical scores. Testing the benefit/effect of showing the lyrics without musical notation with them might not be valid and applicable to practice. This test may be more applicable to those who are not comfortable reading music or learning a song from notated music.

In this study, the song instruction was audio-recorded by a female soprano singer. Since my participants were auditioned choir members, I assumed the instructor’s sex and vocal range would not affect their song learning and singing. However, to rule out the possibility that instructor’s sex and vocal range might affect participants’ learning and recalling of a song (Belin & Zatorre, 2003; Sammler et al., 2010), providing song instruction by a singer of the same sex as the participants in their comfortable range should be considered for future study, especially when studying untrained singers song recall.

**Conclusion**

This study investigated the effect of visually presented lyrics on song recall from a cognitive load perspective. The efficacy of instructional condition depended on the participants’ levels of musical expertise. Overall, non-music majors benefited from seeing the
lyrics while music majors did not. Music majors might process musical and verbal information in a more integrated way and have more processing capacity in the aural channel. Presenting verbal information dually might be redundant for music majors, but the results of this study imply non-music majors could benefit from seeing the lyrics. For non-music majors, the dual processing of verbal information might reduce their cognitive load in the aural channel, providing more capacity to process musical information.

Instructional design should be based on many considerations such as the instructional time, goals, and characteristics of students (e.g., their levels of expertise in the domain). For students with low levels of musical expertise, learning a difficult song only using the aural channel might cause them cognitive overload, which hinders their learning. Given limited instructional time and the need to be more efficient, several strategies should be considered to prevent learners from experiencing cognitive overload while learning a difficult song aurally. Showing the lyrics of the song could be one strategy for that purpose, at least for young adults with lower levels of musical expertise than music majors.
References for Chapter 3


http://dx.doi.org/10.1037/1076-898X.3.4.257


doi:10.5296/ije.v4i3.1785
Appendix. The song learning and recall instructions used in the experiment.

**Song Learning Instruction**

1. **Greetings**
   
   “Hello! I will teach you a song, named April.”

2. **Entire song**
   
   “Let me sing first for you.” (Play the pitches of the first two measures. Sing the entire song)

3. **Short phrases**
   
   “I will teach you the song by short phrases. After I’m done singing a short phrase, I will give you a sign, by singing ‘go’. And then please echo me.” (Sing the song by a short phrase and then “go”)

4. **Long phrases**
   
   “Now I will sing long phrases. After I sing the long phrase, I will give you a sign to echo me.” (Sing the song by a long phrase and then “go”)

5. **Entire song**
   
   “Thanks. Now I will sing the entire song for you. You can sing in your mind but please don’t sing out loud.” (Play the beginning pitches. Sing the entire song)

6. **Participant’s trial**
   
   “Now it’s time for you to try singing the song on your own. After hearing the ready-sing sign, please sing the whole song without stopping. Even if you’re not sure, please keep going.” (Sing the ready-sing sign, “and ready let’s all sing it now”, to the tune of the beginning pitches of the first two measures.)

**Song Recall Instruction**

1. **Entire song**
   
   “I will sing the entire song for you again. You can sing in your mind but please don’t sing out loud.”

2. **Recall test**
   
   “This time we will record your singing. After the “ready-sing” sign, please sing the entire song without stopping. Just like before even if you’re not sure, please do your best and keep going.” (Sing the ready-sing sign)
Chapter 4 Study Two

**Mediating Effect of Cognitive Load in Song Recall with Visually Presented Lyrics**

**Abstract**

The main purpose of this study was to ascertain whether seeing the lyrics while learning a difficult song aurally induces less cognitive load in learners compared to not seeing the lyrics, leading to better recall accuracy of the learned song. Cognitive load was measured through reaction time measures based on a dual-task paradigm. Recall accuracy of the learned song was assessed in terms of lyrics, pitches, and rhythm. Thirty-six non-music majors individually learned two songs through prerecorded aural instruction. During the aural instructional sequence, for one song they saw the lyrics and for the other song they did not see the lyrics. The presentation order of instructional condition (with and without visually presented lyrics) and song were counterbalanced. Results showed instructional condition affected cognitive load but not recall accuracy. A path analysis revealed a mediating effect of cognitive load regarding lyrics and rhythm, suggesting seeing the lyrics indirectly increases recall accuracy of lyrics and rhythm through its positive effect on cognitive load. Implications for teaching practice and future research from this study’s approach and findings were discussed.

*Keywords:* visually presented lyrics, song recall, cognitive load, reaction times, dual-task
Mediating Effect of Cognitive Load in Song Recall with Visually Presented Lyrics

Singing a song is an artistic expression of words through music. A song consists of two conceptually separable but tightly bound types of information - verbal and musical (Sammler et al., 2010). Since music, and hence singing, is fundamentally an aural art, improving learners’ aural capacity to process musical information is an essential task for music educators. Toward that purpose, when teaching a song, a rote teaching procedure - teaching songs by imitation without showing musical notation - has been recommended particularly for younger students (Gordon, 2012; Klinger, Campbell, & Goolsby, 1998). Also, this is how songs are learned in many cultures in that musical learning occurs through oral transmission (Anderson, & Campbell, 2011; Lehmann, Sloboda, & Woody, 2007).

However, in practice, there are times when people learn and sing a song with visually presented lyrics. For example, teachers sometimes show the lyrics to students without the staff notation. In vocal music concerts, program notes sometimes provide lyrics for the audience. Karaoke screens always show the lyrics. People sometimes choose the videos on YouTube that provide the lyrics of the song to learn the song and sing along with it. If we often see the lyrics while learning or singing a song, it seems of interest to determine how seeing the lyrics may affect song learning and singing.

When learning a song aurally, the verbal (i.e., lyrics) and musical (e.g., pitches, rhythm) information is processed in the aural channel. If the song has less musical and textual information, only using the aural channel would likely not impede the students’ capacity to learn the song. However, if the song is more difficult with much information, only using the aural channel may cause cognitive overload, which occurs when the incoming information exceeds the learner’s cognitive capacity (Sweller, 2011; Sweller, Ayres, & Kalyuga, 2011).
Imposing cognitive overload on learners has been shown to hinder learning (Mayer, 2008, 2014; Sweller, 1988, 2011; Sweller, Ayres, & Kalyuga, 2011).

Han (2016) investigated the effect of visually presented lyrics on song recall. It was assumed if the learners saw the lyrics while learning a difficult song aurally, they would better recall the learned song compared to learners who did not see the lyrics. This hypothesis was based on the logic that dual processing of verbal information might reduce the cognitive load in the aural channel, leading to more capacity to process musical information. However, the efficacy of these instructional conditions depended on the participant’s level of musical expertise. Overall, non-music majors benefited from seeing the lyrics whereas music majors did not. Perhaps music majors have more capacity in the aural channel and are able to process both verbal and musical information in a more integrated way than non-music majors (Ginsborg & Sloboda, 2007). The findings were also congruent with the expertise reversal effect (Kalyuga & Sweller, 2014; Sweller, Ayres, Kalyuga, & Chandler, 2003): Presenting verbal information dually – aurally and visually – might be redundant for high knowledgeable learners, but beneficial for low knowledgeable learners. Music majors are considered high knowledgeable learners in the area of music.

In Han’s (2016) study, the efficacy of seeing the lyrics was demonstrated by the learning outcome, song recall accuracy; the underlying mechanism of how seeing the lyrics leads to better recall accuracy was explained from a cognitive load perspective. However, this explanation was based on logical thinking, but not based on empirical evidence: The actual cognitive load induced by each instructional condition was not directly assessed. Therefore, the purpose of this study was to ascertain whether showing the lyrics to non-music majors reduces
their cognitive load in the aural channel and if the reduction of cognitive load leads to better recall accuracy of the learned song.

