REPRODUCTION ACCURACY FOR SHORT RHYTHMS FOLLOWING MELODIC OR MONOTONIC PRESENTATION

ANDREW V. FRANE & MARTIN M. MONTI
University of California, Los Angeles

SOME RESEARCHERS AND STUDY PARTICIPANTS have expressed an intuition that novel rhythmic sequences are easier to recall and reproduce if they have a melody, implying that melodicity (the presence of musical pitch variation) fundamentally enhances perception and/or representation of rhythm. But the psychoacoustics literature suggests that pitch variation often impairs perception of temporal information. To examine the effect of melodicity on rhythm reproduction accuracy, we presented simple nine-note auditory rhythms to 100 college students, who attempted to reproduce those rhythms by tapping. Reproductions tended to be more accurate when the presented notes all had the same pitch than when the presented notes had a melody. Nonetheless, a plurality of participants judged that the melodically presented rhythms were easier to remember. We also found that sequences containing a Scotch snap (a sixteenth note at a quarter note beat position followed by a dotted eighth note) were reproduced less accurately than other sequences in general, and less accurately than other sequences containing a dotted eighth note.

Received: June 23, 2020, accepted December 11, 2020.

Key words: melody, rhythm, tapping, memory, recall

IN A SEQUENCE OF MUSICAL NOTES, PITCH VARIation combines with rhythm to create musical phrases (Boltz, 1998; Jones, 1987; Krumhansl, 2000; Schellenberg, Krysciak, & Campbell, 2000). In music, as in speech (Juslin & Laukka, 2003), many interesting perceptual effects have been documented that involve interaction of the pitch and rhythm information. For instance, Schellenberg et al. (2000) found that the extent to which musical excerpts were rated as "happy" or "scary" was affected by interaction between pitch variation and rhythm. Other researchers found that certain types of "coherence" between pitch and rhythm made musical passages easier to discriminate (Monahan, Kendall, & Carterette, 1987) or made attributes of musical passages easier to recall (Boltz, 1998).

It is not controversial to say that *melodicity* (the presence of musical pitch variation) is important for recognition of musical passages. Even "Happy Birthday To You," which has been called "the most frequently sung song in the English-speaking world" (Guinness, 1998, p. 396), is typically unrecognizable when presented as rhythm alone (Newton, 1990). In fact, in much of Western music, a song's pitch sequence is more distinctive than its rhythm (Bailes, 2010; see also Prince & Pfordresher, 2012). Moreover, multiple studies have found musical passages to be more recognizable by pitch sequence alone than by rhythm alone (Hébert & Peretz, 1997; White, 1960; see also Herff, Olsen, Prince, & Dean, 2018).

Nevertheless, it remains an open question whether melodicity can enhance the perception or encoding of rhythm even when there is no inherent or previously established correspondence between the pitch variation and the rhythm. On the one hand, some researchers expressed an intuition that the answer is yes (Kinney & Forsythe, 2013; Moog, 1979). That intuition was echoed by the metacognitive judgments of one study's participants, who reported that melodicity made rhythmic sequences easier to recall (Kinney & Forsythe, 2013). There are plausible rationalizations for that intuition. For instance, one might speculate that melodicity increases the salience of note onsets, or that melodicity increases attention to the auditory stream in general. Alternatively, one might speculate that the varied pitches serve as "landmarks" that help organize the rhythm into manageable *chunks*, just as rhythm has been proposed to help organize the pitch sequence into chunks (Dowling, 1973; see also Deutsch, 1999).

On the other hand, the psychoacoustics literature suggests that uninformative variation in pitch often impairs, rather than enhances, perception of temporal information, at least for some types of nonmusical stimuli. For example, varying the pitch or spectra of tones or noise bursts has been found to impair anisochrony detection (David, Lavandier, & Grimault, 2014), temporal interval discrimination (Divenyi & Danner, 1977), and temporal gap detection (Grose, Hall, & Bus, 2007; Grose, Hall, Bus, & Hatch, 2001; Heinrich, de la Rosa, & Schneider, 2014). A

Music Perception, VOLUME 38, ISSUE 4, PP. 361–372, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312. DOI: https://doi.org/10.1525/MP.2021.38.4.361

simple explanation for these impairments is that changes in pitch distract attention from temporal information (Grose et al., 2007). But there are other explanations in some cases. For example, consider the *auditory kappa effect*, which is a perceptual distortion in which acoustic events that differ in pitch or spectra are incorrectly perceived as also differing on a temporal variable, such as duration (Crowder & Neath, 1995; Shigeno, 1986, 1993; Yoblick & Salvendy, 1970; see also Gabrielsson, 1973a, 1973b; Hirsh, Monahan, Grant, & Singh, 1990; Sink, 1983, 1984). In such cases, the misperceived temporal differences tend to increase with increasing frequency disparity. That suggests pitch variation does not merely distract attentional resources from temporal information, but in fact systematically distorts temporal information as pitch and timing are integrated into a unified percept. Similarly, in an isochronous tapping task, varying the pitch of the tones triggered by the taps was found to have systematic effects on the intertap intervals, even though participants were instructed to ignore the tones (Ammirante & Thompson, 2012; Ammirante, Thompson, and Russo, 2011).

Few studies have directly examined the question of how melodicity affects the perception or recall of rhythm in musical stimuli. And studies that have examined that question have not yielded very conclusive results, as we discuss in the next two subsections.

Effects of Melodicity on Discrimination of Novel Rhythms

Moog (1979) presented pairs of auditory rhythms to children. The children rated each pair as "same" (by raising a round green card) or "different" (by raising a square red card). The children were categorized into three groups, enumerated as follows in the present paper: (1) children with congenital locomotor impairment, who had HAWIK verbal-IQ scores between 80 and 118, (2) children with no history of locomotor impairment, who had HAWIK verbal-IQ scores between 58 and 81, (3) children with no history of locomotor impairment, who had HAWIK verbal-IQ scores between 87 and 130. Each stimulus was either melodic or monotonic, either one or two bars long, and either "simple" or "complicated."

