Timbre Memory, Familiarity and Dissimilarity

Lap Ian Lou

Music Technology Area
Department of Music Research
Schulich School of Music
McGill University
Montreal, Canada

June 2017

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Arts.

Abstract

This study examines the effect of timbre memory decay on the perception of timbre dissimilarity in tones equalized in pitch, loudness and duration. This study makes two hypotheses: (1) the perception of timbre dissimilarity changes as retention interval (RI) increases; (2) timbre familiarity minimizes the perceptual change of timbre dissimilarity with increasing RI. Two experiments were conducted to test these hypotheses. In the first experiment, participants rated the dissimilarity of synthetic tone pairs with RIs of 0.5 s, 5 s and 10 s. In the second experiment, participants rated the dissimilarity of pairs of synthetic tones, digital transformations of acoustic sounds or acoustic instrument sounds with five RIs between 0.5 s and 5 s. The results of the two experiments show that participants perceive similar pairs as significantly more dissimilar as RI increases, particularly for highly unfamiliar synthetic tones, that RI has less of an effect on more dissimilar pairs, and that timbre familiarity minimizes the RI effect on dissimilarity ratings.

Résumé

Cette étude examine l'effet du déclin de la mémoire du timbre sur la perception de dissemblance entre timbres pour des sons égalisés en termes de la hauteur, de la sonie et de la durée. Cette étude fait deux hypothèses: (1) la perception de dissemblance du timbre change lorsque l'intervalle de rétention (IR) augmente; (2) le caractère familier de timbre minimise le changement perceptif de dissemblance avec une augmentation de l'IR. Deux expériences ont servi à tester ces hypothèses. Dans la première expérience, les participants devaient évaluer la dissemblance entre paires de sons synthétiques avec l'IR de 0.5 s, 5 s et 10 s. Dans la deuxième expérience, les participants devaient évaluer la dissemblance entre paires de sons qui étaient synthétiques, des transformations numériques de sons acoustiques ou provenant d'instruments acoustiques avec cinq IR entre 0.5 s et 5 s. Les résultats des deux expériences montrent que les participants perçoivent des paires acoustiquement semblables comme plus dissemblables lorsque l'IR augmente, surtout pour les sons synthétiques non familiers, que l'IR a un effet moindre sur les paires de sons dissemblables et que le caractère familier des timbres minimise l'effet d'IR sur les jugements de dissemblance.

Acknowledgment

First and foremost, I would like to thank my supervisor Stephen McAdams for his support and advice throughout the completion of the thesis. Second, I would like to thank Bennett Smith for programming the experimental interface and generating synthetic tones in PureData with additive synthesis. Third, I would like to thank Jungkyu Park for showing me how to use the hierarchical linear model to analyze data in SAS and how to report the result of the hierarchical linear model. Finally, I would like to thank the members of the Music Perception and Cognition Lab for providing feedback during piloting phases of the experiment and the participants in the experiments for providing me result data. The Research Ethics Board II of McGill University has reviewed and approved the experiments reported in the thesis (certificate # 67-0905).

Author contributions

Under Stephen McAdams' supervision, I made the hypothesis, designed the experiment, prepared the experiment and stimuli with Bennett Smith, ran the experiments, performed statistical analysis, interpreted the results and wrote the thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the 20^{th} and 21^{st} centuries, composers try to use timbre to create and organize musical syntaxes and musical forms. For example, on the basis of Kaija Saariaho's paper on the synthesis of poetic sounds co-authored with McAdams (1985), she manipulates timbre to create musical syntaxes with tension and release in her vocal work *From the Grammar of Dreams* (1988) (Siegel, 2014). These timbre-based musical syntaxes are based on memorable, non-uniform, discrete and transformable materials, which is similar to Meyer's concept of musical syntax (Nattiez, 2007). Also, with functions of introduction, extension, interruption and conclusion, electroacoustic music composers treat timbre as musical objects and manipulate the objects to create musical forms within which these objects are to be compared over durations of more than 2 s (Roy, 2003; Nattiez, 2007). As timbre can bear musical forms, configurations of timbre can be encoded, organized, recognized and compared in our minds (McAdams, 1989). Therefore, in their minds, people can form relations between different timbres. This study aims to examine how timbre memory decay might affect these mental relations, and how timbre familiarity would interact with the effect. However, studying the mental representations of different timbres is not as straightforward as it seems due to the multidimensional properties of timbre. This chapter will review the framework of timbre and memory research to establish fundamentals of the experimental design.

1.1 Psychophysical Structure of Timbre

According to the American National Standards Institute (ANSI, 1960/1994), timbre is defined as the "attribute of auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness and pitch, are dissimilar." Bregman (1990, p.

92-93) criticized that this definition did not define anything about timbre. The definition simply told us that we did not know much about timbre, but we knew timbre was not loudness or pitch. Since timbre is not defined as what it is, but what it is not, timbre is an umbrella term that describes various perceptual attributes like brightness, nasality, roughness, attack character, temporal fluctuation, and so on. This suggests that timbre has a multidimensional structure as well. In addition, according to McAdams (2013), these perceptual attributes contribute to the recognition, categorization and identification of sound sources. People form long-term mental representations with the perceptual attributes of sound sources from their daily experiences and categorize sound sources in their minds. The perceptual attributes serve as indicators of sound source recognition, and people are able to recognize the physical structures of the sound sources and how the sounds are produced. When they associate lexicons with the categorization of sound sources in their memories, they are able to identify the sound sources.

Multidimensional scaling (MDS) is a technique commonly used in timbre dissimilarity studies for understanding how people perceive the perceptual attributes of timbre and how they form a mental representation of timbre (Grey & Gordon, 1978; McAdams et al., 1995). In these studies, sound events such as recorded acoustic instruments and synthetic tones were first interpreted psychophysically based on their acoustic properties. Participants were then required to listen to different sound pairs and give a dissimilarity rating to each timbre pair. After that, MDS algorithms were used to analyze the matrix of participants' timbre dissimilarity ratings and to construct a multidimensional timbre space to represent the timbre dissimilarities.

MDS algorithms were gradually developed through several studies from MDSCAL (Torgerson, 1958; Gower, 1966) to INDSCAL (Carroll & Chang, 1970), EXSCAL (Winsberg & Carroll, 1989; Krumhansl, 1989), CLASCAL (Winsberg & De Soete, 1993), extended CLASCAL (McAdams et al. 1995), and CONSCAL models (Winsberg & De Soete, 1997; Caclin et al., 2005). MDSCAL is a classical MDS model that uses Euclidean distance to represent timbre dissimilarity. The Euclidean distance between the stimuli *s* and *s'* is presented by

$$d_{ss'} = [\sum_{r=1}^{R} (x_{sr} - x_{s'r})^2]^{0.5}$$

where x_{sr} and $x_{s'r}$ are the coordinates of stimuli *s* and *s'* on the dimension *r* respectively, and *R* is the total number of the timbre dimensions. However, this MDSCAL model has rotational invariance because the model distance is independent of the choice of axes. In order to construct more psychologically meaningful spaces, INDSCAL uses weighted Euclidean distances to model timbre dissimilarity. The weighted Euclidean distance between the stimuli *s* and *s'* is presented by

$$d_{ss'n} = [\sum_{r=1}^{R} w_{nr}(x_{sr} - x_{s'r})^2]^{0.5}$$

where w_{nr} is the weight for dimension r related to the nth participant in the experiment. The weighted Euclidean distance solves the issue of rotational invariance because the dimensions are weighted based on their saliences for the subjects. Besides shared dimensions, the stimuli have their specificities. In order to account for the specificities, an extension of the common dimensions model was developed, known as EXSCAL. This model uses the distance of weighted common dimensions and weighted specificities of the subjects. The distance between stimuli *s* and *s'* yielded by EXSCAL can be presented by

$$d_{ss'n} = [\sum_{r=1}^{R} w_{nr}(x_{sr} - x_{s'r})^2 + v_n(c_s^2 + c_{s'}^2)]^{0.5}$$

where c_s and $c_{s'}$ are the coordinates of stimuli *s* and *s'* along the dimension specific to those stimuli respectively, and v_n is the weight for the whole set of specificities related to the nth participant in the experiment.

MDS algorithms were further improved by the introduction of latent classes. Although the weighted Euclidean model helps to achieve rotational uniqueness, the model simultaneously introduces many individual subject weights w_{nr} and v_n that are not always interpreted individually in practice. The latent-class approach is introduced to ease this issue. This approach assumes that the participants belong to one and only one subpopulation, and the participants within a subpopulation weight the dimensions identically. To which subpopulation a particular participant belongs is not known in advance. The latent-class approach helps to reduce the number of parameters for model choice and does not reintroduce the issue of rotational invariance. Therefore, CLASCAL and extended CLASCAL (a weighted version of CLASCAL) models use this latent-class approach for construction of timbre space. The distance between stimuli *s* and *s'* for latent class *t* yielded by the extended CLASCAL model can be presented by

$$d_{ss't} = [\sum_{r=1}^{R} w_{tr}(x_{sr} - x_{s'r})^2 + v_t(c_s^2 + c_{s'}^2)]^{0.5}$$

where w_{tr} is the weight of the rth dimension for latent class *t* and v_t is the weight for the specificities for latent class *t*. In addition to the latent-class approaches, a more recent constrained scaling (CONSCAL) model used physical dimensions of the stimuli for more accurate representation of timbre dissimilarity. With MDS, mental representations of timbre dissimilarity can be studied and interpreted through correlation between dimensions of the timbre space and acoustic properties of the sound events.

Results of timbre similarity studies consistently suggest that the spectral center of gravity (SCG) and the logarithm of the attack time (AT) are two perceptually salient dimensions of timbre (Grey & Gordon, 1978; Krumhansl, 1989; Krimphoff et al., 1994; McAdams et al., 1995; Marozeau et al., 2003). However, different studies have suggested different dimensions as a third perceptually salient dimension of timbre depending on the stimulus set. For example, Grey and Gordon (1978) proposed spectral flux as a correlate. Other studies like Krimphoff et al. (1994) suggested that other dimensions such as spectral deviation, noisiness and inharmonicity might be salient dimensions of timbre. These different results were due to the use of different stimulus sets. Grey and Gordon (1978) used recorded acoustic instruments and digital transformations of acoustic sounds as stimuli sets, whereas Krimphoff et al. (1994) used Krumhansl's (1989) synthetic frequency modulation tones that were similar to acoustic instruments for their experiments. Different methods of stimulus generation made the studies reach slightly different conclusions. In order to have better control of the stimulus set, Caclin et al. (2005) performed a confirmatory study using synthetic tones to examine salient dimensions of timbre. Their results showed that AT and SCG, like other studies, were perceptually salient. On the other hand, the study showed that attenuation of even harmonics, also known as spectral irregularity (SI) or spectrum fine structure, was significantly salient, but spectral flux was not. Therefore, AT, SCG and SI are three strong candidates of perceptual salience for timbre dissimilarity.

1.2 Timbre and Sound Source Recognition

Giordano and McAdams (2010) carried out a meta-analysis of timbre dissimilarity and musical instrument identification research and found that musical instruments with similar physical structure and manner of excitation were clustered in the same region of the timbre space. Since the timbre space is a model of the mental representation of timbre dissimilarity, this indicates that there is categorization of sound sources in participants' mental representation. The meta-analysis also showed that participants in identification studies normally confused musical instruments with similar physical structure and manner of excitation like the violin and the viola. However, this was not the case for instruments with different physical structure or excitation. All in all, this analysis linked the psychophysical structure of timbre with sound source recognition, categorization and identification. As acoustic properties determine sound source categorization, it may be difficult to separate acoustic properties from sound source categorization.

McAdams (1993) proposed a model for sound source recognition. In this model, processing in

the auditory system first goes through the stage of sensory transduction, which converts physical vibration into neural impulses. The auditory information is then grouped in order to separate auditory representations of various sound sources. After that, the auditory properties or features are analyzed and matched to memory representations. As a result, people can recognize the sound source, or realize that they did not hear the sound source previously. If verbal lexicons can be activated, then people are able to identify the sound source as well.

Under specific situations, there is top-down feedback that involves selection of information on the basis of attentional processes. In a noisy environment, one can easily separate the sound of one's own name from background noises. One can also separate the words of one's native language from noises more easily compared to those of the non-native language. Another example is the phenomenon of phonemic restoration illusion (Warren, 1970). When phonemes of words were cut and replaced with a burst of noise, listeners were still able to hear the whole speech. Nevertheless, if silence replaced the phonemes, then listeners would hear the interrupted speech. This phenomenon happens because when noises replaced the phonemes, listeners were tricked to perceive that the burst of noise masked the phonemes instead of phonemes being cut out of the words. As a result, the mind of the listener automatically restored the missing phonemes, hence the designation of phonemic restoration illusion. All of these examples show that inveterate knowledge helps the top-down process of information.

Many studies have examined how timbre properties contribute to sound source recognition. From spectral factors, people can recognize the shapes, the sizes and the properties of percussion instruments (Lakatos et al., 1997). Spectral and temporal factors taken together can yield the gestures of how an instrument is made to vibrate such as the angle and position of plucking guitar strings (Traube, Depalle & Wanderley, 2003). Spectrotemporal factors contribute to recognition of the materials of struck objects (Freed, 1990; McAdams et al., 2004; Giordano & McAdams, 2006). These results indicate that these associations with physical structures and excitations of sound sources are already stored in people's memories; otherwise, they would not be able to recognize properties and features of the sound events. As timbre familiarity is related to long-term memory, the following sections will review the theoretical framework of memory research, gradually connecting aspects of timbre memory with timbre perception and finally presenting hypotheses concerning the effect of timbre memory decay on the mental representation of timbre and the estimation of timbre dissimilarity.

