IN SEARCH OF

TONAL GROUNDS FOR SHORT TERM MELODY RECOGNITION

ALPEREN KARAN

BOGAZICI UNIVERSITY

2019

IN SEARCH OF

TONAL GROUNDS FOR SHORT TERM MELODY RECOGNITION

Thesis submitted to the

Institute for Graduate Studies for Social Sciences

In partial fulfillment of the requirements for the degree of

Master of Arts

In

Psychology

by

Alperen Karan

Boğaziçi University

2019

DECLARATION OF ORIGINALITY

I, Alperen Karan, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date 08/08/2019

ABSTRACT

In Search of Tonal Grounds for Short Term Melody Recognition

What information is retained after the first hearing of a note sequence? Taylor and Pembrook (1983) investigated for this question with melodies in C major scale using a singing back and a dictation task. They observed that: (1) recall performance for the last notes was better, (2) contour complexity (more direction changes) reduced correct contour recall, (3) fewer errors are made when the melodies started with C4 instead of C5, (4) recall for a note of direction change is better if the preceding interval is larger. Our study reassessed their results with a larger and more controlled stimulus set and a recognition paradigm with non-transposed and contour preserving lures. We found a full J-shaped serial position curve (recency > primacy > center). Participants' correct recognition of melodic intervals decreased with contour complexity. Also, recognition performance was greater when a note of direction change ($\wedge$ or $\vee$) was changed, especially when the preceding interval was larger. These findings indicate that non-transposed comparison melodies did not turn our study into a simple pitch memory test. Furthermore, when filling the retention interval with verbal nonsense stimuli as opposed to melodic stimuli, less interference was observed. Secondly, the higher the similarity between interference and test items, the lower the performances (diatonic > nondiatonic > verbal interference). Finally, we collected memorability and complexity ratings for our note sequences. These ratings were found to be consistent with the recognition rates. Results are discussed with regards to Baddeley's (2000) and Berz' (1995) models of short-term memory.

ÖZET

Kısa Vadeli Melodi Tanımasındaki Tonal Temellerin İncelenmesi

Bir nota dizisini ilk dinleyişten sonra hangi bilgi akılda kalır? Taylor ve Pembrook (1983) bu soruya do majör gamındaki melodileri şarkı gibi söylemek veya onları dikte etmek görevlerini kullanarak cevap aradılar. Araştırmacılar (1) son nota için hatırlama performansının daha iyi olduğunu, (2) kontur karmaşıklığının (daha fazla yön değiştirme sayısı) doğru kontur hatırlamasını düşürdüğünü, (3) C4 ile başlayan melodilerde C5 ile başlayanlara göre daha az hata yapıldığını, (4) melodinin yön değiştirdiği bir notanın hatırlanmasının eğer hemen önceki aralık geniş ise daha kolay olduğunu buldular. Çalışmamız onların sonuçlarını daha geniş ve kontrollü bir uyaran seti ve kontur koruyan transpoze edilmemiş tuzakları tanıma paradigmasıyla yeniden değerlendirdi. Bulgularımız tam bir J şekilli seri konum eğrisi gösterdi (sonralık > öncelik > orta). Katılımcıların melodik aralıkları doğru tanıması kontur karmaşıklığı ile azaldı. Ayrıca, bir yön değiştirme notası ($\wedge$ veya $\vee$) değiştirildiğinde tanıma performansının daha yüksek olduğu bulundu, özellikle hemen önceki aralık geniş ise. Bu sonuçlar, transpoze edilmemiş karşılaştırma melodileri kullanmış olmamızın çalışmamızı basit bir perde hafızası testine dönüştürmediğini göstermektedir. Aralığı sözel tipli anlamsız uyaranla doldurduğumuzda müziksel tipli bir uyaranla kıyasla daha az performans düşüşüne yol açtığını gözlemledik. İkincisi, müdahale ile test maddeleri arasındaki benzerlik arttıkça performanslar düştü (sözel < diyatonik olmayan müziksel < diyatonik müziksel uyaran). Son olarak, melodi dizilerimiz için hatırlanabilirlik ve karmaşıklık dereceleri topladık. Bu derecelerin tanıma performanslarıyla tutarlı olduğu bulundu. Sonuçlarımızı Baddeley'in (2000) ve Berz'in (1995) kısa vadeli bellek modelleriyle kıyaslayacağız.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor Prof. Esra Mungan for her invaluable guidance, encouragement and contributions to this work from beginning to end. This thesis would not have been possible without her continuous unwavering support. I genuinely appreciate all the time she spent on me.

I would like to extend my sincere thanks to my thesis committee members: Prof. Ayşecan Boduroğlu and Prof. Nart Atalay for providing ingenious suggestions and constructive comments.

I would like to acknowledge the help of my friends that I met during my psychology masters for their wonderful and lovely friendship. Harun Yörük, Selen Küçüktaş, Aslıhan İkizoğlu, Nehir Mavioğlu, Didar Karadağ, Mustafa Fırat, Gülşen Güldeste, Bilge Gençoğlu, Nazike Mert, Nihan Keşir, Hanifi Çiftçi, Ece Kaya, Elif Canseza Kaplan and Emre Oral. I am grateful to all of you.

I am also grateful to all faculty members in the department of Mathematics at İTÜ for providing a lovely place to work. I would like thank my colleagues Vedat Kurtay and Gökhan Göksu for giving help whenever I needed.

I thank my mother, Jülide Karan, my father, Erhan Karan, for their thorough love, and best sisters, Aybüke and Alanur for bringing happiness to my life. My friends, Muhammet Memiç and Tuğba Ayan, thank you for your priceless friendship.

Last but not least, I would like to dedicate my thesis to my excellent wife, Şeyda, for being my best friend, motivating me during this tiring process and providing continuous support. Words cannot express how lucky I am to have you in my life.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

When a listener hears a simple melodic motive for the first time, an immediate representation is formed in short term memory. A subject of interest is to explore which particular features of a melody make the exact note sequence more memorable. For example, are melodies with lots of direction changes in their pitch contour more difficult to remember than the ones that have fewer contour changes or are purely ascending or descending? Another curious question is how intervening musical or non-musical events interferes with the representations of the melody. Do some aspects of melodies make them more resistant to interference? In this thesis, we will discuss these problems.

Several studies investigated the elements that affected the memory for musical information. Contour -the arrangement of ups and downs in a sequence of notes- is perhaps one of the most salient cues for short term melodic memory in brief delays. For example, Dowling and Fujitani (1971) found that participants used contour information to recognize transpositions of atonal melodies. When the comparison melodies did not violate contour, participants performed at chance level.

Dowling (1978) extended the previous study to tonal melodies. In a same/different test setup, musically trained and untrained participants performed at chance level in distinguishing exact transpositions (targets) of a standard melody from the "tonal answers" (see Figure 1). The tonal answer differs from the correctly transposed target in that the notes of the tonal answer belong to the scale of the original melody, not the transposed scale. However, correct rejection performances

were higher when the comparison items were atonal. So, not only contour, but also tonality (belonging to a scale) was a clue for short term melody recognition.

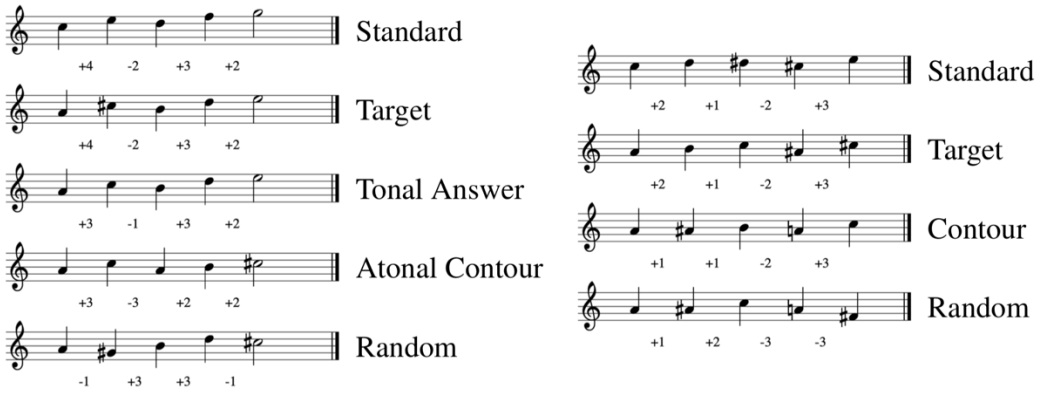


Figure 1 Sample stimuli from Dowling's 1978 (left) and Dowling and Fujitani's 1971 (right) studies. The numbers represent the size of melodic intervals between consecutive notes[1]

Dowling (1991) did a recognition task with strongly tonal, weakly tonal and atonal 7-note sequences in order to detect which particular features of melodies boost recognition performances in short vs. long delays. Participants were asked to discriminate exact transpositions from lures with same or different contours. The paradigm was a continuous running memory test, so, the comparison melodies were not necessarily presented immediately after the targets. For example, T1-C1-T2-T3-C3-C2-… shows a pattern of the ordering of the target and comparison melody pairs. For targets T1 and T3, the comparison melodies came just after the targets (short delay) whereas two melodies interfered between T2-C2 target-comparison pair (longer delays). He observed that tonal melodies yielded greater discrimination performance than atonal ones (Figure 2). This shows the involvement of scale information in short term memory settings such that tonal melodies may be more resistant to interference because of their possible reliance on schematic knowledge

[1] Note that all except the "Random" ones are contour preserving regardless of whether they retain exact pitch intervals or not.

(cf. Agres, 2019). Moreover, when tonal targets were tested against their same and different contour lures, discrimination performances were greater for different contours in short, but not in long delays. This study demonstrates that when tested with tonal stimuli, participants rely on contour information in short delays, and exact interval information in long delay settings. This finding was also replicated in Dowling, Kwak and Andrews' (1995) study. Pembrook (1987) states that longer tonal melodies tend to build a stronger sense of scale, and thus, people use contour information for short and interval information for long melodies.

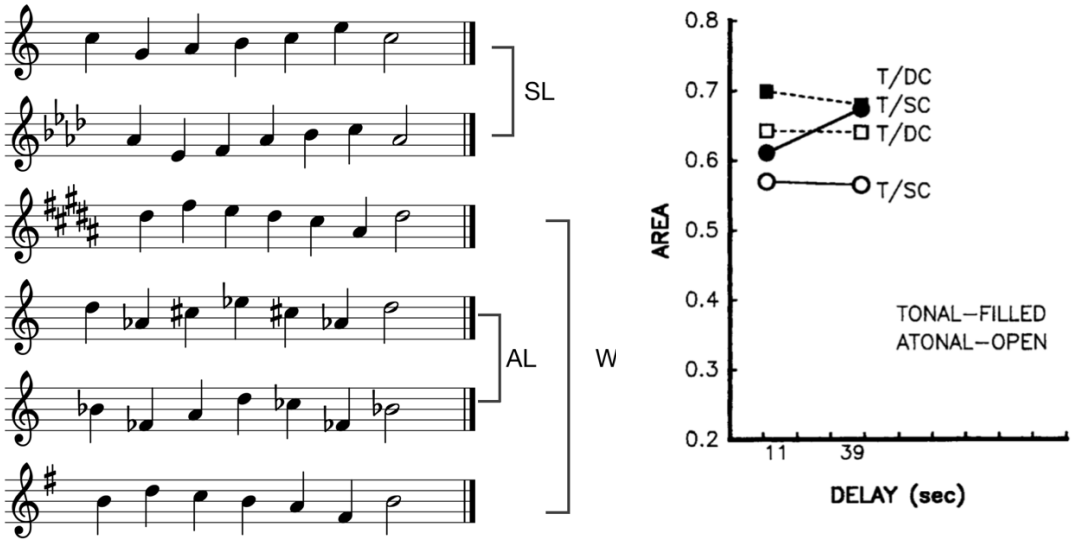


Figure 2 Sample trials and procedure from Dowling's 1991 study. SL and AL refers to strongly tonal and atonal melodies tested with contour preserving lures whereas WT indicates a weakly tonal melody tested with an exact transposition (left). The brackets show target-lure melody pairs in the continuous running paradigm. A brief summary of their results is on the right, image from Dowling (1995). T/DC and T/SC refer to target/same-different contour lures discrimination

Dyson and Watkins (1984) investigated how melodic contour configuration contributed to short-term melody recognition. Participants detected changes that occurred at a "corner note" (i.e., a note at which a melodic direction change occurs, which we will mark as "∧" or "∨" to indicate an ascending-to-descending or descending-to-ascending contour change, respectively) more easily than the changes occurred at a "slope note" (i.e., a note which falls within a given ascending or

descending contour). This finding gives evidence that auditory and visual contours play a somewhat similar function perceptually. However, this "corner" effect was dramatically reduced when the comparison melody was transposed, which might indicate that performance does not singly depend on configuration, i.e. Gestalt abstraction.