Moog (1979) had several hypotheses, one of which was that same/different judgments would be better for melodic stimuli than for monotonic stimuli. But tests of that hypothesis were nominally statistically significant ($.01 < p < .05$) only in three of nine subgroup analyses: (1) one-bar stimuli in group 2, (2) one-bar and two-bar stimuli pooled together in group 2, (3) one-bar stimuli in group 3. Moreover, as Moog rightly noted, although the mean differences for those subgroups were in the hypothesized direction, only about half of the individual participants in those subgroups exhibited a difference in the hypothesized direction. In fact, for one-bar stimuli in group 3, only 11 of 25 participants exhibited a difference in the hypothesized direction, which indicates that the statistical significance was driven by outliers. Thus, the results are difficult to interpret.

Effects of Melodicity on Recall and Reproduction of Rhythms

Kinney and Forsythe (2013) presented two-bar rhythmic sequences to college students who were majoring in music education. Each sequence was presented melodically in one trial and non-melodically in another. In each trial, students attempted to reproduce the presented sequence, using a method of their choosing: clapping, tapping, or vocalizing.

The melodic stimuli were made with a piano sound, whereas the non-melodic stimuli were made with a woodblock sound. Thus, melodicity—the independent variable of primary interest—was confounded with all the acoustic variables that distinguish piano waveforms from woodblock waveforms: amplitude envelope, harmonic spectra, amount of periodicity, etc.

Kinney and Forsythe did not find a main effect of melodicity on reproduction accuracy, but did find a statistically significant interaction between melodicity and bar-number. To describe that interaction, they reported that melodic presentation benefited reproduction accuracy for bar 1, and impaired reproduction accuracy for bar 2. However, no statistical tests of those two effects were reported.

Despite those ambiguous results, participants' metacognitive judgments strongly favored a benefit of melodicity. Of the 40 participants, 28 reported that the melodic trials were easier, whereas only 5 reported that the melodic trials were harder. The remaining 7 participants reported that melodicity had either no effect or an inconsistent effect.

The Present Study

PRIMARY RESEARCH QUESTION

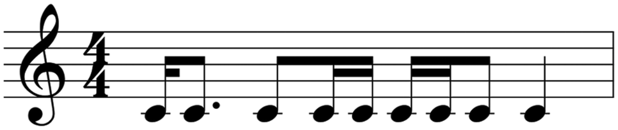
Our primary research question was how melodicity in presented rhythms would affect people's recall and reproduction of those rhythms. Participants listened to novel rhythmic sequences (some melodic, some monotonic), and attempted to reproduce them by tapping. Because of the mixed findings in the Kinney and Forsythe (2013) study, we did not make a prediction regarding the effect of melodicity.

Although our primary research question was essentially the same as that of Kinney and Forsythe (2013), our methods differed from theirs in several important respects. For example, because our pilot participants performed at near-floor accuracy when presented with two-bar sequences, we used shorter sequences for our experiments (unlike the participants in the Kinney and Forsythe study, our participants were not music education majors). Also, to avoid the confounding variables from the Kinney and Forsythe study, our melodic sequences used the same type of sound as our nonmelodic sequences: sine-wave tones of uniform duration. Additionally, whereas Kinney and Forsythe allowed each participant to choose their own reproduction method (clapping, tapping, or vocalizing), we standardized the reproduction method (tapping only).

Another important distinction is that whereas Kinney and Forsythe (2013) used humans both to compose the stimuli and to judge participants' reproduction accuracy, we automated those processes. Consequently, our stimuli were less "inspired" than the Kinney and Forsythe stimuli, but we were able to efficiently generate a large number of sequences and objectively score a large number of participants. Moreover, generating the pitch sequences automatically, and independently of the rhythms, ensured that we would not unconsciously infuse stimuli with derivative phrases or familiar gestalts that could make a given pitch sequence inherently predictive of the corresponding rhythm.

SECONDARY RESEARCH QUESTION

We observed during informal pilot testing that reproduction accuracy tended to be lower for sequences containing a *Scotch snap* (i.e., a sixteenth note at a quarter-note beat position followed by a dotted eighth note; Temperley & Temperley, 2011). Figure 1A shows a rhythmic sequence with a Scotch snap in the first beat.

A. Scotch snap in first beat

B. Strong-position dotted eighth-note in first beat

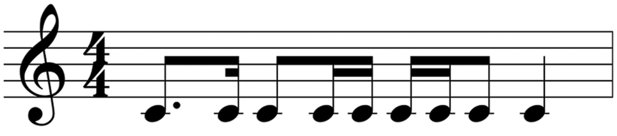



FIGURE 1. Passages with a dotted eighth note in the first beat.

In a Scotch snap, the "weak" metric position of the second note may create an expectation that another note will soon follow. But there is nonetheless a relatively long pause before the next note occurs. Thus, the second note may be considered as an anacrusis whose resolution is delayed, creating a sense of instability (though one could alternatively consider the second note as an extension of the first note, rather than as an anticipation of the third note). This is a form of syncopation at the eighth-note level. Thus, because syncopated rhythms tend to be harder for people to recall and reproduce (Fitch & Rosenfeld, 2007), perhaps it should not be surprising that Scotch snaps appeared to impair reproduction in our pilot testing. Fitch and Rosenfeld suggested that impaired reproduction of syncopated rhythms reflects tension and/or complexity that arise as the listener attempts to reconcile two incongruent streams: (1) the actual notes in the rhythm, and (2) the expected accent structure implied by the underlying pulse.

A dotted eighth note could also be placed at a quarter-note beat position and followed with a sixteenth note (see Figure 1B). This use of a strong-position dotted eighth note (SPDE) creates an example of what Temperley and Temperley (2011) called a "regular dotted pair": a strong-position dotted note followed by a note that has one-third the value of the dotted note. In the case of an SPDE (and regular dotted pairs more generally), the metric and agogic accents are aligned, so there is no delayed resolution of an anacrusis. Thus, the sense of instability may be subtler for an SPDE than for a Scotch snap. Indeed, strength of syncopation has been defined as a function of the interonset interval following, rather than preceding, a weak-position note (Longuet-Higgins & Lee, 1984; see also Fitch & Rosenfeld, 2007). We therefore expected that reproduction of rhythms containing a Scotch snap would be impaired not only relative to other rhythms in general, but also relative to rhythms with an SPDE. In other words, we expected that difficulty reproducing rhythms with a Scotch snap would not be fully explained as simply the presence of a dotted eighth note.