1.3 Framework of Memory Models

Based on empirical evidence, Broadbent's memory model (1958) and the Atkinson-Shiffrin multicomponent model of memory (1968) describe three types of memory stores in a serial order: a large-capacity but brief sensory memory, a small-capacity short-term memory, and a permanent long-term memory. After sensory information is processed by the nervous system, the information is first stored in sensory memory. Compared to short-term memory and long-term memory, information in sensory memory decays the fastest. The capacity of sensory memory in this model is claimed to be unlimited as well. Then, the information goes through an attentional filter where unattended information is filtered out. After that, the information reaches the short-term memory store. After rehearsal in short-term memory, the information goes to the permanent memory store. If the information is not rehearsed, then the information is lost, and people will forget the information. In the Atkinson-Shiffrin model, long-term memory indefinitely holds the information that has been rehearsed in short-term memory. One cannot reactivate long-term memories because one loses the connections, cues or associations to the memories. The long-term memories are still there, but they are just unreachable. Information in long-term memory is retrieved to short-term memory regularly for refreshing the connections and cues in order to keep the long-term memory accessible. In contrast to Broadbent's model, the Atkinson-Shiffrin multi-store model emphasizes voluntary and effortful transfers of information among the memory storages.

Neisser (1967) introduced the term echoic memory, which was the synonym for auditory sensory memory, and according to Nees (2016), recent memory research (Näätänen et al., 1989; Winkler & Cowan, 2005) agrees that echoic memory specifically stores auditory information and decays over time. Echoic memory also seems to store episodic information because the auditory sensory memory is high in resolution. Echoic memory can passively organize acoustic parameters within the time limit of memory decay, which allows people to compare just-heard sounds (Massaro, 1972; Crowder, 1982). This process is independent of attention as well (Näätänen et al., 1989; Winkler & Cowan, 2005).

The Atkinson-Shiffrin multi-store model, however, has received much criticism. First of all, the fixed stages describe a bottom-up memory process but ignore the top-down influence of long-term memory (Broadbent, 1984). Shepard (1967) also found that participants in his experiment were able to remember and recognize around 600 to 700 visual stimuli like words, sentences and pictures of magazine advertisements in a short period of time effortlessly. Rehearsal in short-term memory is not necessarily required in order to store information in long-term memory (Balota,

1983; Dawson & Schell, 1982). The memory process of each store can be parallel as well (Nees & Walker, 2013; Nees, 2016). Therefore, the serial storage described in Atkinson-Shiffrin's model does not completely explain the memory process. Furthermore, the multi-store model is heavily criticized for lacking explanation of the cognitive processes.

Baddeley and Hitch (1974) proposed a multicomponent model of working memory that explained the cognitive process involved in short-term memory in order to overcome one of the weaknesses of Atkinson-Shiffrin's model. Baddeley and Hitch's model consists of three components for a description of manipulation and temporary storage of information in short-term memory: a phonological loop, a visual-spatial sketchpad and a central executive. The phonological loop holds verbal and acoustic information and rehearses them to prevent informational decay. However, this process is interrupted when one speaks something irrelevant out loud, a phenomenon known as articulatory suppression. The visual-sketchpad holds visual and spatial information for temporary information maintenance. The central executive is responsible for various tasks such as coordination of the phonological loop and visual-spatial sketchpad, directing attention to relevant information and integration of the information. Neuroradiological evidence (Baddeley, 2000) suggests that the temporary store and the articulatory rehearsal system of the phonological loop are associated with Brodmann areas 40 and 44, respectively. The visual-spatial sketchpad can be separated into visual, spatial and kinesthetic elements that are associated with the right hemisphere of the brain. The central executive seems to be associated with the frontal lobe areas, but since the central executive has many functions, the association is less understood compared to the two other components. Later, Baddeley (2000) added an extra component, the episodic buffer, to the multicomponent model of working memory in order to include the memory process of binding information and the role of conscious awareness on memory processes. The buffer is episodic in the way that it integrates information from other components and over time. The episodic buffer is responsible for the storage and integration of visual, spatial and phonological information from the central executive and serves as an interface between short-term memory and long-term memory. At the same time, the central executive can retrieve this buffer through consciousness.

Both the Atkinson-Shiffrin and Baddeley-Hitch models of working memory are developed on the basis of visual and semantic memory research, so the model does not necessarily account for musical information such as pitch and timbre. Nevertheless, research shows that cognitive processing of non-verbal information is different from that of verbal information. For example, the effect of articulatory suppression is not found in memory of timbres of synthetic sounds (McKeown, Mills & Mercer, 2011; Soemer & Saito, 2015). Consequently, an extended memory model is needed

for describing memory processes of non-verbal information.

According to Cowan (2015), his embedded-process model (1984, 1988) is a further extension of the Atkinson-Shiffrin multi-store model and of the Baddeley-Hitch working memory model. Cowan's model simplifies sensory memory, short-term memory and long-term memory into two phases of sensory storage: a brief afterimage and a recollection of sensation. First of all, the phase of brief afterimage mainly processes sensory information including color, smell, loudness and pitch, whereas the phase of sensation recollection processes sensory information, and more abstract semantic information such as recognized phonemes, shapes and meanings. Second, Cowan's embedded-process model considers the short-term store described in the Atkinson-Shiffrin multi-store model as an activated long-term memory that is capable of processing previously stored and new sensory information. For this reason, the embedded-process model groups short-term and long-term memory as the second phase of sensory storage.

The brief sensory storage, according to Cowan (1984), was the cause of forward and backward masking that lasted several hundred milliseconds. On the other hand, the second phase of the sensory storage, according to Cowan (2015), is a recollection of sensations that allows comparisons such as between two slightly different words in a foreign language or between two slightly different shades of paint. The mismatch negativity (MMN), which is a response of event-related potentials to mismatched components presented in a train of similar stimuli, is useful for examining the duration of the second phase of the sensory storage. During the examination, participants are required to be engaged in a visual task like reading a book or watching a silent movie, and they are asked to ignore a series of identical sounds with some deviant tones. At the same time, electroencephalography (EEG) is used to monitor the electrical activities of the brain. In memory research, delay of the deviant stimuli is presented. When the delay increases, the participants lose the representation of the previous stimuli and are not able to recognize the deviances clearly. As a result, MMN does not occur anymore as the delay increases to a certain point, and this delay time is the estimated duration of the longer sensory storage (Cowan et al., 1993). Besides EEG, a two-tone discrimination task is used to examine the length of the second phase. In this task, participants are required to listen to a first tone and then a second tone after a retention interval (RI). In this case, the participants are required to determine whether the first and second tones are the same or different, and a decline in task performance can estimate the duration of the second phase. However, memory studies have not reached a common estimation of the duration of the second phase. According to Nees (2016), different studies have estimated several durations for the second phase of sensory memory: less than 2 s (Huron & Parncutt, 1993), 3.5 s (Mcevoy et al., 1997), 4-5 s (Glucksberg & Cowen, 1970;

Darwin et al., 1972), 10 s (Sams et al., 1993), 10-15 s (Winkler & Cowan, 2005), 20 s (Watkins & Todres, 1980; Schoröger, 2007), at least 30 s (Winkler et al., 2002) and 60 s (Engle & Roberts, 1982). Furthermore, Winkler, Schröger and Cowan (2001) found that this time delay was not absolute but relative to the RIs of the standard stimuli. For instance, when the interstimulus interval (ISI) between the standard stimuli is 1 s, and the delay of the deviance is 7 s, there is no occurrence of MMN. On the other hand, when the ISI is 7 s, and the delay is also 7 s, there is occurrence of MMN because, in this situation, the delay is not long enough to prevent the deviance from being grouped with the standard stimuli, compared to the ISI.

Cowan's memory model shares similar components with Baddeley and Hitch's working memory model, which is to keep the central executive, phonological loop and visual-spatial sketchpad as components of working memory. As reviewed by Baddeley (2012), Cowan's embedded-process model places more emphasis on the connection between the episodic buffer and the central executive in Baddeley's model, but uses different terminology than Baddeley, although the two models are very similar. The central executive described in Cowan's memory model can direct attention to activate the recollection of sensations or to the new sensory input for rehearsal of information. The phonological loop and visual-spatial sketchpad are components of the activated sensory and semantic features of the recollection of sensations. These components can rehearse spatial arrangements of tones, as well as pitch and timbre information. Another similar portion between the Baddeley-Hitch and Cowan models is the aspect of conscious awareness. Focus of attention in Cowan's model is considered as a subset of the recollection of sensations. Voluntarily attended sensory information will be rehearsed and encoded more for memory storage. For example, semantic processing does not occur much for unattended information (Conway, Cowan & Bunting, 2001; Cowan & Wood, 1997). This subset is related to the limited capacity of working memory as well. Several studies have shown that auditory working memory is able to hold around three to four items (Chen & Cowan, 2005, 2009; Cowan et al., 2004; Cowan, 2001; Pashler, 1988; Rouder, Morey, Morey & Cowan, 2011). However, according to Baddeley (2012), these studies did not manipulate the stimuli in such as way that participants would not bind the items into chunks or episodes. As a result, the actual capacity might be smaller.

Both echoic memory and working memory are parts of the recollection of sensations, but they have different effects on the memory process. For instance, in serial recall tasks, participants are required to hear a list of stimuli like words and digits and then report what they heard. The active rehearsal of working memory helps the participants to recall the first few items, which is known as the primacy effect. The echoic sensory memory trace, on the other hand, contributes

to the recall of the last few items, which is known as auditory recency (Jones et al., 2004, 2006; Nees, 2016). The primacy effect and auditory recency can occur simultaneously in the second phase described in Cowan's memory model (Nees & Walker, 2013). Furthermore, according to Nees (2016), cognitive neuroscience research (Buchsbaum et al., 2005) provides evidence for the parallel processing of echoic memory and working memory as well. As mentioned above, MMN is a neurological marker for identifying the process of echoic memory, whereas oscillations within the alpha range of frequencies recorded by EEG are associated with the active rehearsal of auditory working memory (Lim et al., 2015). The simultaneous appearance of these neurological markers supports the parallel processing of echoic memory and auditory working memory.

The parallel processing of these two memory types raises issues for estimating the duration of working memory presented in the second phase of the sensory storage described in Cowan's memory model. According to Nees (2016), experiments that try to estimate the duration of the second phase do not usually effectively separate the effect of active maintenance from the effect of echoic memory. Consequently, different experiments have estimated different durations of working memory. Also, this parallel process explains various estimations of the duration of the second phase. Experiments like the one conducted by Glucksberg and Cowen (1970) did not completely prevent the attentional process of active maintenance from prolonging the memory decay. For this reason, estimation of the second phase of the sensory storage varies in different studies.

1.4 Features of Timbre Memory

Active maintenance of synthetic timbre information involves imagery rehearsal, or attentional refreshing as termed by Siedenburg and McAdams (2016). When attentional refreshing is prevented, memory for timbres of synthetic sounds decays faster (McKeown, Mills & Mercer, 2011; Soemer & Saito, 2015; Siedenburg & McAdams, 2016). However, active maintenance of memory for timbres of synthetic sounds does not involve verbal rehearsal because there is no effect of articulatory suppression on it (McKeown, Mills & Mercer, 2011; Soemer & Saito, 2015). In order to examine decay characteristics of memory for timbres of synthetic sounds, Demany et al. (2008) used a change detection task. In this task, participants were required to listen to two complex tones constructed by several pure tones and detect whether there were frequency shifts of one of the pure tones between two given complex tones. The result showed that between 0 ms and 750 ms, timbre memory decayed rapidly, whereas between 750 ms and 2 s, the decay slowed down. Also, working memory capacity for storing synthetic timbre information is suggested as around 1 to

2 items (Golubock & Janata, 2013), which is smaller than the capacity of working memory for sensory information suggested by several memory studies. The smaller capacity of timbre memory indicates that larger storage might be required for processing the multidimensional properties of timbre.

Timbre familiarity allows timbre to serve as an indicator for source recognition, categorization and identification. Memory of an unfamiliar timbre does not benefit from verbal labeling. Research mentioned in the previous paragraph mainly used synthetic tones as unfamiliar timbres because verbally labeling the stimuli as synthesized tones did not provide for effective active maintenance of the timbre information. Therefore, the research could eliminate the effect of verbal labeling and examine imagery rehearsal. In contrast, active maintenance of familiar timbre information can involve both verbal labeling and attentional refreshing due to long-term familiarity and categorization. Through serial recall tasks, Siedenburg and McAdams (2016) found that articulatory suppression influenced recognition of timbre information that participants were familiar with and were able to categorize or associate with verbal labels. As attentional refreshing and verbal labeling are allowed for active maintenance, timbre memory would intuitively last longer.

In cognitive processes like grouping and streaming, there are interaction effects between timbre and pitch. For memory process, things are quite complicated. Based on the general principle that similar acoustic features would interfere with each other (Cowan, 2015), Semal and Demany (1991, 1993) reported that the process of pitch working memory was independent from intensity and dynamic aspects of timbre using Deutsch's (1970, 1972) interpolated tones paradigm. In this paradigm, participants are required to compare two tones separated by various RIs, and within the RIs, there is at least one unrelated tone. Participants are required to ignore the interpolated tones and determine whether the two tones are the same or different. With a similar paradigm, Starr and Pitt (1997) reported that timbre memory is independent of pitch. However, if subjects are required to perform a timbre discrimination task on tones that combine pitch and timbre information, when the pitch difference between stimuli increases, timbre discrimination performance gets worse (Melara & Marks, 1990; Krumhansl & Iverson, 1992; Pitt, 1994; Caruso & Balaban, 2014). Allen and Oxenham (2014) showed that with a discrimination paradigm, the interaction effect between pitch and timbre could be observed. The reason why different studies had different results is due to different experimental designs. The interpolated-tones tasks require listeners to resist and filter out unrelated information, whereas discrimination tasks require subjects to separate pitch information and timbre information from each other in a single tone in order to focus judgments on the criterial dimension. Because the two tasks require different processes, the studies reach

different conclusions.

There is an ongoing debate on whether the cause of informational loss is due to memory limit in duration or informational interference. Although one of the positions claims that the information held in memory fades away naturally, another position claims that the information remains in memory storage until it is replaced by other new external or internal information (Broadbent, 1957; Cowan & AuBuchon, 2008; Cowan, Saults & Nugent, 1997; Lewandowsky & Oberauer, 2009; McKeown, Mills & Mercer, 2012; Neath & Nairne, 1995; Schweickert & Boruff, 1986). No matter what the cause of the informational loss is, this study will focus solely on the effect of RI on the perception of timbre dissimilarity and the role of long-term timbre familiarity, which contributes to internal timbre categorization and sound source recognition and identification. This study will not be able to provide any evidence for either position of the ongoing debate on informational loss.

