Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

In Further Search of Tonal Grounds in Short Term Memory of Melodies

Alperen Karan[1] and Esra Mungan[2]

[1] *Psychological Sciences Graduate Program, Boğaziçi University, Turkey*
[2] *Psychology Department, Boğaziçi University, Turkey*

[1]alperen.karan@boun.edu.tr, [2]esra.mungan@boun.edu.tr

Abstract

Taylor & Pembrook (1983) proposed several factors to affect short-term memory for melodies. We reassessed their findings using a more controlled stimulus set and a 2-alternative forced choice (Experiment 1) or same/different test (Experiment 2) instead of a dictation or singing-back task. Nonmusicians listened to a total of 158 isochronous 5-tone melodies. Each melody was followed by a same-length retention interval filled with silence, nonsense syllables, a nondiatonic melody, or a diatonic melody, and a subsequent test with same-contour lures. In both experiments and across all conditions listeners showed above-chance short-term recognition performance. We replicated Taylor & Pembrook's recency effect for the 5th note of the sequences but also found a full J-shaped serial position curve (recency>primacy>center). Secondly, listeners performed better for tone sequences that were either fully ascending or descending than those with melodic direction changes. Thirdly, listeners were better in noticing a changed note that occurred at a point of melodic direction change (e. g., ∧ or ∨). In Experiment 1 but not 2, we furthermore found that this "corner note effect" was even more pronounced if that note was preceded by a "skip" (3 or more semitones) instead of a "step" (2 or less semitones) pitch interval. The latter finding was somewhat similar to Taylor and Pembrook's finding that listeners were more accurate in their reproduction of skip as opposed to step intervals when they marked a point of change in melodic direction. Fourthly, type of retention interval had a major effect on participants' performances. When tested with a 2AFC setup, the silence group performed best, when tested with a same/different setup, both the silence and nonsense syllable groups performed best. Results are discussed in reference to related STM studies using short tone sequences and Berz' (1995) working memory model for music.

Introduction

Given that most simple songs are built on fairly short melodies, an interesting question arises as to how well listeners retain exact note sequences after first hearing in short-term memory. More specifically, how precise is the immediate configural representation of a short and simple tone sequence, and how much is this precision jeopardized by intervening musical or non-musical events? Are there specific melodic characteristics that lead to more precise and more resistant configural representations? The present study aims to address these questions by expanding upon earlier, related research.

Although encoding of musical information is affected by several organizational factors, which are stored in long-term memory, such as contour, tonality and range (Berz, 1995), several studies showed that around 7-11 notes can be held in short-term memory (Pembrook, 1987). More recent studies with musically untrained participants suggest an average span of 5-7 notes (e. g., Benassi-Werke, Queiroz, Germano, & Oliveira, 2010; Schulze, Dowling, & Tillmann, 2012).

Probably one of the earliest studies to investigate short-term memory for tone sequences was Ortmann's 1933 experiment in an attempt to identify tonal characteristics that boost or reduce short-term memory retention. He used twenty 5-tone sequences as stimuli and asked participants to recall them in a dictation task after each was played to them on a piano. Taylor and Pembrook (1983) did a follow-up on Ortmann's work with the same stimulus set to ensure comparability. But they added a singing-back task as a less demanding memory task and corrected some of the methodological shortcomings of the 1933 study. The simpler singing-back task allowed them to include musically untrained participants to their study. The researchers replicated Ortmann's recency effect finding (i. e., better recall for the last note) for musically trained and untrained participants. Like Ortmann, they also found worse performance when the tone sequences had two or more melodic direction changes compared to sequences with one or none. This was again observed in both musically trained and untrained participants. They also replicated Ortmann's finding of better recall when a melodic direction change occurred at a larger pitch interval ("skip") than a small one ("step", defined as a semitone or whole tone interval). This difference was even greater for musically untrained participants. Taylor and Pembrook explained this interesting "skip-effect" as resulting from attentional processes caused by unexpected skip intervals. Finally, just as in Ortmann's 1933 findings, all participants performed better with ascending than descending sequences.

Even though there were improvements in Taylor and Pembrook's 1983 study, some factors jeopardize the generalizability of their findings. Some sequences, for instance, had full melodic symmetry which possibly served as a strong mnemomic in itself, an issue they acknowledge as a potential pitfall in the Ortmann melody set. Moreover, the fact that all sequences not only started but also ended with a C, makes the recency effect potentially meaningless. Finally, since their corpus had only 20 stimuli, we cannot be sure how generalizable their findings are.