EXPERIMENTAL DESIGN

Two within-subjects experiments were conducted. In each experiment, simple nine-note rhythmic sequences—some melodic, some monotonic—were presented to college students. After each presentation, the participant attempted to reproduce the rhythm by tapping. In Experiment 1, the rhythms and melodic sequences were independently generated for each trial in each condition (melodic/monotonic) for each participant. Experiment 2 replicated the results of Experiment 1 on a different sample, using a single set of rhythms that was presented both melodically and monotonically to all participants. Thus, we were able to observe whether a melodicity effect was evident not only in a majority of participants, but also in a majority of rhythms. The experiments were approved by the college's Institutional Review Board in the Office of the Human Research Protection Program (approval #19-000787). Data, stimulus, and analysis files are available at https://osf.io/qcf7p/.

Experiment 1

METHOD

Participants. Participants consisted of 50 volunteers (36 female, 14 male, nearly all 18–25 years old, median age = 20). They were recruited from a participant pool of college students, most of whom were enrolled in introductory psychology courses. They received course credit for their participation. All participants reported normal hearing and normal (or corrected-to-normal) vision.

On a post-experiment questionnaire, 40% of participants reported that they played a musical instrument. In that subset of participants, the median self-rated musical skill level was 6.5 on a scale from "1" (*beginner*) to "10" (*very high*). Only two participants reported playing a percussion instrument. In open-ended responses to "What style of music do you listen to the most," the vast majority of participants responded either "pop," "hip hop," "R&B," "indie," "alternative," or a genre similar to one of those categories (e.g., "rap"), or a combination of two or more of those categories (e.g., "pop and alternative").

Stimuli. For each participant, 80 rhythmic sequences (not including practice-trial sequences) were generated: 40 melodic sequences (20 of which were in C major, 20 of which were in C natural-minor) and 40 monotonic sequences (in which the pitch of every note was C4, commonly known as middle C). Each sequence consisted of nine notes. Tempo was fixed at 56 quarter-note beats per minute, which our pilot testing had suggested was comfortable for tapping. Each note was a 200-ms sine wave tone with a linear fadeout applied to the last 40 ms, and thus was fairly staccato. Tone frequencies were based on standard 12-tone equal-temperament tuning in the A-440 system (e.g., C4 corresponded to approximately 261.6 Hz).

For each participant, the rhythm of each sequence was independently randomly generated using a Matlab program, under the following constraints:

- There was a note onset at each quarter-note beat position from 1 to 4, in order to ensure a strong, coherent pulse.
- Each of the remaining five note-onsets occurred at a sixteenth-note grid position that was later than the beat 1 onset, earlier than the beat 4 onset, and not occupied by any other note. This constraint ensured that all sequences had the same absolute length, the same mean number of notes per second, and (due to the strong metric positions of the first and last notes) a sense of "completeness" (Palmer & Krumhansl, 1987).
- There were not note onsets at more than five adjacent sixteenth-note grid-positions in a row. This constraint, in conjunction with the next constraint, prevented rhythms from having a highly "lopsided" rhythmic density and from being overly simplistic.
- At no point before beat 4 were there more than two adjacent sixteenth-note grid-positions in a row without a note onset.

Both sequences in Figure 1 satisfy all the above constraints.

For each participant, the melody of each melodic sequence was independently randomly generated using a Matlab program, under the following constraints:

- The pitch of the first and last notes was C4, in order to ensure a strong sense of tonicity and completeness.
- For notes 2 through 8, each pitch was either one scale-tone below, one scale-tone above, or the same scale-tone as, the preceding note. This constraint ensured that all melodies were fairly simple and mostly stepwise. An exception to this constraint was that for major melodies, the pitch of note 2 was allowed to be two scale-tones below note 1. This exception was allowed because we judged it not to disrupt the simple, conventional feel of the melodies.
- No pitch was below G3 or above G4. This constraint prevented unusually large pitch-intervals between the penultimate note and the last note.
- No pitch occurred more than twice in a row. This constraint ensured that the melodic sequences were sufficiently melodic.

For each participant, four additional sequences (one minor melodic, one major melodic, two monotonic) were generated to use in the practice trials. These practice sequences were generated under the same constraints that were used for the sequences in the experiment proper.

Procedure. Each participant wore Sony MDR-7506 headphones, and was seated in front of a computer in a quiet room. The participant was given the following instructions, both verbally and on the screen: “In this study, you will hear a series of tones, and then you will copy the rhythm of those tones by tapping the spacebar. Press <Return> to practice.” The participant then completed the four practice trials in the presence of the experimenter, in the following order: monotonic, minor melodic, monotonic, major melodic. No feedback was given regarding accuracy, and the tapping produced no sound other than the unamplified mechanical sounds of the keystrokes themselves.

The participant was then given the following instructions, both verbally and on the screen: “Now we’ll start the experiment. It works just like the practice that you did. Press <Return> to begin.” The experimenter then left the room, and the participant completed 80 trials (one for each generated sequence), with a 30-second break after the 40th trial. The trial order was independently randomized for each participant, under the following constraints: (1) In each 40-trial half of the experiment, there were 10 major-melodic, 10 minor-melodic, and 20 monotonic trials, (2) there were never more than two melodic or two monotonic trials in a row. As in the practice trials, no feedback was given regarding accuracy, and the tapping produced no sound other than the unamplified mechanical sounds of the keystrokes themselves. Figure 2 shows a diagram of the trial procedure.