1.5 Hypotheses of the Study

Most of the previous research on timbre memory has focused on how the decay of timbre memory affects timbre discrimination. As timbre is a multidimensional set of auditory attributes, it would also be interesting to examine the effect of timbre memory decay on its mental representation. Based on previous findings, the author hypothesizes that when timbre information starts to decay, timbre dissimilarity ratings will change accordingly. Nonetheless, there is not enough evidence to determine the way in which participants will perceive timbres as more or less dissimilar as a function of the time delay between them. Therefore, one of the purposes of this study is to observe the tendency of this perceptual change. As timbre familiarity allows verbal rehearsal of information, the second hypothesis of this study is that timbre familiarity will minimize the effect of timbre informational decay on timbre dissimilarity ratings.

Chapter 2

Experiment 1—Effects of Retention Interval on Timbre Dissimilarity Ratings with Synthesized Sounds

This experiment combined the paradigms of a timbre dissimilarity experiment (Caclin et al., 2005) and a two-tone discrimination experiment (Deutsch, 1972; Cowan, 1984; Semal & Demany, 1991; Semal & Demany, 1993; Starr & Pitt, 1997) in order to test the hypothesis that when timbre information starts to decay, timbre dissimilarity ratings will change accordingly when the RI between the to-be-compared tones increases. Given that previous studies consistently observed that AT and SCG were perceptually salient dimensions for timbre dissimilarity (Grey & Gordon, 1978; Krumhansl, 1989; Krimphoff et al., 1994; McAdams et al., 1995; Marozeau et al., 2003; Caclin et al., 2005), this experiment used these two dimensions to examine how the decay of timbre in working memory might change the perception of timbre dissimilarity.

2.1 Method

2.1.1 Participants

There were 27 participants (18 females and 9 males) in the experiment with an average age of 25 years (SD = 6.23, range: 19-49 years old). All participants were recruited from either a mailing list of the Schulich School of Music of McGill University or web-based McGill classified advertisements. All of them reported normal hearing, and this was confirmed by a standard pure-tone audiogram

just before the experiment, which examined their hearing thresholds at octave-spaced frequencies from 125 to 8000 Hz. All participants had hearing thresholds of 20 dB HL or better (ISO 398-8, 2004; Martin & Champlin, 2000).

2.1.2 Stimuli

The stimuli used in the experiment were harmonic complexes with a fundamental frequency of 311 Hz (Eb4). They were generated in Pure Data (Pd-extended version 0.43.4, http://puredata.info) with additive synthesis. They all contained 40 harmonics for manipulation of SCG in harmonic rank:

$$SCG = \frac{\sum_{n=1}^{40} n \times A_n}{\sum_{n=1}^{40} A_n}$$

where A_n is the amplitude of the nth harmonic in decibels (dB). The amplitudes of the 40 harmonics conformed to a linearly decreasing slope of the spectral envelope in dB. In order to maintain the same loudness for all stimuli, a level correction was applied:

$$\Delta dB = -6.35 \ln(SCG) + 4.76$$

This correction was based on the loudness matching results from Wood's (2015) experiment whose stimuli were generated using the same method. The amplitude envelope of the stimuli comprised a linear attack portion, a steady-state plateau, and an exponential decay. In order to maintain the same perceived duration for all the stimuli, the durations of the plateaus were adjusted according to the following formula (Caclin et al., 2005):

$$t2 = 412 - 0.8 \times t1$$

where $t2$ is the duration of the plateau in ms, and $t1$ is the AT of the stimuli in ms. After the plateau, there was a 100-ms exponential decay to -60 dB relative to the maximum of the amplitude envelope. Consequently, all of the stimuli had the same loudness and same perceived duration. Within each synthetic tone, the SCG did not vary over time. Only the AT and SCG of the stimuli were varied.

The ranges of AT and SCG were chosen from those used in Caclin et al. (2005). In their study, the range of AT was from 15 ms to 199 ms, whereas the range of SCG was from 3.0 to 4.5 in harmonic rank. As a result, a two-dimensional space with axes of AT and SCG was constructed.

When the ranges of AT and SCG were trisected, the two-dimensional space was equally divided into nine regions, and the centers of each region were the main points shown with stars in Figure 2.1. However, contrary to Caclin et al.'s (2005) study, our experiment required participants to memorize stimuli for dissimilarity ratings. It would be undesirable if participants could strategically label the stimuli rather than trying to memorize the stimuli in each trial. In order to prevent this, the stimuli were set to roam within one of the nine regions, and the resulting positions were uniformly distributed, as shown in Figure 2.1. Also, there were gaps separating the nine roaming regions in order to prevent the stimuli from intruding into other roaming regions. The width of the gaps along each dimension was equal to the width of a roaming region along the same dimension, as shown in Figure 2.1. The stimuli in this experiment were considered as maximally unfamiliar to the participants.

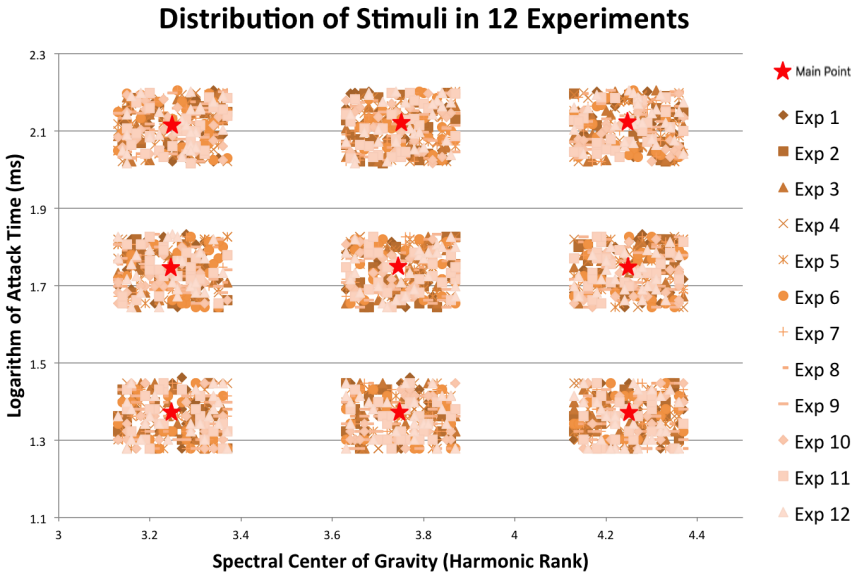

Fig. 2.1 Random distribution of stimuli in 12 simulated experiments. The two-dimensional space is constructed by SCG in harmonic rank and AT in ms. The stars indicate the centers of the roaming regions. Stimuli are uniformly distributed within the regions, and the roaming regions do not overlap with other roaming regions.

2.1.3 Procedure

At the beginning of the experiment, 18 tones distributed across the roaming regions were picked and played. The purpose was to let participants know how the stimuli would sound, so that they

got a sense of relative differences among the stimuli. As shown in Figure 2.2, there were three blocks in the experiment. In each trial, participants were required to listen to two stimuli separated by an RI of 0.5 s, 5 s or 10 s, with RIs presented in three separate blocks. The first stimulus in each trial was picked from one of the nine roaming regions, and the second one was picked from other regions or within the same region so that acoustic distances of SCG and AT between the two tones were always in 0, 1 or 2 units, keeping the acoustic distance between points constant in linear harmonic rank and log attack time. All combinations of roaming areas appeared only once in each block. Therefore, there were 45 trials in each block, including nine identical pairs. In each trial, the two stimuli were presented once, and participants used a standard mouse to interact with the interface given in Figure 2.3 in order to rate the dissimilarity between the two tones by clicking on the scale with an "identical" label on the left hand side (coded as 0) and a "very dissimilar" label on the right hand side (coded as 1). After they clicked on the scale, a yellow cursor would appear. Then, they could drag the cursor for adjustment of the rating. The order of blocks and that of trials in each block was randomized.

There were four practice trials before each block. In these practice trials, four pairs of stimuli were specifically picked: a pair of stimuli with the same AT and SCG (labeled as A0C0), a pair with two-unit (maximum) differences of AT and SCG (A2C2), a pair with a two-unit difference of AT but the same SCG (A2C0), and a pair of stimuli with the same AT but a two-unit difference of SCG (A0C2). Participants were not instructed that stimuli in the practice trials were picked specifically because the purpose was to help them to familiarize themselves with the task of that block. Although results of the practice trials were not recorded, participants were required to finish all of them, and the order of the practice trials was randomized. During the practice trials, participants could replay the stimuli as many times as they wanted, but in the real trials, the stimuli were played once only. The number of times people played the stimuli in the practice trials, however, was not recorded. As the stimuli in the practice trials were also roamed, replaying the stimuli for many times would not help the participants to familiarize themselves with the stimuli. In order to avoid participants not paying attention when stimulus presentation began, there was a 0.5-s visual cue starting 0.5 s before the presentation of the first and second stimuli in each trial. As shown in Figure 2.3, there were two visual cues above the scale that blinked before the presentation of tones. The left one cued the presentation of the first tone, and the right one cued the second tone.

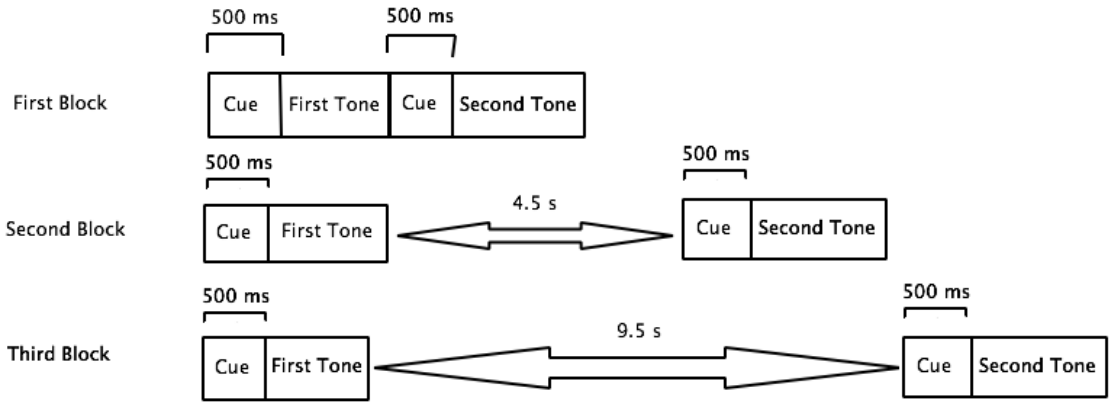

Fig. 2.2 There were three blocks in the whole experiment. RIs in the three blocks were 0.5 s, 5 s and 10 s, respectively. There was a 0.5-s visual cue just before the presentation of each tone

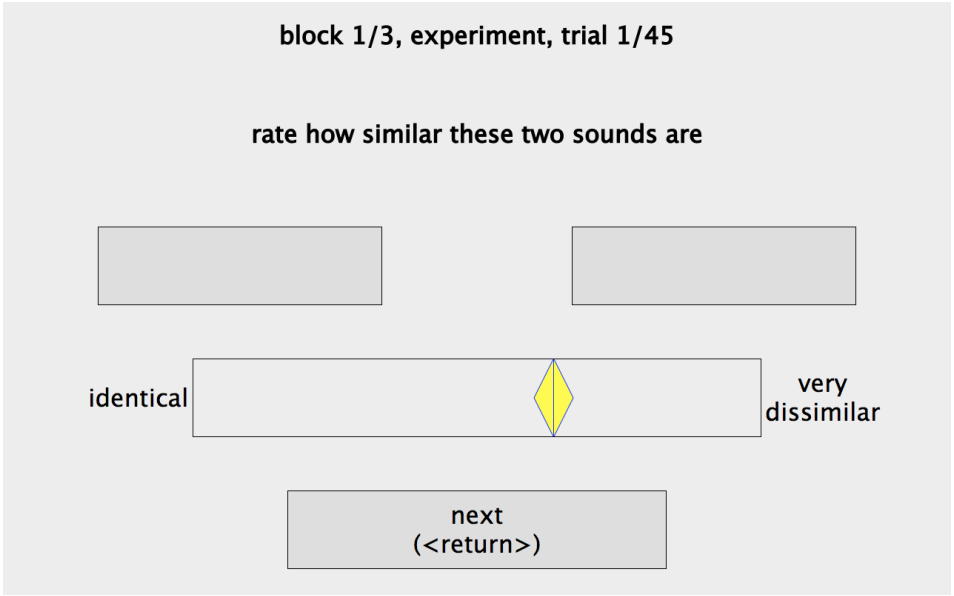

Fig. 2.3 The interface used in the experiment. The left and right rectangles above the scale were illuminated 0.5-s prior to presentation of the first and second tones in a trial, respectively.

2.1.4 Apparatus

The experiment was carried out in a double-walled sound-isolation chamber (IAC Acoustics, Bronx, NY). Participants used a Mac Pro 5 computer running OS 10.6.8 (Apple Computer, Inc, Cupertino, CA) to do the experiment. All stimuli were synthesized with Pure Data (Pd-extended version 0.43.4). The audio signals were amplified through a Grace Design m904 monitor (Grace Digital Audio, San Diego, CA). The signals were transmitted through Sennheiser HD280 Pro earphones (Sennheiser Electronic GmbH, Wedemark, Germany). Sound pressure output levels of all possible stimuli ranged from 62.7 to 64.8 dB SPL as measured with a Brüel & Kjær Type 2205 sound-level meter with the setting of Level A-weighting equalized (LAeq) and Brüel & Kjær Type 4153 Artificial Ear (Nærum, Denmark) to which the headphone was coupled. The interface for the experiment was created with the PsiExp computer environment (Smith, 1995).