Many studies have investigated the effect of certain instructional designs based on learning outcomes, interpreting the outcome scores as an indicator of cognitive load experienced by the learners (e.g., Kalyuga, Chandler, & Sweller, 1999; Mayer & Moreno, 1998, 1999; Mousavi, Low, & Sweller, 1995; Tabbers, Marten, & van Merriënboer, 2004). However, learners’ cognitive load under certain instructional conditions can be directly measured through a dual-task approach, as demonstrated by Brünken and his colleagues (2002, 2004). The dual-task approach is considered a direct and objective measurement of cognitive load (Brünken, Plass, & Leutner, 2003). The dual-task method, also known as a secondary task technique, has been used in the field of experimental psychology (Britton, Glynn, Meyer, & Penland, 1982) and applied in studies in which the learning occurred through multimedia (Brünken, Steinbacher, Plass, & Leutner, 2002; Brünken, Plass, & Leutner, 2004). Given the limited capacity of working memory, the dual-task approach assumes that performance of a secondary task depends on the amount of cognitive resources induced by a primary task if both tasks are processed at the same time and require the same cognitive resources (Brünken et al., 2002). Hence, it is assumed if a primary task induces less load, then participants will perform better in a secondary task and vice versa (Britton et al., 1982; Brünken et al., 2003).

Brünken et al. (2002) employed a visual monitoring task as a secondary task to measure cognitive load in the visual channel during multimedia learning: Participants were asked to press the space bar on the computer keyboard as soon as they detected the color of the letter ‘A’ changed from black to red on the computer screen during the learning process. They found that the participants performed better on the secondary task when the participants learned the
textual and pictorial information in an audiovisual format (as pictures and narrations) than in a visual-only format (as pictures and text). This finding indicated when the learning materials were presented in an audiovisual format, participants experienced less cognitive load in the visual channel compared to when only the visual format was used. Brünken et al. (2002) claimed, “the utility of dual-task methodology as a promising approach for the assessment of cognitive load induced by complex multimedia learning systems” (p. 109).

In follow-up study, Brünken and his colleagues (2004) investigated auditory load using a detection task of a simple auditory tone in relation to three instructional design conditions: learning from 1) on-screen text and pictures (no auditory information), 2) on-screen text and pictures with background music (background music only), and 3) pictures and narration with background music (background music + narration). Similar to the visual monitoring task in their previous study (Brünken et al., 2002), participants were asked to tap the space bar on the computer keyboard as soon as they heard the tone. Interestingly, “the auditory cognitive requirements of background music alone did not differ from the auditory cognitive requirements of the materials without any auditory stimuli.” (Brünken et al., 2004, p. 129). However, when background music was played with narration, participants’ reaction time was significantly slower indicating more cognitive load induced in the aural channel.

As used by Brünken et al. (2004), in this current study, reaction time in a simple auditory monitoring task under the dual-task paradigm was considered an indicator of the aural cognitive load induced by the song-learning task. A faster reaction indicates more free cognitive capacity in the learner. For non-music majors, I assumed if the learners see the lyrics while learning a song, the dual processing of the verbal information would induce less
cognitive load in the aural channel, leading to faster reaction time compared to when they do not see the lyrics.

It is noteworthy that although Brünken and his colleagues (2002, 2004) measured the learners’ cognitive load and learning outcomes, they did not design their studies to examine the relationship between cognitive load and learning outcomes based on different instructional conditions. However, since this current study aimed to ascertain whether seeing the lyrics reduces cognitive load in the aural channel, ultimately leading to better recall accuracy of the learned song, the mediating effect of cognitive load on causal relationships between instructional condition and song recall were examined.

Hence, the following research questions guided the study.

1. Do non-music majors recall songs more accurately if they learn difficult songs with or without visually presented lyrics?
2. Do non-music majors experience less aural cognitive load if they learn difficult songs with or without visually presented lyrics?
3. Does cognitive load mediate the relationship between instructional condition (learning difficult songs with or without visually presented lyrics) and song recall accuracy in non-music majors?

**Method**

**Design**

This study employed a within-participants design with repeated measures to reduce the possibility of confounding factors such as the individual differences in singing ability, working memory, and reaction times. The independent variable was the instructional condition consisting of two levels: learning a song (1) with visually presented lyrics (With VPL), and (2)
without visually presented lyrics (Without VPL). While learning two difficult songs aurally, participants saw the lyrics of one song and did not see the lyrics of the other song. The dependent variables were song recall accuracy and cognitive load, operationalized as a reaction times (RTs) measure. However, in the final analysis to answer the third research question, cognitive load was used as a mediator. According to MacKinnon (2011), “a mediating variable is relevant whenever a researcher wants to understand the process by which two variables are related, such that one variable causes a mediating variable which then causes a dependent variable” (p. 2). To lower the order effect, the presentation order of instructional condition and song were counterbalanced.

Participants

Undergraduate students who were not majoring in music but enrolled in a music course at a South-Atlantic University were invited to participate in this study. Thirty-six undergraduate students (8 male and 28 female, mean age 19.28 years old; SD = 1.085) participated in this study. All participants were native English speakers, and none was familiar with the stimulus songs. Twenty-two participants had previous experience singing in a choir for 4.26 years on average (SD = 2.866). Twelve participants played in a band or orchestra for 5.17 years on average (SD = 2.725). Nine participants reported they took a voice lesson from one time up to two years. Although 20 participants indicated that they were engaging in some musical activities such as playing the ukulele, singing, and writing music, none were taking voice lessons or participating in a choral or instrumental ensemble at the time of data collection.

Instructional Materials

The songs and song instruction used in this study are described below.
**Songs.** Two songs were selected to compare the difference between two instructional conditions. One song, “April”, composed by Mary Ellen Pinzino with lyrics written by Sara Teasdale, was used in a previous study (Han, 2016). The other song was “February Twilight”, written by the same composer and lyricist. “February Twilight” is similar to “April” in that it is in triple meter, consists of 16 measures, and is indicated as a difficult song in the song library of “Come Children Sing Institute” site (http://comechildrensing.com/). Like “April”, there is no repetition of any phrase – the melody varies throughout the song. “February Twilight” was originally in G Mixolydian with a vocal range from F4 to G5 whereas “April” was in A Aeolian with a vocal range from E4 to E5. These songs seemed too high for the participants, so I transposed “April” to F Aeolian and “February Twilight” to D Mixolydian, matching the tessitura of two songs. I also reduced one macrobeat in the 14 measure of “February Twilight” so the song had 32 macrobeats like “April.” The revised scores of “April” and “February Twilight” are presented in Figure 4-1. The songs were used by permission from the copywriter, Mary Ellen Pinzino.
**Song instruction.** All song instructions were audio-recorded by trained singers who are native English speakers. Male participants heard the song instruction recorded by a male singer while female participants heard the female version.

The song instruction procedures used the phrase-by-phrase approach, basically following those from a previous study (Han, 2016). In the instruction for each song, the instructor sang the song (tune and lyrics) in its entirety. Second, the instructor sang the song by short phrases (two measures of the song at a time) and the participant echoed. This procedure was repeated two times. During the first singing by short phrases, a distinct click sound was added several times over the aural instruction to measure participant’s reaction times. In the second singing by short phrases, the aural instruction was presented without the click sounds. The instructor then sang long phrases of the entire song (four measures) and the participant echoed. After the instructor sang the whole song again, s/he sang a “ready-sing”, a short singing cue, so that participants could sing the whole song by themselves. In this study the “ready-sing” included the beginning pitches of the song sung with appropriate tempo and meter.

**Data Collection Tools**

Several data collection tools were employed and are described below.

**Demographic questionnaire.** Using an online survey tool, Google Forms, participants’ demographic information such as age and sex was collected. In addition, participants’ musical experiences including private voice lessons, participation in a choir, band or orchestra and current musical activity were asked.

**Song recall test.** All instructions for the song recall test were audio-recorded by the same person who recorded the song instruction. In the recall test stage, the lyrics were not
shown. In the test, the entire song was sung again before asking the participant to sing the song from memory. The participant was prompted to sing the song with a singing cue (“ready-sing”) and the performance was audio-recorded through a Shure SM 58 microphone and Cubase 9, an audio sequencing program.