The above studies investigated how different characteristics of melodies alter the memory performances. Yet, the span of short-term memory of melodies is limited. Earlier studies proposed a short-term memory capacity of 7-11 notes (Pembrook, 1987). However, more recent studies indicate a smaller capacity of around 5-7 notes (e.g., Benassi-Werke, Queiroz, Germano, & Oliveira, 2012; Schulze, Dowling, & Tillmann, 2012). Baddeley (2000), in his working memory model, suggests a span of 2 seconds for the phonological loop which is responsible for holding acoustic information. However, Berz (1995) points out that a phonological loop for holding all sound information is insufficient to explain certain empirical findings with musical stimuli. One would expect that acoustic tasks be equally affected by different types of unattended sounds if there is a single acoustic store in working memory. However, previous studies suggest otherwise. For example, Martin, Wogalter, and Forlano (1988) observed that unattended[2] speech had a greater interfering effect on reading comprehension than unattended music. Furthermore, they also found that the effect of unattended music is greater than unattended speech on music identification. In another study, Salame and Baddeley (1989) did a serial recall of visually presented digits task. They found that verbal music interferes with performance more than instrumental music. To account for

[2] Participants were told that while reading they should ignore (one of the several types of) background noise and focus on reading because they were going to answer questions about each passage. The dependent variable was the normalized rate of correct responses for the questions about passages.

such findings, Berz proposed a separate “music memory loop” within working memory. He states that even if there might be some connection between the phonological and musical loops it is one of negligible significance.

Mikumo (1992) conducted a same/different test with tonal and atonal melodies lasting for 6 seconds. The significance of this study was that the author used different retention interval manipulations in the experiment. That is, between the target and comparison melodies, there was a 12-second retention interval either unfilled (pause) or filled with an interfering melody, nonsense syllables or note names. In the latter two conditions, listeners had to shadow a set of nonsense syllables or note names till test. All participants performed better with tonal than atonal melodies. Although musicians’ performance for tonal melodies was affected differently by different types of interference with worst performance for the note names shadowing task, non-musicians’ performance was equally impaired across interference conditions. The author concludes that non-musicians did not use any strategy for the task since no significant difference was observed between different interference conditions. Thus, she claims that non-musicians listen to melodic sequences in terms of global configuration (such as contour).

Thompson and Yankeelov (2011) did same/different tests with 3-note sequences and three-syllable words with a retention interval of 5 seconds. They found that both musicians and nonmusicians performed better when the retention interval was filled with verbal rather than musical material. Atherton et al. (2018) replicated this finding by conducting same/different tests with both monosyllabic 3 letter words (like cap and cab) and major chords (like C major and D major). The duration of the retention interval was 6-seconds long, and it could either be silent, or filled with 4 chords or 4 monosyllabic 3 letter words. They observed that the

participants performed better when the retention interval was silent. Moreover, performances were more impaired when the interference was the same type of the test (i.e., chord interference in chord test and letter interference in letter test).

All studies mentioned above investigated memory for musical materials in terms of interfering events (such as verbal or musical) or features of the musical items (such as tonality). To our knowledge, Ortmann's 1933 study was the first work trying to identify the tonal characteristics of short isochronous melodies that increase or decrease short term memory recall. The melodies in his experiment consisted of twenty 5-note sequences (see Figure 3) and he asked musically trained participants to write down the melody after each of them was played on the piano.

Taylor and Pembrook (1983) did an extension of the above study using the same stimulus set but adding a singing-back task. This allowed them to test non-musicians as well. Moreover, they improved some methodological issues in Ortmann's study such as randomizing presentation order of the melodies. Taylor and Pembrook replicated most of Ortmann's findings. There was a serial position effect such that the last note was remembered best compared to the other ones. When tested for correct recall of contour (instead of exact intervals) participants performed better for melodies with no or one than two or three direction changes. Taylor and Pembrook link the absence of a difference between contour recall of no and one direction change melodies to the fact that all melodies with one direction change have a configural symmetry since the direction change always occurred at the third note (e.g., up-down or down-up).

Ortmann claimed that participants would notice a melody direction change better if the interval before that note is large enough. To test this claim, Taylor and Pembrook classified each interval in the melody as a "step" (a semitone or whole

tone interval) or a “skip” (a three semitones or larger interval). They verified Ortmann’s claim: fewer errors were made at a note of direction change if the note was preceded by a skip rather than a step.

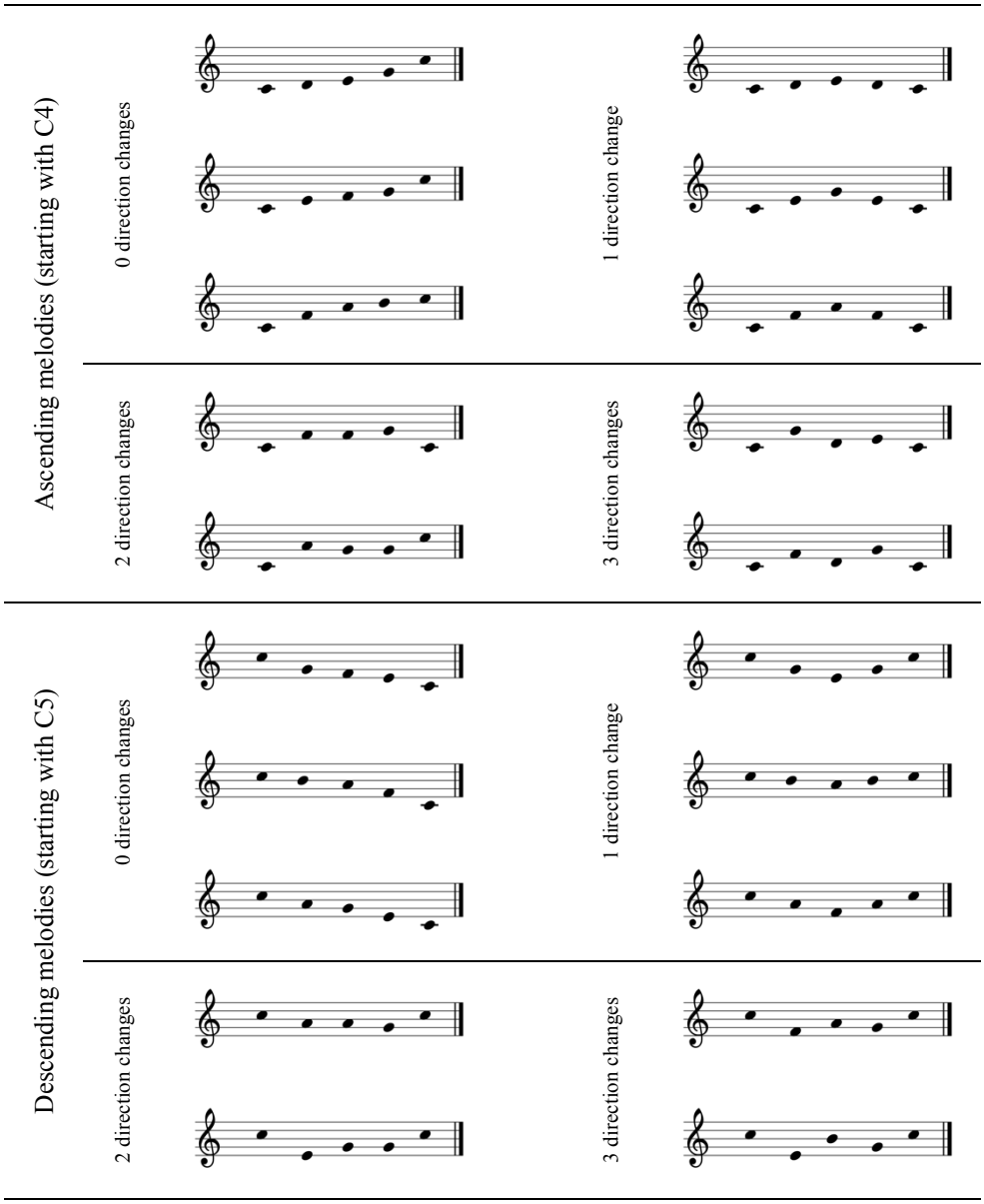


Figure 3 The stimulus set used by Ortmann (1933) and Taylor & Pembrook (1983)

Ortmann also suggested a better memory for ascending rather than descending melodies since Western music is built on ascending scales and chords. Note that the authors refer to melodies that start with C4 as ascending and those that start with C5 as descending. They found that participants made fewer errors with

ascending melodies than descending ones. They did not run any additional analyses concerning the difference between fully ascending and fully descending melodies.

Although Taylor and Pembrook extended Ortmann's findings in a more controlled manner, there are still some factors that make their findings questionable. For example, some sequences had a full melodic symmetry with triadic structure (e.g., C-E-G-E-C). One can see these sequences as arpeggios of chords rather than melodies. Furthermore, the initial and the final note of each melody was a C. As also suggested by Dowling (1973), varying initial notes would have made the memory task too difficult. Yet, there was no need to keep the last note constant, which makes it impossible to know whether Taylor and Pembrook's recency effect finding reflected genuine episodic memory or simply an awareness of a common last note. Finally, their stimulus set was quite limited, and thus, we cannot really know whether their conclusions can be extended to other sequences.

The way Taylor and Pembrook (1983) identified the correctness of a response was also problematic. A note was defined to be incorrect only if the preceding interval and the exact pitch were wrong. For instance, assume that the participant responds to C-F-D-G-C with C-G-E-G-C (see Figure 4).

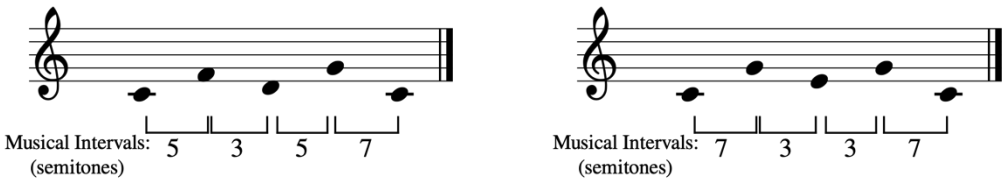


Figure 4 A melody from Taylor and Pembrook's stimulus set (left) and a possible response from a participant (right)

In this case, they regarded only the second note to be incorrect. As the musical interval between second and third notes is correct, they identified the third note as correct. Also, the fourth note was also regarded as correct because the pitch is

correct even though the interval was wrong. Thus, we cannot really know whether a specific finding of theirs is due to participants' memory for exact interval size or exact pitch information, or both.