2.2 Result

In each block, each category of the acoustic intervals contained two or more pairs. There were nine identical pairs (A0C0), six A1C0 pairs, six A0C1 pairs, eight A1C1 pairs, three A2C0 pairs, three A0C2 pairs, four A2C1 pairs, four A1C2 pairs and two A2C2 pairs. Therefore, mean of the ratings by each subject was calculated for each category, and these means constituted the dependent variable. Mean ratings across the 27 participants are shown in Figure 2.4. The data were analyzed with a repeated-measures ANOVA design: RI (3) x AT Intervals (3) x SCG Intervals (3). Mauchly's sphericity test is first used to validate the assumption that the variances of the differences between all combinations of the conditions, which are RI, AT Intervals and SCG Intervals in this analysis, are equal. If the assumption of sphericity is violated, corrections to the number of degrees of freedom were made with either the Greenhouse-Geisser (GG) or Huynh-Feldt (HF) epsilon. When the GG epsilon was less than 0.75, it was used. Otherwise, the HF epsilon was used (Girden, 1992).

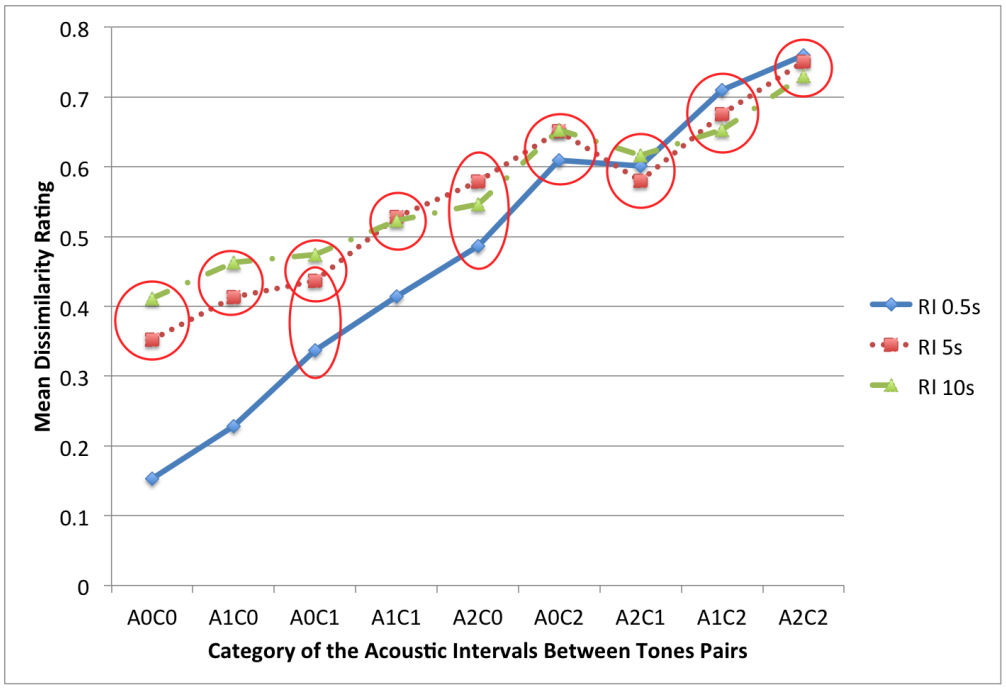

Fig. 2.4 Mean dissimilarity ratings by 27 subjects as a function of the category of the acoustic intervals between the pairs of tones. Different lines represent blocks with different RIs. Circled mean dissimilarity ratings are not significantly different according to post-hoc comparisons with Bonferroni-Holm correction.

Mauchly's sphericity test revealed violations for AT intervals, $\chi^2(2) = 12.2$, $p = 0.002$, SCG intervals, $\chi^2(2) = 6.27$, $p = 0.04$, the interaction between AT intervals and SCG intervals, $\chi^2(9) = 20.2$, $p = 0.017$, and the interaction between RI, AT intervals and SCG intervals, $\chi^2(35) = 52.1$, $p = 0.034$.

The repeated-measures ANOVA yielded significant main effects of RI, $F(2, 52) = 8.22$, $p = .001$, $\eta_p^2 = .24$, AT intervals, $F(1.44, 37.5) = 45.3$, $p < .001$, $\epsilon = .72$, $\eta_p^2 = .64$, and SCG intervals, $F(1.73,45.0) = 166.5$, $p < .001$, $\epsilon = .87$, $\eta_p^2 = .87$, on the participants' timbre dissimilarity ratings. However, AT intervals, SCG intervals and RI were not completely independent from each other because there were also significant interactions between RI and AT intervals, $F(4,104) = 4.81$, $p = .001$, $\eta_p^2 = .16$, RI and SCG intervals, $F(4,104) = 12.2$, $p < .001$, $\eta_p^2 = .32$, and AT intervals and SCG intervals, $F(2.92,75.9) = 6.92$, $p < .001$, $\epsilon = .73$, $\eta_p^2 = .21$. In addition, for the interaction between RI, AT intervals and SCG intervals, there was marginal effect, $F(5.13,133) = 1.95$, $p = .089$, $\epsilon = .64$, $\eta_p^2 = .07$. These interactions indicated that the strength of the RI effect on dissimilarity ratings was marginally related to the acoustic intervals of sound pairs including AT intervals and SCG

intervals.

There was no difference of overall mean dissimilarity ratings between the 5-s and 10-s RIs because the post hoc test with Bonferroni-Holm (B-H) correction did not yield a significant difference between 5-s and 10-s RIs ($p = 0.53$). In contrast, the post hoc test yielded a significant difference between 0.5-s and 5-s RIs ($p = 0.006$), as well as between 0.5-s and 10-s RI ($p = 0.01$). This result showed that the RI effect on the perception of synthetic timbre dissimilarity reached its maximum within 5 s. Simultaneously, as there were significant interactions between RI and the timbre intervals, Figure 2.4 showed that the differences of mean dissimilarity ratings between 0.5-s and 5-s RIs were not equal across the categories of the acoustic intervals. A pairwise sample-t test with B-H correction yielded no rating difference between 0.5-s, 5-s and 10-s RIs for the categories of the larger acoustic intervals such as A2C0, A0C2, A2C1, A1C2 and A2C2, whereas there were significant rating differences for the smaller intervals like A0C0 between 0.5-s and 5-s RIs ($p < .001$) and between 0.5 s and 10 s ($p < .001$), A1C0 between 0.5 s and 5 s ($p < .001$) and between 0.5 s and 10 s ($p < .001$), and A1C1 between 0.5 s and 5 s ($p = .006$) and between 0.5 s and 10 s ($p = .033$). A0C1 was an outliner as a small timbre interval because the post hoc test yielded a significant difference between 0.5 s and 10 s ($p = .028$), but not between 0.5 s and 5 s ($p = .28$). The result suggested that the strength of the RI effect on the perception of timbre dissimilarity depended on the distance of the timbre intervals.

2.3 Discussion

It was assumed that participants in the experiment could easily identify identical pairs with 0.5-s RI as well, but less than half of the participants (10 out of 27) rated all the identical sound pairs with 0.5-s RI under 0.1 on average on a scale from 0 to 1. This situation did not occur in previous timbre dissimilarity studies such as Caclin et al. (2005). One possible reason for nearly two-thirds of the participants overestimating the dissimilarity of identical pairs at the shortest RI is that in the previous study, participants could play as many times as they needed, whereas participants in this experiment could play each pair only once, and the synthetic sounds were unfamiliar to them. When participants can play the stimuli once only, overestimating the dissimilarity of identical pairs at the shortest RI suggests a rapid decay of memory for synthetic sounds. All the results in this experiment were kept for data analysis because Demany et al. (2008) have shown that timbre working memory decay can occur as rapidly as 0.5 s.

Increases of both AT and SCG intervals had a significant effect on the dissimilarity ratings. The

result agreed with previous studies (Grey & Gordon, 1978; Krumhansl, 1989; Krimphoff et al., 1994; McAdams et al., 1995; Marozeau et al., 2003; Caclin et al., 2005) suggesting that AT and SCG are salient perceptual attributes of timbre. At the same time, the significant interaction between AT and SCG demonstrates that the two perceptual attributes were not completely independent of each other. Second, when the acoustic intervals of SCG or AT within the pairs were small or zero, participants rated the pairs as more dissimilar in timbre when the second tone was presented with a delay. The RI effect on the perception of timbre dissimilarity, however, decreased as the acoustic distance between sounds increased, and there was no rating difference between 0.5-s, 5-s and 10-s RIs when the timbre intervals became relatively large (see Fig. 2.5). This observation suggests that as timbre working memory of a first synthetic tone decays, its mental representation gets noisier in such a way that acoustic intervals of identical or similar synthetic tone pairs are overestimated, whereas the noisiness of the timbre representation of a first synthetic tone has less effect on the perception of more dissimilar synthetic tone pairs. In other words, if the acoustic distance of a synthetic tone pair is not larger than the variability of the representation of a first tone induced by timbre memory decay, the distance will often be overestimated.

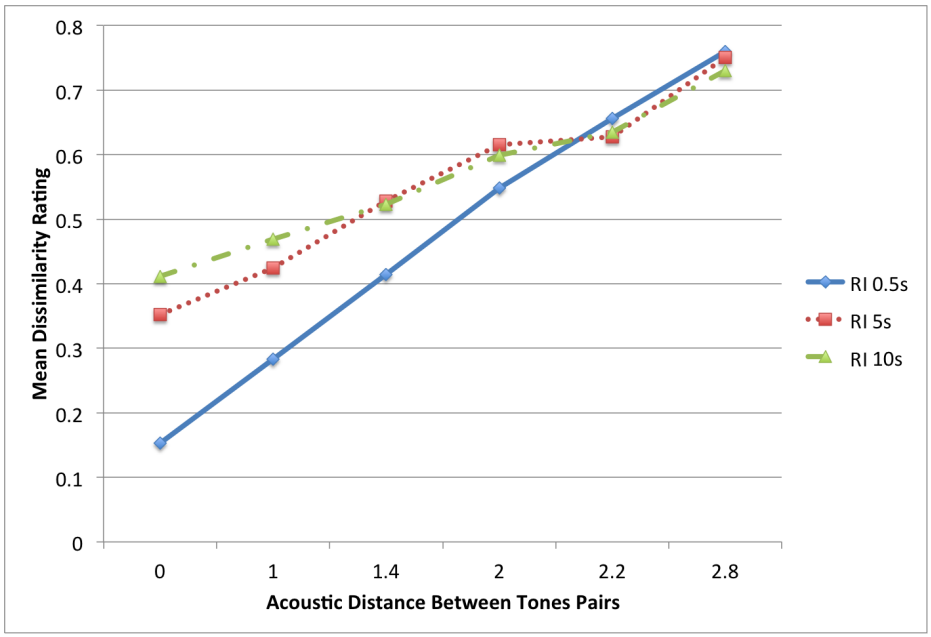

Fig. 2.5 Mean dissimilarity ratings by 27 subjects as a function of the acoustic distance between the pairs of tones. Different lines represent blocks with different RIs.

The result showed that there was no significant difference between dissimilarity ratings of the sound pairs with 5-s and 10-s RIs, so the effect of RI on the dissimilarity of synthetic sounds

appears to have reached its maximum within 5 s. At the same time, the fact that nearly two-thirds of the participants in the experiment did not rate the identical pairs properly indicates that the decay in memory of timbre had already begun by 0.5 s, which is consistent with a previous study by Demany et al. (2008). Therefore, the second experiment is designed to examine the RI effect on the perception of timbre dissimilarity between 0.5 s and 5 s. Furthermore, in addition to the synthetic tones, the next experiment also introduces digital transformations of recorded acoustic instrument sounds and recorded acoustic instruments in order to examine decay of timbre representation in terms of timbre familiarity.

Chapter 3

Experiment 2—Effects of Retention Interval and Familiarity of Sounds

The first experiment showed that the effect of RI on timbre dissimilarity ratings reached its maximum within 5 s, and the effect might already occur within 0.5 s. The purpose of the second experiment is to examine in detail the curve of the RI effect between 0.5 s and 5 s. Also, the first experiment only examined memory for the timbre synthetic sounds. As previous research has shown that articulatory suppression influenced recognition of timbre information that participants were familiar with and were able to categorize or associate with verbal labels (Siedenburg & McAdams, 2016), Experiment 2 examines how timbre familiarity interacts with timbre memory decay on the perception of timbre dissimilarity. The second experiment uses synthetic tones, digital transformations of recorded acoustic instrument sounds and recorded acoustic instruments as stimuli of low, moderate, and high timbre familiarity, respectively. In addition, compared to synthetic tones, digital transformations of recorded acoustic instruments sounds and recorded acoustic instruments had higher timbre complexity. The second hypothesis, that timbre familiarity will minimize the effect of timbre informational decay on dissimilarity ratings, is tested by comparing the memory decay curves of the three categories between 0.5 s and 5 s.

3.1 Method

3.1.1 Participants

Twenty participants (4 males and 16 females) with an average age of 25 years (SD = 4.52, range: 18-35 years old) were recruited from either a mailing list of the Schulich School of Music at McGill University or web-based McGill classified advertisements. Five of them reported that they considered themselves as professional musicians, seven of them considered themselves as amateur musicians, and eight of them considered themselves as non-musicians. Among the non-musicians, seven out of eight considered themselves as music lovers and reported that they actively listened to western music like pop, jazz and classical music. One non-musician did not consider herself a music lover, but she reported that she usually listened to pop, jazz or classical music while working. Similar to the first experiment, all of the participants reported normal hearing, and this was confirmed by a standard pure-tone audiogram just before the experiment as in Experiment 1.

3.1.2 Stimuli

There were three sets of stimuli in the experiment: synthetic tones, recorded acoustic instruments and acoustically transformed sounds. All synthetic tones were generated in Pure Data through additive synthesis, using the same process as in Experiment 1. Experiment 2 used five recorded acoustic instruments as stimuli of high timbre familiarity and five acoustically transformed sounds as stimuli of moderate timbre familiarity. Therefore, for the synthetic tones, instead of nine roaming regions of the two-dimensional spaces constructed by AT and SCG, as shown in Figure 3.1, five roaming regions were picked so that there were five sound pairs of A0C0, two pairs of A0C2, two pairs of A2C0, four pairs of A1C1 and two pairs of A2C2 as stimuli of low timbre familiarity. The five regions did not overlap with each other, and the width of the gaps along each dimension was equal to the width of a roaming region along the same dimension. No familiarity rating was obtained for the synthetic tones, but it was assumed that they were maximally unfamiliar.