**Reaction time measure.** A simple, continuous monitoring task of a single tone (Brünken et al., 2004) was used as a secondary task. Reaction time was considered an indicator of the amount of cognitive resource available. A faster reaction time indicates less cognitive load is induced. A clave timbre (pitch: B6, duration: 125 ms) in Cubase 9 was used for the distinct click sound. The basic task was to tap a key on the computer keyboard (marked by a green sticker on the key) as soon as the distinct click sound was heard. For each song, eight click sounds were randomly placed over the aural instruction during the first learning phase of short phrase echoes: Each short phrase of the song contained one click sound, but the click sounds were not overlapped with any starting point of the words of singing and matched the beat. Participants’ reaction times were recorded in milliseconds through Cubase 9, no auditory/visual feedback on tapping was provided.

**Procedures**

Data were collected for each participant individually in a quiet room. After giving an overview of the research procedure, I informed the participants that they would not see the lyrics for the song recall test even if they saw the lyrics during instruction for one song. Participants began with practicing the click sound detection task (Reaction time measure). Participants learned the first song through the pre-recorded aural instruction on computer using a headphone. As stated earlier, the order of songs and instructional conditions were counterbalanced, thus, half of the participants learned the song “April” first, the other half
learned “February Twilight” first. Half of the participants saw the lyrics while learning the first song; the other half saw the lyrics while learning the second song. While taking a short break, they completed the demographic questionnaire on computer. And then they took the recall test for the learned song. Participants learned the second song in the same manner. But if they saw the lyrics for the first song, they did not see the lyrics for the second song and vice versa. After taking a brief (about 2 minutes) break, they took the recall test for the second song. During the break, I asked simple questions about participants’ musical experiences to prevent them from rehearsing the song. I was present during the entire data collection, which took about 35 minutes for each participant.

Ratings

Two experienced music teachers rated the recall accuracy of the two songs based on the written instructions provided by the researcher. Recall accuracy was rated by a half bar unit (Ginsborg & Sloboda, 2007) for the components of lyrics, pitches, and rhythm. For “April,” 32 points were possible for each component. For “February Twilight,” only 31 points were possible for each component because the second beat of the second measure of “February Twilight” does not contain new information (see Figure 4-1), thus this unit was excluded in the rating. Since the total possible scores were different for each song, I transformed the raw scores of recall accuracy to percentage scores. The percentage scores were used for all analyses.

Interrater reliability was calculated through three different measures: percentage agreement, Kappa agreement, and Pearson correlation coefficient (Han, 2016; McHugh, 2012; Liao, Hunt, & Chen, 2010) and was presented in Table 4-1.
Table 4-1.

*Interrater Reliability based on Percentage Agreement, Kappa Agreement, and Pearson Correlation*

<table>
<thead>
<tr>
<th></th>
<th>April</th>
<th></th>
<th></th>
<th>February Twilight</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lyrics</td>
<td>Pitches</td>
<td>Rhythm</td>
<td>Lyrics</td>
<td>Pitches</td>
<td>Rhythm</td>
</tr>
<tr>
<td>Percentage Agreement</td>
<td>94.46</td>
<td>86.07</td>
<td>88.84</td>
<td>96.75</td>
<td>87.46</td>
<td>94</td>
</tr>
<tr>
<td>Kappa Agreement</td>
<td>.818</td>
<td>.613</td>
<td>.705</td>
<td>.877</td>
<td>.624</td>
<td>.72</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.986</td>
<td>.926</td>
<td>.969</td>
<td>.995</td>
<td>.930</td>
<td>.982</td>
</tr>
</tbody>
</table>

Percentage agreement seemed high considering the total possible scores were only 32 for “April” and 31 for “February Twilight.” The Kappa statistics indicated very good agreement (ranged from .80 to 1) for lyrics and good agreement (ranged from .60 to .80) for pitches and rhythm. The Pearson correlation coefficients also yielded high coefficients. Hence, the averages of the two raters’ scores for each component (i.e., lyrics, pitches, and rhythm) were used for data analysis.

**Results**

Before running the statistical analysis, test assumptions for Repeated Measures Multivariate Analysis of Variance (RM-MANOVA) were examined. Box-plots were generated for each instructional condition in terms of lyrics, pitches, and rhythm to detect outliers. Although there was an outlier for pitches in both instructional conditions, since these were minor outliers all statistical analyses were undergone without removing them. Also, the dependent variables were considered normally distributed since the values for skewness and kurtosis for dependent variables were far less than the acceptable range, -2 and +2 (George & Mallery, 2010). Since there were only two levels for each variable, sphericity was not a concern.
Research Question 1. Do non-music majors recall songs more accurately if they learn difficult songs with or without visually presented lyrics?

Descriptive statistics of recall accuracy and the results of a one-way RM-MANOVA in terms of instructional condition, song, and presentation order are presented in Table 4-2. A one-way RM-MANOVA revealed no main effect for instructional condition ($p = .24$). This analysis indicates participants performed similarly on the recall tasks regardless of whether they saw the lyrics while learning the song. However, in additional analyses, a main effect for song ($p = .01$) and a main effect for presentation order ($p = .04$) were found, respectively. Univariate Analysis of Variance revealed the effect of song for lyrics ($p = .03$) and rhythm ($p = .003$) not for pitches ($p = .35$). Participants recalled better “February Twilight” than “April in terms of lyrics and rhythm. Similarly, a presentation order effect was found for lyrics ($p = .03$) and rhythm ($p = .01$), but not for pitches ($p = .54$). Participants recalled the lyrics and rhythm better for the second song than for the first song regardless of the instructional condition and song.

Although the instructional condition, song, and presentation order were within-participants variables, since participants did not experience all the possible combination of those variables (the variables were counterbalanced), any possible interactions were not examined through RM-MANOVA. Instead, after parceling out the data by song and presentation order, I compared two instruction conditions in each data set (i.e., April & 1st order, April & 2nd order, February Twilight & 1st order, February Twilight & 2nd order) using MANOVA. The means and SDs of recall accuracy by song, presentation order, and instructional condition are presented in Table 4-3. However, no statistical significant effect for instructional condition was found in all analyses perhaps due to small sample size and high variances.
Table 4-2

Descriptive Statistics of Recall Accuracy (%) in terms of Instructional Condition, Song, and Presentation Order

<table>
<thead>
<tr>
<th>Instructional Condition</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>62.44</td>
<td>25.22</td>
<td>1, 35</td>
<td>.78</td>
<td>.39</td>
<td>.02</td>
</tr>
<tr>
<td>Without VPL</td>
<td>58.30</td>
<td>27.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>33.53</td>
<td>17.21</td>
<td>1, 35</td>
<td>.05</td>
<td>.83</td>
<td>.001</td>
</tr>
<tr>
<td>Without VPL</td>
<td>34.29</td>
<td>21.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhythm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>70.32</td>
<td>21.76</td>
<td>1, 35</td>
<td>.04</td>
<td>.85</td>
<td>.001</td>
</tr>
<tr>
<td>Without VPL</td>
<td>71.25</td>
<td>22.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Song</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>55.47</td>
<td>26.06</td>
<td>1, 35</td>
<td>4.85</td>
<td>.03</td>
<td>.12</td>
</tr>
<tr>
<td>February Twilight</td>
<td>65.28</td>
<td>25.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>32.29</td>
<td>16.39</td>
<td>1, 35</td>
<td>.9</td>
<td>.35</td>
<td>.03</td>
</tr>
<tr>
<td>February Twilight</td>
<td>35.53</td>
<td>21.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhythm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>64.11</td>
<td>20.28</td>
<td>1, 35</td>
<td>10.59</td>
<td>.003</td>
<td>.23</td>
</tr>
<tr>
<td>February Twilight</td>
<td>77.46</td>
<td>21.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation Order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Order</td>
<td>53.23</td>
<td>26.96</td>
<td>1, 35</td>
<td>5.41</td>
<td>.03</td>
<td>.13</td>
</tr>
<tr>
<td>2nd Order</td>
<td>67.52</td>
<td>23.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Order</td>
<td>32.33</td>
<td>17.89</td>
<td>1, 35</td>
<td>.39</td>
<td>.54</td>
<td>.01</td>
</tr>
<tr>
<td>2nd Order</td>
<td>35.49</td>
<td>20.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhythm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Order</td>
<td>64.45</td>
<td>22.79</td>
<td>1, 35</td>
<td>7.41</td>
<td>.01</td>
<td>.18</td>
</tr>
<tr>
<td>2nd Order</td>
<td>77.12</td>
<td>18.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-3