Our goal was to take Taylor and Pembrook's research one step further by using a larger corpus and a short-term memory *recognition* test setup. As such we aimed to bridge Taylor and Pembrook's endeavors with the extensive amount of findings coming from the short-term recognition literature with tonal sequences (e.g., Deutsch, 1970; Dewitt & Crowder, 1986; Dowling, 1973, 1978, 1991; Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling, Kwak, & Andrews, 1995). A carefully constructed corpus of hundred fifty eight 5-tone sequences was used. All sequences were isochronous and started either with C4 or C5, just like Ortmann's sequences. Further facts such as number of melodic direction changes and note repetitions within each tune were also controlled for. Participants were tested after each 5-tone sequence with either a 2-alternative forced choice (2AFC) test (Experiment 1) or a same/different test (Experiment 2).

In addition, we included a retention interval manipulation. For one group of participants, the retention interval was not filled, for second group it was filled with a sequence of five nonsense syllables, for a third group with a sequence of five nondiatonic tones and for a fourth group with a sequence of five diatonic tones. The purpose of this manipulation was to attempt a comparison between two competing models of short-term memory for music. Whereas Baddeley's model (2000) proposes that both verbal and musical materials are processed in the phonological loop, Berz (1995) proposes a new "slave system" exclusively for musical material. Based on Baddeley's model, we should expect equally reduced short-term recognition performance under all retention interval conditions except the unfilled one. Berz' model, on the other hand, would predict reduced performance for the intervening *musical* sequences but not for the intervening nonsense syllable or silence conditions. We were also curious about how our different retention interval manipulations would affect the expected J-shaped serial position curve. Would they only wipe out mostly the recency portion or create an overall depression of the curve? Earlier, Dowling (1973) had shown that musical material just like verbal material produces J-shaped serial position curves in short-term memory testing. But, to our knowledge, little is known about how the different sections of the curve respond to manipulations such as ours.

Methods

Experiment 1: 2AFC Test

Participants. Seventy-two Boğaziçi University undergraduate students with no to negligible musical background (M= .96; SD=1.33; range 0-4 yrs) participated in return for course credit in introductory psychology courses. None of them had absolute pitch or a hearing impairments.

Stimuli and Apparatus. We created 158 5-tone target-lure melody pairs using MATLAB. All melodies were in the C major scale within the range of C4 and C5. All melodies started with either C4 or C5. Note repetitions occurred in a melody at most twice, and never in succession. Lure items were obtained by changing the second, third, fourth or fifth tone of the target melody by one diatonic note. All lure items preserved the contours of their target melodies. Target-lure identity as well as target-lure position at test were counterbalanced across participants and trials.

We also controlled for the number of melodic direction changes by including a sufficient number of target items for all possible four conditions: zero (in 30 melodies), one (in 48 melodies), two (in 48 melodies), or three (in 32 melodies) direction changes (see Figure 1, for an example with two direction changes). Among sequences without melodic direction changes, half were ascending and half descending. For those with one, two or three direction changes, half started on C4 and half on C5. The base rates of pitch intervals were distributed as 31% "steps" (1 or 2 semitones) to 69% "skips" (> 2 semitones) across all pitch intervals of 158 5-tone sequences. Finally, the likelihood of whether a note was at a point where a melodic direction change occurred or not was 51% to 49%, respectively, across all tunes.

Figure 1: A target and its lure with two direction changes. 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.

For the intervening nonsense syllable condition, 5-syllable sound sequences were created from a set of 23 nonsense syllables via an online text to speech converter (texttospeech.org), and the length of each syllable was adjusted to 0.25s using Adobe Audition 3.0 to match the note durations of the 5-tone sequences. For the intervening melody condition, 158 5-tone diatonic and nondiatonic sequences were created with same pitch averages across five tones. None of these intervening tone sequences contained repetitions, and all were different from the experimental target-lure melodies.

All 5-tone sequences were played with the Steinway grand piano in Logic Pro X software at a speed of 0.25s per note. Over- and on-ear padded headphones were used (Philips SHP 1900, Urbanears Plattan 2, Sennheiser Momentum On-Ear) in cubicles of the Cognitive Processes lab at Boğaziçi University.

Procedure. Participants were tested individually or in groups of up to four. Instructions were followed by four closely monitored practice trials to ensure comprehension of the task. Participants listened to a target melody, which was followed by a retention interval filled with either silence, nonsense syllables, a nondiatonic or diatonic tone sequence. After the retention interval, participants received a 2AFC test where they heard two comparison melodies (Figure 2) from which to pick the target. They were also asked to provide a 3-point confidence rating ("very sure", "sure", "not sure") each time they made their choice. Each experimental session lasted about 35 minutes.