Recorded acoustic instruments and acoustically transformed sounds were chosen from Siedenburg et al.'s (2016) study. Their study examined timbre familiarity ratings and timbre dissimilarity ratings of recorded acoustic instruments and acoustically transformed sounds. One of the two experiments in their study required their participants to identify and give familiarity ratings (from 1 for highly unfamiliar to 5 for highly familiar) of the acoustic instruments and the transformed sounds. Another experiment required their participants to give timbre dissimilarity ratings of the

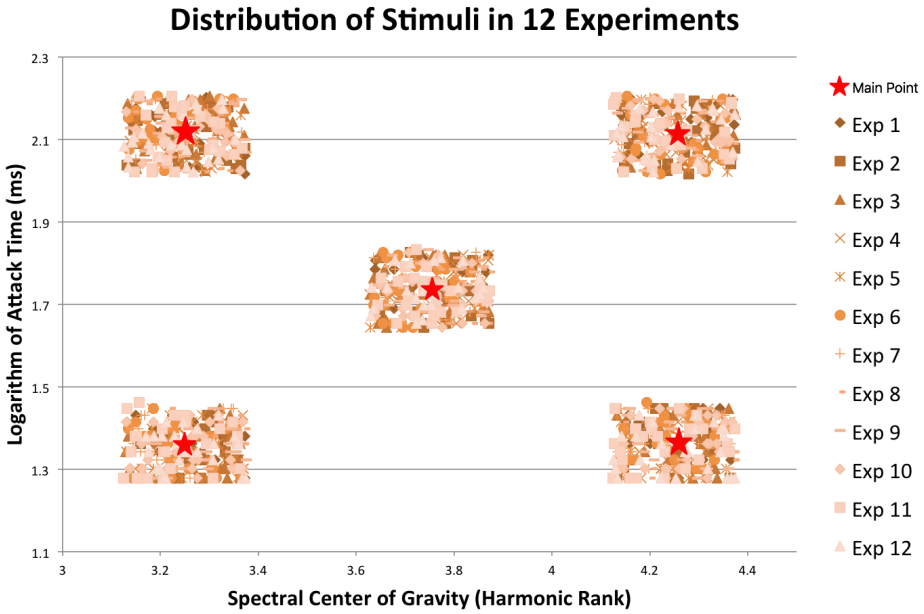

Fig. 3.1 Five of the nine roaming regions in the first experiment were selected for the second experiment.

sets of the acoustic instruments and transformed sounds, and the mixed set of both. Based on familiarity ratings and timbre dissimilarity ratings in their two experiments, five acoustic instruments and five transformed sounds were picked for this second experiment. In general, the familiarity ratings of the five acoustic instruments were higher than the five digitally transformed sounds. The dissimilarity ratings of the sound pairs of the five acoustic instruments and the five transformed sounds ranged from identical to very dissimilar. Compared to synthetic tone pairs, the acoustic instrument pairs and the transformed sound pairs sounded more dissimilar overall.

The chosen acoustic instruments were the bass clarinet (BCL), the flute (FLT), the harp (HRP), the piano (PNO) and the vibraphone (VIB). Their mean familiarity ratings in the original study were between 4.1 and 4.3 on a scale of 5. According to Siedenburg et al. (2016), the piano stimulus was a recorded sample from Logic Professional 7. Other instruments were taken from Vienna Symphonic Library (http://vsl.co.at, last accessed December 17, 2016). The original sample files were in stereo, and their experiment took the left channel of the files only so that the stimuli were in mono. The sampling rate of the files was 44.1 kHz, and the fundamental frequency was 311 Hz (Eb4). The duration of these five stimuli was 500 ms with 20 ms of barely noticeable fadeout. The peak amplitudes of the five stimuli were normalized as well.

The original study described acoustically transformed sounds as "chimeric" because the trans-

formation method was quite similar to source-filter synthesis and the perceptual transformation of speech signals used by Smith et al (2002). First of all, besides the five stimuli selected for this second experiment, there were nine more instruments used in the original study: bassoon, harpsichord, horn, marimba, trumpet, bowed violoncello, pizzicato violoncello, bowed violin, and pizzicato violin. All of these instruments were also from Vienna Symphonic Library, except for the harpsichord, which was from Logic Professional 7. These 14 samples of acoustic instruments served as either source or filter. For instance, if the piano was picked as the source and the trumpet as the filter, the two audio signals would be decomposed by a linear 24-band Gammatone-filterbank (Patterson et al., 1992) provided in the MIRtoolbox (Lartillot & Toiviainen, 2007) in MATLAB version R2013a (The MathWorks, Inc., Natick, MA). Amplitude envelopes of the piano and the trumpet in each filterband were extracted through low-pass filtering and half-wave rectification (Lartillot & Toiviainen, 2007). The value of the amplitude envelopes of the trumpet was then used to normalize the envelopes of the piano so that the fine structure of the piano could be obtained. Finally, the fine structure of the piano was point-wise multiplied with the time-varying envelope magnitudes of the trumpet. As a result, the product contained the fine structure of the source and the spectrotemporal envelope of the filter.

Besides the acoustic instruments, the original study generated less perceptually familiar sounds through a process called filterband scrambling, and these sounds were also served as the source and the filter in the original study. In this process, 24 Gammatone filterbands of the 14 acoustic instruments were generated through decomposition (Patterson et al., 1992). Then, four of the 14 acoustic instruments were randomly chosen, and six different filterbands of each of the four instruments were distributed in such a way that the 24 filterbands could form a new filterband-scrambled sound. Six of these scrambled sounds (FBS1~6) were picked for the acoustic transformation of stimuli. FBS1 and FBS2 had slow ATs, FBS3 and FBS4 had fast ATs, and FBS5 and FBS6 had medium ATs.

Based on the familiarity and dissimilarity ratings of the transformed sounds in the original study, five stimuli were selected for the second experiment: the horn as the source and FBS6 as the filter (HRN-FBS6), the violoncello as the source and the vibraphone as the filter (VCE-VBS), the flute as the source and FBS1 as the filter (FLT-FBS1), FBS2 as the source and the violoncello as the filter (FBS2-VCE), and FBS3 as the source and FBS2 as the filter (FBS3-FBS2). Their familiarity ratings were between 2.0 and 2.1 on a scale of 5.

3.1.3 Procedure

There were three randomly ordered blocks presenting stimulus sets of synthetic tones, acoustically transformed sounds and recorded acoustic instruments separated by RIs of 0.5 s, 1 s, 2 s, 3.5 s or 5 s. Similar to the first experiment, the first synthetic tone in each trial of the synthetic tone block was picked from one of the five roaming regions, and the second one was picked from other regions or within the same region so that the acoustic distance of the timbre pair was A0C0, A1C1, A2C0, A0C2 or A2C2. In the transformed sound and acoustic instrument blocks, two tones were randomly picked from the five sounds of the corresponding set. The combinations of the stimuli and RIs appeared only once in the three blocks. The experimental instructions given to participants were identical to those of Experiment 1. As the stimuli were played only once in all the blocks, the same visual cues in the first experiment were presented 0.5 s before each tone. Because participants needed to rate the relative differences among pairs in each block, at the beginning of each block, all tones presented in that block were played in random order to give participants a sense of the range of dissimilarities.

Before the real trials, there were four practice trials in each block. In the synthetic tone block, the practice trials were drawn from the sound pairs with the acoustic distances of A0C0, A0C2, A2C0 and A2C2. Similar to Experiment 1, the parameters of the synthetic tone stimuli were roamed in the practice trials as well. In the transformed sound and acoustic instrument blocks, the four practice trials contained four conditions: an identical pair, a slightly dissimilar pair, a dissimilar pair and a very dissimilar pair. Based on the timbre dissimilarity ratings of the original study, the transformed sound pairs and the acoustic instrument pairs were assigned to one of the four categories. Tables 3.1 and 3.2 show the assigned categories of the transformed sound pairs and the acoustic instrument pairs, respectively. The four pairs were randomly picked from the four categories. The RI in all the practice trials in the three blocks was 0.5 s. Participants were not instructed that the stimuli in the practice trials were picked specifically. The results of the practice trials were not recorded, and their order was randomized. During the practice trials, participants could also replay the stimuli as many times as they wanted, and they were required to finish all of them. The number of times people played the stimuli in the practice trials was not recorded.

3.1.4 Apparatus

The apparatus was identical to that in Experiment 1. Sound pressure levels of synthetic, transformed and acoustic sounds ranged from 62.7 to 64.8 dB SPL, from 57.0 to 63.0 dB SPL and from 59.0 to

Table 3.1 Dissimilarity categories for each transformed sound pair. Mean dissimilarities are from Siedenburg et al. (2016) (scale of 1-7).

Pairs	Category	Mean Dissimilarity
VCE-VIB and FLT-FBS1	Very Dissimilar	6.5
VCE-VIB and HRN-FBS6	Very Dissimilar	6.8
VCE-VIB and FBS2-VCE	Very Dissimilar	7.0
FBS2-VCE and FLT-FBS1	Very Dissimilar	5.7
FBS3-FBS2 and HRN-FBS6	Very Dissimilar	5.0
FBS2-VCE and HRN-FBS6	Very Dissimilar	5.2
VCE-VIB and FBS3-FBS2	Very Dissimilar	6.9
FBS3-FBS2 and FLT-FBS1	Dissimilar	4.3
FBS3-FBS2 and FBS2-VCE	Dissimilar	4.8
FLT-FBS1 and HRN-FBS6	Slightly Dissimilar	3.4
HRN-FBS6 and HRN-FBS6	Identical	1.0
VCE-VIB and VCE-VIB	Identical	1.0
FLT-FBS1 and FLT-FBS1	Identical	1.0
FBS2-VCE and FBS2-VCE	Identical	1.0
FBS3-FBS2 and FBS3-FBS2	Identical	1.0

Table 3.2 Dissimilarity categories for each acoustic instrument pair. Mean dissimilarities are from Siedenburg et al. (2016) (scale of 1-7).

Pairs	Category	Mean Dissimilarity
HRP and VIB	Very Dissimilar	6.8
PNO and BCL	Very Dissimilar	6.1
VIB and BCL	Very Dissimilar	8.0
PNO and VIB	Very Dissimilar	6.0
VIB and FLT	Very Dissimilar	8.5
HRP and FLT	Dissimilar	4.6
PNO and HRP	Dissimilar	4.6
HRP and BCL	Dissimilar	5.3
PNO and FLT	Dissimilar	5.4
FLT and BCL	Slightly Dissimilar	3.1
BCL and BCL	Identical	1.0
FLT and FLT	Identical	1.0
HRP and HRP	Identical	1.0
PNO and PNO	Identical	1.0
VIB and VIB	Identical	1.0

64.0 dB SPL, respectively, as measured with a Brüel & Kjær Type 2205 sound-level meter with the setting of Level A-weighting equalized (LAeq) and Brüel & Kjær Type 4153 Artificial Ear (Nærum, Denmark) to which the headphone was coupled. The interface for this experiment was created with the PsiExp computer environment (Smith, 1995), which was similar to the first experiment as shown in Figure 2.3.

3.2 Results

Mean dissimilarity ratings of the 20 participants are plotted in Figures 3.2-3.4 for the three sound sets.

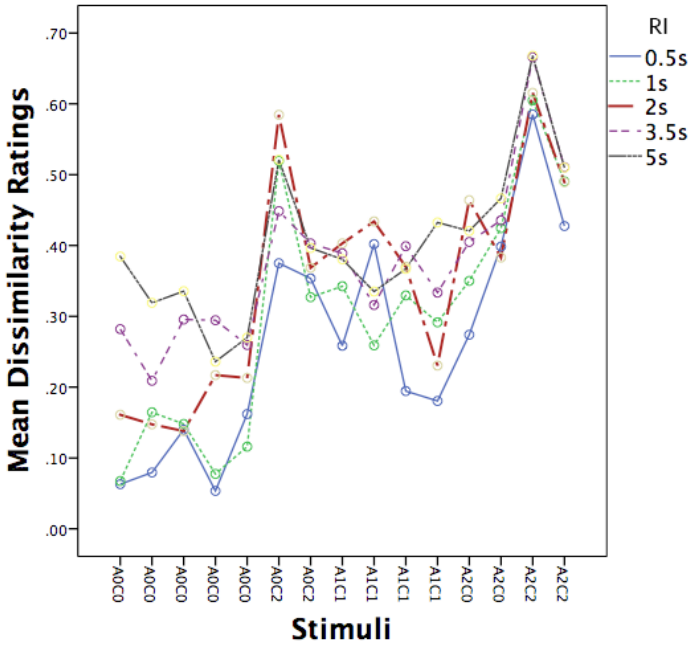

Fig. 3.2 Mean dissimilarity ratings of synthetic tone pairs by 20 subjects as a function of stimulus pair. The lines represent different RIs.

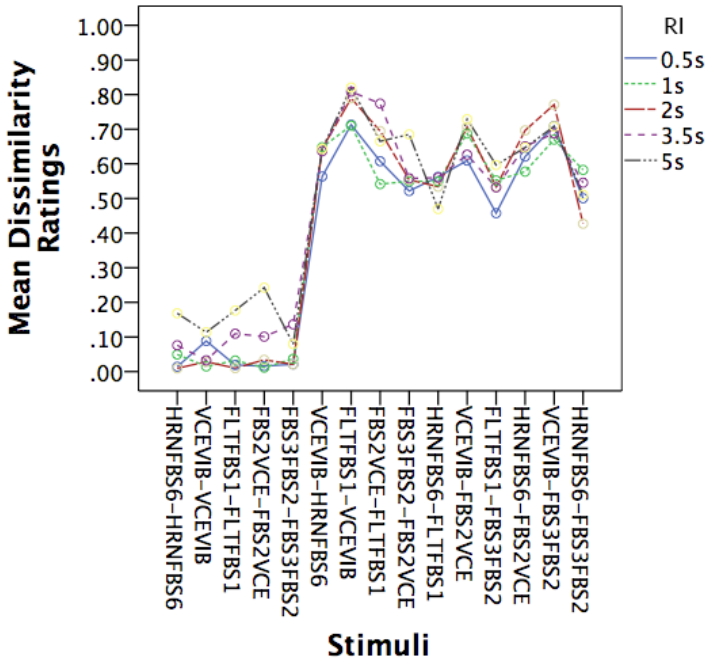

Fig. 3.3 Mean dissimilarity ratings of transformed sound pairs by 20 subjects as a function of stimulus pair. The lines represent different RIs.