CHAPTER 2

THE PRESENT STUDY

We wanted to expand on Taylor and Pembrook's study by using a larger stimulus set and a short-term memory recognition instead of recall test setup. The recognition test setup allowed us to eliminate the correct interval/correct pitch dilemma by forcing the listener to respond positively only to exactly matching test tunes. Furthermore, by using only same-contour lures we were able to test the resolution of the representation of a melody after first hearing. Earlier studies have shown that contour information seems to be a more dominant heuristic in brief delay recognition and interval information more so in long delay recognition (e.g., Dowling, Kwak, & Andrews, 1995). Yet, we were curious to see how much of exact interval information was retained even after brief delays. Since Dowling (1991) showed that recognition performance drops to chance level when using transposed same-contour targets and lures, we decided to keep our tunes untransposed at test and thus allow listeners to make use of their pitch memory capacity. We believe that this also better mimics natural settings which often require an immediate recall of a just heard musical motive. Nees (2016) proposed that the effect of auditory sensory memory is little or none when the retention interval is longer than 2 seconds. We reduced the possibility for that effect by assigning a brief yet long enough duration (2.25 sec) to the retention interval between the target and test items.

Our purpose was to connect Taylor and Pembrook's findings with the ones coming from the broad literature of short-term recognition for tonal melodies (e.g., Deutsch, 1970; Dewitt & Crowder, 1986; Dowling, 1973, 1978, 1991; Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling et al., 1995). We created 158

five-tone target-lure melody pairs in the C major scale similar to Ortmann's sequences. All melodies were isochronous and their initial note was C4 or C5. Unlike Ortmann's sequences, the last note varied across melodies. We also controlled for several properties of the melodies such as the number of direction changes, note repetition, the position of the note that is changed to create the lure item. Three recognition experiments were conducted: a 2-alternative forced choice (2AFC) test (Experiment 1), a same/different test (Experiment 2) and a same/different test with feedback and silent retention interval (Experiment 3). The purpose for using two types of recognition tests was to see whether these tests are indeed interchangeable as is often assumed in the literature. Musical short-term memory studies either use one or the other, and when surveying empirical findings, authors more often than not address this difference as potentially tangential for observed results (e.g., Dowling, 1978, Dowling & Fujitani, 1971). Yet, many studies in verbal and visual working memory show that test setups do have considerable effects on performance (cf. Makovski, Watson, Koutstaal, & Jiang, 2010).

In addition, we were interested in looking at interference effects when a retention interval of fixed duration was a silent or filled interval. In Experiments 1 and 2, there were four types of manipulation: for one group of participants, the retention interval was unfilled. For the other three groups, it was filled with a sequence of five nonsense syllables, a 5-tone nondiatonic melody, or a 5-tone diatonic melody, respectively. By using musical versus nonmusical interference, we attempted to put the two different models of short-term memory for melodies to test. Whereas Baddeley's (2000) model proposes that any kind of acoustic information, be it verbal or musical, is processed in the phonological loop subsystem, Berz' (1995) model proposes a separate "slave system" specifically for musical information. If

participants performed equally bad in the nonsense syllables, nondiatonic and diatonic sequence conditions, this would be evidence for Baddeley's rather than Berz' model. If, however, they performed worse in the musical compared to nonmusical interference conditions, this would constitute support for Berz' rather than Baddeley's model. To our knowledge, no one has so far put these to test in a direct manner.

Our last research problem was whether the serial position effect is affected by different types of retention interval manipulations. It could be possible that when the retention interval is filled, the recency effect is diminished (cf. Atkinson & Shiffrin, 1968; 1971). Although Dowling (1973) found a J-shaped serial position curve for musical stimuli similar to the ones found with verbal material, how this curve is affected by different types of interference has not been researched yet in the music memory literature.

2.1 Methods

2.1.1 Experiment 1: 2AFC test

Participants. Seventy-two Boğaziçi University undergraduate non-musician students with an average of 1.33 years of musical training (ranging from 0 to 4 years) participated in the study in return for course credit in introductory psychology courses. None reported having absolute pitch or a hearing impairment.

Target and lure melodies. Using MATLAB, we created all possible 5-tone sequences in C major scale starting with C4 or C5 within the range of C4 and C5. Occasionally, these sequences will be referred as melodies, although they might not give a strong sense of musicality. After the creation, we selected the sequences which have note repetitions at most twice, and never in succession. We divided these

sequences according to the number of direction changes and starting note. Then, 30, 48, 48 and 32 melodies (with a total of 158) were randomly sampled from the set of melodies with zero, one, two and three direction changes, respectively. The number of melodies starting with C4 and C5 for a particular number of direction changes were equal. We also controlled for the position of the notes of direction changes. For example, the direction change occurred at the second, third or fourth note with equal probability for all melodies with only one direction change.

For every melody, we also created a contour preserving lure. The lures were obtained by changing the second, third, fourth or the fifth note of the target melody by one diatonic note while controlling for the number of direction changes, the position of notes of direction change, starting note and whether the changed note is higher or lower (see Figure 5). Also, of 39 melodies with a repeated note, in only 3 trials the changed note was a note of repetition. Finally, we made sure that none of the lures was the same as a target melody. Thus, the creation process of the lures was pseudo-random.

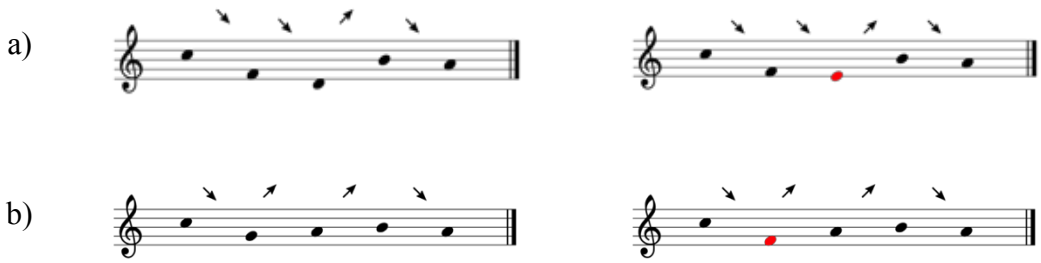


Figure 5 (a) A target and its lure with two direction changes. (b) A target with note repetition and its lure[3]

Intervening Stimuli. In each trial of the experiment, there was a brief retention interval. This retention interval was either silent (empty), or filled with

[3] In Figure 5a, the first note of the melodies is C5. The lure was created by changing the 3rd note at which a direction change occur in the melody. In Figure 5b, the lure was created by changing the second note of the target melody.

nonsense syllables, or diatonic sequences, or nondiatonic sequences across different groups of participants.

For the intervening nonsense syllables condition, a set of 23 nonsense syllables was created via an online text to speech converter (texttospeech.org). The duration of each syllable was adjusted to 0.25s using Adobe Audition 3.0. For every target-lure melody pair, we assigned a randomly created 5-syllable sequence. Note that the durations of target, lure and the 5-syllable sequence were equal.

For the intervening diatonic sequences, we created 158 novel diatonic 5-note sequences (not necessarily starting with C4 or C5) and assigned each diatonic sequence to a target-lure melody pair. The intervening diatonic sequences never contained the critical notes that were changed. For example, for the target-lure melody pair in Figure 5a, the diatonic sequence could not contain D or E. Also, the diatonic sequences were never a transposition of a target, lure or another intervening melody.

The nondiatonic sequences were created by moving at least three notes of the corresponding diatonic sequence to one of the nearest nondiatonic notes. The pitch averages of the 158 diatonic and nondiatonic sequences were equal.

All target, lure, diatonic and nondiatonic sequences were played with the real sound samples of Steinway Grand Piano in Logic Pro X software at a speed of 4 notes per second. Thus, each melody lasted exactly for 1.25 seconds.

Apparatus. The experiment was conducted in cubicles of the Cognitive Processes lab and the Social Psychology lab at Boğaziçi University. Over- and on-ear headphones with comparable quality and characteristics (Philips SHP 1900, Urbanears Plattan 2, Sennheiser Momentum On-Ear) were used. We used the PsychoPy 3.0 (Peirce, 2007) or E-prime software to create the experiments.

Procedure. Participants were tested in groups of up to four. Each participant was randomly assigned to either the “silence”, “nonsense syllables”, “nondiatonic sequences” or “diatonic sequences” conditions. The experiment consisted of 4 training trials followed by 158 experimental trials in random order. The doors of the cubicles were closed during the experiment.

In each trial, participants listened to a target melody followed by a retention interval which was either unfilled (silence), or filled with nonsense syllables, or a nondiatonic or diatonic sequence. After each trial, participants received a 2-alternative forced-choice test (2AFC), in which they decided which of the comparison melodies the same as the target (see Figure 6) and provided a confidence rating on a 3-point scale as “not sure”, “sure” or “very sure”. Target-lure position at test was counterbalanced across trials. No corrective feedback was given. The experiment lasted about 35 minutes.

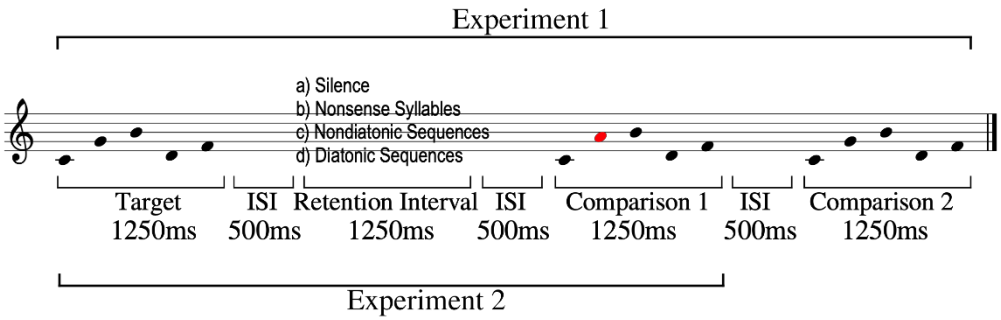


Figure 6 The experimental designs of Experiment 1 and 2

2.1.2 Experiment 2: Same/different test

Participants. Fifty-eight Boğaziçi University undergraduate non-musician students (who did not participate in Experiment 1) with an average of 1.25 years of musical training (ranging from 0 to 4 years) participated in the study in return for course

credit in introductory psychology courses. None reported having absolute pitch or a hearing impairment.

Apparatus and stimuli. The apparatus and the stimuli were the same as the first experiment.

Procedure. The procedure was similar to that of Experiment 1 except that only one comparison melody was presented after the retention interval. The participant decided whether the target and the comparison melody were the same or different. Same and different trials appeared in random order and equal likelihood across trials. Participants again rated their confidences on a 3-point scale. The experiment lasted about 25 minutes.

2.1.3 Experiment 3: Same/different test with feedback and silent retention interval

Participants. Twenty Boğaziçi University undergraduates who did not participate in the earlier experiments were recruited in return for course credit in psychology courses. The average duration of musical training was 1.30 years. Again, no participants had an absolute pitch or a hearing impairment.

Apparatus and stimuli. The apparatus and the stimuli were the same as in the previous experiments.

Procedure. The design of this experiment was identical to Experiment 2 except that the retention interval was always unfilled, and at each trial after each decision, participants received a “correct” or “incorrect” feedback screen that appeared for 1 s. The duration of the experiment was about 30 minutes.

2.1.4 Memorability and complexity ratings

Participants. Another 19 Boğaziçi University students participated in the study in return for course credit in psychology courses. The average duration of musical training was 1.32 years. Again, no participants had an absolute pitch or a hearing impairment.

Apparatus and stimuli. The apparatus and the stimuli were the same as the previous experiments, but without the memory test components.

Procedure. Participants were tested in the same environment as the previous experiments. Before the experiment, participants were informed that they were going to rate melodies based on complexity and memorability. Complexity was defined to them as how unusual ('complexity' was explained in Turkish using the terms "karmaşık" for "complex" and "sıradışı" for "unusual") or not ('noncomplex') the consecutive notes and intervals sounded to them. Memorability, on the other hand, was simply referred to as is and none of the participants asked for a clarification. All instructions also appeared on the computer screen before the experiment.

The experiment consisted of four training trials followed by 158 experimental trials with no interruption between the training and experiment. In each trial, participants listened to a target melody and rated both the memorability and complexity of the melody over a 5-point scale. One group of participants rated memorability first, while the other group rated complexity first.