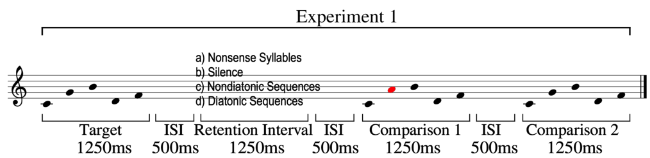

Figure 2: The experimental designs.

Experiment 2: Same/Different Test

Participants. Fifty-eight Boğaziçi University undergraduate students with no to negligible musical background (M= .75; SD=1.25; range 0-4 yrs) participated in return for course credit in introductory psychology courses. None of them had absolute pitch or a hearing impairments.

Stimuli and Apparatus. The apparatus and the stimuli were the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1 except for the recognition test which this time was a same/different test (see Figure 2). Participants were explicitly instructed that the comparison melody which followed the retention interval would either be same or different with an equal likelihood. They were again asked to provide a 3-point confidence rating ("very sure", "sure", "not sure") each time they made their choice. Each experimental session lasted about 25 minutes.

Results

We calculated average performance scores using the proportion of correct responses. The data of seven participants, all from Experiment 2, who showed an average recognition performance of .52 or less but rated 120 or more of their 158 responses as 3 ("very sure") were excluded from the study. We also excluded one participant from Experiment 1 who performed around chance level (.54) but rated all his 158 responses as 3. We took this as an indication that those participants were not using the 3-point confidence rating scale as they should have.

Experiment 1: 2AFC Test

Participants performed above chance level in silence (M=.67, SD=.08, $t(15)$=8.44, p<.001), nonsense syllables (M=.63, SD=.12, $t(19)$=5.00, p<.001), nondiatonic (M=.61, SD=.09, $t(17)$=5.29, p<.001) and diatonic (M=.58, SD=.09, $t(17)$=3.70, p<.01) conditions.

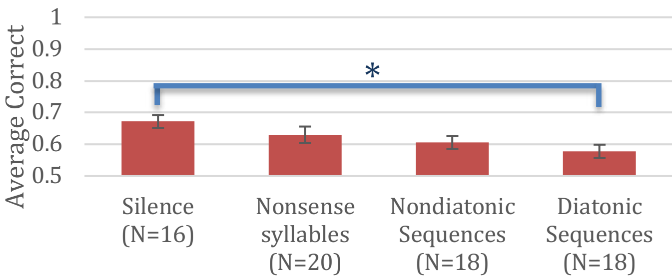

Figure 3. Average correct recognition across groups. Error bars (as in all remaining graphs) indicate standard errors.

A one-way ANOVA revealed a significant effect for type of retention interval, F(3,68)=3.02, p<.05, η_p^2=.12. A Tukey's HSD post-hoc test showed that recognition performance was significantly higher for the silence condition than diatonic sequences condition, p<.05 (Fig. 3).

Although many participants stated written or verbally that the experiment lasted quite long, no fatigue or practice effects were observed. We calculated performances for the first and second half of the trials. A 2 (order) x 4 (group) mixed ANOVA showed that order had neither a main nor an interactive effect on performance, p>.10[1].

Taylor and Pembrook (1983) found that note repetition did not change recall performance. Our findings confirmed their results. A 2 (repetition) x 4 (group) mixed ANOVA revealed neither a main nor an interactive effect for note repetition, all ps>.10.

The lure items contained a tone that was one diatonic tone higher than the target in half of the trials or lower in the other half of the trials. A 2 (repetition) x 4 (group) mixed ANOVA revealed a marginal main effect for type of lure, indicating slightly better performance with lures that had a one diatonic tone up deviation (p=.06).

Participants had a metacognitive understanding of the task. We calculated average performances for low, moderate and high confidence trials (Fig. 4). A 3 (confidence) x 4 (group) mixed ANOVA showed that performances differed between different levels of confidences, F(1.37,93.21)=14.60, p<.001, η_p^2=.18. Post-hoc analyses revealed that participants' performances were significantly greater for the high confidence trials than low or moderate confidence trials. No interaction was observed, p>.10.

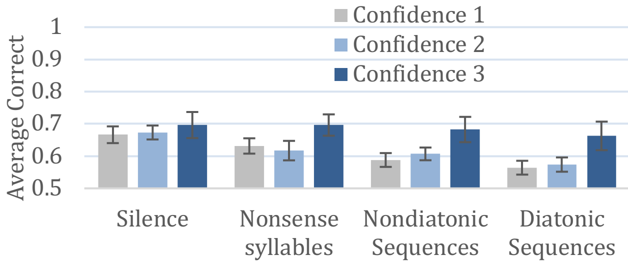

Figure 4. Average correct recognition by confidence across groups.