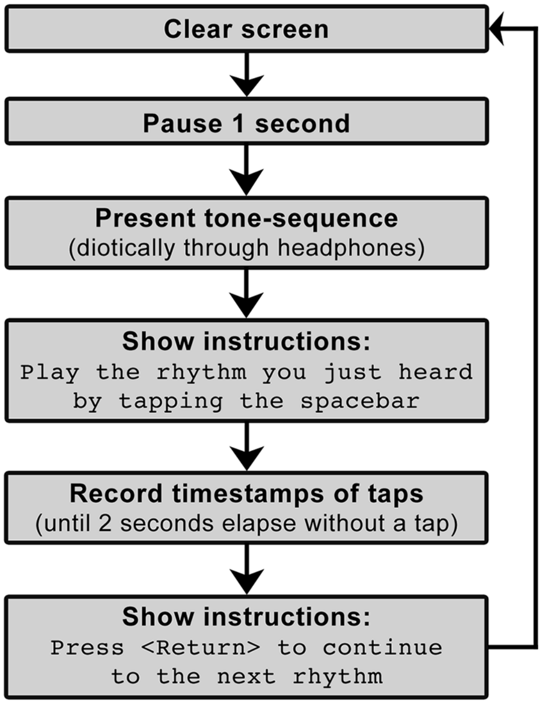


FIGURE 2. Trial procedure.

After the experiment, the participant completed a questionnaire using the computer keyboard. The text of the first questionnaire item was as follows: “In some of the rhythms that you heard, the tones all had the same pitch. In other rhythms, there was a melody, meaning that the tones had different pitches. Which of the following statements do you agree with most? (A) Rhythms with melody were EASIER to remember than rhythms without melody, (B) Rhythms with melody were HARDER to remember than rhythms without melody, (C) Rhythms with melody were NEITHER EASIER NOR HARDER to remember.” The questionnaire also collected the data summarized in the Participants section.

Computations. So that mere tempo variation and timing “looseness” in participants’ tapping could not cause intertap intervals to be judged as incorrect, the recorded intertap intervals were quantized at the sixteenth-note level, based on the presentation tempo. This made accuracy computations more efficient and straightforward. But it should be acknowledged that analyzing the quantized intertap intervals, rather than the absolute intertap intervals, does not consider microtiming nuances, such as the extent to which taps at incorrect metric positions were biased toward the correct position, and the extent to which taps at nominally correct metric positions were biased slightly early or late.

Accuracy of rhythm reproduction was computed as the mean of the following two percentages: (1) percentage of “first-half” intertap intervals (i.e., of the first four quantized intertap intervals in each trial) that were equal to the corresponding interonset intervals in the presented sequence, (2) percentage of “second-half” intertap intervals (i.e., of the last four quantized intertap intervals in each trial) that were equal to the corresponding interonset intervals in the presented sequence. Thus, theoretically possible accuracy scores range from 0 to 100. All mean differences in accuracy that are reported in this paper are on that unitless scale.

Computing accuracy using this “averaged-halves” approach prevented extra or omitted notes in the first half of a tapped sequence from causing correct intertap intervals in the second half to be mislabeled as incorrect (as might otherwise occur due to misalignment between the indexes of the intertap intervals and the indexes of the presented interonset intervals). Distributions of accuracy using this metric appeared relatively normal (Shapiro–Wilk $W > .96$ in both conditions in both experiments).

Additional metrics of accuracy could have been used, but any reasonable metric of accuracy would likely be highly correlated with the metric we used, at least for sequences as short as the ones in this study. An alternative analysis (not reported in detail here) that scored each trial in a binary correct/incorrect manner produced scores that were less normally distributed (when averaged across trials), but resulted in the same significance decisions.

Accuracy comparisons were conducted using two-sided paired-samples *t*-tests. But an alternative analysis using two-sided Wilcoxon signed-rank tests (not reported in detail here) produced the same significance decisions. Each statistically significant test ($p < .05$) in Experiment 1 was replicated in Experiment 2, and would remain statistically significant even using a Bonferroni-adjusted alpha level of .007 to account for all seven comparisons tested in this study. The CIs that are reported without corresponding *p* values are for comparisons that were not in the original analysis plan, and thus should essentially be taken as descriptive.

RESULTS

Effects of Melodicity. Accuracy was higher in monotonic trials than in melodic trials; $t(49) = 4.71$, $p < .0001$, mean difference = 2.78, 95% CI = [1.60, 3.97]). This is illustrated in Figure 3, which shows monotonic versus melodic trial accuracy for each participant. The diagonal line is the *line of equality*, meaning that points above the line represent participants who performed better on monotonic trials, and points below the line represent participants who performed better on melodic trials. Note that the vast majority of points are above the line.

Interestingly, although most participants performed worse on the melodic trials, they tended not to realize it. In fact, only 20 of the 50 participants reported that the melodic sequences were harder, whereas 25 reported that the melodic sequences were easier, and 5 reported that the melodic sequences were neither easier nor harder. The shading of the points in Figure 3 indicates participants' metacognitive judgments regarding the relative difficulty of melodic and monotonic trials. Note that the graph does not suggest a strong relationship between metacognitive judgment and actual relative performance on melodic and monotonic trials. We estimated the interaction of melodicity and metacognitive judgment as the monotonic-minus-melodic accuracy difference among participants who said "harder" subtracted from the monotonic-minus-melodic accuracy difference among participants who said "easier." The Welch's 95% CI for that interaction included zero, though it nominally favored the congruent direction: [−3.60, 1.40]. Thus, there was not substantial evidence that the perceived accuracy-disparity between melodic and monotonic trials was related to actual accuracy-disparity between melodic and monotonic trials.

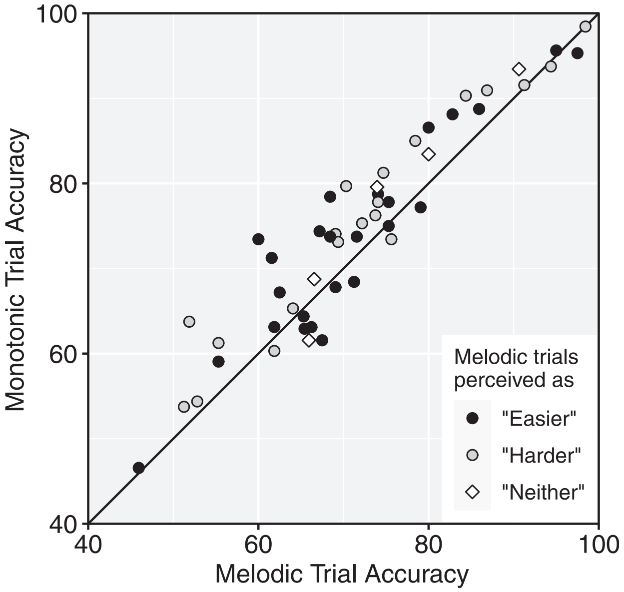


FIGURE 3. Monotonic versus melodic accuracy for each subject in Experiment 1.