2.2 Results

First, we calculated the correct response rate for all participants. The data of seven participants (all from Experiment 2) whose performance fell below .52 and rated at least 120 of 158 their responses as "very sure" were excluded. We also excluded the

data of a participant from Experiment 1 who rated all responses as "very sure" yet performed nearly at chance level (.54). Throughout the results section, interactions between variables will be reported only if $p < .10$.

2.2.1 Experiment 1: 2AFC test

All analyses for Experiment 1 used the average correct responses as the dependent variable. Student's *t* tests showed that participants performed above chance level in silence ($M = .67$, $SD = .08$, $t(15) = 8.44$, $p < .001$, Cohen's $d = 2.11$), nonsense syllables ($M = .63$, $SD = .12$, $t(19) = 5.00$, $p < .001$, Cohen's $d = 1.12$), nondiatonic ($M = .61$, $SD = .09$, $t(17) = 5.29$, $p < .001$, Cohen's $d = 1.25$) and diatonic sequences ($M = .58$, $SD = .09$, $t(17) = 3.70$, $p < .01$, Cohen's $d = .87$) conditions.

The type of retention interval had a significant effect on recognition. A one-way ANOVA with Tukey's HSD post-hoc test revealed that participants in the silence condition performed significantly higher than those in diatonic sequences condition, $F(3,68) = 3.02$, $p < .05$, $\eta_p^2 = .12$ (See Figure 7).

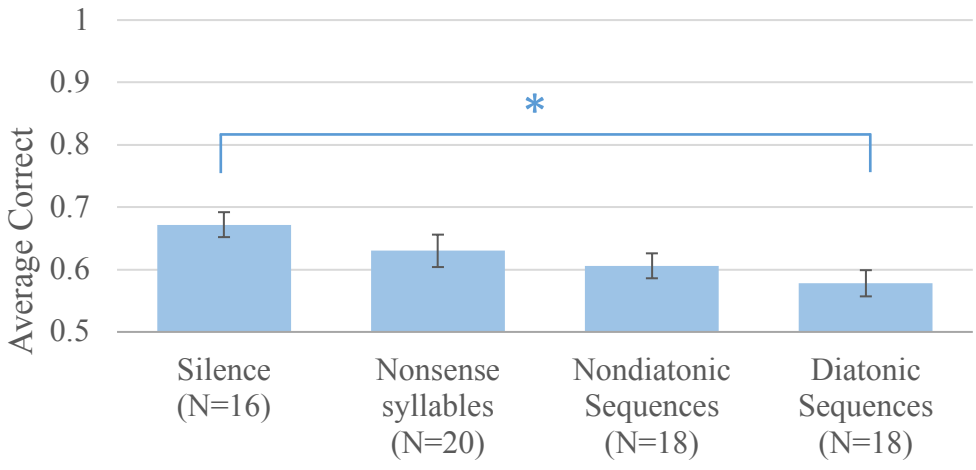


Figure 7 Average correct recognition across groups. Error bars (as in all remaining graphs) indicate standard errors

Many participants stated in writing or verbally that the duration of the experiment was long. We therefore tested whether a fatigue or practice effects

occurred. A 2 (order) x 4 (groups) mixed ANOVA showed that participants performed equally well in the first and the last half of the trials, $p > .10$.

Taylor and Pembrook's set of melodies contained successive and non-successive note repetitions, and they investigated the error rates at these notes. On the other hand, our stimulus set contained only non-successive note repetitions. A 2 (repetition)[4] x 4 (groups) mixed ANOVA indicated that note repetition did not have a significant main effect, $p > .10$, showing that melodies that had note repetitions were not better (or worse) remembered than those without.

We created the lure items by moving any of the notes except for the first note in the target melody by one diatonic note upwards or downwards. A 2 (higher vs. lower) x 4 (groups) mixed ANOVA showed that participants detected changes more easily when the lure melody had a diatonic upwards than downwards change, $F(1,74) = 4.73$, $p < .05$, $\eta_p^2 = .06$. In other words, discrimination performances were better when the lure item contained a higher note.

We were also curious about listeners' metacognitive awareness when giving their yes responses. A 3 (confidence) x 4 (groups) mixed ANOVA indicated that participants' performances differed across levels of confidence, $F(1.37,93.21) = 14.60$, $p < .001$, $\eta_p^2 = .18$. In particular, performances for trials rated as "very sure" were significantly greater than other trials, $p < .001$ (see Figure 8).

We also plotted recognition accuracy and response confidence (see Figure 9). There was a strong positive correlation between correct recognition rates and average confidences for each question, $r = .704$, $p < .001$.

[4] The two levels of the within subjects variable "repetition" were the average recognition performances for the 39 trials with repetition and 119 trials without repetition.

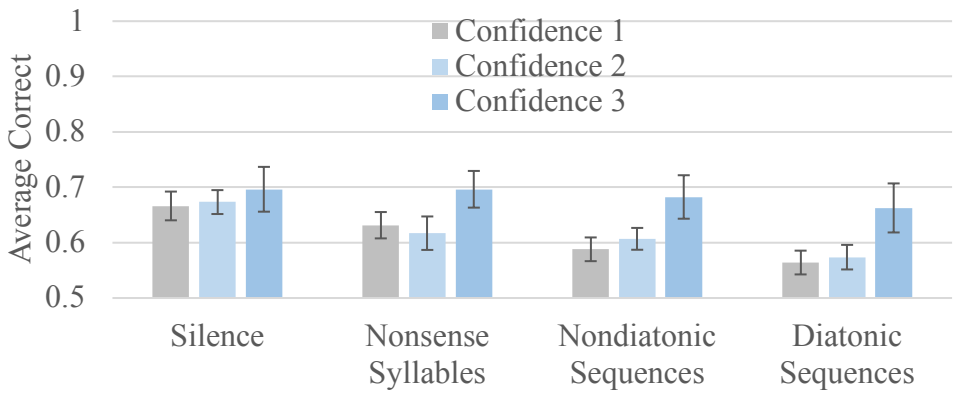


Figure 8 Average correct recognition by confidence across groups

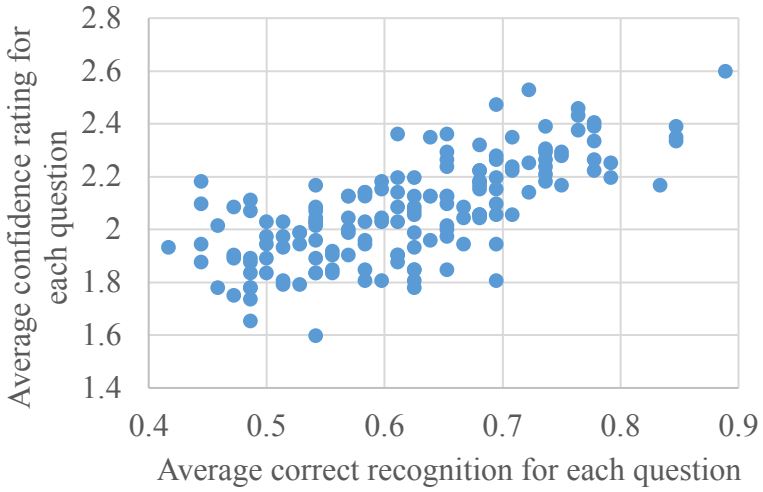


Figure 9 Scatterplot of average recognition rate and average confidence rating for each question

To check for serial position effects, we conducted a 4 (serial position of the changed note) x 4 (groups) mixed ANOVA which showed that the serial position of changed note had an effect on performance $F(3,204) = 19.29$, $p < .001$, $\eta_p^2 = .22$. We predicted to find a J-shaped serial position curve (recency > primacy > center) and tested this with planned comparisons at an overall $\alpha = .05$. We found that changes in the last note were detected better than changes in the second note; and changes in the second note were detected better than changes in the center notes (see Figure 10), which reflects a J-shape curving.

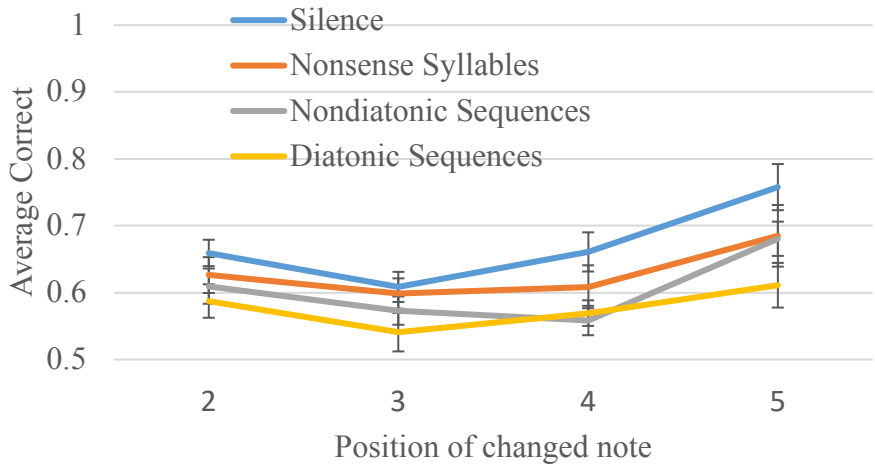


Figure 10 Average correct recognition based on position of changed note within 5-tone sequence across groups

Taylor and Pembrook (1983) stated that participants recalled the contour of a melody more correctly if the melody has zero or one direction changes rather than two or three direction changes. As our lure items never violated the contour of the target melody, our analyses regarding the number of direction changes will directly measure the correct recognition of intervals. If Taylor and Pembrook's findings are limited only to contour recall, we might not find any difference between recognition of melodies with different number of direction changes. We tested this effect with a 4 (number of direction changes) x 4 (groups) mixed ANOVA. Results indicated that number of direction changes of a melody had a main effect on recognition, $F(3,213) = 2.80$, $p < .05$, $\eta_p^2 = .04$. Moreover, post-hoc analyses showed that participants recognized melodies with no direction changes better than those with one or three direction changes, $p < .05$ (see Figure 11). Performance did not differ for fully ascending and descending sequences, $p > .10$.

Taylor and Pembrook classifies melodies as ascending when they start with C4 (i.e. start with an ascending movement) and as descending when they start with C5. Based on Ortmann's hypothesis, they expected and found that there are fewer

errors in ascending sequences. We confirmed their findings. A 2 (C4 vs C5) x 4 (groups) ANOVA indicated that participants performed better when the target melody started with C4 rather than C5, $F(1,68) = 7.69$, $p < .01$, $\eta_p^2 = .10$. We also found a significant interaction, $p < .05$. In particular, for all groups except the diatonic sequences, performance for melodies starting with C4 was higher, $p < .01$.

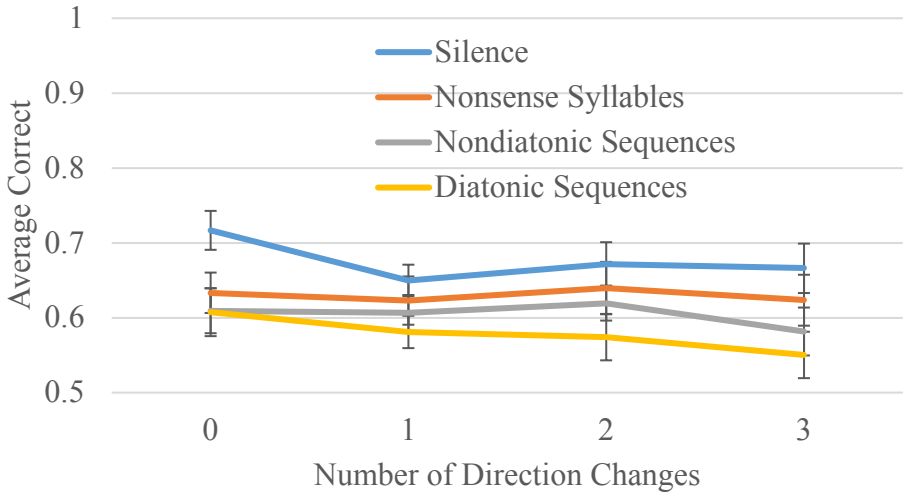


Figure 11 Average correct recognition by number of melodic direction changes across groups

We also looked at whether changes in "corner" as opposed to "slope" notes (cf. Dyson and Watkins, 1984) affected recognition performance differently. The number of trials where the changed note was a "corner" and "slope" was 60 and 58, respectively. The remaining 40 trials were those where the change occurred at the last note. A 2 ("corner" vs "slope") by 4 (groups) mixed ANOVA[5] revealed a significant effect of the direction of the melody at the changed note, $F(1,68) = 12.50$, $p < .01$, $\eta_p^2 = .16$. Participants' performances were better when the changed note was a "corner" rather than a "slope" note, $p < .05$ (see Figure 12).

