Spring 2019

The effects of timbre on perceptual grouping in a melodic sequence

Thomas Rohaly

Recommended Citation

Rohaly, Thomas, "The effects of timbre on perceptual grouping in a melodic sequence" (2019). *Masters Theses*. 599.
https://commons.lib.jmu.edu/master201019/599

The Effects of Timbre on Perceptual Grouping in a Melodic Sequence

Thomas Rohaly

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts

Department of Graduate Psychology

May 2019

FACULTY COMMITTEE

Committee Chair: Michael D. Hall, Ph.D.

Committee Members/Readers:

Jeff Dyche, Ph.D.

Kethera Fogler, Ph.D.

Acknowledgements

I would like to express my gratitude to my advisor, Dr. Michael Hall, for his support and encouragement throughout these past two years. It has been a pleasure to work with and learn from him throughout my time here and my thesis would not have been possible without his guidance. I would also like to thank my thesis committee, Dr. Kethera Fogler and Dr. Jeff Dyche, who have constantly given me their full support and have been very patient with the ups and downs of this thesis project. Both have been very important to my professional development over the past couple of years and I am very grateful for all of the assistance they have given me. Additionally, I would like to thank the students in my cohort, as I would probably not have made it through this program without their support and friendship these past two years. Finally, I would like to thank my parents, Chris and Cindy Rohaly, and my siblings, Matt and Anna, for all of the love and encouragement I have received from them throughout my life.

Table of Contents

List of Figures

Abstract

The current investigation sought to examine the effects of timbre on perceptual grouping in melodic sequences. While past research has shown that timbre shifts influence listeners' pitch perception on a note-to-note basis (e.g., see Pitt, 2004; Russo & Thompson, 2005, & Creel, Newport, & Aslin, 2004), the current investigation extended this to timbre's influence on pitch perception in the context of a melodic phrase. In Experiment 1, participants were presented with melodic sequences, made of sawtooth-like waves. Sequences, consisting of 6 tones, were followed by a target tone that had a static, dull, or bright timbre shift through the use of low-pass filters in order to shift the spectral centroid. Target tones were equally presented at ascending and descending interval sizes of a minor 3rd, perfect 4th, and minor 6th. These target tones were paired with timbre conditions, equally, to create timbre shifts that were static, where timbre did not change at all, congruent, where timbre and the pitch of the target moved in the same direction, and incongruent, where timbre and pitch moved in opposite directions. Participants were tasked with rating how well the target tone belonged to the sequence before it. Experiment 2 extended a similar approach to instrumental stimuli. Cello samples were filtered so that the corresponding impact on spectral centroids was similar to the timbre manipulation in Experiment 1. Contrary to hypotheses, participants rated target tones as being less likely to continue the initial melody if any form of timbre shift was present, regardless of interval size. This effect of timbre suggests that it was not subsumed by high-order processes of melody perception. As hypothesized, this effect was negatively related to musical training. Additionally, as expected, interval size influenced ratings regardless of timbre shifts, where larger intervals were less likely to be perceived as belonging to the initial melody. Thus, participants also appear to have used expectations about pitch intervals to make judgments. Finally, the direction of the initial interval within the sequence also influenced target judgments when the target tone constituted a shift in

timbre, indicating that participants used directional information to create expectations for the target pitch. Taken together, the findings from the current investigation minimally indicate that, at least under conditions reflecting a single change in instrument source, timbre has the capacity to drastically impact the perception of melodic phrase structure.

The Effects of Timbre on Perceptual Grouping in a Melodic Sequence

The perception of melody is a complicated process that includes the combined input of both high-level and low-level processes. Determination of melodic structure is based upon the grouping of tones according to perceptual/cognitive principles and consists of many factors in order to create this grouping. These factors include pitch proximity (e.g., see Anta, 2013), generation of expectations (e.g., Cenkerova' & Parncutt, 2015), and statistical probabilities (e.g., Creel, Newport, & Aslin, 2004). In contrast, the potential influence of timbre on grouping of melodic structure has not been researched as frequently. While timbre is generally defined by what it is not, it can broadly be considered the quality of a sound that a listener uses to identify the sound source (e.g., a cello vs. a piano).

Factors that influence melodic structuring can be broken down into those that reflect both low and high-level processing. An example of a low-level factor is pitch proximity, which is the frequency-distance between two tones. When interwoven sequences are separated by pitch, or have different timbres, they will be processed as two different streams of information, whereas if there is no separation in pitch they are perceived as a single stream (Creel, Newport, & Aslin, 2004). For example, in any musical sequence, strong pitch proximity would have small interval sizes, whereas weak proximity would have large interval sizes. While looking at the effects of pitch proximity on melodic expectations, some support has been found to indicate that proximity for both the last and the penultimate tones in a sequence influence the expectation of a target tone (Anta, 2013). Although both the penultimate tone and last tone of a sequence influence expectations, the penultimate tone has less of an influence than the final tone. While the penultimate tone’s influence is weak compared to the final tone of a sequence, this could indicate that listeners do not solely take individual tones into account when creating expectations, but might use the context of the entire sequence presented. The primary goal of the current investigation was to evaluate the potential influence of timbre on grouping judgments of a target tone within a sequence. Specifically, the aim was to investigate whether timbre will have an influence on grouping when an entire sequence is present to give context to the listener. This kind of influence has been found in intervals (Russo & Thompson, 2005), but has not been tested in the presence of a sequence of tones, sequences that consist of more than two tones.