Effects of Scotch Snaps. Approximately 36% of the presented sequences contained a Scotch snap. As expected, accuracy was considerably lower for those sequences than for other sequences; $t(49) = 5.41$, $p < .0001$, mean difference = 4.12, 95% CI = [2.59, 5.65]. Approximately 37% of the presented sequences contained an SPDE. Accuracy was lower in trials with a Scotch snap than in trials with an SPDE; $t(49) = 3.28$, $p = .002$, mean difference = 2.62, 95% CI = [1.02, 4.23]. Only about 3% of presented sequences contained both a Scotch snap and an SPDE.

Effects of Mode. There was not substantial evidence for an effect of mode on accuracy; $t(49) = 0.78$, $p = .437$, major minus minor mean difference = 0.53, 95% CI = [−0.82, 1.87].

Accuracy of First Four Versus Last Four Intertap Intervals. We compared accuracy in the first four intertap intervals to accuracy in the last four intertap intervals. This comparison was not part of our original analysis plan, but was suggested by a reviewer. Mean accuracy in the first four (81.12 for monotonic, 78.29 for melodic) was much higher than mean accuracy in the last four (68.41 for monotonic, 65.69 for melodic). The 95% CI for the mean difference was [10.60, 14.82] for monotonic, and [10.21, 14.99] for melodic. It is not clear to what extent that accuracy drop-off reflects limitations of the accuracy assessments themselves. But it is notable that Kinney and Forsythe (2013) also found that rhythm reproduction accuracy was lower in the second halves of sequences.

Experiment 2

METHOD

Participants. Participants consisted of 50 volunteers (31 female, 19 male, nearly all 18–24 years old, median age = 20). These participants had not participated in Experiment 1, but were recruited from the same pool that was used for Experiment 1. They received course credit for their participation. All participants reported normal hearing and normal (or corrected-to-normal) vision.

Fifty percent of participants reported that they played a musical instrument. In that subset of participants, the median self-rated musical skill level was 6 on a scale from "1" (*beginner*) to "10" (*very high*). As in Experiment 1, only two participants reported playing a percussion instrument. Responses to "What style of music do you listen to most" were similar to those in Experiment 1.

Stimuli. The stimulus generation method was the same as in Experiment 1, except in the following three respects: (1) rather than independently generating a unique set of 80 sequences for each participant, a single set of 80 sequences (40 melodic, 40 monotonic) was used for all participants, (2) rather than independently generating the rhythms for all 80 sequences, a single set of 40 rhythms was used for both the melodic and monotonic sequences, (3) all 40 melodic sequences were in C major because the results of Experiment 1 suggested that mode was not an important factor. There were only 35 unique rhythms, because we neglected to include a constraint that each of the 40 rhythms be unique. However, each of the 40 pitch–rhythm combinations in the melodic trials was unique. In addition to the 80 sequences that were used in the experiment proper, four practice sequences were generated: two major melodic, two monotonic.

Procedure. The procedure was essentially the same as in Experiment 1. The trial order was independently randomized for each participant, under the following three constraints: (1) in each 40-trial half of the experiment, there were 20 melodic trials and 20 monotonic trials, (2) each of the 40 rhythms was presented melodically in one half of the experiment and monotonically in the other half, (3) there were never more than two melodic or two monotonic trials in a row.

We added a fourth response option ("Not sure") to the questionnaire item about the relative difficulty of melodic and monotonic trials. We included this option to avoid pressuring participants into making arbitrary judgments that they did not have confidence in.

Computations. Accuracy computations were analogous to those in Experiment 1. Each statistically significant comparison in Experiment 2 was a replication of a statistically significant comparison in Experiment 1, and would remain statistically significant even using a Bonferroni-adjusted alpha level of .007 to account for all seven comparisons tested in this study. The CIs that are reported without corresponding p values are for comparisons that were not in the original analysis plan, and thus should essentially be taken as descriptive.

RESULTS

Effects of Melodicity. As in Experiment 1, accuracy was higher in monotonic trials than in melodic trials; $t(49) = 5.26$, $p < .0001$, mean difference = 2.88, 95% CI = [1.78, 3.97]. Figure 4 shows monotonic versus melodic trial accuracy for each participant. As in Figure 3, the vast majority of points are above the diagonal line of equality, indicating that the vast majority of participants performed better on monotonic trials than on melodic trials.

As in Experiment 1, although most participants performed worse on the melodic trials, they tended not to realize it. In fact, only 14 of the 50 participants reported

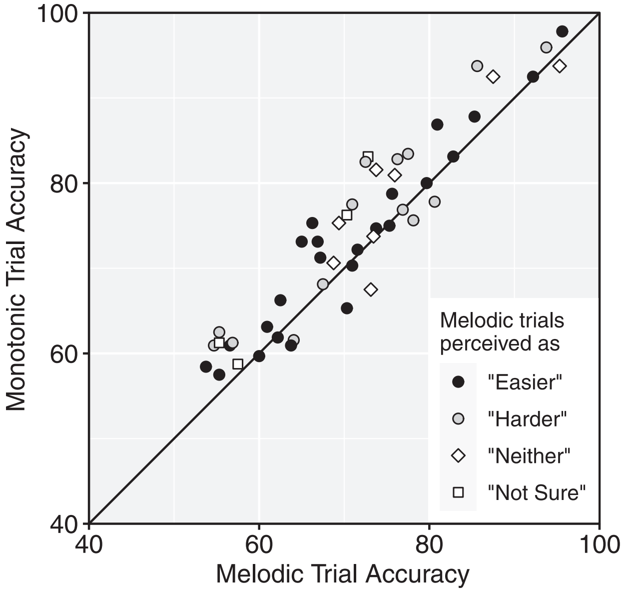


FIGURE 4. Monotonic versus melodic accuracy for each subject in Experiment 2.

that the melodic sequences were harder, whereas 24 reported that the melodic sequences were easier, 8 reported that the melodic sequences were neither easier nor harder, and 4 reported "not sure." The shading of the points in Figure 4 indicates participants' metacognitive judgments regarding the relative difficulty of melodic and monotonic trials. Note that as in Figure 3, the graph does not suggest a strong relationship between metacognitive judgment and actual relative performance on melodic and monotonic trials. As in Experiment 1, we estimated the interaction of melodicity and metacognitive judgment. The Welch's 95% CI for the interaction included zero, though it nominally favored the congruent direction: [−4.20, 1.36]. Thus, there was not substantial evidence that the perceived accuracy-disparity between melodic and monotonic trials was related to actual accuracy-disparity between melodic and monotonic trials.