Descriptive Statistics of Recall Accuracy (%) by Song, Presentation Order, and Instructional Condition

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Lyrics</th>
<th></th>
<th>Pitches</th>
<th></th>
<th>Rhythm</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>April</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>9</td>
<td>50.17</td>
<td>23.47</td>
<td>31.43</td>
<td>16.47</td>
<td>58.85</td>
<td>21.97</td>
</tr>
<tr>
<td>Without VPL</td>
<td>9</td>
<td>43.57</td>
<td>26.41</td>
<td>32.29</td>
<td>20.43</td>
<td>54.34</td>
<td>21.87</td>
</tr>
<tr>
<td>2nd Order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>9</td>
<td>73.61</td>
<td>27.14</td>
<td>32.64</td>
<td>16.28</td>
<td>76.56</td>
<td>13.76</td>
</tr>
<tr>
<td>Without VPL</td>
<td>9</td>
<td>54.51</td>
<td>20.59</td>
<td>32.81</td>
<td>14.82</td>
<td>66.67</td>
<td>18.15</td>
</tr>
<tr>
<td>February Twilight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>9</td>
<td>60.57</td>
<td>23.21</td>
<td>30.47</td>
<td>12.56</td>
<td>68.1</td>
<td>21.95</td>
</tr>
<tr>
<td>Without VPL</td>
<td>9</td>
<td>58.6</td>
<td>34.43</td>
<td>35.12</td>
<td>23.26</td>
<td>76.52</td>
<td>22.38</td>
</tr>
<tr>
<td>2nd Order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>9</td>
<td>65.41</td>
<td>25.18</td>
<td>39.61</td>
<td>23.31</td>
<td>77.78</td>
<td>25.64</td>
</tr>
<tr>
<td>Without VPL</td>
<td>9</td>
<td>76.52</td>
<td>17.6</td>
<td>36.92</td>
<td>28.13</td>
<td>87.46</td>
<td>12.33</td>
</tr>
</tbody>
</table>

Research Question 2. Do non-music majors experience less aural cognitive load if they learn difficult songs with or without visually presented lyrics?

In this study, cognitive load was measured by Reaction Times (RTs). A faster reaction indicates more cognitive capacity is available. Before analyzing the RTs data, any RTs that happened before the target stimuli or slower than 3000 ms (only one case) were eliminated. To detect any outliers, a box-plot was generated based on individual participant’s RTs for each song. Only extreme outliers (April: 11 cases, February Twilight: 3 cases among all participants’ RTs (288 cases for each song)) were deleted. The remaining RTs of each participant for each song were aggregated and the average score was used for data analysis. The RTs for each song were normally distributed.

A paired samples \( t \)-test revealed a main effect for instructional condition on RTs, \( t(35) = 2.18, p = .036 \). In the instructional condition with VPL, participants detected the click sounds
faster compared to without VPL ($M_D = 45.31$ ms). In additional analyses, the effect of song and the effect of presentation order were also revealed. Participants’ RTs were faster for “February Twilight” than for “April” ($M_D = 45.62$), $t(35) = 2.2$, $p = .035$ and for the second song than for the first song ($M_D = 66.88$), $t(35)= 3.51$, $p = .001$. Table 4-4 shows the descriptive statistics and t-test results. The means and SDs of RTs for each song by instructional condition and presentation order were represented in Figure 4-2.

Table 4-4

| Reaction Times (RTs) for Instructional Condition, Song and Presentation Order |
|---|---|---|
| Instructional Condition | $n$ | $M$ | $SD$ |
| With VPL | 36 | 396.3 | 118.29 |
| Without VPL | 36 | 441.61 | 103.23 |
| Song | | | |
| April | 36 | 441.76 | 125.77 |
| February Twilight | 36 | 396.15 | 93.89 |
| Presentation Order | | | |
| $1^{st}$ | 36 | 452.39 | 101.36 |
| $2^{nd}$ | 36 | 385.52 | 114.59 |

*Figure 4-2.* Bar graphs of reaction times for “April” and “February Twilight” by instructional condition and presentation order.
Although main effects for song and presentation order existed, the instructional condition significantly influenced the participants’ RTs. It seems counterbalancing in the study design did balance the effect of song and presentation order in terms of RTs. According to the theoretical assumption in this study, the main effect of instruction condition indicates that seeing the lyrics while learning a difficult song induces less cognitive load in the aural channel than learning a difficult song without seeing the lyrics.

Research Question 3. Does cognitive load mediate the relationship between instructional condition (learning difficult songs with or without visually presented lyrics) and song recall accuracy in non-music majors?

To investigate whether cognitive load mediates the effect of visually presented lyrics on song recall, a path analysis, MEMORE (MEdiation and MOderation analysis for REpeated measures designs) developed by Montoya and Hayes (2016), was utilized. This statistical mediation analysis allows researchers to estimate the indirect effect in the two-condition within-participants design even without establishing evidence of a direct effect (Montoya & Hayes, 2016). This model uses Bootstrap Confidence Intervals (CI) to indicate the indirect effect. If the CI excludes zero, the indirect effect is considered significant.

A MEMORE based 5,000 Bootstrap samples revealed indirect effects of instructional condition on recall accuracy through cognitive load for lyrics (95% CI [.252, 9.887]) and rhythm (95% CI [.263, 9.362]), but not for pitches (95% CI [-.856, 6.393]) Yet, no direct and total effects were found. This mediation analysis suggests that seeing the lyrics indirectly increased the recall accuracy of lyrics and rhythm through its positive effect on cognitive load in the aural channel than the counterpart condition, which in turn increased the recall accuracy of lyrics and rhythm.
Summary of Findings

This study investigated how seeing the lyrics while learning a difficult song affects non-music majors’ song recall. As a result of different instructional conditions, recall accuracy of the learned song and cognitive load induced during learning were examined. The central interest of this study was whether seeing the lyrics induces less cognitive load in the aural channel compared to not seeing the lyrics, which in turn leads to better recall accuracy of the learned song.

First, the results showed the instructional condition (with or without visually presented lyrics (VPL)) did not affect participants’ recall accuracy of the learned song, whereas the song and the presentation order had an impact on their recall accuracy: Participants better recalled “February Twilight” than “April”; they better recalled the lyrics and rhythm of the second song they learned than the first song they learned. Second, the instructional condition affected participants’ cognitive load, which was measured by reaction times (RTs). When seeing the lyrics, participants reacted faster to the click sound over the aural instruction than when not seeing the lyrics. This finding suggests learning a difficult song with VPL induced less cognitive load in the aural channel than learning the song without VPL. Finally, a mediation analysis revealed the indirect effect of instructional condition on recall accuracy of the lyrics and rhythm through cognitive load. That is, seeing the lyrics indirectly increased the recall accuracy of the lyrics and rhythm through its positive effect on RTs producing less cognitive load, which in turn increased the recall accuracy of lyrics and rhythm. This finding suggests the cognitive load mediates the association between instructional condition and recall accuracy of lyrics and rhythm.
Discussion

The finding of no main effect of instructional condition (learning a difficult song with or without VPL) contradicts the finding of a previous study (Han, 2016). In the former study, non-music majors better recalled the pitches and rhythm when seeing the lyrics compared to when not seeing the lyrics. However, in this study, the instructional condition did not affect the participants’ recall accuracy.