We also tested the effect of number of melodic direction changes on performance. A 4 (number of direction changes) x 4 (groups) mixed ANOVA showed that number of melodic direction changes had a main effect on performance, F(3,204)=2.94, p<.05, η_p^2=.04. In particular, participants performed higher for melodies without direction changes (which were melodies that were either fully ascending or descending) than melodies with one or three direction changes, p<.05. No interaction was found, p>.10 (Fig. 5). No performance difference was observed between ascending and descending sequences, p>.10.

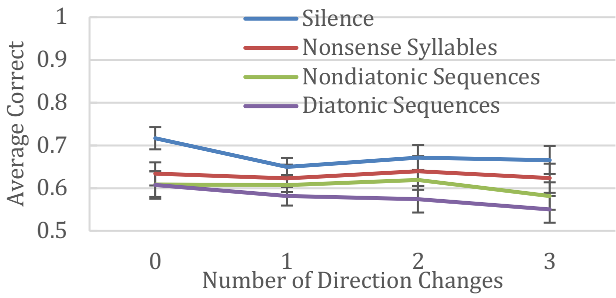

Figure 5. Average correct recognition by number of melodic direction changes across groups.

A 4 (serial position of the changed tone) x 4 (groups) mixed ANOVA revealed a main effect for serial position (F(3, 204)=19.29, p<.001, η_p^2=.22) but no interaction effect (p>.10). Planned comparisons at an overall α-level of .05 confirmed the expected J-shaped curve (recency > primacy > center): changes in the last tone were detected better than changes in the second tone; changes in the second tone were detected better than changes in the middle (third and fourth) tones (Fig. 6).

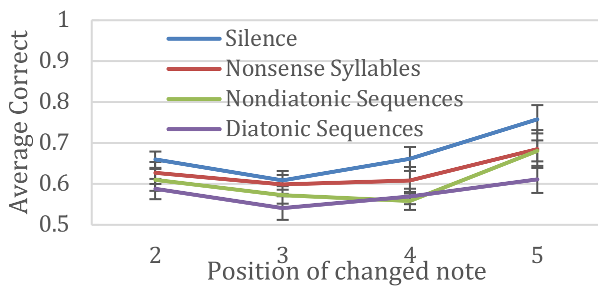

Figure 6. Average correct recognition based on position of changed note within 5-tone sequence across groups.

[1] We also found no main effect for order when splitting trials into three blocks.

In 60 trials, lure melodies differed from target melodies at a note of melodic direction change (see Fig. 1 as opposed to Fig. 2). Note that it is conceptually wrong to talk about the melody's direction at the 1st and the 5th tones. Dyson and Watkins (1984) coined these notes as "corners" and the remainders as "slopes". A 2 (change at "corner" or "slope") x 4 (groups) ANOVA showed that participants were better at detecting note changes at "corners" than "slopes" ($F(1,68)=12.50$, $p<.01$, $\eta_p^2=.16$). No interaction was observed, $p>.10$ (Fig. 7).

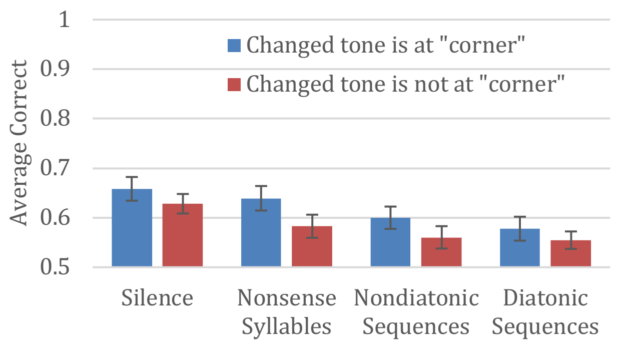

Figure 7. Average correct recognition rates when the change did/did not occur at a "corner" note.

We also replicated Taylor and Pembrook's (1983) "skip effect". For trials where the change occurred at a "corner" note, participants' correct recognition rates were higher if the interval before that note was a "skip" rather than a "step", $F(1,68)=9.08$, $p<.01$, $\eta_p^2=.12$.

Experiment 2: Same/Different Test

We conducted our analyses using area under the receiver operating characteristics curve (AUC). When the research question of interest was about the lure trials, we used correct rejection rates as the dependent variable. For each group, the AUCs were significantly greater than chance level .50, $p<.001$.

A one-way ANOVA showed that discrimination differed between groups, $F(3,54)=7.66$, $p<.001$. Tukey's post-hoc HSD analyses showed that the silence and nonsense syllable groups performed significantly better than the nondiatonic and diatonic sequence groups, $p<.05$ (Fig. 8).