Another low-level surface feature in melody is tempo, the speed at which notes are played. An individual's perception of music is often dependent upon the temporal relationship that tones have between each other in a sequence. As such, tempo is commonly used as a manipulation in melody recognition. A series of experiments examined the influence of pitch proximity and temporal cues on perceptual grouping within a sequence (Hamaoui & Deutsch, 2010). Participants were exposed to sequences and were asked to report whether they perceived three or four groups in each sequence. While pitch proximity always influenced grouping, very small changes in inter-tone intervals were also found to influence how participants grouped tones within a sequence. As the inter-tone intervals increased, from 0-60 ms, the percent of judgements based on pitch decreased. These findings reveal that tempo is a strong cue in perceptual grouping within a sequence. However, tempo has been shown to be a surface feature of melody that fades in memory more quickly than other features (Schellenberg & Habashi, 2015). Participants were exposed to melodies and then, after a delay of 10 minutes, one day, or one week, were presented the same melodies again. These melodies were either unchanged, had a tempo change, or had a key change, and participants completed a recognition task. When melodies were presented a second time with a different key or tempo, there were negative effects towards recognition when the delay was 10 minutes or a day, but not a week. A second experiment repeated the same methodology but extended the manipulation to included timbre changes, going from piano to saxophone or vice versa, as a condition. When melodies were presented a second time, timbre changes negatively impacted recognition in all time delays. It appears that, as the time delay increases, information for key and tempo decrease, and only relational information for them is preserved. While tempo information fades more quickly over time, it is still a strong cue during initial judgements of perceptual grouping.

One final example of a low-level factor is good continuation, which is the principle that things with a smooth contour, where the contour does not suddenly change directions (e.g., see Prinzmetal & Banks, 1977), are more likely to be seen as continuous items compared items with erratic contours, where there are random changes in direction. A sequence with a smooth contour, like a group of descending tones, has good continuation and is more likely to be grouped together. To investigate this, Balch (1981) exposed participants to two sequences and asked to rate how well the second sequence followed the first sequence (Balch, 1981). Participants gave higher ratings when it was an inversion of the first sequence or was the first melody played backwards. When the second sequence had good continuation in relation to the first, participants gave higher ratings. Good continuation is also related to pitch proximity and statistical learning in a sequence. If, for example, a listener hears a sequence that descends in pitch, and the tones that have a small step sizes, then the sequence should have good continuation. Also, since this sequence is made of tones that all go in the same direction and have small step sizes, it should also have strong pitch proximity and each tone in the sequence will have high statistical probabilities.

While low-level factors are processes that build up perception from initial input, high-level factors break the perception of something down into subcomponents in order to see how they react with each other to create the perceptual outcome. When comparing high and low-level factors in music, low level factors will focus on basic principles like speed. Knowledge of how different components of music interact with each other is not needed in order to know what the tempo is. High-level factors, which focus on how components of music relate to each other, that influence melodic structure include rhythm, which is the pattern of pulses caused by harmonic beats. The duration of tones in relation to each other and the meter, or pattern of beats in musical measures, are both different elements of rhythm. Tones are systematically arranged based upon these temporal relationships. When you have tones that are of similar duration to each other and fall on the pattern of beats, as determined by the meter, then you have a strong metric rhythm. If the duration of tones are more variable and tones do note fall on this pattern of beats, then the sequence has a weak metric rhythm (i.e., a weak internal clock). For example, if you have a string of shorter notes, like eighth notes, then they are more likely to be grouped together because they are all the same duration. If there were a group of 7 notes, 6 of which were eighth notes, and the tone in the middle of the sequence was a half note, listeners would detect the change in rhythm at the half note and be more likely to perceptually begin a new group at the start of the next tone.

Different aspects of rhythm in melodic sequences were studied in infants to determine how sensitive they were to these features (Bergeson & Trehub, 2006). It was found that 9-month old infants could better detect changes when the context of rhythms had strong metrics compared to those that had weak metric rhythms. Weak metric rhythms were ones where the distribution of accented tones did not match the beats of the meter, making it hard for listeners to internally represent those beats (internal clock). Infants could also better detect pitch changes in the sequences in a duple meter compared to a triple meter. Strong versus weak metric rhythms indicate that infants are more sensitive to accented tones in a sequence. The difference in duple versus triple meters could indicate that infants perceive these sequences in a binary fashion and need more musical experience in order to successfully represent triple meters. This indicates that rhythm plays a large roll in detecting acoustic changes within a sequence.