It is noteworthy that there might have been another critical factor when comparing results of these two studies. In the first study, participants included members from auditioned choirs so their singing accuracy was not an issue. On the contrary, the second study participants were not members of an auditioned choir. In fact several were poor-pitch singers who “are deficient with respect to vocal imitation” (Pfordresher & Halpern, 2013, p. 747) based on my informal observation. Further, a few participants changed the tonality of the song (e.g., Mixolydian to Dorian) when learning and recalling the song. Although the second study used within-participants design to control for individual difference in singing ability along with other possible confounding variables, participants’ singing accuracy might be a critical factor to be considered to better understand the effect visually presented lyrics may have on song learning. If participants were not able to match the pitches and learn the song accurately, their recall accuracy would be seriously affected by their singing accuracy that is not a direct outcome of the learning condition. Singing accuracy is often measured through a pitch-matching task (Dalla Bella, 2015; Nichols, 2015; Salvador, 2010), which is correlated with pitch-discrimination ability (Watts, Moore, & McCaghren, 2005) as well as the perceived vividness of auditory imagery (Pfordresher & Halpern, 2013). Also, Rutkowski (2015) reported a correlation between children’s use of vocal registers and singing accuracy in a pattern-matching
task. Hence, the variables related to singing accuracy such as pitch-matching ability, pitch-discrimination ability, auditory imagery, and vocal registers should be considered in further study when examining the VPL effect.

Although not a variable of interest in this study, the song and presentation order did affect the song recall accuracy (lyrics and rhythm) and cognitive load. “February Twilight” turned out to be an easier song to remember than “April.” When simply comparing the number of components in the two songs, “April” has seven more words and ten more notes than “February Twilight”; “April” consists of nine rhythm patterns based on a half bar unit including subdivided notes (sixteen notes) whereas “February Twilight” has five rhythm patterns without subdivision. Perhaps, the less amount of information in “February Twilight” induced less cognitive load than “April” as demonstrated in RTs (faster RTs in “February Twilight”), leading to better recall accuracy.

Although participants were informed about the song learning procedure, the novelty of the learning situation seemed to cause more cognitive load when learning the first song than when learning the second. In Gault’s (2002) study investigating effects of pedagogical approaches (whole-song vs. phrase by phrase), text condition (singing with vs. without lyrics), and developmental music aptitude on children’s song learning, the results depended on the song taught. Gault (2002) speculated that the experience gained while learning the first song might have carried over to the learning the second song since the songs were not counterbalanced. Similarly, the song itself and the presentation order of instructional condition could influence song learning, which was measured by song recall accuracy in this study.

It is noteworthy that the effects of song and presentation order existed only for the recall of lyrics and rhythm, but not pitches. Interestingly, the correlation between lyrics and rhythm in
the recall accuracy was higher \( (r_{lyrics \& rhythm} = .798) \) than other two combinations \( (r_{lyrics \& pitches} = .478, r_{pitches \& rhythm} = .575) \) although all the correlations were significant at the .001 level. According to Purnell-Webb and Speelman (2008), familiar rhythm patterns can facilitate text recall by furnishing the learner with a schematic frame to which the words can be attached. It seems that words and rhythm are encoded and retrieved in association with each other (Cason, Astésano, & Schön, 2015; Peretz, Radeau, & Arguin, 2003; Purnell-Webb & Speelman, 2008), but the precise nature of their interaction, for example which aspect is psychologically primary, is not known and can be another topic in further research.

The results from the reaction times measure suggest learning a difficult song while seeing the lyrics induces less cognitive load in the aural channel compared to learning it without seeing the lyrics. Furthermore, cognitive load mediates the causal relationship between instructional condition and recall accuracy of lyrics and rhythm. The underlying assumptions of these findings were that seeing the lyrics during aural instruction allows learners to process the verbal information dually – visually and aurally (Baddeley, 1992; Penney, 1989); the dual processing of verbal information reduces the cognitive load in the aural channel leading to more capacity to process musical information.

This logic stands on the empirical findings called modality effect in the cognitive load literature (Low & Sweller, 2014; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995; Tabbers, Martens, & van Merriënboer, 2004; Tindall-Ford, Chandler, & Sweller, 1997) and verbal redundancy effect in the working memory literature (Lewandowski & Kobus, 1993; Montali & Lewandowski, 1996; Moreno & Mayer, 2002, Penney, 1989). Given the limited capacity of working memory (Baddeley, 1992; Cowan, 2000; Miller, 1956), attempting to maximize the learner’s capacity by employing two channels – aural and visual – is the basic
idea of these effects. Specifically, the modality effect means that when learning from pictures and words, if the verbal information is presented aurally as narration instead of visually as text, the visual channel might be offloaded with regard to the verbal information, having more capacity to process the pictures and in turn enhances the learning (Mayer & Moreno, 1998). The verbal redundancy effect in the literature of working memory refers to presenting identical words in two modalities as written and spoken words helps learning compared only using one modality (i.e., written or spoken), but if there is another mode of information in the visual channel such as diagrams or graphs, the verbal redundancy is not effective (Kalyuga, Chandler, & Sweller, 1999; Kalyuga & Sweller, 2014).

In this study, when showing the lyrics with aural instruction, participants learned lyrics aurally and visually, which means the verbal information remained in the aural channel. This condition does not satisfy the premise that each channel should have one mode of information to establish the modality effect and the verbal redundancy effect. However, despite the study condition that the aural channel holds two modes of information (verbal and musical), the research findings indicate showing the lyrics while aurally teaching a difficult song induces less cognitive load in learners than not showing it, implying dual processing of verbal information reduces cognitive load in the aural channel. Perhaps, the verbal redundancy effect found in this study without satisfying the condition is because the relationship between lyrics and tunes is different from the relationship between graphs (or diagrams) and words in that lyrics and tunes in singing are temporally integrated, while the graphs and words are not.

The possible concern about showing the lyrics might be whether the learners pay more attention to lyrics than other musical information as raised in the studies investigating the effect of singing with/without lyrics on singing achievement (Andress, 1986; Goetze et al, 2010;
Levinowitz, 1987). This study did not reveal the effect of instructional condition on song recall, which means participants remembered the songs similarly regardless of whether they saw the lyrics or not; learners’ recall accuracy was not reduced due to seeing the lyrics. On the other hand, seeing the lyrics reduced the aural cognitive load in the learners, at least young adults who were untrained singers. However, this finding should be tested with learners of different ages and different levels of musical expertise.

Among various possible factors, this study investigated the effect of visually presented lyrics on song recall. While various types of visual and kinesthetic aids can be used in music class, the musical elements mostly tested in the research studies were melodic contour (Apfelstadt, 1984; Dunn, 2008; Gomko & Poorman, 1998; Hughes, 1991; Pautz, 1988; Persellin, 1993; 1994; Zikmund & Nierman, 1992; Tarnowski, 1986) or melodic rhythm (Dunn, 2008; Pautz, 1988; Persellin, 1988, 1992, 1994; Zikmund & Nierman, 1992; Tarnowski, 1986). If we often see the lyrics while learning or singing a song, how seeing the lyrics affect song learning seems worthy to examine.

The basic idea of this study followed the dual-task investigation conducted by Brünken and his colleagues (2002, 2004), measuring cognitive load through reaction times. While they did not intend to statistically examine the relationship between cognitive load and learning outcome in their studies, this study utilized a mediation analysis to understand how learning a difficult song with visually presented lyrics affects song recall through cognitive load. The dual-task approach seems a promising method to assess cognitive load induced under various instructional conditions and with a mediation analysis, researchers could better understand the underlying mechanisms of learning process with certain learning conditions.
Although this study revealed the mediating effect of cognitive load on the relationship between instructional condition (with and without VPL) and recall accuracy of the learned song in terms of lyrics and rhythm, the instructional condition did not directly affect song recall. The outcome measure, song recall, might not be explained only with the cognitive factor. As mentioned above, the learners’ vocal production skills such as pitch-matching ability and the use of vocal registers might affect their song learning. To control for these possible confounding factors, in a previous study, Han (2016) recruited auditioned choir members. However, they are not representative of who most music educators teach. It seems important to better understand the issue of showing the lyrics to learners representing a variety of singing abilities. These factors should be considered in future studies. Also, learners of different ages and levels of musical expertise might yield different results.

This study utilized a within-participants design to control for individual difference across groups. Although presentation order of the instructional condition and song were counterbalanced, their effects were salient. Perhaps, a matched-participants design would solve this issue, employing one song with one instructional condition but still controlling for individual difference in pitch-matching ability and vocal registers.