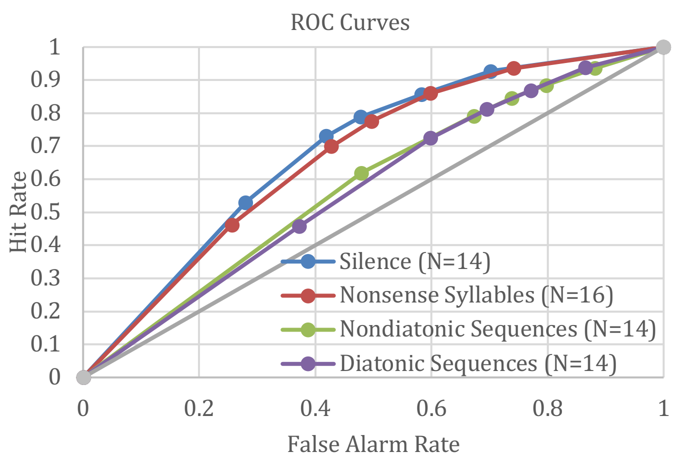

Figure 8. Receiver operating characteristics (ROC) curves across groups.

Participants' performances did not differ in the first and second half of the trials, $p>.10$. We observed a marginal advantage for melodies with a note repetition compared to those without, $p=.07$. Participants performed equally well in trials where the lure item contained a higher or lower tone, $p>.10$. None of these 2 x 4 mixed ANOVAs revealed an interaction effect with type of retention interval.

A 4 (number of melodic direction changes) x 4 (groups) ANOVA showed that number of direction changes had a significant effect, $F(2.55,137.41)=2.89$, $p<.05$, $\eta_p^2=.05$. Participants were better in discriminating melodies that had either no or two direction changes compared to melodies with three direction changes, $p<.05$. No interaction was found, $p>.10$ (Fig. 9). We also did not find any performance difference between ascending and descending sequences, $p>.10$.

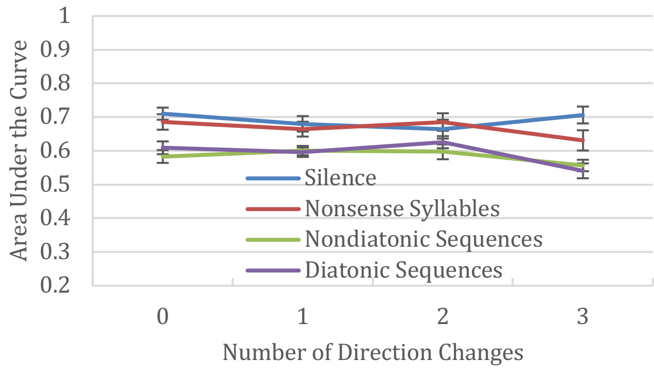

Figure 9. Average AUC values by number of melodic direction changes across groups.

A 4 (serial position of the changed tone) x 4 (groups) mixed ANOVA revealed a main effect for serial position ($F(2.47,133.26)=34.48$, $p<.001$, $\eta_p^2=.39$). Planned comparisons at an overall α-level of .05 confirmed the expected J-shaped curve (recency > primacy > center): changes in the last tone were detected better than changes in the second tone; changes in the second tone were detected better than changes in the middle (third and fourth) tones (Fig. 10).

However, serial position of the changed tone interacted with type of retention interval group, $F(7.40,133.26)=4.83$, $p<.001$, $\eta_p^2=.21$. To better understand the interaction effect, we decided to do a planned comparison in the form of a 2 (silence/nonsense syllable groups vs. non-diatonic/diatonic groups) x 3 (primacy vs. center vs. recency position) ANOVA. We found a main effect for group, such that the combined silence/nonsense syllable group performed significantly better than the combined non-diatonic/diatonic group. This points to an overall depression of performance when the retention interval was filled with *musical* material. Moreover, musical material also curbed the recency effect, as can be seen in Fig. 10.

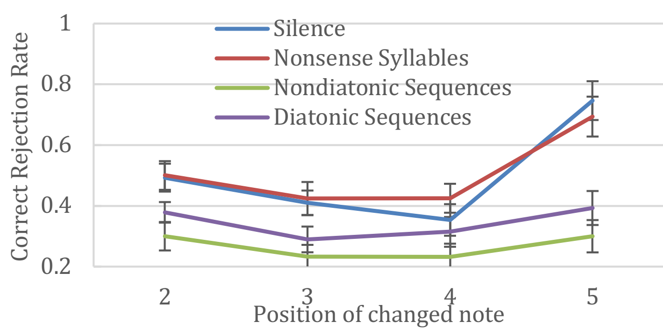

Figure 10. Average correct rejection rates based on position of changed note within 5-tone sequence across groups.