[5] The 40 trials with the last note change were excluded from this analysis since these notes cannot be classified as a "corner" or "slope".

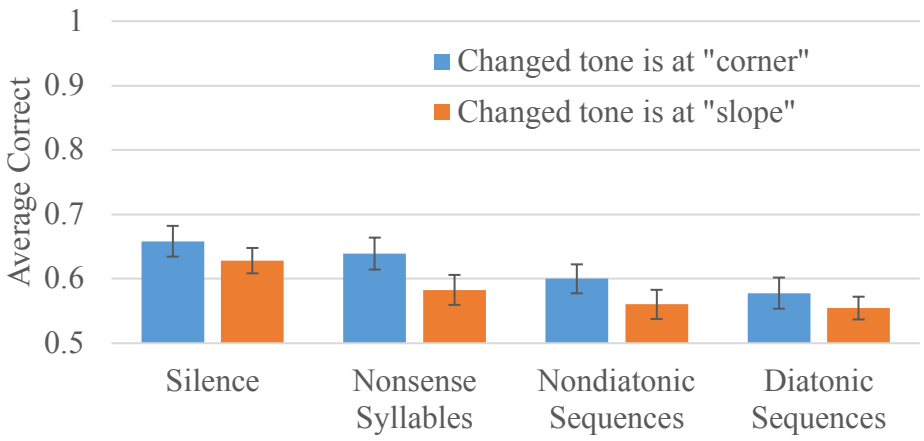


Figure 12 Average correct recognition rates when the change did/did not occur at a "corner" note

Finally, we replicated Taylor and Pembrook's (1983) findings concerning the interval before a note of direction change. Specifically, we considered the trials where the change occurred at a note of direction change and compared the effect of the interval size before that note. A 2 ("step" vs "skip") x 4 (groups) ANOVA revealed that participants performed better when the interval before a note of direction change is a "skip" rather than a "step" interval, $F(1,68) = 9.08$, $p < .01$, $\eta_p^2 = .12$ (see Figure 13).

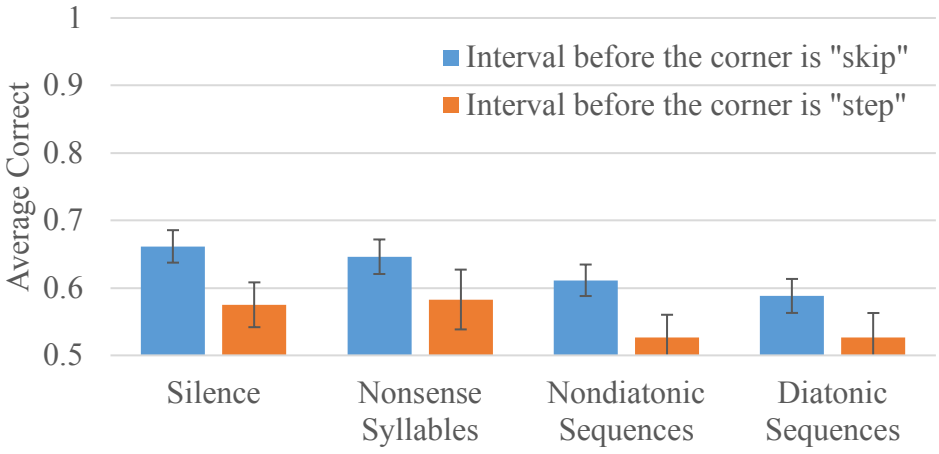


Figure 13 Average correct recognition rates when the changed note is a "corner" and the preceding interval is a "step" vs a "skip"

2.2.2 Experiment 2: Same/different test

Since same/different tasks are likely to be more vulnerable to bias effects compared to simple 2AFC tasks, we transformed all participant data into AUC data (Figure 14 shows a clear bias to say "same" for any given trial).

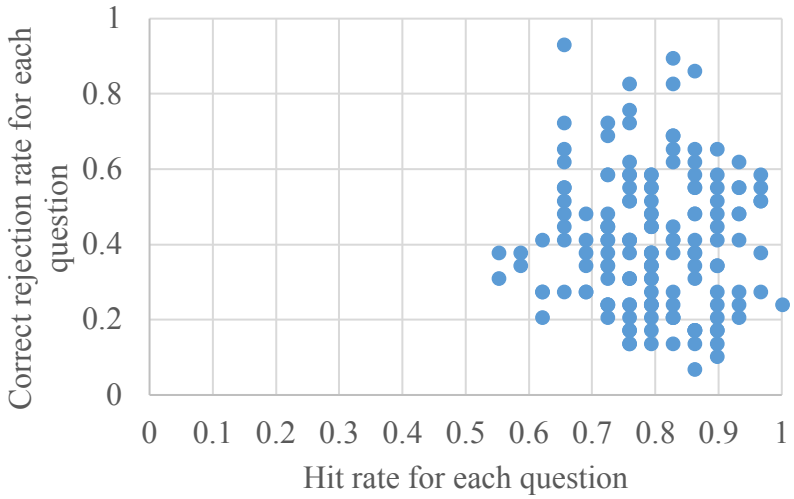


Figure 14 Scatterplot of hit and correct rejection rates for each question

One sample *t*-tests showed that all four groups performed above the chance level (AUCs were significantly greater than .50), $p < .001$. Moreover, a one-way ANOVA indicated that there was a significant discrimination difference across groups, $F(3,54) = 7.66$, $p < .001$. Tukey's post-hoc test showed that silence and nonsense syllables groups performed greater than the nondiatonic and diatonic sequences groups (see Figure 15). In particular, interference with musical material caused more performance decrease than interference with verbal material.

A 2 (first vs second half) x 4 (groups) mixed ANOVA indicated that participants' performances as measured in AUCs did not differ between the first and the last half of the trials, $p > .10$. Melodies with note repetition had a marginal advantage over melodies without note repetition, $p = .07$. However, this marginal effect appears to be created by an interaction with groups. In particular, only the

group that received interfering diatonic sequences performed better with melodies with repetition than other melodies, $p < .01$.

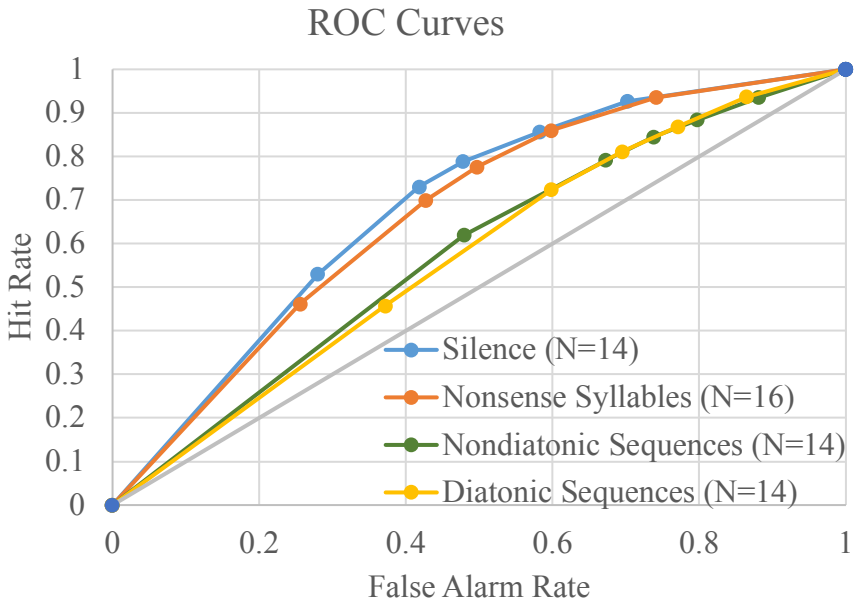


Figure 15 Receiver operating characteristics (ROC) curves across groups

A 2 (higher vs. lower) x 4 (groups) mixed ANOVA showed that in "no signal" trials (when the answer is "different"), participants' performances did not differ between the trials where the lure item had a heightened or lowered note, $p > .10$.

A 4 (serial position of the changed note) x 4 (groups) mixed ANOVA, on the other hand, showed that the serial position of the changed note has an effect on recognition, $F(2.47,133.26) = 34.48$, $p < .001$, $\eta_p^2 = .39$. To check for the expected J-shaped curve, we conducted planned comparisons at an overall α level of .05. The results confirmed our hypothesis: changes in the last note were detected better than changes in the second note which were detected better than changes in the center (third and fourth) notes (see Figure 16). We also found an interaction between the serial position of the changed note and groups, $F(7.40,133.26) = 4.83$, $p < .001$, $\eta_p^2 =$

.21. In particular, musical interference caused a reduction in the recency effect (see Figure 16).

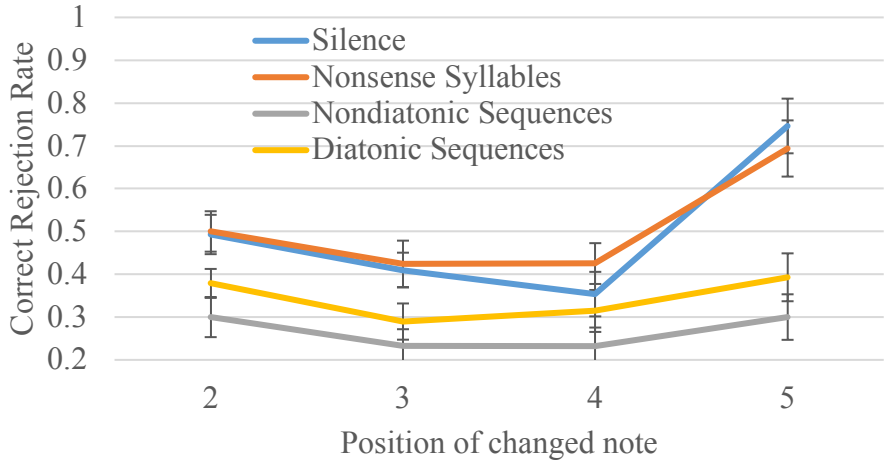


Figure 16 Average correct rejection rates based on position of changed note within 5-tone sequence across groups

The number of direction changes also had an effect on discrimination performance. A 4 (number of direction changes) x 4 (groups) mixed ANOVA with Tukey's post-hoc test indicated that participants performed better in recognizing melodies with zero or two direction changes than with three direction changes, $F(2.55,137.41) = 2.89$, $p < .05$, $\eta_p^2 = .05$ (see Figure 17).