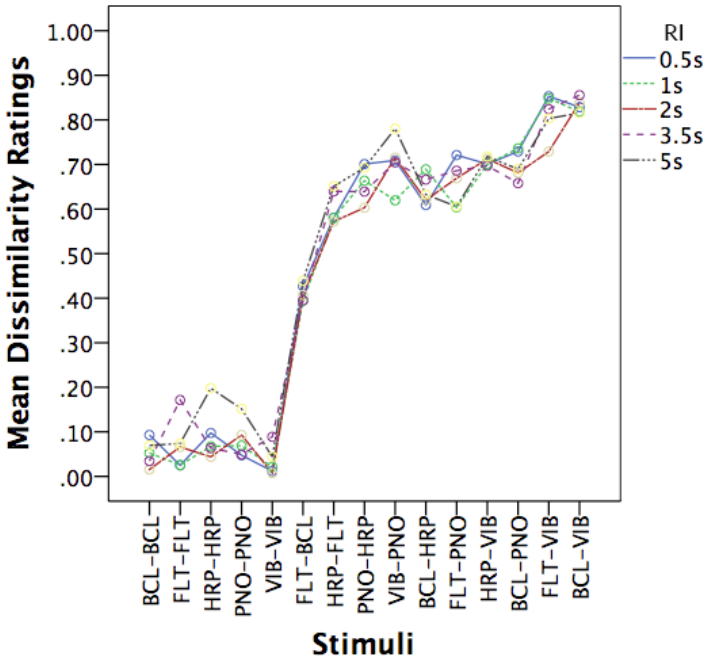

Fig. 3.4 Mean dissimilarity ratings of acoustic instrument pairs by 20 subjects as a function of stimulus pair. The lines represent different RIs.

As stimulus sets were in a nesting structure with stimulus pair within stimulus set within subject, the mean results were analyzed with a three-level hierarchical linear model, as shown in Table 3.3. According to the experimental design, the first level of the maximum model was stimulus pair nested within stimulus set, which included RIs and stimulus pairs as fixed effects of predictors. Also, the first level included a priori distances as covariates, which were calculated by the normalized acoustic distances of the synthetic tone pairs and the normalized dissimilarity ratings of the transformed sounds and the acoustic instruments from Siedenburg et al. (2016). The second level was stimulus set nested within subjects, which included timbre familiarity as a covariate. The third level was the subject level. In addition, the maximum model included random effects for levels 2 and 3, so the remaining variation at level 1 was presented in the estimate of the residual component. Hierarchical linear modeling starts from the simplest null model, which included random effects for levels 2 and 3 only in this analysis. Using estimates of the random effects, the intraclass correlation coefficient (ICC) can be calculated, which indicates how much the levels account for the total variation in dependent variables. If the null model justifies the use of hierarchical linear modeling, fixed effects of predictors and random slopes are inserted to the model, and the model is increased in complexity. Estimates and standard errors of the predictors, covariates and error variances determine statistical significance of the fixed effects and the random effects. Model selection was based on the Aikake Information Criterion (AIC, Aikake, 1973) and the Bayesian Information Criterion (BIC, Schwarz, 1978), which measure the quality of statistical models through the model fit and complexity. Models with smaller values of AIC and BIC are considered better.

First of all, the null model in Table 3.3 (Model 0) justified the use of the three-level hierarchical linear model because there were sufficient variations that could accommodate additional predictors at level 2 ($p < .001$) and level 3 ($p < .001$). Also, the estimate of the residual component suggested that there was a significant remaining variation at level 1 ($p < .001$) for which the predictors at levels 2 and 3 could not account.

Model 1 in Table 3.3 included RI, timbre familiarity and stimulus pair as predictors. As AIC and BIC decreased in Model 1, the introduction of these predictors improved the model. Model 1 revealed that these predictors had significant effects on the perception of timbre dissimilarity. In Model 2, the a priori distance of the sound pairs was introduced as a predictor, and the model was improved. Model 2 showed that the a priori distance had significant effects on timbre dissimilarity ratings, whereas the effect of timbre familiarity became insignificant.

Model 3 added the random slope for timbre familiarity, but the variance estimate of this random slope was not significant. Therefore, Model 4 added the random slopes for RI and a priori distance

Table 3.3 Model estimates and standard error (in parentheses) for hierarchical linear models of timbre dissimilarity ratings.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effect						
Intercept	-	.53* (.04)	–.04 (.044)	–.04 (.044)	.092* (.041)	.067 (.041)
RI	-	.02* (.003)	.02* (.003)	.02* (.003)	.02* (.005)	.051* (.011)
Familiarity	-	–.6* (.021)	.025 (.029)	.025 (.029)	–.037 (.028)	.024 (.028)
Stimulus Pair	-	.044* (.001)	.003* (.002)	.003* (.002)	.003 (.002)	.003 (.002)
Priori Distance	-	-	.58* (.019)	.58* (.019)	.59* (.038)	.59* (.038)
RI x Familiarity	-	-	-	-	-	–.015* (.005)
Error Variance						
Level 1 (Residual Component)	.12* (.002)	.077* (.002)	.064* (.001)	.064* (.001)	.053* (.001)	.053* (.001)
Intercept(Familiarity)	.012* (.003)	.008* (.002)	.009* (.002)	.009* (.002)	.006* (.002)	.006* (.002)
Intercept(Subject)	.18* (.058)	.009* (.004)	.009* (.004)	.009* (.004)	.003 (.002)	.004 (.002)
Slope(Familiarity$_{Subject}$)	-	-	-	0	-	-
Slope(Delay$_{Familiarity}$)	-	-	-	-	.001* (< .001)	.001* (< .001)
Slope(A Priori Distance$_{Familiarity}$)	-	-	-	-	.068* (.014)	.068* (.013)
Model Fit						
AIC	3272.2	1428.0	577.8	577.8	–26.4	–25.7
BIC	3266.2	1422.0	571.8	571.8	–36.4	–35.7

Note: *Statistically significant, $p < .05$; $ICC_{Familiarity} = .94$; $ICC_{Subject} = .62$

instead. The variance estimates of the random slopes were significant, and the effect of stimulus pair became insignificant in the Model 4. Also, the variance estimate of random intercept for subject became insignificant (p = .068), and this indicated that the introduced predictors accounted for a substantial amount of variation among subjects. Still, Model 5 added the interaction between RI and timbre familiarity to test the second hypothesis that timbre familiarity will minimize the effect of timbre informational decay on timbre dissimilarity ratings. Model 5 revealed that the interaction between RI and timbre familiarity was significant (p = .005). Furthermore, decreases of the residual component in Models 4 and 5 suggest that the introduced predictors accounted for some of the variance at level 1.

As the result of the first experiment revealed that the RI effect on the perception of timbre dissimilarity also depended on the timbre intervals, it is worth examining the identical pairs and the non-identical pairs separately. As the data of the identical pairs alone were not within a nested structure, dissimilarity ratings of the 20 participants for the identical pairs of synthetic tones, acoustically transformed sounds and acoustic instruments were analyzed with a repeated-measure RI (5) x Sound Type (3) ANOVA, as shown in Figure 3.5.

There were violations of sphericity for RI, $\chi^2(9) = 55.4$, $p < .001$, sound type, $\chi^2(2) = 14.9$, $p = .001$, and interaction between RI and sound type, $\chi^2(35) = 89.8$, $p < .001$. The result showed that as RI increased, the participants rated the identical pairs as significantly more dissimilar, $F(1.59, 30.2) = 14.8$, $p < .001$, $\epsilon = .40$, $\eta_p^2 = .44$. At the same time, timbre familiarity had a significant effect on the participants identifying the identical pairs, $F(1.28, 24.3) = 14.3$, $p < .001$, $\epsilon = .64$, $\eta_p^2 = .43$. The effect of timbre familiarity interacted with the RI effect, $F(3.64, 69.2) = 2.96$, $p = .029$, $\epsilon = .46$, $\eta_p^2 = .14$. In Figure 3.5, it can be seen that an increase in timbre familiarity significantly minimized the change in the mean dissimilarity ratings of the identical pairs.

The overall RI effect on the identical pairs did not occur until around 3.5 s. Post-hoc *t*-tests with B-H correction revealed that there were significant mean rating differences for 3.5-s RI (p = .025) and 5-s RI (p = .002), compared to the mean rating with 0.5-s RI. In terms of timbre familiarity, the identical synthetic tones were rated differently from those of the transformed sounds (p = .001) and the acoustic instruments (p = .003) in the experiment due to the acoustic complexity. However, the identical transformed sounds were not rated differently from the identical acoustic instruments (p = .915), which suggested that both the transformed sounds and the acoustic instruments were fairly complex acoustically.

Finally, interaction between the RI effect and the timbre familiarity effect was presented. When RI was less than 1 s, the mean rating differences between the identical synthetic tones, the identical

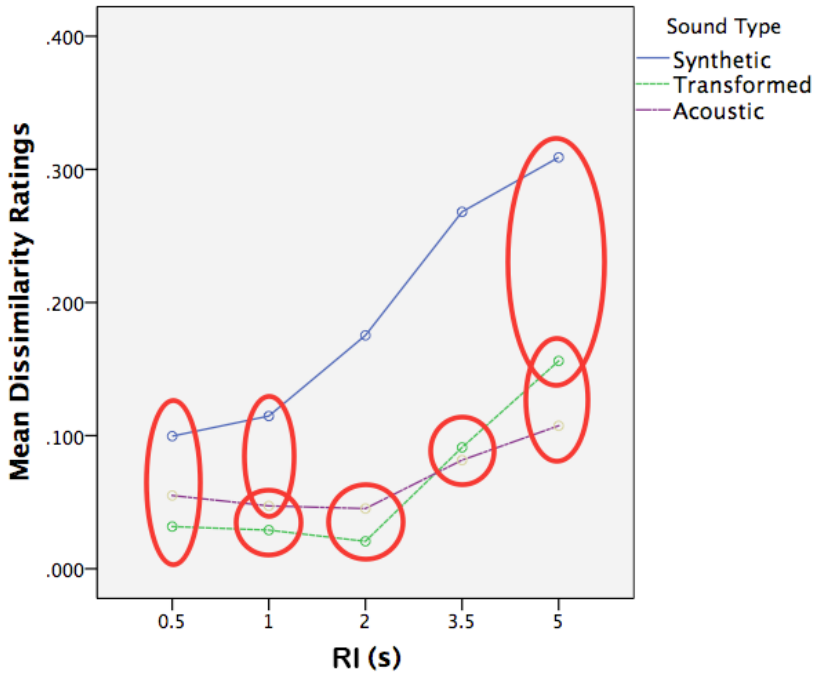

Fig. 3.5 Mean dissimilarity ratings of the identical pairs of synthetic tones, transformed sounds and acoustic instruments by 20 subjects as a function of RI. The lines represent different sound types. Circled mean dissimilarity ratings are not significantly different by B-H corrected post-hoc comparisons.

transformed sounds and the identical acoustic instruments were not significant. When RI became around 1 s, the interaction between RI and timbre familiarity began, as the mean rating difference between the identical synthetic tones and the identical transformed sounds became significant ($p = .036$). As RI increased to 2 s, the difference between the identical synthetic tones and the identical acoustic instruments also became marginally significant ($p = .06$), as well as between the synthetic tones and the identical transformed sounds ($p = .014$). When RI became 3.5 s, the mean rating differences between the synthetic tones and the transformed sounds ($p = .026$), and between the synthetic tones and the acoustic instruments ($p = .015$) were significant. However, when RI reached 5 s, the standard deviation of the ratings of the synthetic tones (0.26) covered a quarter of the rating scale. As the variability of the rating increased, the mean rating difference between the synthetic tones and the transformed sounds became insignificant ($p = .12$). The difference between the synthetic tones and the transformed sounds also became marginally significant again ($p = .055$). Since the participants perceived the identical synthetic tone pairs as more similar than either the identical transformed sounds or the identical acoustic instruments when RI became 2 s or more, timbre familiarity helped to minimize changes of mean dissimilarity ratings of identical pairs induced by the noisiness of the timbre representation.

Since the non-identical pairs were within a nested structure and contained different timbre intervals, mean dissimilarity ratings of the non-identical pairs were analyzed with a three-level hierarchical linear model, as shown in Table 3.4. The null model (Model 0) justified the use of the three-level hierarchical linear model because there were sufficient variations that could accommodate additional predictors at level 2 ($p < .001$) and level 3 ($p = .001$). Also, the estimate of the residual component suggested that there was a significant remaining variation at level 1 ($p < .001$) for which the predictors at levels 2 and 3 could not account.