We conducted a 2 (change at "corner" note or "slope") x 4 (groups) ANOVA to see whether the contour-wise position of the changed note had an effect on correct rejection performance. Participants performed better when the change occurred at a "corner" note, $F(1,54)=6.34$, $p<.05$, $\eta_p^2=.11$. No interaction was found, $p>.10$ (Fig. 11).

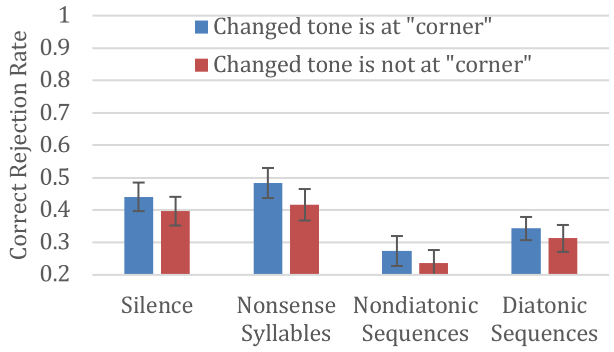

Figure 11. Average correct rejection rates when the change did/did not occur at a "corner" note.

However, we could not replicate the "skip effect" from Experiment 1. When the changed tone was a "corner" note, participants performed equally well regardless of whether the interval before the changed tone was a step or a skip, $p>.10$.

Conclusions

The main motivation of this research was to continue Taylor and Pembrook's (1983) line of research which was a follow-up on Ortmann's 1933 study. The goal of these studies was to detect potential tonal features that improve memorability of short, isochronous 5-tone sequences in a short-term memory (serial) recall setting. Taylor and Pembrook found better recall for (1) melodies that had fewer melodic direction changes (e. g., melodies with up-up-down-down contours were better recalled than melodies with up-down-up-down contours); (2) ascending than descending sequences; (3) notes that were immediately preceded by skips (pitch intervals of 3 or more semitones); and (4) notes in the final position (which was not an meaningful finding since all 5-tone sequences started and ended on a C hence recalling the ending C could have been a simple knowledge-based inference rather than a unique act of remembering).

Our goal, in turn, was threefold. Firstly, we wondered how much Taylor and Pembrook's findings would replicate in a short-term *recognition* test setting. Secondly, we wanted to see how well musically untrained listeners would detect one-diatonic-tone deviations in same-contour lures, which are known to be harder to detect compared to contour-violating lures (e. g., Dowling, 1991). This would also provide us with some clues about the "resolution" of the representation immediately formed after listening to a short isochronous tone sequence. Given that children's songs or lullabies tend to have quite overlapping melodic contours (Trehub, Unyk, & Trainor, 1993), it seems worth pursuing how well a "layperson" attends to, codes, and retains melodic information beyond basic contour. Thirdly, we were curious whether type of intervening stimuli would affect short-term recognition performances as predicted by Baddeley's (2000) working memory model as opposed to Berz' (1995) model. We will start with a summary of our main findings regarding our first goal then move on to the next two points.

Listeners showed above chance short-term recognition after listening to isochronous, diatonic, 5-tone melodies which were tested against same-contour lures with minute one-diatonic-note deviations. This was so regardless of the type of test (2AFC in Experiment 1 and same/different in Experiment 2) and regardless of whether the retention interval between exposure and test was filled with silence, nonsense syllables, nondiatonic or diatonic tone sequences. Second, we not only replicated Taylor and Pembrook's recency effect but also found a full J-shaped serial position curve (recency > primacy > center). Participants were best in detecting a change in the 5th note, second best in detecting a change in the 2nd note, and worst in detecting changes in the 3rd and 4th notes. This was so for all groups and regardless of type of test. Third, in Experiment 2 but not 1, we observed a general depression of the J-shaped curve and a curbing of the recency effect when the retention intervals were filled with musical sequences but not when filled with silence or nonsense syllables. Fourth, we observed that note changes at "corners" (i. e., points of melodic direction change) were better detected than note changes within "slopes" (cf. Dyson & Watkins, 1984). We also replicated Taylor and Pembrook's interesting "skip effect" such that a note change that occurred at the end of a *larger* pitch interval ("skip") was detected (or reproduced, in Taylor & Pembrook's case) better than one that occurred at the end of a *smaller* pitch interval ("step"), provided it marked a point of melodic direction change and tested in a 2AFC test. In other words, when notes formed "corners" in terms of melodic contour, they appeared to be better encoded when they followed a "skip" than a "step" interval. Fifth, we found better short-term memory performance for tone sequences that did not have melodic direction changes, i. e., that were fully ascending or descending. We did not replicate Taylor and Pembrook's (1983) advantage for ascending over descending tone sequences. Finally, whether notes appeared repeatedly (though never in unison) or not did not affect listeners' performance in a 2AFC or same/different test set up (though there was a marginal trend in the latter in favor of note repetition).