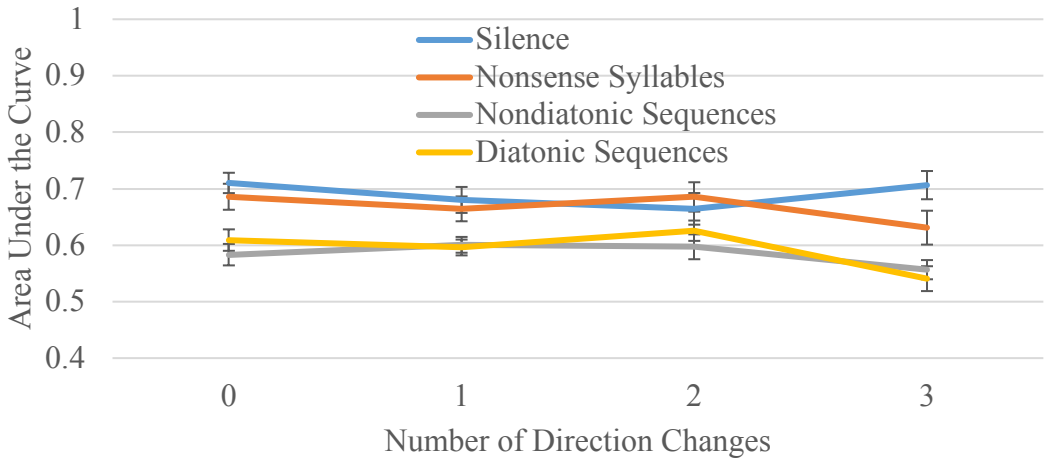


Figure 17 Average AUC values by number of melodic direction changes across groups

A 2 (C4 vs C5) x 4 (groups) mixed ANOVA showed that the discrimination performance for melodies starting with C4 was significantly greater than melodies

starting with C5, $F(1,54) = 11.86$, $p < .01$, $\eta_p^2 = .18$. There was again no significant difference between the performance for fully ascending and fully descending melodies, $p > .10$.

A 2 ("corner" vs "slope") x 4 (groups) mixed ANOVA showed that there was a main effect of the contourwise position of the changed note on performance. When the changed note was at a "corner" rather than a "slope", participants performed better, $F(1,54) = 6.34$, $p < .05$, $\eta_p^2 = .11$ (see Figure 18).

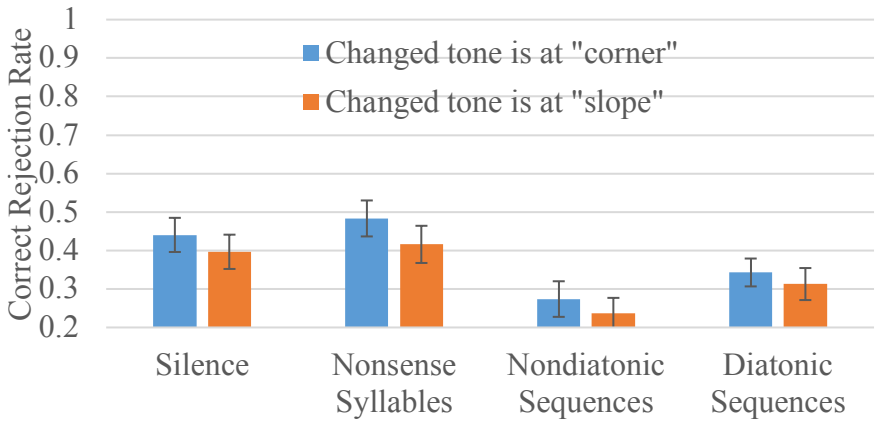


Figure 18 Average correct rejection rates when the change did/did not occur at a "corner" note

Contrary to Experiment 1, we failed to find the "skip effect". Given that the changed note is a "corner", participants' performances did not change according to the interval before the "corner" was a step or a skip, $p > .10$. In other words, a changed corner note was not better detected if it was preceded by a larger (skip) interval as was found in Experiment 1.

2.2.3 Experiment 3: Same/different test with feedback and silent retention interval

In the previous experiment, participants were mostly biased to give "same" responses whenever they were not sure whether the two melodies were identical or not. We wanted to increase participants precision (and hence decrease their biased response tendencies) by providing corrective feedback after each trial. Figure 19 shows a

scatterplot distribution of hit and correct rejection rates. All subsequent analyses were performed on AUC and correct rejection rates.

In a same/different test, a "same" response is either a hit, or a false alarm. So, a decrease in "same" responses will either increase correct rejection rate, or reduce the hit rate (or both). With feedback, average hit rates decreased from .79 to .73 although this did not reach statistical significance, $t(32) = 1.50$, $p > .10$. On the other hand, correct rejection rates were marginally higher with feedback ($M = .61$, $SD = .16$) than without feedback ($M = .50$, $SD = .14$), $t(32) = 1.98$, $p = .057$. There was no significant effect of feedback on recognition performances measured by AUCs, $p > .10$.

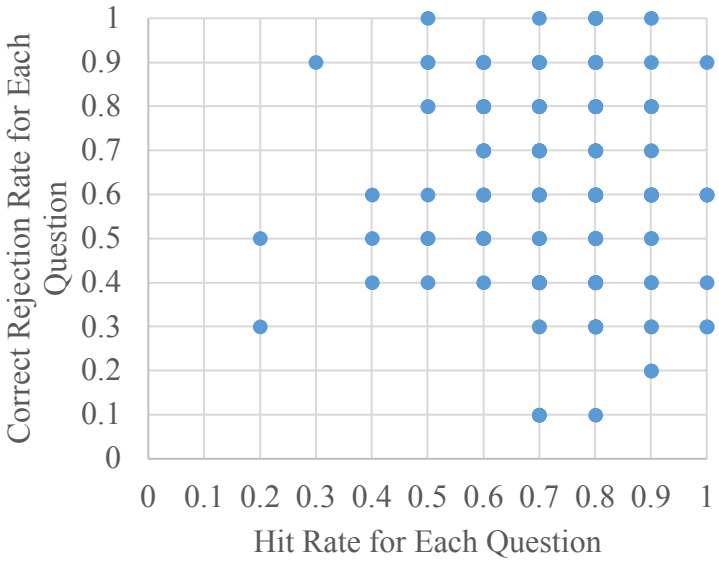


Figure 19 Scatterplot of hit and correct rejection rates for each of 158 trials in Experiment 3

A one-sample *t*-test on AUC scores showed that participants performed above chance level ($M = .70$, $SD = .11$, $t(19) = 8.04$, $p < .001$). Moreover, participants' "same" responses still exceeded the actual .50 rate ($t(19) = 3.17$, $p < .01$). So, even with feedback, participants did not use the two options equally often. However, when comparing Experiment 3 with the silence group of Experiment 2 descriptively, we see fewer "same" responses in the feedback than the no-feedback group, M_{fdback} =

.56, SD_{fdback} = .09, $M_{no\text{-}fdback}$ = .64, $SD_{no\text{-}fdback}$ = .12, respectively. In other words, the magnitude of the "same bias" appeared to decrease when feedback was provided.

We then looked at the effect of corrective feedback across two blocks of trials to see whether feedback improved performance and whether this improvement was more pronounced in the second compared to first block of trials. A 2 (with/without feedback in silence condition) x 2 (blocks) mixed ANOVA showed that participants performed equally well in each block of trials, $F(1,32) = .52$, $p > .10$. No interaction between feedback and block was observed, $p > .10$. When plotting our data for each stimulus manipulation we found more or less the same effects we observed in Experiment 2. Except for serial position (Figure 20)[6], none of our manipulations showed effects different from those observed earlier (e.g., Figure 21). In other words, corrective feedback improved overall recognition performance.

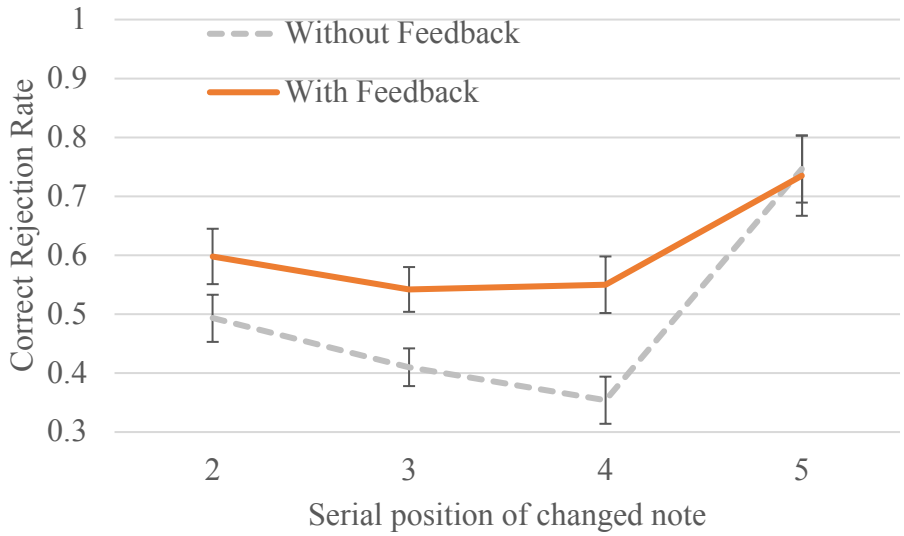


Figure 20 Average correct rejection rates based on position of changed note within 5-tone sequences across groups

[6] A 4 (serial position of the changed note) x 2 (feedback) yielded a main effect of feedback as well as a significant interaction. Participants performed better with feedback when the 2nd, 3rd or the 4th note was changed. However, no difference was observed when the 5th note was changed possibly because of the recency effect.

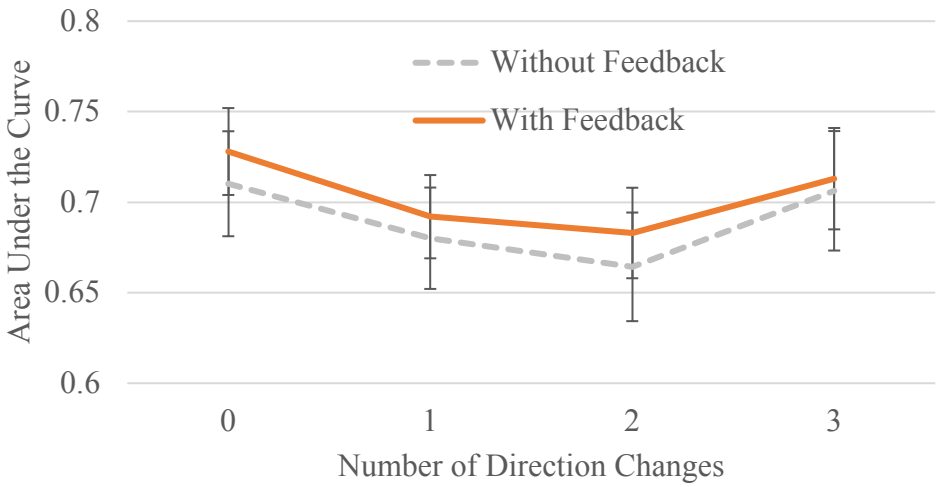


Figure 21 Average AUC values by number of melodic direction changes across groups

2.2.4 Memorability and complexity ratings

Our results so far clearly showed that certain melody characteristics had an influence on recognition performance. To better understand whether listeners had some metacognitive awareness of these differences, we decided to collect memorability and complexity ratings for the target melodies, hence independent of their particular lure versions.

First, we tested for whether rating memorability or complexity first made a difference. Two independent samples *t*-tests for participants' average memorability and complexity ratings, respectively, showed that question order did not matter, *ps* > .10. Then we wanted to see whether memorability and complexity correlate or not, and it did[7] (Spearman correlation coefficient r_s = -.70, p < .001). An item- instead of participant-wise analysis confirmed this relationship, r = -.69, p < .001 (Figure 22).

[7] On the other hand, the memorability and complexity ratings did not correlate with AUCs or hit rates. i.e. they were not good in predicting which note sequences they would remember better.