**Conclusion**

The purpose of this study was to ascertain whether showing the lyrics to non-music majors induces less cognitive load in the aural channel compared to not showing the lyrics, which in turn leads to better recall accuracy of the learned song. Although learning a difficult song with visually presented lyrics did not affect recall accuracy of the learned song, seeing the lyrics induced less cognitive load in the aural channel compared to not seeing the lyrics. Also, the mediation analysis suggests seeing the lyrics indirectly increased recall accuracy of the
lyrics and rhythm through its positive effect on reaction times producing less cognitive load compared to not seeing the lyrics.

When teaching a song to young adults without formal musical training given limited instructional time, if the song is difficult – has much musical and verbal information - showing the lyrics should be considered to reduce their cognitive load in the aural channel, which might lead to better song learning in terms of the lyrics and rhythm.
References for Chapter 4


https://doi.org/10.1177/102986499800200204


http://dx.doi.org/10.1037/h0043158


https://doi.org/10.1177/002221949602900305


http://dx.doi.org/10.1037/0022-0663.91.2.358


http://dx.doi.org/10.1037/0022-0663.94.1.156


http://dx.doi.org/10.1037/0022-0663.87.2.319


https://doi.org/10.1177/8755123315576764


http://dx.doi.org/10.1037/1076-898X.3.4.257


https://doi.org/10.1177/0305735692201005
Chapter 5 Discussion

When learning a song aurally, verbal information (i.e., lyrics) and musical information (e.g., pitches and rhythm) are processed through the aural channel. However, if a song has much verbal and/or musical information, only using the aural channel might cause cognitive overload, thus hindering one’s ability to learn the song. Given the limited capacity of working memory (Baddeley, 1992; Cowan, 2000; Miller, 1956), maximizing the learner’s capacity by employing two channels - aural and visual - has been suggested (Low & Sweller, 2014; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995; Tabbers, Martens, & van Merriënboer, 2004; Tindall-Ford, Chandler, & Sweller, 1997). Based on these ideas, I assumed presenting the lyrics dually - visually and aurally - might reduce cognitive load while learning a difficult song. That is, if participants see the lyrics while learning a song through aural instruction, the verbal information (i.e., lyrics) might be processed dually. The dual processing of verbal information might reduce cognitive load in the aural channel, leading to more capacity to process musical information. In turn, more aural cognitive capacity would lead to better recall of the learned song.

Summary of the Studies

Based on these assumptions, the effect of Visually Presented Lyrics (VPL) on song recall was examined through two experimental studies in this dissertation. As a result of two different instructional conditions (learning a difficult song with or without VPL), the recall accuracy of the learned song (Study 1 & 2) and cognitive load (Study 2) were examined. Recall accuracy was measured by a half bar unit for the component of lyrics, pitches, and rhythm. Cognitive load was measured through a simple sound monitoring task, and the Reaction Times
(RTs) were used as an indicator of cognitive load available in the aural channel at that point: A faster RTs indicates the less cognitive load induced from the particular instructional condition.

In the first study (chapter 3), members from auditioned choirs were invited to participate. Twenty-six study volunteers were randomly assigned to one of two instructional conditions. Participants in one group saw the lyrics while learning a difficult song, “April,” through prerecorded aural instruction via headphone. Participants in the other group did not see the lyrics but saw a blank screen while learning the song with the same instruction. When controlling for participants’ phonological working memory, I found non-music majors benefited from seeing the lyrics in that they better recalled the pitches and rhythm when seeing the lyrics while learning a difficult song, “April,” than when not seeing the lyrics. However, this was not the case for music majors. Music majors outperformed non-music majors in recalling the pitches and rhythm of “April” only when learning the song without VPL. Perhaps, music majors have more capacity in the aural channel and process the verbal and musical information in a more integrated way than non-music majors (Ginsborg & Sloboda, 2007): Presenting additional verbal information visually might be redundant for singers with high level of musical expertise (Kalyuga & Sweller, 2014; Kalyuga, Chandler, & Sweller, 1998; Sweller, Ayres, Kalyuga, & Chandler, 2003). On the other hand, for singers with lower levels of musical expertise, showing the lyrics can be one way to facilitate their song learning given limited instructional time.

In the second study (chapter 4), non-music majors, but not auditioned choir members, were invited to participate. Thirty-six participants learned two difficult songs, “April” and “February Twilight,” through prerecorded aural instruction via headphone. For one song, they saw the lyrics and for the other song, they did not see the lyrics. To control for the order effect
of instructional condition and song, the order of instructional condition and song was counterbalanced. While learning the song (short phrase echoes), they took a distinct sound detection task. Their reaction times (RTs) in the detection task were recorded and used as an indicator of cognitive load. The instructional condition did not affect recall of the learned song, but affected cognitive load: When seeing the lyrics, participants reacted faster than when not seeing the lyrics, which suggests less cognitive load induced from the learning task. A mediation analysis called MEMORE (Montoya & Hayes, 2016) revealed instructional condition indirectly affected recall accuracy of lyrics and rhythm through cognitive load. I concluded when teaching a song to young adults without formal musical training if the song is difficult due to much musical and verbal information, showing the lyrics should be considered, particularly if instructional time is limited, to reduce their cognitive load in the aural channel, which might lead to better song recall of lyrics and rhythm.

Comparison of the Two Studies

Prior to a comparison of the results of these two studies, it should be noted that several differences were inherent in the research methods used in the two studies. In the first study, all participants, including non-music majors, were members of auditioned choirs, whereas in the second study, although participants had some musical experiences, none of them participated in a choir during college. In the second study, some participants were poor-pitch singers who “are deficient with respect to vocal imitation” (Pfordresher & Halpern, 2013, p. 747); this was not the case in the first study. Second, the first study utilized a between-participants design: Participants were randomly assigned to one instructional condition of two and learned only one song “April.” Also, participants’ phonological working memory was used as a covariance to control for individual differences among groups. On the other hand, the second study utilized a
within-participants design in order to control for any confounding factors such as individual differences in working memory, singing ability and reaction times. Participants learned two songs “April” and “February Twilight.” Although the presentation order of instructional condition and song were counterbalanced, it turned out the presentation order and song were critical factors that affected the recall accuracy of the learned songs. Lastly, the time gap between the instruction and recall test was different for the two studies. In the first study, between the learning and recall stage, participants took a phonological working memory test, which lasted about 15 minutes, so the test can be considered a delayed recall test. On the other hand, the time gap between learning and recall in the second study was about 2 minutes.

Notwithstanding, I attempted to compare the results of two studies.

Discussion of Recall Accuracy

Regarding recall accuracy, the findings from these two studies were incongruent. In study 1, non-music majors better recalled the pitches and rhythm when seeing the lyrics compared when not seeing the lyrics. In study 2, instructional condition did not affect song recall. For a comparison of the results of these two studies, I only used the recall accuracy scores of “April,” only non-music majors’ results from the first study, and the first order results in the second study because song, the level of musical expertise, and presentation order did matter. As shown in Table 5-1, participants in the second study better recalled the lyrics than the first study participants. Since participants experienced a 15-minute delay between learning and recall phases in the first study, but had only a 2-minute delay between these phases in the second study, the higher scores of recall of lyrics in the second study seems reasonable. However, participants in the first study better recalled pitches and rhythm than the second study participants even with the longer delay. This result might reflect the study participants’ musical
ability, including singing skills, since the first study participants all were in an auditioned college choir while the second study participants were not experienced singers. A two-way (Study × Instructional Condition) MANOVA revealed a main effect for study, \( F(3, 26) = 6.83, p = .002 \), indicating difference between the first study and the second study. However, the univariate analysis did not reveal any difference for lyrics (\( p = .11 \)), pitches (\( p = .12 \)), and rhythm (\( p = .14 \)) regarding study, perhaps due to the small sample size and high variance. Also, no main effect were found for instructional condition, \( F(3, 26) = 1.84, p = .18 \) and for interaction, \( F(3, 26) = 2.49, p = .08 \).