The fact that we found above-chance performance in two different short-term recognition setups with only same-contour lures provides evidence for a higher representational "resolution" that incorporates specific pitch interval information. Since we used a recognition instead of recall paradigm our findings also need to be compared to those of Dowling (1978; 1991) and his colleagues (e. g., Dowling, Kwak, & Andrews, 1995), who used comparable isochronous 5- or 7-tone sequences in partly comparable recognition test setups. The closest experiment to ours would be Experiment 1 in Dowling and Fujitani (1971), which showed that when using untransposed same-contour lures in a 2AFC setup with isochronous 5-tone sequences, listeners with mixed musical background were successful in discriminating targets from lures [2] (with a mean area-under-the-memory-operating-

[2] Although Dowling and Fujitani's participants performed close to ceiling in the untransposed condition, this was not the case for our participants. This could be due not only to their musical background-wise mixed participant profile but also their use of only 60 tone sequences (whereas we used 158 tone sequences).

characteristics-curve of M_{AUC}=.92). In all other studies, Dowling and colleagues used transposed targets only, to exclude any "help" from pitch memory. When using transposed targets only, musically trained and untrained participants dropped to chance performance in a same-different task (Dowling, 1978). Above chance performance in discriminating transposed targets against transposed same-contour lures was obtained in later studies with 7-tone sequences when changing *two* instead of a single note in a continuous-running-memory task (e. g., Dowling, 1991; Dowling, Kwak, & Andrews, 1995; also cf. Dewitt & Crowder, 1986). However, because there are many parametric differences between their studies and ours (e. g., their tone sequences had only small pitch intervals, all their ending notes were twice as long, their pauses between two consecutive melodies were 7-sec long, and their shortest unfilled delays between study and test were 7 s, and 12 s and more if filled with intervening melodies), caution is needed when comparing our results. The most major difference is that in all of the Dowling studies, contour change versus pitch-interval-change-only are pitted against each other. When exposed to *both* types of lures, listeners showed an interesting trend of ever improving recognition sensitivity with same-contour lures the longer the delay (filled with other melodies) between exposure and test. In contrast, our main interest was to look specifically at the "resolution" of one's *exact* short-term memory representation of isochronous tone sequences as used by Taylor and Pembrook (1983). This is why we intentionally allowed for the use of pitch memory as it would occur in a natural setting. We made it impossible for participants to use any contour-based proxies to differentiate targets from lures, and instead, forced them to attend to exact pitch interval information, which they were indeed able to do beyond chance, despite being musically untrained. Byron and Stevens (2006) found similar sensitivity to same-contour pitch interval changes with non-isochronous 8-note sequences (derived from unknown folk tunes) and a silent 2-sec delay between exposure and test in a same/different task with musically untrained participants. Their participants could detect differences in pitch interval magnitude with same contours even in transposition hinting at some pitch interval magnitude retention independent of contour information[3]. Cuddy and Cohen (1976) found evidence for similar pitch interval extraction even for 3-tone sequences in a 2AFC task with a 2-sec unfilled retention interval.

The question of what makes a melody more discriminable is beautifully addressed in Harrison, Musil, and Müllensiefen (2016) who modeled responses from same/different tests looking at a large range of melodic (and rhythmic) features. In their model, melody complexity is defined on the basis of number of notes (the more notes, the higher the working memory load) as well as contour and pitch interval related aspects (also cf. Cuddy, Cohen, & Mewhort, 1981). Since our study controlled for a variety of melodic aspects we believe that an assessment of our corpus would render a rather stable complexity level caused by the fact that *all* melodies had to be discriminated against contour-wise *same* and pitch interval-wise *very similar* lures. Given that listeners received little help from lure dissimilarity, their recognition responses had to rely more on the quality of their initial encoding. In that sense, it is worth looking at which features in such an otherwise very homogenous set of targets and lures still help listeners in their discrimination performance. We plan to do an analysis on the best and worst discriminated melodies to see whether we can detect certain shared saliency increasing or decreasing features. If we cannot find such shared characteristics we might think that at least the ones with top discriminability might be tone sequences reminiscent of sections of well-known tunes which can be found out by asking our or a new group of participants to rate those top and worst tunes based on criteria such as "sounds well-known", "reminds me of tune X".