Figure 5 shows monotonic versus melodic trial accuracy for each unique rhythm. It is analogous to Figure 4, but each point represents a rhythm (averaging accuracy across participants) rather than a participant. The vast majority of points are above the diagonal line of equality, indicating that the vast majority of rhythms were reproduced more accurately when presented monotonically than when presented melodically.

Effects of Scotch Snaps. In 30% of trials, the presented sequence contained a Scotch snap. As in Experiment 1, accuracy was considerably lower in trials with a Scotch snap than in trials without a Scotch snap; $t(49) = 7.29$, $p < .0001$, mean difference = 5.36, 95% CI = [3.88, 6.84]. In 37.5% of trials, the presented sequence contained an SPDE. As in Experiment 1, accuracy was lower in trials with a Scotch snap than in trials with an SPDE; $t(49) = 4.71$, $p < .0001$, mean difference = 3.95, 95% CI = [2.27, 5.64]. The shading of the points in Figure 5 indicates whether given rhythms contained a Scotch snap, an SPDE, or neither. None of the rhythms contained both a Scotch snap and an SPDE.

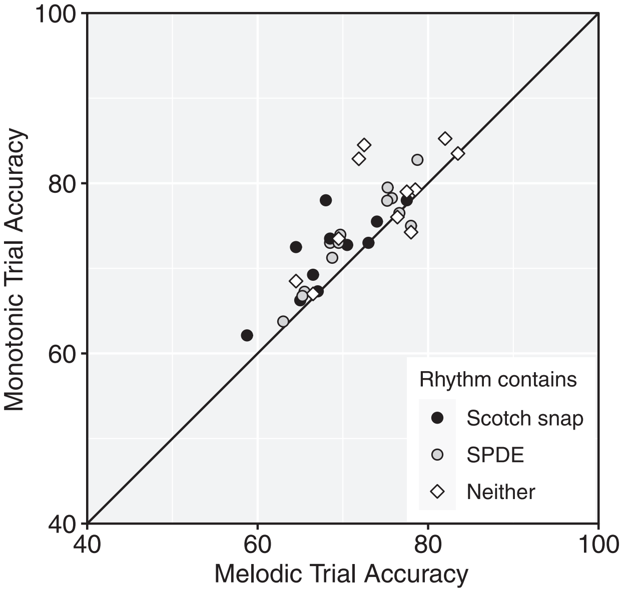


FIGURE 5. Monotonic versus melodic accuracy for each unique rhythm in Experiment 2.

A common error when reproducing Scotch snaps was to play the second note of the snap as an eighth note, instead of as a dotted eighth note. In fact, when a presented rhythm had a Scotch snap in the first beat (which was the case in 5 of the 40 rhythms), participants tapped the second note as an eighth note 30% of the time on average. The same error also frequently occurred for Scotch snaps at other beat-positions, but exact frequencies could not be determined because in poorer reproductions, correspondences could not be definitively established between a beat-position in the presented rhythm and a particular temporal position in the reproduction.

Accuracy of First Four Versus Last Four Intertap Intervals. As in Experiment 1, mean accuracy in the first four intertap intervals (81.21 for monotonic, 78.34 for melodic) was much higher than mean accuracy in the last four (67.65 for monotonic, 64.78 for melodic). The 95% CI for the mean difference was [11.30, 15.83] for monotonic, and [11.14, 15.98] for melodic.

General Discussion

The primary purpose of this study was to evaluate how melodicity in presented rhythms affects reproduction accuracy. We found that melodicity impaired reproduction. The secondary goal of this study was to evaluate how Scotch snaps affect reproduction accuracy. We expected Scotch snaps to impair reproduction, and that hypothesis was supported by the results.

IMPAIRED REPRODUCTION OF MELODICALLY PRESENTED RHYTHMS

Experiments 1 and 2 both showed that reproduction was definitively less accurate for melodically presented rhythms than for monotonically presented rhythms. Thus, it appears that melodicity does not fundamentally enhance rhythm recall and reproduction. In fact, melodicity appears to impair immediate reproduction of rhythm, at least for the type of stimuli used in this study.

The simplest explanation for this impairment is that the pitch variation distracts attentional resources from the temporal information, resulting in less accurate representation (though there could also be impairment at the reproduction stage, due to the inherent incongruence between the one-dimensional rhythm-only reproduction and the two-dimensional pitch-and-rhythm presentation). The same attentional explanation has been proposed for the disruptive effects of pitch variation on temporal interval judgments in psychoacoustic experiments (Grose et al., 2007). Another explanation is that the unfamiliar combination of pitch-sequence and rhythm in a given stimulus can cause the listener to unconsciously distort the rhythm, so it more closely resembles familiar pitch–rhythm combinations (such as a known song or a previous stimulus). In future studies, comparing impairment for the type of stimuli used in this study to impairment when pitch sequences are less musical (i.e., when tone-frequencies do not closely correspond to conventional scale tones) may help disambiguate purely psychoacoustic effects from music-specific effects.

The possibility remains that melody enhances rhythm reproduction in some cases. Note that the sequences in this study were quite short, and the reproductions were performed immediately after a single exposure. Future studies can examine longer, repeatedly presented sequences, with varied amounts of delay between presentation and reproduction. Presumably, repeated presentation will avoid the floor effect that we observed in our pilot testing with longer sequences. Using longer sequences would also allow different levels of coherence between pitch-contour and rhythm to be examined (see Boltz, 1998). Additionally, a reviewer suggested that blocked presentation of monotonic and melodic trials might yield different results than interleaved presentation.