Table 5-1

Means and SDs of the First (results from non-music majors) and Second (results from April presented in the first order) Studies

<table>
<thead>
<tr>
<th></th>
<th>Study 1 (only non-music majors)</th>
<th>Study 2 (April presented in the first order)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With VPL</td>
<td>Without VPL</td>
</tr>
<tr>
<td>( n )</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>runemeLyrics</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Pitches</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Rhythm</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>With VPL</td>
<td>5</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without VPL</td>
<td>9</td>
<td>12.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With VPL</td>
<td>9</td>
<td>16.06</td>
</tr>
<tr>
<td>Second Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(April presented in the first</td>
<td></td>
</tr>
<tr>
<td>order)</td>
<td>order)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The level of musical expertise is a critical factor to understanding the relationship between tune and lyrics in memory (Besson, Chobert, & Marie, 2011; Besson, Schön, Moreno, Santos, & Magne, 2007; Ginsborg & Sloboda, 2007; Kilgour, Jakobson, & Cuddy, 2000). In other words, the level of musical expertise affects how lyrics and tune are processed and retrieved in association each other. Musical training improves auditory temporal processing (Besson et al., 2007; Jakobson, Cuddy, & Kilgour, 2003), which in turn increases verbal memory (Chan, Ho, & Cheung, 1998; Jellison & Miller, 1982; Kilgour et al., 2000). Also, persons with a higher level of musical expertise would better encode and represent music since
they better understand musical structure than those with a lower level of musical expertise. Their superior understanding of musical structure and characteristics would facilitate their recall of music, serving as retrieval cues (Racette & Peretz, 2007). Since a retrieval of one component in a song (i.e., lyrics or tune) facilitates the retrieval of the other components (Crowder et al., 1990; Ginsborg & Sloboda, 2007, Serafine et al, 1984, 1986), the better recall of musical components leads to the higher likelihood of an increased recall of words (Ginsborg & Sloboda, 2007).

In the first study, music majors better recalled the pitches of the song when they did not see the lyrics. Seeing the lyrics seemed to help the non-music majors (but trained accurate singers) process the verbal information, yielding comparable recall accuracy between non-music majors and music majors. However, in the second study, participants were not only non-music majors but also non-members of auditioned college choir. Their level of musical expertise might not be comparable even with non-music majors in the first study. For example, in the second study, some participants changed the tonality of songs (e.g., from Mixolydian to Dorian) when learning and singing the song. Hence, to better understand the effect of visually presented lyrics, the learners’ level of musical expertise should be considered.

Discussion of Cognitive load

From a cognitive load perspective, several theoretical assumptions guided this dissertation. In the first study, it was assumed seeing the lyrics while learning a song aurally would allow learners to process verbal information dually – aurally and visually. The dual processing of verbal information may reduce cognitive load in the aural channel, providing more cognitive resources to process musical information. In turn, it would increase recall accuracy of the learned song. The findings of non-music majors in the first study seemed to be
in line with these theoretical assumptions. However, music majors did not benefit from seeing the lyrics. Music majors have more capacity in the aural channel (Chan, Ho, & Cheung, 1998; Jellison & Miller, 1982; Kilgour et al., 2000) and might process verbal and musical information in a more integrated way (Gingsborg & Sloboda, 2007).

In the second study, this theoretical assumption was assessed through a dual-task approach (Britton, Glynn, Meyer, & Penland, 1982; Brünken, Steinbacher, Plass, & Leutner, 2002; Brünken, Plass, & Leutner, 2004). The performance level in the secondary task indicates cognitive capacity available while performing the primary task. In the second study, following Brünken and his colleagues (2004), reaction times in the single sound monitoring task (secondary task) while learning the song (primary task) were used as an indicator of cognitive load in the aural channel. When seeing the lyrics, participants’ reaction times were faster than when not seeing the lyrics. These results support the theoretical assumption that seeing the lyrics indirectly increases the recall accuracy of lyrics and rhythm through its positive effect on cognitive load, which in turn increase the recall accuracy of lyrics and rhythm. This result seems to reflect the close relationship between lyrics and rhythm in memory and neural processing (Cason, Astésano, & Schön, 2015; Peretz, Radeau, & Arguin, 2003; Purnell-Webb & Speelman, 2008) but it is not clear whether the memory for words facilitates the memory for rhythm or the other way round: It can be bidirectional (Asaridou & McQueen, 2013; Peretz, Radeau, & Arguin, 2003).

**Limitations**

Song learning may be affected by a wide array of factors (Goetze, Cooper, & Brown, 2010; Hedden, 2002; Nichols, 2016; Pfordresher et al., 2015). Hedden (2012) categorized the affecting variables into internal and external factors. The internal factors include pitch-
matching ability, vocal range, and sex difference (Hedden, 2012) as well as developmental music aptitude (Gault, 2002; Hornbach & Taggert, 2005). The external factors include pedagogical approach -phrase-by-phrase or holistic (Klinger, Campbell, & Goolsby, 1998; Gault, 2002; Persellin & Bateman, 2009), vocal modeling (Persellin, 2006; Rutkowski & Miller, 2003), singing with a neutral syllable or with lyrics (Gault, 2002; Goetze; 1985; Levinowitz, 1987, 1989; Smale, 1987), accompaniment (Atterbury & Silcox, 1993; Guilbault, 2004; Hale, 1977; Hedden & Baker, 2010; Simeon & Ku, 2015) and individual or group singing (Goetze & Horii, 1989, Rutkowski, 1996; Rutkowski & Miller, 2003).

To investigate the effect of visually presented lyrics on song learning, several external factors were controlled and not included in this dissertation. For example, the song instruction basically followed the phrase-by-phrase approach instead of the holistic approach. For the untrained singers, female and male vocal models were provided to match the gender of the participant. All songs were presented with lyrics not a neutral syllable. No accompaniment was provided and individual singing was assessed. In this dissertation, visually presented lyrics was examined as a possible external factor for song learning.

Internal factors such as participants’ pitch-matching skills and vocal register might play a crucial role to understand the phenomenon. According to Loui and her colleagues (2015), “successful singing requires perceptual skills (pitch matching, interval reproduction, and fine-grained pitch discrimination ability), cognitive abilities (working memory, attention, and learning processes), and motor skills (motor planning, motor selection, and motor execution)” (p. 263). Thus, learning a song and singing it in tune involves not only the cognitive process but also perceptual and motor skills. In this dissertation, only a cognitive factor was examined under the assumption that dual processing of verbal information may reduce cognitive load in
working memory, thus facilitating song recall. In the second study of this dissertation, the reaction time results supported this underlying assumption, however, instructional condition did not directly affect the song recall accuracy. As noted earlier, in the second study, participants’ singing abilities were varied. Some participants were not able to match the pitches and sing in tune consistently and some participants’ singing ranges were limited. Although an attempt to control for the individual difference in singing ability was made using within-participants design, participants’ use of vocal registers and pitch-matching skills seem critical factors to better understanding the effect of visually presented lyrics. Similarly, their level of reading comprehension might also affect the results, although this is unlikely with college students.

In both studies, for the recall test, participants did not see the lyrics regardless of whether they saw the lyrics or not during instruction. Although participants were made aware of this recall condition at the beginning of experimentation, a change in format between the learning and recall phases may introduce a bias in text recall in favor of the “without VPL” condition. Because participants who saw the lyrics might be used to seeing the lyrics while singing the song and then experience additional difficulty during the recall test when the lyrics were not present. However, for consistency, if I showed the lyrics to the participants who saw the lyrics while learning the song, another bias in text recall in favor of the “with VPL” condition may exist. The visually presented lyrics would function as an influential retrieval cue for song recall while the counterparts (the participants who did not see the lyrics in both – learning and recall phases) do not have any retrieval cue. To balance out this potential bias situation, if all participants saw the lyrics on the recall test regardless of instructional condition, the participants who did not see the lyrics while learning might generate additional cognitive load to process and integrate the newly presented visual information with aural information,
which is not an ideal situation, either. Thus, in this dissertation, the potential bias toward the “without VPL” condition in song recall due to format change might be a limitation of the study methods.

Due to the different research design, the elapse of time between learning phase and recall phase were different in both studies: In the first study, it was about 15 minutes whereas, in the second study, it was about 2 minutes. Although in the studies investigating the effect of tune on text recall (Kilgour et al., 2000; Wallace, 1994), the effects of certain instructional conditions did not differ between immediate and delayed recall. When an efficacy of certain instructional condition is demonstrated through both types of recall – immediate and delayed, the more persuasive results would be yielded than using one type of recall.