Another interesting and quite "single" study in the short-term recognition literature for melodies is Mikumo's (1992) where she looked at the effects of different types of intervening stimuli in a same-different recognition task. Among others, she found that with tonal (rather than atonal) 6-tone sequences, musically less trained participants performed significantly better when there was a silent 12-sec delay between the standard and comparison stimulus than when the delay was filled with another melody or with shadowing tasks. In her setup, where targets were untransposed and melodies were played at a slow rate (1 s per note), same-contour melodies triggered high false alarms when collapsed over interference conditions. Despite its discrepancies from ours, this study could have been an interesting first study to look at the differential effects of melodic versus nonsense syllabic interference in short-term recognition memory for isochronous melodies. However, since intervening nonsense syllables were shadowed whereas intervening tone sequences were not, we are unable to make clear inferences with respect to the vulnerability of a melody representation to verbal versus musical interference. Any detrimental effect in the nonsense syllable condition could have been due to the motor process of shadowing rather than nonsense syllables per se. We believe that our study provides a unique opportunity to look at the differential effects of intervening verbal (nonsense syllables) versus musical (diatonic or nondiatonic) stimuli on short-term memory performance in melody recognition. This, in turn, may allow us to compare the predictions of Baddeley's (2000) working memory model which proposes that the phonological loop is responsible for the processing of both verbal and musical material and Berz' (1995) proposal of a separate "slave system" for musical material only. Interestingly, when using a 2AFC test, our findings seem more in line with Baddeley's model (though we see a trend for a stepwise worsening when going from nonsense syllables to non-diatonic to diatonic intervening sequences). But when using a same/different test, our findings seem more in line with Berz' model. In a same/different test setup, we were also able to observe a general depression of the J-shaped serial position curve and a curbing of its recency section for the intervening *melodic* sequence groups only.

Last but not least, we found some differences in overall performance as well as in the effects of our variables on performance depending on whether participants were tested with a 2AFC or a same/different test. The same/different test led to a bias to say "same" which was likely driven by the high similarity between target and lures. Given that we used exactly

[3]But unlike our minute one-diatonic pitch interval changes at test, theirs pitted "step" changes against "leap" (>3 semitones) changes.

the same stimulus corpus, this shows that the particular way of testing directly affected the decision making process which, in turn, affected the results (cf. Jang, Wixted, Huber, 2009). Benassi-Werke et al. (2010) showed that when using a forward melodic span test, which is a relatively bias-free test due to being recall instead of recognition, musically untrained listeners had an average span of 5.9 notes for diatonic sequences with max. four semitone leaps, and 4.9 notes for those with unrestricted pitch intervals. This confirms that even in the more challenging same/different setup our listeners were dealing with a task that was fully within their short-term memory capacity for tunes. In other words, it is unlikely to think that their inferior performance in the same/different setup was due to capacity limits.

In conclusion, we believe that further work has to be done on the instantaneous resolution of short-term memory representations for melodies. Given the extensive amount of research on visual-spatial short-term and working memory representations (e. g., Luck & Vogel, 2013; Sligte, Vandenbrouke, Sholte, & Lamme, 2010), we think the musical domain has to follow suit. This will also provide us with a better understanding of how musical material is processed in immediate memory and which models better predict observed findings. Certainly, future research also has to look at short-term memory for melodies with rhythmic elements, such as the ones used by Byron & Stevens (2006; also cf. Byron, 2008).

References

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.

Byron, T. (2008). The processing of pitch and temporal information in relational memory for melodies. (Unpublished doctoral dissertation). University of Western Sidney, Penrith, New South Wales, Australia.

Byron, T., & Stevens, C. (2006). Steps and leaps in human memory for melodies: The effect of pitch interval magnitude in a melodic contour discrimination task. In M. Baroni, A. R. Addessi, R. Caterina, M. Costa (Eds.) *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9)*, Bologna, Italy.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *The Quarterly Journal of Experimental Psychology, 28(2),* 255-270.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7(4),* 869.

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). The perception of interleaved melodies. *Cognitive psychology, 5(3),* 322-337.

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

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.

Harrison, P. M., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: descriptive and explanatory approaches. *Journal of New Music Research, 45(3),* 265-280.

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General, 138(2),* 291.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences, 17(8),* 391-400.

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

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

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

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.

Sligte, I. G., Vandenbroucke, A. R., Scholte, H. S., & Lamme, V. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in psychology, 1*, 175.

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),* 16-35.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant behavior and development, 16(3),* 285-295.