Note also that because the sequences in this study were randomly generated under strict constraints, they were likely not very interesting to the participants. Perhaps for musical passages that are not only longer, but also more distinctive (Bailes, 2010; Müllensiefen & Halpern, 2014) and more emotive, there is more opportunity for melody-based semantics to enhance the listener's encoding and recall of the rhythm. Future studies can explore this possibility by using both randomly generated sequences and expertly composed sequences. Additionally, even randomly generated sequences can likely be made more interesting by making the constraints more sophisticated and less restrictive.

ILLUSORY BENEFIT OF MELODICITY

Although participants tended to perform worse on melodic trials, a plurality of them reported that the melodic trials were easier, as did most of the participants in the Kinney and Forsythe (2013) study. The reason for this apparent metacognitive error is not clear. Perhaps melodicity can help legitimize an incorrect reproduction in a participant's mind, thereby inflating confidence on melodic trials. For instance, if the participant unconsciously alters the rhythm to fit the pitch sequence in a way that is subjectively "better," then the altered rhythm's seeming congruence with the melody might validate the alteration—making it "feel right." This mistaking of fluency for accuracy would be roughly analogous to the "illusions of familiarity" (Whittlesea, 1993) that have been observed in recognition memory tasks.

On the other hand, because perceived accuracy-disparity between melodic and monotonic trials did not appear strongly related to actual accuracy-disparity between melodic and monotonic trials, perhaps a better explanation is that many participants were "giving the right answer to the wrong question." That is, participants may have based their judgment on an intuition about whether melodicity benefits memory for musical passages in general, rather than reflecting on their actual experience with the novel stimuli in the experiment. After all, for a familiar musical passage, melodic presentation would of course produce superior recognition than rhythm-only presentation would. Recall that even well-known songs, such as "Happy Birthday To You," tend to be unrecognizable in rhythm-only form (Newton, 1990). And it seems unlikely that one could attempt to tap the rhythm of "Happy Birthday To You," or any other familiar song, without hearing/singing the melody in one's mind to some extent. Moreover, in everyday music recall scenarios (e.g., when asking one's self "How does that song go?"), pitch-sequence and rhythm seem to be inextricably linked. Indeed, one does not typically find that the rhythm of a song is recalled, either effortfully or spontaneously, without an associated pitch sequence also coming to mind (though that pitch sequence may not be exactly correct).

It should also be acknowledged that the question of whether something is easier/harder "to remember" is somewhat vague, and that there is not necessarily a one-to-one correspondence between reproduction accuracy and stimulus memorability per se. In theory, one sequence might be less accurately represented than another, and thus less accurately reproduced in the short term, yet still be more likely to "stick" in one's mind (albeit in inaccurate form) in the long term.

In future studies of this type, it will likely be more informative to avoid mentioning melodicity to the participants at all, and instead have participants report a confidence judgment immediately after each trial. That way, mismatches between accuracy and confidence can be more directly analyzed, and participants will not be explicitly speculating about the effect of the independent variable. It may also be useful to have participants evaluate playback of their tapping, as well as report how comfortable they were with using the spacebar (or whatever instrument was being tapped), to help disambiguate execution errors from recall errors per se.

IMPAIRED REPRODUCTION OF RHYTHMS CONTAINING A SCOTCH SNAP

As expected, Experiments 1 and 2 both showed that reproductions were less accurate for rhythms containing a Scotch snap than for other rhythms. That extends the findings of Fitch and Rosenfeld (2007), who found that syncopated rhythms in general were more difficult for people to recall and reproduce. As expected, we also found that reproductions were less accurate for rhythms containing a Scotch snap than for other rhythms containing dotted eighth notes. That supports the view that there is greater instability when the weak-position note precedes, rather than follows, a relatively long interonset interval (Fitch & Rosenfeld, 2007; Longuet-Higgins & Lee, 1984). As noted in our introduction, this instability can be conceptualized as the delayed resolution of an anacrusis.

Scotch snaps are fairly common in the drum patterns and vocal cadences of contemporary rap and pop music (Marshall, 2019; Neely, 2019). However, when Scotch snaps appear, they often appear repeatedly, e.g., on every beat in a measure (as occurs in Bruno Mars' vocals in "24 K Magic" starting at 1:17; Mars, Lawrence, & Brown, 2016). In contrast, Scotch snaps in the present study's stimuli were unpredictable. Future studies can explore whether Scotch snaps remain harder to reproduce than other phrases when rhythms are more repetitive.

Conclusions

Previous speculations by some researchers and by some study participants suggested that melodicity might benefit recall and reproduction of novel musical rhythms. But our findings suggest the opposite: Melodicity impairs recall and reproduction. We have suggested some future directions to help explain this effect and verify the extent to which it generalizes beyond the type of stimuli used in this study. The illusory benefit of melodicity remains mysterious, but we have suggested future directions to help explore that phenomenon as well. More generally, how pitch and rhythm information combine in perception and memory continues to be an intriguing area of inquiry.

Author Note

Correspondence concerning this article should be addressed to Andrew Frane, UCLA Department of Psychology, 502 Portola Plaza, 1285 Psychology Building, Los Angeles, CA 90095. E-mail: avfrane@ucla.edu

References

Ammirante, P., & Thompson, W. F. (2012). Continuation tapping to triggered melodies: Motor resonance effects of melodic motion. *Experimental Brain Research, 216*, 51–60. DOI: 10.1007/s00221-011-2907-5

Ammirante, P., Thompson, W. F., & Russo, F. A. (2011). Ideomotor effects of pitch on continuation tapping. *The Quarterly Journal of Experimental Psychology, 64*(2), 381–393. DOI: 10.1080/14740218.2010.495408

Bailes, F. (2010). Dynamic melody recognition: Distinctiveness and the role of musical expertise. *Memory and Cognition, 38*(5), 641–650. DOI: 10.3758/MC.38.5.641