Model 1 in Table 3.4 included RI, timbre familiarity and stimulus pair as predictors and revealed that these predictors had significant effects on the perception of the non-identical pairs. In Model 2, the priori distance of the sound pairs was introduced as a predictor, and the model was improved. Model 2 showed that the a priori distance had significant effects on timbre dissimilarity ratings of the non-identical pairs, whereas the effect of stimulus pair became insignificant. Model 3 and Model 4 added the random slopes for RI and a priori distance and the random slope for timbre familiarity, respectively, and the model was improved. After the introduction of the random slope for timbre familiarity in Model 4, the variance estimate of random intercept for subject became insignificant, which indicated that the introduced predictors accounted for a substantial amount of variation among subjects. Still, Model 5 added the interaction between RI and timbre familiarity

Table 3.4 Model estimates and standard error (in parentheses) for hierarchical linear models of timbre dissimilarity ratings of non-identical pairs.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed Effect						
Intercept	-	.32* (.052)	–.056 (.055)	–.08 (.078)	–.08 (.069)	–.11 (.07)
RI	-	.013* (.003)	.013* (.003)	.013* (.003)	.013* (.005)	.038* (.012)
Familiarity	-	–.053* (.021)	.15* (.033)	.18* (.042)	–.18* (.043)	–.2* (.043)
Stimulus Pair	-	.012* (.001)	–.001 (.002)	–.002 (.002)	–.002 (.002)	–.002 (.002)
Priori Distance	-	-	.43* (.026)	.44* (.037)	.44* (.037)	.44* (.037)
RI x Familiarity	-	-	-	-	-	–.012* (.006)
Error Variance						
Level 1 (Residual Component)	.068* (.002)	.067* (.002)	.061* (.002)	.057* (.002)	.057* (.002)	.057* (.002)
Intercept(Familiarity)	.032* (.007)	.015* (.004)	.013* (.003)	.044* (.014)	.024* (.009)	.024* (.009)
Intercept(Subject)	.33* (.11)	.013* (.006)	.013* (.006)	.001* (< .001)	0	0
Slope(Familiarity$_{Subject}$)	-	-	-	-	.004* (.002)	.004* (.002)
Slope(Delay$_{Familiarity}$)	-	-	-	.001* (< .001)	.001* (< .001)	.001* (< .001)
Slope(A Priori Distance$_{Familiarity}$)	-	-	-	.044* (.014)	.046* (.014)	.045* (.014)
Model Fit						
AIC	727.3	598.1	333.5	272.5	269.3	273.4
BIC	721.3	592.1	327.5	262.5	259.3	263.4

Note: *Statistically significant, $p < .05$; $ICC_{Familiarity} = .91$; $ICC_{Subject} = .84$

to examine whether timbre familiarity would minimize the effect of timbre informational decay on timbre dissimilarity ratings of the non-identical sound pairs. Model 5 revealed that the interaction between RI and timbre familiarity was significant (p = .036).

Mean dissimilarity ratings of the non-identical pairs were analyzed with a paired-samples t-test with B-H correction as well. The mean dissimilarity ratings with 0.5-s RI were compared to those with 1-s, 2-s, 3.5-s and 5-s RIs for the three types of sounds. Nevertheless, as shown in Figure 3.6, the post hoc test showed that the 12 comparisons were not significant. Also, Figure 3.6 showed that the line of synthetic tones was approximately parallel to the line of acoustic instruments. The line of the transformed sounds was not parallel to the lines of synthetic tones and acoustic instruments, which was the reason why the hierarchical linear modeling yielded significant interaction effect between RI and timbre familiarity. For this reason, the result did not support the hypothesis that timbre familiarity helped to minimize the change of mean dissimilarity ratings of non-identical sound pairs induced by RI.

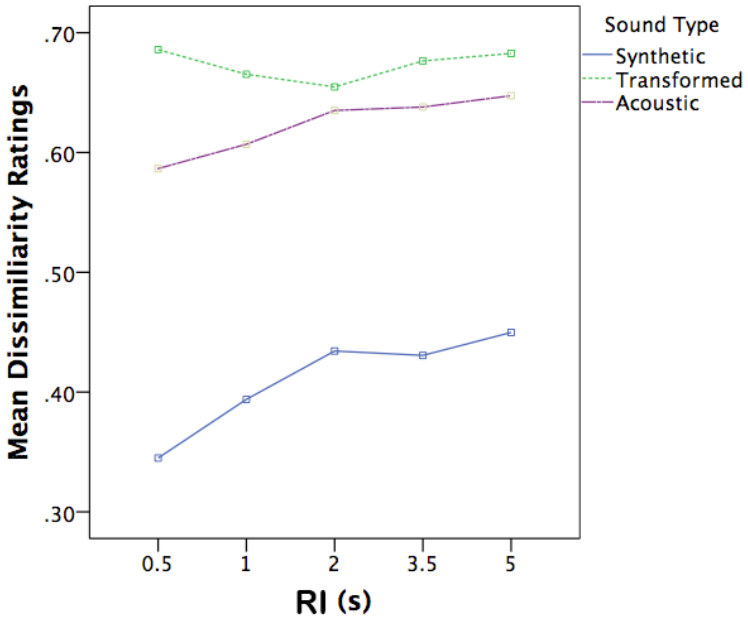

Fig. 3.6 Mean dissimilarity ratings across all non-identical pairs of the synthetic tones, the transformed sounds and the acoustic instruments by 20 subjects as a function of RI.

3.3 Discussion

3.3.1 Global Result

As the first experiment showed that the overall effect of RI on dissimilarity ratings of synthetic sounds reached its maximum within 5 s, the second experiment examined the RI effect at several RIs within 5 s. The result revealed that the overall RI effect changed the mean dissimilarity ratings significantly, from which this study inferred changes in the timbre representation. In addition to the various durations of the RIs, the second experiment added five digital transformations of acoustic instruments and five recorded acoustic instruments that were categorized as stimuli of moderate and high timbre familiarity, respectively. The result showed that timbre familiarity alone did not have a significant overall effect on the perception of timbre dissimilarity but interacted with RI. In other words, the RI effect on changes of mean dissimilarity ratings depended on timbre familiarity and acoustic complexity, which supports the aspect of the second hypothesis that there is a relationship between timbre familiarity effect and RI effect on changes of timbre representation.

3.3.2 Result of Identical Pairs

The second experiment showed that increasing RI to larger than around 3.5 s led to the participants perceiving the identical sound pairs as more dissimilar. Also, compared to synthetic tones, the participants perceived the transformed sounds and the acoustic instruments as less dissimilar in general. One of the possible explanations for this observation was that the participants got confused over the identical pairs and the similar pairs of the synthetic tone pairs as the synthetic tones sounded very similar more generally. In other words, varying the AT and SCG alone did not make the synthetic tones sound very different from each other. On the other hand, the five transformed sounds and the five acoustic instruments were acoustically complex, and this timbre complexity led to their distinct timbre representations. As a result, the participants did not get confused over the identical pairs and the similar pairs of the transformed sounds or the acoustic instruments, and on average they rated the identical synthetic tone pairs differently from the identical transformed and acoustic instrument sounds.

The result supports the notion that timbre familiarity minimizes the effect of timbre informational decay on dissimilarity ratings of the identical sound pairs in general. Within about 2 s, the participants in the experiment maintained the information of the three types of sounds fairly well. When the RIs were in the range of around 2 to 3.5 s, the representation of the identical

synthetic tone pairs may have become significantly noisier than the transformed sound information and the acoustic instrument information. This suggests that timbre complexity and familiarity of the transformed and the acoustic instrument sounds helped to minimize the RI effect on timbre representation. For RIs of 3.5 s and 5 s, the mean rating difference between the synthetic tone pairs and the transformed sound pairs was not significant. There are at least two potential explanations for this observation. The first one is that active maintenance of the transformed sound information starts to lose its effect. Another one was that with an increased RI, the variability of the ratings for the identical synthetic tone pairs increased, making the rating difference insignificant. On the other hand, although the effect of active maintenance of the acoustic instrument pairs decreased, the participants were able to maintain the acoustic instrument information marginally better than the synthetic tone information.

3.3.3 Result of Non-Identical Pairs

Hierarchical linear modeling yielded significant main effects of RI and timbre familiarity on mean dissimilarity ratings of the non-identical sound pairs, from which we inferred that RI and timbre familiarity are predictors of changes in timbre representation in the model. At the same time, interaction between RI and timbre familiarity suggested that the effects of the predictors depended on each other. On the other hand, pairwise comparison with B-H correction revealed that the mean rating differences between the conditions of the RIs were not significant for the three types of sounds. This result suggested that when the mean ratings of the non-identical synthetic tones, transformed sounds and acoustic instruments were examined separately, the mean ratings did not differ significantly as RIs increased from 0.5 s to 5 s, which was consistent with the result of the first experiment that there was no rating difference for sound pairs with large acoustic intervals within 10 s. Furthermore, although hierarchical linear modeling revealed that RI and timbre familiarity depended on each other to change the mean dissimilarity ratings of the non-identical sounds pairs, Figure 3.6 did not give any evidence that timbre familiarity helped to minimize the RI effect on the mean dissimilarity ratings.

Chapter 4

General Discussion

In this study, two experiments were conducted to examine the effect of retention interval (RI) on the mental representation of timbre. Based on the framework of the memory models and the previous timbre memory research, it was hypothesized that when timbre information started to decay, there were changes in the mental representation of timbre, on which dissimilarity ratings were based. The results of the two experiments partially supported this hypothesis. Identical synthetic tone pairs in both experiments were perceived as significantly dissimilar, even at RIs of only 0.5 s. This observation was quite consistent with the result of Demany et al. (2008) that timbre information for synthetic sounds decayed rapidly within 0.75 s. Therefore, the results of the two studies suggested that the RI effect on the perception of identical and similar sound pairs was due to the decay of timbre information. On the other hand, participants in the first experiment did not perceive dissimilar synthetic tone pairs like A0C2, A2C1, A1C2 and A2C2 as more similar or dissimilar as the RIs increased up to 10 s. In other words, within 10 s, memory decay of synthetic timbre information does not have an effect on the perception of timbre dissimilarity when the presented timbre intervals are large.

The second experiment contained digital transformations of acoustic instruments and recorded acoustic instruments as stimuli in addition to synthetic tones in order to test the second hypothesis that timbre familiarity would minimize the effect of timbre informational decay on dissimilarity ratings. These two types of stimuli were different from the synthetic tones in two ways. First, the transformed sounds and the acoustic instruments contained not only varied attack characteristics and spectral centroid but also other acoustic information for timbre comparisons. As mentioned in the first chapter, spectral, temporal and spectraltemporal factors can help people to recognize

the physical structures and excitations of sound sources, and the gestures of how an instrument is made to vibrate (Freed, 1990; Lakatos et al., 1997; Traube, Depalle & Wanderley, 2003; McAdams et al., 2004; Giordano & McAdams, 2006). The richness of acoustic information presented in the transformed sounds and the acoustic instruments allows the participants to recognize properties and features of the sound events, which was beneficial for the active maintenance of the timbre information and minimizing the noisiness of timbre representation as suggested in the second experiment. Second, compared to the synthetic tones, the acoustic distances of non-identical transformed sound pairs and non-identical acoustic instrument pairs were relatively large, which made these stimuli sound distinct from each other. The participants in the experiments continuously reported that the synthetic tone pairs sounded so similar that the dissimilarity-rating task became difficult, whereas they could easily distinguish the transformed sounds and the acoustic instruments. Detectability of the synthetic tone differences was, however, not the issue because the participants reported that they could still discriminate the synthetic tone pairs. The main issue was that the participants had difficulty telling what exactly the differences between the synthetic tones were. Their informal reports suggest that small overall acoustic differences of the sound pairs might make comparisons of the timbre information more difficult. As a result, even if the RI was small, the participants were more likely to confuse over the identical pairs and the similar pairs of the synthetic tones than those of the transformed sounds and the acoustic instruments.

In the two experiments, it was observed that small acoustic intervals of sound pairs made the perception of timbre dissimilarity more vulnerable to the noisiness of timbre representation. One of the potential explanations for this observation is that the participants often overestimated the distance of the sound pairs when the presented acoustic intervals were smaller or equal to the variability of the representation of a first tone induced by timbre memory decay. This explanation needs to be tested objectively with discrimination tasks in a signal-detection theory paradigm to remove response bias. Nevertheless, the tendency of perceiving similar sound pairs as more dissimilar cannot be explained because the participants in the two experiments could have rated the similar sound pairs as more similar or identical instead of more dissimilar. Further investigation is needed to explain this tendency.

The acoustic distance of the sound pairs could be a confounding factor for examining timbre memory, which is illustrated as one of the factors for the different results between this study and the study of Demany et al (2008). These authors aimed to examine the RI effect on timbre discrimination, and the stimuli in their study were varied through the frequency shifts of harmonic components, so the overall acoustic distances between the sound pairs in their study are smaller than

the overall acoustic distances of the synthetic tone pairs in this study. The different experimental design of the stimuli and the tasks led to different experimental results. Their study observed that the RI effect on timbre discrimination reached its maximum around 2 s, while this study observed that the RI effect on the perception of timbre dissimilarity reached its maximum at more than 2 s. For this reason, the design of sound pairs has a great effect on the experimental conclusion.

Although the purpose of the study is not to examine the *mechanism* of working memory for timbre, it is still worth mentioning the strategies used by participants to maintain the information of a first tone. For synthetic tones, participants reported that they replayed the first tone in their minds as the strategy for maintaining the timbre information. At the same time, participants with some sound engineering backgrounds reported that they noticed differences in the attack characteristics and the tone brightness. Therefore, they labeled the first tone like "bright tone with sharp attack" in order to retain the timbre information. For the recorded acoustic instruments, participants reported that they used words like "piano" to label the tones they heard in order to memorize the tones. For the transformed sounds, some participants reported that they were able to associate with some sci-fi sound effect, but many of them reported that they took an approach that was similar to how they approached the synthetic tones. Compared to synthetic tones, the participants reported that the transformed sounds were more distinct and easier to distinguish. Therefore, it was easier for them to retain the information of the transformed sounds in their minds, and the experimental results support their claim. These reports raise questions about the effectiveness of mental rehearsal for minimizing the noisiness of timbre representation. Regarding mental rehearsal, studies have proposed various mechanisms like the phonological loop (Baddeley & Logie, 1992), a music memory loop (Berz, 1995) and attention (Siedenburg & McAdams, 2016). In contrast, Demany and Semal (2008) have suggested that mental rehearsal is in fact not effective, and non-verbal information cannot be rehearsed. Nevertheless, this study does not provide any evidence to agree or disagree with either side. Further investigation is needed to account for the mechanism behind the mental rehearsal of timbre information.

As mentioned before, Nees (2016) argued that echoic memory, which was equivalent to the brief sensory store described in Cowan's memory model (1984, 1988), could be processed in parallel to mental rehearsal, and the parallel process allowed the echoic memory to contribute to the rehearsal. The two experiments in this study were not designed to separate the contribution of the echoic memory and the working memory on retaining timbre information; therefore, the task of the second experiment might involve both types of memory.