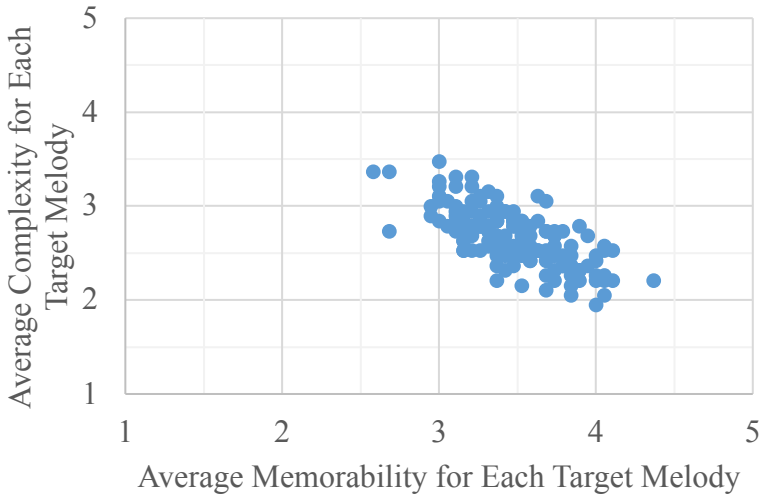


Figure 22 Scatterplot of average memorability and complexity for each target melody ratings

A one-way repeated measures ANOVA indicated that the number of direction changes of a melody had an effect on memorability ratings, $F(1.37,24.71) = 15.75$, $p < .001$, $\eta_p^2 = .47$ as well as complexity ratings $F(1.36,24.41) = 7.88$, $p < .01$, $\eta_p^2 = .31$. Post-hoc analyses showed that melodies with no direction changes were rated as significantly more memorable and less complex than melodies with one direction change, $p < .05$ (Figure 23). Also, memorability ratings of melodies with one direction change were significantly higher than melodies with two or three direction changes, $p < .001$. Similarly, melodies with one direction change were rated as less complex than melodies with two or three direction changes, $p < .05$. Participants' memorability and complexity ratings did not differ between melodies with two or three direction changes, $p > .10$. When melodies without any direction changes were considered, the memorability and complexity ratings did not differ between fully ascending and fully descending sequences, $p > .10$.

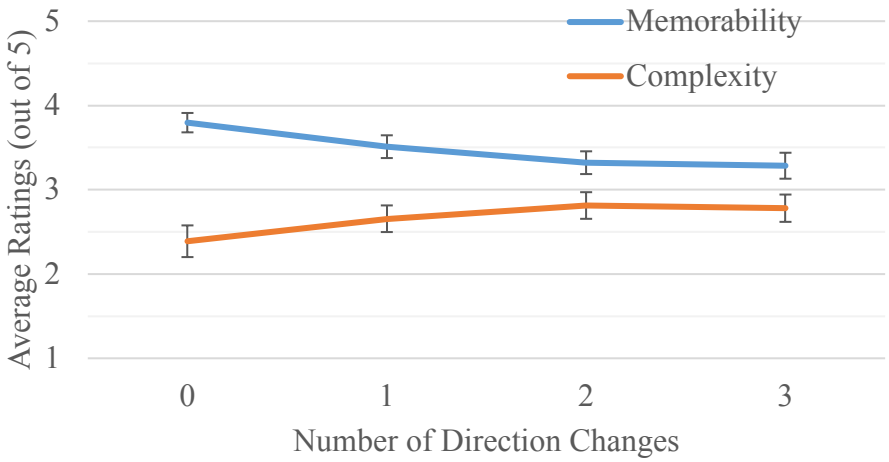


Figure 23 Average memorability and complexity ratings by number of direction changes

An interesting finding is that although participants' memorability ratings for melodies that start with C4 or C5 did not differ ($p > .10$), participants rated melodies that started with C4 to be less complex than melodies starting with C5, $t(18) = 2.25$, $p < .05$, $\eta_p^2 = .22$ (Figure 24). This is parallel to our previous findings since participants' recognition performances were found be better for melodies starting with C4. Finally, note repetitions did not affect memorability or complexity ratings, $p > .10$.

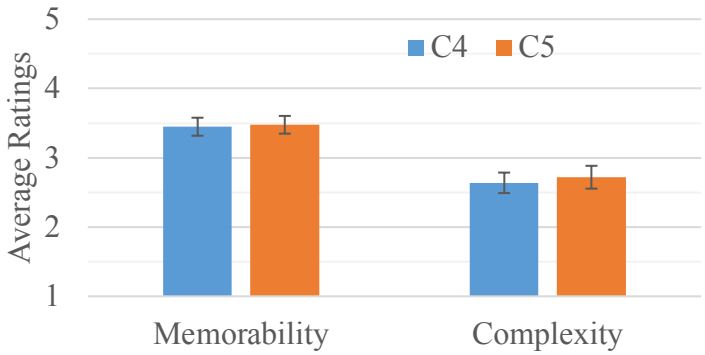


Figure 24 Melody ratings based on initial note

2.2.5 Best-worst melodies

As a last step in our exploration, we wanted to take a closer look at the melodies which yielded the highest (Figure 25) and lowest (Figure 26) recognition rates in Experiment 1 (silence), Experiment 2 (silence) and Experiment 3.

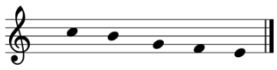
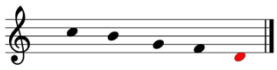
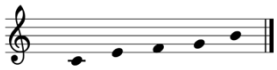
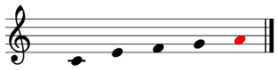
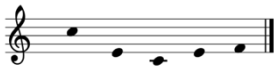
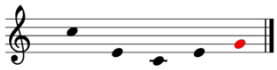
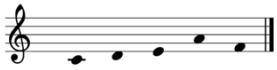
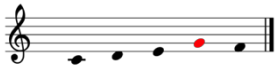
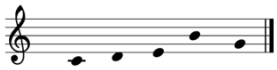
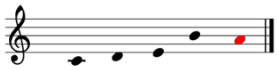
Target-Lure Pair	Target Melody	Lure	Expt 1. (silence) Hit Rate	Expt. 2 (silence) AUC	Expt. 3 AUC	Average Memorability	Average Complexity
15			.94	.86	.95	3.79	2.37
23			94	1.00	.82	3.63	2.53
45			.88	.94	.84	3.21	2.68
76			.88	.87	.90	3.47	2.58
77			1.00	.89	.87	3.47	2.47

Figure 25 The top five target-lure melody pairs that yielded the highest recognition rates in all experiments at silence conditions

The target and the lure differed at the 5th note in four of the five melody pairs presented in Figure 25. The other melody had its changed note at a "corner" whose preceding interval was a "skip". Also, all melodies had zero or one direction change. The lure of pair 45 was an arpeggio of the C-major chord and the lure of pair 76 was a permutation of C-D-E-F-G notes.

For four of the five pairs presented in Figure 26, the changed note was at the center (3rd or 4th notes). Three melodies had two or three direction changes. Although three melodies had their changed notes at "corners", for two of them, the interval before the changed note was a "step".

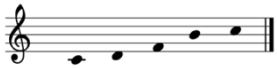
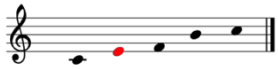
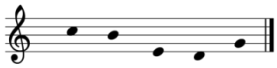
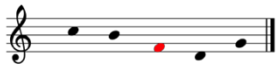
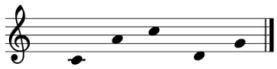
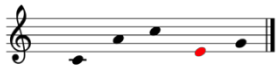
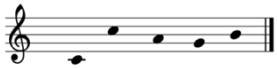
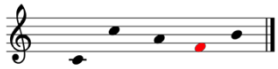
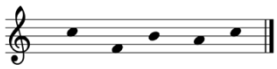
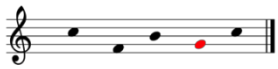
Target-Lure Pair	Target Melody	Lure	Expt 1. (silence) Hit Rate	Expt. 2 (silence) AUC	Expt. 3 AUC	Average Memorability	Average Complexity
16			.25	.49	.54	4.11	2.53
49			.44	.52	.51	3.11	2.84
107			.44	.18	.51	3.21	2.79
116			.44	.53	.40	3.11	2.95
140			.31	.54	.54	3.26	2.89

Figure 26 The bottom five target-lure melody pairs that yielded the lowest recognition rates in all experiments at silence conditions

CHAPTER 3

DISCUSSION

Taylor and Pembrook (1983), as a follow up of Ortmann's (1933) study, aimed to explain tonal factors that affect short term memory for isochronous and diatonic 5-note melodies. We wanted to elaborate their research by using a larger and more controlled stimulus set and altering their free recall test setting to a recognition test setting. The main findings of their research that relates to our study were as follows: (1) correct contour recall performance was greater when the melody had fewer direction changes, (2) participants did fewer mistakes with ascending (starting with an ascending movement) than descending sequences, (3) recall for a note of direction change was higher if it was preceded by a larger (which is an interval of size at least 3 semitones) than a smaller (2 semitones or less) interval, (4) fewer errors were made at the last note (which was always a C).

Although our stimuli were basically similar to theirs in the sense that they were isochronous 5-tone diatonic sequences in C major scale, there were some significant qualitative differences. First, our sequences were constructed in a random but controlled way, independent of melodicality and tonality. Secondly, we controlled for the number of direction changes as well as the serial position of the notes of direction changes. For instance, Taylor and Pembrook's melodies with one direction change had full melodic symmetry (e.g. C-F-A-F-C). Thus, those melodies had two repeated notes and the note of direction change was always the third. We also made sure that note repetitions appeared at most once and never consecutively to keep the stimuli more or less comparable in terms of complexity. A C-F-A-F-C sequence would, for instance, easily match with an arpeggio-like schematic

representation hence making a note deviation much more detectable compared to a C-F-A-G-D sequence. Finally, unlike Taylor and Pembrook's melodies which always ended on the C note, our last notes were random to ensure that observed recency effects really measure retention rather than ad hoc knowledge.

The first aim of this study was to see whether their findings replicate in a recognition test setup (2AFC in Experiment 1, Same/different in Experiments 2 and 3). Secondly, we wanted to know how "fine-grained" the resolution of the melodic representation was for listeners with no musical training. As previous studies suggest that listeners mainly rely on contour information in brief delay settings (e.g., Dowling, 1991), we tested our targets against same-contour lures to address how well participants would correctly recognize exact intervals when contour information was no longer discriminatory. The third goal of our study was to know how different types of intervening events (silence, nonsense syllables, nondiatonic sequences, diatonic sequences) in the retention interval affect recognition performances. Then we will discuss whether our findings can be explained by Baddeley's (2000) or Berz' (1995) working memory model. Finally, we were curious to see how much listeners' memorability and complexity ratings for the melodies overlap with the actual recognition performances.

One of our main findings was that participants performed above chance level in all experiments and all types of retention interval or presence of feedback manipulations. This finding indicates that the "resolution" of the representations after a brief delay contains not only general contour but also quite exact pitch interval information. There are previous studies (mostly by Dowling and colleagues) which can partially be compared with our study. The first experiment in Dowling and Fujitani's (1971) study is perhaps the most analogous experiment in the literature as

they used untransposed lures in a same/different test with 2 second retention interval. The fundamental difference from our study was that their target melodies consisted of atonal sequences, and contour preserving lures were created by altering an interval rather than a note. They found high recognition accuracy when 5-tone atonal sequences were tested against contour preserving untransposed lures (AUC = .91). In Experiment 2a (same/different test with silent retention interval) of our study, the corresponding AUC value was .69. This demonstrates that the presence of tonality actually reduced recognition performance by reducing the saliency of the difference between the target and the lure.

All other relevant studies from Dowling and colleagues used transposed lures to focus on the retention of the melody as a Gestalt rather than the retention of its individual pitches. Dowling (1978) found that musically inexperienced listeners performed at chance level when 5-tone tonal targets were tested against transposed contour preserving tonal alternatives in a same/different test. So, if we had chosen our lure items to be transposed, performances would likely decline to chance level.

However, it was not always the case that same contour transposed lures yielded a near chance level performance. For example, Dowling (1991) and Dowling, Kwak and Andrews (1995) observed that participants showed above chance level performance in distinguishing tonal targets from contour preserving transposed lures. However, these studies are not really comparable to ours. For example, their targets were 7-tone sequences with small pitch intervals (Taylor and Pembrook suggest that such sequences are recalled easier) and the lure items were created by changing two instead of one note in the target melody. Moreover, they used a continuous running memory paradigm and the duration of the shortest retention interval was 7 seconds which is much longer than ours. An interesting

finding of these studies is that when both same- and different-contour lures are presented, participants' sensitivity to exact interval information for contour preserving lures becomes increased when delays were longer. Given that their shortest retention interval was much longer than ours, this may explain the above-chance performance they observed for same contour transposed lures.