Boltz, M. G. (1998). The processing of temporal and nontemporal information in the remembering of event durations and musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1087–1104. DOI: 10.1037/0096-1523.24.4.1087

Crowder, R. G., & Neath, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception, 12*(4), 379–386. DOI: 10.2307/40285672

David, M., Lavandier, M., & Grimault, M. (2014). Room and head coloration can induce obligatory stream segregation. *Journal of the Acoustical Society of America, 136*(5), 5–8. DOI: 10.1121/1.48833871

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 299–348). Cambridge, MA: Academic Press. DOI: 10.1016/B978-012213564-4/50010-X

Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception and Psychophysics, 21*(2), 125–142. DOI: 10.3758/BF0398/BF03198716

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics, 14*(1), 37–40. DOI: 10.3758/BF03198614

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception, 25*(1), 43–58. DOI: 10.1525/MP.2007.25.1.43

Gabrielsson, A. (1973a). Similarity ratings and dimension analyses of auditory rhythm patterns. I. *Scandinavian Journal of Psychology, 14*(1), 138–160. DOI: 10.1111/j.1467-9450.1973.tb00105.x

Gabrielsson, A. (1973b). Similarity ratings and dimension analyses of auditory rhythm patterns. II. *Scandinavian Journal of Psychology, 14*(1), 161–176. DOI: 10.1111/j.1467-9450.1973.tb00106.x

Grose, J. H., Hall III, J. W., & Bus, E. (2007). Gap duration discrimination for frequency-asymmetric gap markers: Psychophysical and electrophysiological findings. *Journal of the Acoustical Society of America, 122*(1), 446–457. DOI: 10.1121/1.2735106

Grose, J. H., Hall III, J. W., Bus, E., & Hatch, D. (2001). Gap detection for similar and dissimilar gap markers. *Journal of the Acoustical Society of America, 109*(4), 1587–1595. DOI: 10.1121/1.1354983

Guinness Books. (1998). *The Guinness book of world records.* New York: Bantam.

Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory and Cognition, 25*(4), 518–533. DOI: 10.3758/BF03201127

Heinrich, A., de la Rosa, S., & Schneider, B. A. (2014). The role of stimulus complexity, spectral overlap, and pitch for gap-detection thresholds in young and old listeners. *Journal of the Acoustical Society of America, 136*(4), 1797–1807. DOI: 10.1121/1.4894788

Herff, S. A., Olsen, K. N., Prince, J., & Dean, R. T. (2018). Interference in memory for pitch-only and rhythm-only sequences. *Musicae Scientiae, 22*(3), 344–361. DOI: 10.1177/1029864917695654

Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception and Psychophysics, 47*(3), 215–226. DOI: 10.3758/BF03204997

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception and Psychophysics, 41*, 621–634. DOI: 10.3758/BF03210494

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814. DOI: 10.1037/0033-2909.129.5.770

Kinney, D. W., & Forsythe, J. L. (2013). Does melody assist in the reproduction of novel rhythm patterns? *Contributions to Music Education, 39*, 69–85. Retrieved from https://www.jstor.org/stable/24127245

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin, 126*(1), 159–179. DOI: 10.1037/0033-2909.126.1.159

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1*(4), 424–441. DOI: 10.2307/40285271

Mars, B., Lawrence, P., & Brown, C. B. (2016). 24K magic. [Recorded by Bruno Mars]. On *24 K magic* [Digital album]. New York: Atlantic.

Marshall, W. (2019, April 1). Ariana Grande was accused of copying '7 Rings,' again and again . . . and again. But did she actually do anything wrong? *Vulture*. Retrieved from https://www.vulture.com/2019/04/did-ariana-grande-copy-7-rings.html

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception and Psychophysics, 41*(6), 576–600. DOI: 10.3758/BF03210491

Moog, H. (1979). On the perception of rhythmic forms by physically handicapped children and those of low intelligence in comparison with non-handicapped children. *Bulletin of the Council for Research in Music Education, 59*, 73–78. Retrieved from https://www.jstor.org/stable/40317547

Müllensiefen, D., & Halpern, A. R. (2014). The role of features and context in recognition of novel melodies. *Music Perception, 31*(5), 418–435. DOI: 10.1525/mp.2014.31.5.418

Neely, A. (2019, March 11). *Scotch snaps in hip hop* [Video]. YouTube. Retrieved from https://www.youtube.com/watch?v=i7cG9QIvIWo

Newton, E. L. (1990). *The rocky road from actions to intentions* (Doctoral dissertation). Stanford University. Retrieved from https://creatorsvancouver.com/wp-content/uploads/2016/06/rocky-road-from-actions-to-intentions.pdf

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 116–126. DOI: 10.1037/0096-1523.13.1.116

Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psyhologica, 141*(2), 184–198. DOI: 10.1016/j.actpsy.2012.07.013

Schellenberg, E G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception, 18*(2), 155–171. DOI: 10.2307/40285907

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception and Psychophysics, 40*(1), 9–19. DOI: 10.3758/BF03207588

Shigeno, S. (1993). The interdependence of pitch and temporal judgments by absolute pitch possessors. *Perception and Psychophysics, 54*(5), 682–692. DOI: 10.3758/BF03211792

Sink, P. E. (1983). Effects of rhythmic and melodic alterations on rhythmic perception. *Journal of Research in Music Education, 31*(2), 101–113. DOI: 10.2307/3345214

Sink, P. E. (1984). Effects of rhythmic and melodic alterations and selected musical experiences on rhythmic processing. *Journal of Research in Music Education, 32*(3), 177–193. DOI: 10.2307/3344837

Temperley, N., & Temperley, D. (2011). Music–language correlations and the "Scotch snap." *Music Perception, 29*(1), 51–63. DOI: 10.1525/MP.2011.29.1.51

White, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology, 73*(1), 100–107. DOI: 10.2307/1419120

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1235–1253. DOI: 10.1037/0278-7393.19.6.1235

Yoblick, D. A., & Salvendy, G. (1970). Influence of frequency on the estimation of time for auditory, visual, and tactile modalities. *Journal of Experimental Psychology, 86*(2), 157–164. DOI: 10.1037/h0029935