**Implications**

In music instruction, including song teaching, various types of visual and kinesthetic aids can be used, yet the musical elements mostly tested by other researchers were melodic contour (Apfelstadt, 1984; Dunn, 2008; Gomko & Poorman, 1998; Hughes, 1991; Pautz, 1988; Persellin, 1993; 1994; Zikmund & Nierman, 1992; Tarnowski, 1986) or melodic rhythm (Dunn, 2008; Pautz, 1988; Persellin, 1988, 1992, 1994; Zikmund & Nierman, 1992; Tarnowski, 1986) using visual and kinesthetic representation. Some researchers have investigated the effect of staff notation on music learning (Kendall, 1988; Korenman & Peynircioglu, 2007; Owens & Sweller, 2008; de Stwolinski, Faulconer & Schwarzkopf, 1988) but not visually presented lyrics. In practice, if we often see the lyrics while learning a song, how seeing the lyrics affects the song learning seems an important question to be explored further.
**Implications for Future Research**

In the area of educational psychology, a number of cognitive load effects have been demonstrated across various instructional domains, yet only Owens and Sweller (2008) examined the applicability of the cognitive load effects to music instruction. Specifically, they demonstrated split-attention and modality effects in music theory instruction (e.g., filling a blank with a correct note in a bar based on time signature and identifying the type of meters): When learning from spatially and temporally integrated materials, participants better performed in the knowledge and transfer tests than when learning from separated materials. While Owens and Sweller (2008) tested the cognitive load effects through participants’ theoretical understanding of music, I explored whether the cognitive load effects, specifically modality effect, can be applied to song learning. Although the modality effect (using two modalities - visual and aural - instead of one modality yields better performance) was only demonstrated for the non-music majors but experienced singers in the first study.

In this dissertation, when showing the lyrics with aural instruction, participants learned lyrics aurally and visually, which means the verbal information remained in the aural channel. This condition does not satisfy the condition that each channel should have one mode of information to establish the modality effect and the verbal redundancy effect. However, despite the study condition that the aural channel holds two modes of information (verbal and musical), the research findings indicate showing the lyrics while aurally teaching a difficult song induces less cognitive load in learners than not showing it, implying dual processing of verbal information reduces cognitive load in the aural channel. Perhaps, the verbal redundancy effect found in this study without satisfying the condition is because the relationship between lyrics
and tunes is different from the relationship between graphs (or diagrams) and words in that lyrics and tunes in singing are temporally integrated, while the graphs and words are not.

As visual aids are commonly used in music instruction (Sheldon, 1991; Tarnowski, 1986; Zikmund & Nierman, 1992), the effectiveness of visual representation in addition to aural information could be examined from a cognitive load perspective. Using a dual-task approach I demonstrated dual presentation of verbal information induces less cognitive load in the learner than single presentation, and cognitive load mediates song learning with or without visually presented lyrics. Dual-task approach seems a promising method to assess cognitive load induced under various instructional conditions. With dual-task approach and a mediation analysis, researchers could better understand the underlying mechanisms of the learning process with certain learning conditions.

**Implications for Music Learning and Teaching**

When the processing demands induced by certain learning task and material exceed the learner’s cognitive capacity, it causes cognitive overload. Mayer and Moreno (2003) suggested nine ways to alleviate cognitive load in multimedia learning. In their contexts, multimedia learning means learning from words and pictures. One of the nine strategies was the modality principle: Dividing the essential processing into visual and auditory channels. Besides, other strategies included segmenting, pretraining, weeding, signaling, aligning, eliminating redundancy, synchronizing, and individualizing. Table 5-2 shows the description of each strategy. Note that the first strategy, off-loading, is the modality principle, but described differently since the learning context is different: These ideas were based on the studies tested the instructional condition containing verbal (i.e., words) and visual (e.g., pictures) information. However, in music learning, mostly aural information is involved and remains in the aural
channel. A way of applying this modality principle to music learning may be employing the visual channel in addition to the auditory channel. Regarding eliminating redundancy, if there is no competing visual information, even redundant information could be helpful for the learner (Moreno & Mayer, 2002). The descriptions of nine strategies are based on multimedia learning, but the principles could be applied to music learning with careful consideration. As some strategies such as segmenting, pretraining, and signaling, are already used in music instruction. The phrase-by-phrase song teaching method can be one example of using the segmenting strategy in music instruction.

Table 5-2
Nine Strategies to Alleviate Cognitive Load in Multimedia Learning (from Mayer & Moreno, 2003)

<table>
<thead>
<tr>
<th>Load-Reducing Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off-loading</td>
<td>Move some essential processing from visual channel to auditory channel</td>
</tr>
<tr>
<td>Segmenting</td>
<td>Allow time between successive bite-size segments</td>
</tr>
<tr>
<td>Pretraining</td>
<td>Provide pretraining in names and characteristics of components</td>
</tr>
<tr>
<td>Weeding</td>
<td>Eliminate interesting but extraneous material to reduce processing of extraneous material</td>
</tr>
<tr>
<td>Signaling</td>
<td>Provide cues for how to process the material to reduce processing of extraneous material</td>
</tr>
<tr>
<td>Aligning</td>
<td>Place printed words near corresponding parts of graphics to reduce need for visual scanning</td>
</tr>
<tr>
<td>Eliminating redundancy</td>
<td>Avoid presenting identical streams of printed and spoken words</td>
</tr>
<tr>
<td>Synchronizing</td>
<td>Present narration and corresponding animation simultaneously to minimize need to hold representations in memory</td>
</tr>
<tr>
<td>Individualizing</td>
<td>Make sure learners possess skill at holding mental representations</td>
</tr>
</tbody>
</table>
Instructional design should be based on many considerations such as the instructional time, goals, and characteristics of students (e.g., their levels of expertise in the domain). Based on results of these studies, when learning a difficult song only using the aural channel students with low levels of musical expertise might experience cognitive overload, which hinders their learning. Given limited instructional time and the need to be more efficient, several strategies should be considered to prevent learners from experiencing cognitive overload while learning a difficult song aurally. Showing the lyrics of the song could be one strategy for that purpose, at least for young adults with low levels of musical expertise.
References for Chapter 5


Hale, M. R. (1977). An experimental study of the comparative effectiveness of harmonic and melodic accompaniments in singing as it relates to the development of a sense of


https://doi.org/10.1177/8755123314540660


http://dx.doi.org/10.1037/1076-898X.3.4.257

http://dx.doi.org/10.1037/0278-7393.20.6.1471


https://doi.org/10.1177/0305735692201005s
Vita
Yo-Jung Han

Academic Preparation

2018  Ph.D., Music Education, the Pennsylvania State University, PA, USA
2013  M.M.E., Music Education, the Pennsylvania State University, PA, USA
2008  M.A., Music Education, Seoul National University, Seoul, South Korea
2002  B.M., Music Education, Konkuk University, Seoul, South Korea

Professional Experience

2016-2018  Visiting Assistant Professor, Hayes School of Music
            Appalachian State University, Boone, NC
2012-2016  Teaching Assistant, School of Music
            The Pennsylvania State University, University Park, PA
2013-2014  General Music Teacher
            Friends School, State College, PA
2008-2011  Music Teacher
            Public Middle Schools, Seoul, South Korea

Publications


Research Presentations at the National and International Conferences

NA/ME Music Research and Teacher Education National Conference
2018  Attentional flexibility in pianists.
2016  An exploration of recent approaches to research in music education
      Building preservice elementary teachers’ self-efficacy to use music

Symposium on Music Teacher Education
2013  Democracy in public school ensembles: A descriptive survey of current practice

International Conference on Music Perception and Cognition
2016  The effect of visually presented lyrics on song recall.
      An investigation of multimodal imagery employed by ear players without absolute pitch
2014  Analysis of verbal descriptors for timbre: Based on conceptual metaphor theory
2008  Steady Beat Production (SBP) with various external stimuli by 7th grade students

International Conference on Music Learning Theory
2015  Viewing the stages of audiation through information processing models lens