One work which also found an above-chance performance in discriminating transposed targets from contour preserving lures was Byron and Stevens' (2006) study. In a same/different test, participants distinguished alterations between small (step) and large (leap) musical intervals when 8-tone sequences based on some Western folk tunes were tested against their contour- and implied harmony-preserving lures. We believe that the big magnitude of the alterations in the interval size of the changed note made the difference quite salient. Cuddy and Cohen (1976) did a 2AFC task with 3-note sequences based on arpeggios of the C major chord (like G-C-E) and tested them against transposed lures which were created by changing one note by one semitone, thus turning the major chord to a minor or augmented chord. In their case, above-chance performance could have been due to the emotions induced by different chords (cf. Parncutt, 2014).

The reason why we particularly wanted the lures to be untransposed is to mimic a natural setting (for instance, recognizing a freshly heard commercial melody with the additional help of instant pitch memory) and inspect the resolution of the exact short-term memory representations of the melodies. As contour is perhaps the most salient cue in short-term melody memory (e.g., Dowling, 1991), scale violating lures (even when contour is preserved) are detected very easily (e.g., Agres, 2019), and listeners' mistakes for incorrectly recalled intervals are within a range of 2

semitones (Taylor & Pembrook, 1983) we specifically had chosen our lures to be contour preserving, diatonic, and deviating by only one diatonic note.

Harrison, Musil and Müllensiefen (2016) tried to model the potential factors that affect melody discrimination performance. As our lure items were created in a very controlled way with the purpose of making all of them as challenging as possible, we believe that any difference in discrimination performance aside from the recency effect be caused by the complexity or memorability of the target item itself.

Before getting into details of what made target melodies more memorable and less complex, we will discuss the saliency of the changed note in the target-lure melody pairs. Taylor and Pembrook (1983) found a recency effect for the last note, but given that it was always a C, this effect was potentially meaningless. In all experiments, we replicated their findings even though our last notes were no longer predictable. However, we also found a primacy effect for the 2nd note. This gave us a classic J-shaped serial position curve such that changes in the fifth note were detected better than changes in the second note, and changes in the second note were detected better than changes in the third and fourth notes. However, in Experiment 2, when the retention interval was filled with musical material (nondiatonic or diatonic melodies) rather than nonsense syllables or silence, the magnitude of the recency effect decreased significantly. These findings are compatible with Oberauer et al.'s (2018) study where they define benchmarks for models of working memory. In that study, they classify primacy and recency effects as type A (strong) benchmarks, and it is stated that a filled distractor at the end of the to-be-memorized list reduces the recency effect. Furthermore, such detrimental effects of same modality interference in short-term storage have been observed since the time of the model proposed by

Atkinson and Shiffrin (1968, 1971). Similar effects are also predicted by Baddeley's working memory model (2000; also cf. Salame & Baddeley, 1989).

Our retention interval manipulation was quite uncommon. Although silent retention intervals as brief as ours are typical in Dowling's studies, to our knowledge, there are no other studies that put several types of interference in such a short duration. A somewhat related study is by Mikumo (1992). She did a same/different test with 6-note tonal and atonal sequences. The duration of the retention interval was 12-seconds and it was either unfilled, or filled with an interfering melody or a verbal shadowing task. She found that the performances of musically highly trained participants were more negatively affected by the "shadowing with note names" task than the "shadowing with nonsense syllables" task or silence. Although it is stated in the article that no effect of different types of interfering stimuli was observed for the musically less trained participants, we believe that this null finding is due to the atonal sequences. However, since the nonsense syllables condition in her study contained a shadowing task but the interfering melody condition did not, and the retention interval was about 5 times longer than ours, it would be conceptually wrong to compare her findings with our study.

Our findings give us a chance to compare different models of working memory. Baddeley's (2000) model suggested that all types of auditory information (including verbal and musical sounds) are stored in a "phonological buffer". On the other hand, Berz' proposed a slave system à la Baddeley specifically for musical information. Our findings from Experiment 2 (same/different test) are quite compatible with Berz' hypothesis: recognition performances decreased when the retention interval was filled with musical material. Results from the 2AFC test showed a similar a stepwise decrease from silence to diatonic sequences conditions,

even if the only significant difference was observed between these two. Baddeley's model would have predicted a difference between silence and all other conditions. In that sense, results from Experiment 1, too, appear more in line with Berz' model. Oberauer et al. (2018) suggest that filled retention intervals that prevent rehearsing cause a decrease in performance. On the other hand, they state that studies on the effect of silent retention intervals are contradictory. More statistical power may be needed to see more clear effects when comparing silent conditions to non-silent ones.

Harrison et al., for instance, define melodic complexity in terms of number of notes, frequency of unique intervals etc. However, since all our sequences had a fixed number of notes and comparable intervals, their measures of complexity would produce the same value for all target melodies. Likewise, statistical models such as IDyOM (Information Dynamics of Music; Pearce, 2005, 2018) might also shed some light as to why those specific melodies turned out to be the best and worst recognized ones. Finally, the worst recognized melodies might need further inspection since they were all recognized below chance-level. This implies that their lure counterpart somehow appeared more "correct". Again, statistical models might help us understand why this is so, i.e., what exactly makes a 5-tone sequence sound more like a just-heard target, even if falsely so.

To address the melodic complexity problem, we collected memorability and complexity ratings over a 5-point scale for the target melodies. We observed a large effect size with respect to number of direction changes of the melody. Participants rated fewer direction changes as less complex and more memorable[8]. This is a partial confirmation of our findings from the experiments. In all of our experiments,

[8] One might think that melodies with higher number of direction changes may be parsed as more than one unit. However, since our melodies contained only five notes, it was impossible to create smaller chunks in nearly all of them.

participants performed better when the target melody was fully ascending or fully descending (i.e., with no direction changes).

Taylor and Pembrook observed that correct contour recall performance was greater when the melody had a simple contour (fewer direction changes). As our lure items never violated the contour of the target melody, any response in our experiments was actually a correct recognition of contour. However, our participants' performances were better when melodies had no direction changes. This indicates that simpler contours yield a better memory not only for contours but also musical intervals.

Dyson and Watkins (1984) observed that in a same/different test setting, participants' performances were greater if the lure item is obtained by changing a "corner" note (a note of direction change) rather than a "slope" note (note of direction no-change). In Experiments 1 and 2, we replicated their findings. Taylor and Pembrook proposed an additional factor to boost the recall of "corner" notes. They state that the recall performance for a "corner" note is greater if it is preceded by a "skip" (an interval greater than or equal to 3 semitones) rather than a "step" (an interval less than or equal to 2 semitones). We replicated this finding in Experiments 1 and found a marginal effect in Experiment 3: when the changed note was a "corner", recognition performances were greater if the interval before that particular note was a "skip" rather than a "step".

These results taken together might suggest that "corner" notes increase melodic complexity and require more attention. When there is no "corner" note, the target melody may remain within the limits of working memory capacity, increasing the memory performance. In contrast, one might expect a "corner" note to draw more

attention, particularly when preceded by a larger interval, and hence be better recognized when a change occurred right there.

There are also other determinants for memory of melodies. Taylor and Pembrook state that “ascending” (starting with an ascending movement, i.e. starting with C4) melodies were recalled better than “descending” (starting with C5) ones. We replicated this finding in all of our experiments. Moreover, participants found “ascending” melodies to be significantly less complex than “descending” ones. On the other hand, no such effect was observed in terms of memorability.

To increase participants’ sensitivity to “no signal” trials in a same-different test setup, we conducted Experiment 3 (same/different test with feedback and silent retention interval). Although feedback did not increase performances as measured by AUCs, participants responded with “same” significantly fewer times when feedback was provided. That is, participants responded more cautiously in Experiment 3. Overall, providing feedback boosted recognition performance (especially correct rejection performance) but otherwise confirmed all of our findings from Experiment 2.

In conclusion, we believe our study to be a step forward in the search of tonal determinants in immediate melodic memory. We believe that future research should address how music is stored in the short-term memory and which models of working memory best explain obtained results. This research focused exclusively on spectral characteristics of melodies. Our melodies were all within the 5-7 note limits of melodic short-term memory capacity (e.g., Benassi-Werke, Queiroz, Germano, & Oliveira, 2012; Schulze, Dowling, & Tillmann, 2012) Nonetheless, even in such short, isochronous sequences listeners might be using chunking processes, particularly when these contain melodic direction changes. Such processes are even

more likely to occur when rhythmic elements are added. Hence, future research needs to scrutinize the combined effects of spectral and temporal aspects on musical short-term memory performance.

REFERENCES

Agres, K. R. (2019). Change detection and schematic processing in music. *Psychology of Music, 47*(2), 173-193.

Atherton, R. P., Chrobak, Q. M., Rauscher, F. H., Karst, A. T., Hanson, M. D., Steinert, S. W., & Bowe, K. L. (2018). Shared processing of language and music. *Experimental Psychology,* 65(1), pp.40-48.

Atkinson, R., & Shiffrin, R. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation,*89-195. doi:10.1016/s0079-7421(08)60422-3.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225(2),* 82-91.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in cognitive sciences*, *4*(11), 417–423.

Benassi-Werke, M. E., Queiroz, M., Germano, N. G., & Oliveira, M. G. M. (2012). Melodic memory and its dependence on familiarity and difficulty. In S. Serafin (Ed.) *Proceedings of the 7th Sound and Music Computing Conference* (pp. 1-8). Spain: Universitat Pompeu Fabra.

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception: An Interdisciplinary Journal*, *12*(3), 353–364.

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

Dewitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception: An Interdisciplinary Journal*, *3*(3), 259–274.

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, *14*(1), 37–40.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, *50*(4), 305–313.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology: A Journal of Research in Music Cognition*, *1*(1), 30.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2B), 524–531.

Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics*, *57*(2), 136–149.

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception & Psychophysics*, *35*(5), 477–488.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*(4), 334.

Makovski, T., Watson, L. M., Koutstaal, W., & Jiang, Y. V. (2010). Method matters: systematic effects of testing procedure on visual working memory sensitivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1466.

Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382-398.

Mikumo, M. (1992). Encoding strategies for tonal and atonal melodies. *Music Perception: An Interdisciplinary Journal*, *10*(1), 73–81.

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in Psychology, 7*, 1892.

Nees, M. A., Corrini, E., Leong, P., & Harris, J. (2017). Maintenance of memory for melodies: Articulation or attentional refreshing? *Psychonomic Bulletin & Review, 24*(6), 1964-1970.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D., Conway, A., Cowan, N., ... & Ma, W. J. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*(9), 885.

Ortmann, O. (1933). Some tonal determinants of melodic memory. *Journal of Educational Psychology*, *24*(6), 454.

Parncutt, R. (2014). The emotional connotations of major versus minor tonality: One or more origins? *Musicae Scientiae, 18*(3), 324-353.

Pearce, M. T. (2005). The construction and evaluation of statistical models of melodic structure in music perception and composition. London: City University of London.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences, 1423*(1), 378–395.

Pembrook, R. G. (1987). The effect of vocalization on melodic memory conservation. *Journal of Research in Music Education*, *35*(3), 155–169.

Salamé, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology Section A, 41*(1), 107-122.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception: An Interdisciplinary Journal*, *29*(3), 255–267.

Taylor, J. A., & Pembrook, R. G. (1983). Strategies in memory for short melodies: An extension of Otto Ortmann's 1933 study. *Psychomusicology: A Journal of Research in Music Cognition*, *3*(1), 16.

Thompson, L., & Yankeelov, M. (2011). Music and the phonological loop (Doctoral dissertation, Belmont University).