4.1 Future Directions

This study leaves some questions open for future studies to explore.

4.1.1 Instrumental Expertise and the Effect of Timbre Ear Training

Both musicians and non-musicians were recruited for the two experiments. As previous research (McAdams et al, 1995; Siedenburg & McAdams, 2016) showed that musicians, in general, did not perform differently in timbre dissimilarity rating tasks from non-musicians, this study ignored the musicianship factor. However, instrumental expertise and timbre ear training might help people to perform the rating task differently. Although the two experiments did not include any extended techniques as stimuli, musicians having extensive experience with contemporary music might recognize and identify extended techniques of their own instruments more easily than others. Consequently, the musicians might better resist the effect of RI on the perception of timbre dissimilarity for their own instruments. Also, composers are usually trained to recognize and identify different instruments including extended techniques. As a result of the training, the composers are usually able to play the instruments in their mind easily, which might make them rate the acoustic instruments more consistently as RIs increase. Sound engineers, on the other hand, might rate the synthetic tones more consistently as RIs increase. They are professionally trained to detect subtle changes of spectral information and attack characteristics in order that they can correct potential issues in audio applications, such as darkening a bright recording, attenuating the nasality of a recording, and changing transients of percussion instruments. This specialty might make sound engineers more accurately detect the subtle timbre changes of synthetic tone pairs, and consequently, they might be less likely to perceive identical synthetic tone pairs as dissimilar as RIs increase. For these reasons, further studies on how well composers and sound engineers could resist the RI effect on the perception of timbre dissimilarity compared to others would be worth conducting.

4.1.2 Timbre Interval Effect

In the two experiments, small acoustic intervals of the synthetic tone pairs resulted in the participants perceiving the identical pairs and similar pairs as more dissimilar as the RI increased, and this indicates the interaction effect between RI and timbre intervals. However, as the synthetic tones are generated with additive synthesis, the tones are not complex in terms of acoustic information,

which raises the question whether the timbre interval effect also applies to sounds with higher timbre complexity. For example, when a piano sample is manipulated in terms of attack characteristics and spectral centroid alone, the resulting sounds will be very similar to the original piano sample. When participants are required to rate the timbre dissimilarity between the original piano sound and the manipulated sounds with varied RIs, would the rating curves for the stimulus set be similar to the curves of the synthetic tones observed in the two experiments? The present study does not answer this question and leaves the question open for future investigations.

4.1.3 Interference Effect

One of the examples to which this study contributes is timbre-based theme and variations. This study showed that without interference, the representation of familiar timbre information changed less than the representation of unfamiliar timbre information, and the large timbre intervals could minimize the effect of noisy timbre representation on the perception of timbre dissimilarity. Based on these fundamental findings, composers will be able to write a theme and variations by varying timbre intervals and contours in such a way that listeners are able to recognize and make sense of the timbre variations. Nevertheless, in reality, variations along other musical dimensions are always present as well. These other dimensions have the potential of interfering the listeners' timbre memory for the timbre-based motives, and this study does not explore this potential interference effect. Future studies can combine the paradigm of the two experiments with the interpolated tones paradigm to examine this interference effect. To illustrate, in an experiment, participants will hear a piano sound followed by a violin sound after a few seconds. Depending on conditions, within the RIs, the participants will or will not hear interpolated tones like a trumpet sound or a harpsichord sound. After that, the participants are required to give a dissimilarity rating for the piano-violin pair. By comparing the mean rating differences between different conditions, future studies could examine interference effects on timbre representation.

References

Aikake, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (eds.), *2nd International Symposium on Information Theory:* 267-81. Budapest: Akademiai Kiado.

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*(3), 1371-1379.

ANSI. (1960/1994). *Psychoacoustic Terminology; Timbre*. New York, NY: American National Standards Institute.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, *2*, 89-195.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation*, *8*, 47-89.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in Cognitive Sciences*, *4*(11), 417-423.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Baddeley, A., & Logie, R. H. (1992). "Auditory imagery and working memory," in *Auditory Imagery*, ed. D. Reisberg (Hillsdale, NJ: Lawrence Erlbaum Associates), 179-197.

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning & Verbal Behavior*, 22, 88-104.

Berz, W. L. (1995). Working memory in music: a theoretical model. *Music Perception*, *12*, 353-364.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Broadbent, D. E. (1957). "Effects of noise on behavior," in *Handbook of Noise Control*, C. M. Harris, Ed. (New York: McGraw-Hill), 10-34.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Broadbent, D. E. (1984). The Maltese cross: A new simplistic model for memory. *The Behavioral & Brain Sciences*, *7*, 55-94.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron*, *48*(4), 687-697.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, *118*(1), 471-482.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457-469.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*(3), 283-319.

Caruso, V. C., & Balaban, E. (2014). Pitch and timbre interfere when both are parametrically

varied. *PloS ONE*, *9*(1), e87065.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: a reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1235.

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *The Quarterly Journal of Experimental Psychology*, *62*(7), 1420-1429.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: the importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331-335.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*(2), 341.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163-191.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-185.

Cowan, N. (2015). Sensational Memorability: Working Memory for Things We See, Hear, Feel, or Somehow Sense. *Mechanisms of Sensory Working Memory: Attention and Performance XXV*, 1.

Cowan, N., & AuBuchon, A. M. (2008). Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin & Review*, *15*(1), 230-235.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial-recall task: A logical sequel to Miller (1956). *Psychological Science*, *15*(9), 634-640.

Cowan, N., Saults, J. S., & Nugent, L. D. (1997). The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons. *Psychonomic*

Bulletin & Review, *4*(3), 393-397.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of the mismatch negativity in the auditory event-related potential (ERP*). Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 909-921.

Cowan, N., & Wood, N. L. (1997). Constraints on awareness, attention, and memory: some recent investigations with ignored speech. *Consciousness and Cognition*, *6*, 182-203.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*(2), 255-267.

Dawson, M. E., & Schell, A. M. (1982). Electrodermal responses to attended and unattended significant stimuli during dichotic listening. *Journal of Experimental Psychology: Human Perception & Perfromance*, *8*, 315-324.

Demany, L., Clément, S., & Semal, C. (2001). Does auditory memory depend on attention? *Physiological and Psychophysical Bases of Auditory Function*, edited by D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs & R. Schoonhoven. Shaker Publishing BV, Maastricht (The Netherlands), 461-467.

Demany, L., Trost, W., Serman, M., & Semal, C. (2008). Auditory change detection: simple sounds are not memorized better than complex sounds. *Psychological Science*, *19*(1), 85-91.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*(3939), 1604-1605.

Deutsch, D. (1972). Mapping of interactions in the pitch memory store. *Science*, *175*(4025), 1020-1022.

Engle, R. W., & Roberts, J. S. (1982). How long does the modality effect persist?. *Bulletin of the Psychonomic Society*, *19*, 343-346.

Freed, D. J. (1990). Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *Journal of the Acoustical Society of America*, *87*, 1236-1249.

Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, *119*, 1171-1181.

Giordano, B. L., & McAdams, S. (2010). Sound source mechanics and musical timbre perception: Evidence from previous studies. *Music Perception*, *28*, 155-168.

Girden, E. (1992). *ANOVA: Repeated Measures*. Newbury Park, CA: Sage.

Glucksberg, S., & Cowen, G. N. (1970). Memory for nonattended auditory material. *Cognitive Psychology*, *1*(2), 149-156.

Golubock, J. L., & Janata, P. (2013). Keeping timbre in mind: Working memory for complex sounds that can't be verbalized. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 399.

Gower, J. C. (1966). Some distance properties of latent root and vector methods using multivariate analysis. *Biometrika*, *53*, 325-338.

Grey, J. M., & Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, *63*(5), 1493-1500.

Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. Psychomusicology, *12*, 154-171.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and language*, *54*(2), 265-281.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory:

Is it phonological and is it a store?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 656.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Le Journal de Physique IV*, *4*(C5), C5-625.

Krumhansl, C. L. (1989). "Why is musical timbre so hard to understand?," in *Structure and Perception of Electroacoustic Sound and Music*, Vol. 846, eds S. Nielzén and O. Olsson (Amsterdam: Excerpta Medica), 43-53.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre.*Journal of Experimental Psychology: Human Perception and Performance*,*18*(3), 739.

Lakatos, S., McAdams, S., & Caussé, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics*, *59*, 1180-1190.

Lartillot, O., & Toiviainen, P. (2007). "A matlab toolbox for musical feature extraction from audio," in *Proceedings of the 10^{th} International Conference on Digital Audio Effects (DAFx)* (Bordeaux), 237-244.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1545.

Lim, S. J., Wöstmann, M., & Obleser, J. (2015). Selective attention to auditory memory neurally enhances perceptual precision. *Journal of Neuroscience*, *35*(49), 16094-16104.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America*, *114*(5), 2946-2957.

Martin, F. N., Champlin, C. A., et al. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*(2), 64-66.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*, 124-145.

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, *4*(1), 181-198.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand, (Eds.), *Thinking In Sound: The Cognitive Psychology of Human Audition* (pp. 146-198). Oxford: Oxford University Press.

McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The Psychology of Music* (3^{rd} ed., pp. 35-67). San Diego, CA: Academic Press.

McAdams, S., Saariaho, K. (1985). Qualities and functions of musical timbre. *Proceedings of the 1985 International Computer Music Conference* (pp. 367-374). San Francisco: The Computer Music Association.

McAdams, S., Chaigne, A., & Roussarie, V. (2004). The psychomechanics of simulated sound sources: Material properties of impacted bars. *Journal of the Acoustical Society of America*, *115*, 1306-1320.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177-192.

Mcevoy, L., Levänen, S., & Loveless, N. (1997). Temporal characteristics of auditory sensory memory: neuromagnetic evidence. *Psychophysiology*, *34*, 308-316.

McKeown, D., Mills, R., & Mercer, T. (2011). Comparisons of complex sounds across extended retention intervals survives reading aloud. *Perception*, *40*(10), 1193-1205.

McKeown, D., & Mercer, T. (2012). Short-term forgetting without interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1057.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, *48*(2), 169-178.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1989). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neuroscience Letters, 98*, 217-221.

Nattiez, J. –J. (2007). Le timbre est-il un paramètre secondaire. *Cahiers de la Société Québécoise de Recherche en Musique*, *9*(1-2), 13-24.

Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review*, *2*(4), 429-441.

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology*, *7*.

Nees, M. A., & Walker, B. N. (2013). Flexibility of working memory encoding in a sentence-picture-sound verification task. *Journal of Cognitive Psychology*, *25*, 800-807.

Neisser, U. (1967). *Cognitive Psychology*. New York, NY: Meredith Publishing Company.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M. (1992). Complex sounds and auditory images. *Auditory physiology and perception*, *83*, 429-446.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369-378.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working-memory capacity in the change-detection paradigm. *Psychonomic Bulletin & Review*, *18*, 324-330.

Roy, S. (2003). *L'analyse des musiques électroacoustiques: Modèles et proposition*. Paris, France: L'Harmattan.

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trance persists about 10 sec: neuromagnetic evidence. *Journal of Cognitive Neuroscience*, *5*, 363-370.

Schröger, E. (2007). Mismatch negativity: a microphone into auditory memory. *Journal of Psychophysiology*, *21*, 138-146.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Schweickert, R., & Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 419.

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *The Journal of the Acoustical Society of America*, *89*(5), 2404-2410.

Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *The Journal of the Acoustical Society of America*, *94*(3), 1315-1322.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1). 156-163.

Siedenburg, K., Jones-Mollerup, K., & McAdams, S. (2016). Acoustic and categorical dissimilarity of musical timbre: Evidence from asymmetries between acoustic and chimeric sounds. *Frontiers in Psychology*, *6*(1977).

Siedenburg, K., & McAdams, S. (2016). The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory*, 1-15.

Siegel, K. (2014). *Timbral Transformations in Kaija Saariaho's From the Grammar of Dreams* (Doctoral dissertation). Retrieved from CUNY Academic Works.

Smith, B. K. (1995). "Psiexp: An environment for psychoacoustic experimentation using the ircam musical workstation," in *Proceedings of Meeting of the Society for Music Perception and Cognition*, ed D. Wessel (Berkeley, CA: University of California Berkeley).

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chiemaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*, 87-90.

Soemer, A., & Saito, S. (2015). Maintenance of auditory-nonverbal information in working memory. *Psychonomic Bulletin & Review*, *22*(6), 1777-1783.

Starr, G. E., & Pitt, M. A. (1997). Interference effects in short-term memory for timbre. *The Journal of the Acoustical Society of America*, *102*(1), 486-494.

Traube, C., Depalle, P., & Wanderley, M. (2003). Indirect acquisition of instrumental gesture based on signal, physical and perceptual information. In F. Thibault (Ed.), *Proceedings of the 2003 Conference on New Interfaces for Musical Expression, Montreal* (pp. 42-47). Montreal, Canada: McGill University.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392-393.

Watkins, M. J., & Todres, A. K. (1980). Suffix effects manifest and concealed: further evidence for a 20-second echo. *Journal of Verbal Learning and Verbal Behavior*, *19*, 46-53.

Winkler, I., & Cowan, N. (2005). From sensory to long-term memory evidence from auditory memory reactivation studies. *Experimental Psychology, 52*, 3-20.

Winkler, I., Korzyukov, O., Gumenyuk, V., Cowan, N., Linkenkaer-Hansen, K., Ilmoniemi, D. R., Alho, K., & Näätänen, R. (2002). Temporary and longer term retention of acoustic information. *Psychophysiology*, *39*, 530-534.

Winkler, I., Schröger, E., & Cowan, N. (2001). The role of large-scale memory organization in the mismatch negativity event-related brain potential. *Journal of Cognitive Neuroscience*, *13*, 59-71.

Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, *58*(2), 315-330.

Winsberg, S., & Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British Journal of Mathematical and Statistical Psychology*, *50*(1), 55-72.

Wood, C. (2015). *Perception of Timbre Intervals* (Master thesis). Retrieved from McGill Library. (PID: 135659)