Memory is another example of a high-level factor that can influence grouping decisions in a sequence. For example, many people are brought up listening to music that is almost always in a Western diatonic frame, where musical elements relate to harmonic modes and transpositions. This could influence a listener's expectations when listening to a musical piece that is written in a Japanese traditional tonal frame, which focuses on the pitch interval that the whole melody is played with. Consistent with this possibility, it has been found that a Western musical expert used harmony-oriented strategies to process melodies while a Japanese musical expert used contour-oriented strategies (Abe & Hoshino, 1990). Additionally, most likely due to the fact that Western musical styles are also common in Japan, the Japanese expert was capable of switching between harmony and contour based strategies when completing the melodies. In-so far as this task reveals their expectations about what is required to complete the presented melodic phrase, this could also be taken as an indication of how expectations would likely impact perceptual grouping decisions.

Memory for melodic structure can also be influenced by changes in the surface features of the melody, like timbre and tempo (e.g, see Halpern & Muellenseifen, 2008). Furthermore, some surface features, including timbre, are retained in memory for longer periods of time compared to others, like tempo (e.g., see Schellenberg & Habashi, 2015). This could indicate that timbre has more salience over some other features. Also, 6-month old infants form memories of timbre, as well as tempo, for songs that were played to them every day for a week (Trainor, Wu,

& Tsang, 2004). This indicates that people learn and remember surface features in complex acoustical sounds very early in life.

Timbre research

How might timbre additionally influence the melodic structure we perceive? There are many studies of lower-level perceptual processes that have demonstrated an influence of timbre on pitch perception. One example comes from the Garner speeded classification task (Pitt, 1994). On every trial, participants were exposed to two tones. Two levels of pitch mixed with two levels of timbre were used to give four unique conditions where stimuli could vary. After the two tones were played participants could respond with no change, pitch change, instrument change, or both changes. Three conditions were measured in each dimension: baseline, where participants classified stimuli on one dimension while the other was held constant, selective attention, where one dimension varied orthogonally with the other, and correlated, where one dimension varied predictably with the other. Non-musicians had high baseline accuracy and short reaction times, but when a dimension, especially timbre, varied unpredictably their accuracy dropped and reaction times increased significantly. Musicians also showed high baselines and short reaction times. Additionally, while accuracy never dropped, reaction times increased when one dimension was allowed to vary orthogonally. Neither group could completely ignore the irrelevant dimension, but non-musicians' performance was effected more than musicians' performance.

Another example of the influence of timbre on pitch perception is the tritone paradox (e.g, see Deutsch, 1986a; also see Deutsch, 1986b; also see Deutsch, 1991). This paradox occurs when you play a pair of Shepard tones (Shepard, 1964), tones made out of multiple harmonics that are one octave apart and have a normal distribution of amplitude, sequentially with an interval size of a tritone, or half an octave. Some listeners will perceive this as an ascending interval and some people will perceive it as a descending interval. This paradox follows the chromatic circle where at some point in the chromatic circle the pattern shifts so that roughly half of the intervals on the circle are perceived as ascending and half as descending. Deutsch argues that this pattern of responses is due to a form of absolute pitch. In contrast, others argue that listeners are able to detect shifts in the spectral envelope when comparing the tone pairs and that musical training could reflect individual differences seen in listeners' perceptions. Spectral envelope shape refers to how intensity changes as a function of frequency within a tone (e.g., see Warren, Jennings, & Griffiths, 2005). This later argument, that changes in spectral envelope shape can explain the pattern of response, has been supported by previous research (e.g., Repp 1997; also see Becker & Hall, 2014).

Timbre's influence on pitch can also be seen in the phenomenon of the missing fundamental, in which a tone is presented with the fundamental (F_0), the lowest frequency of a waveform, removed and yet some listeners will perceive the same pitch (e.g., see Goldstein, 1973; also see Licklider, 1951; also see Seither-Preisler, et al., 2007). In order to investigate the missing fundamental phenomenon, participants took the Auditory Ambiguity Test (AAT) and had to determine whether each tone pair was a rising or falling pitch sequence (Seither-Preisler, et al., 2007). The AAT consists of 100 ambiguous tones which have a rise in its spectrum that is due to a missing F_0 that is either falling or rising in pitch. Based on performance on the AAT, musicians classified pitch changes according to the F_0, but non-musicians tended to classify based on the harmonics and amateurs responded between the two. While there is some debate on why this happens, one main argument is that musicians are trained to determine F_0 from the other harmonics in the tone since they are all integer multiples of the F_0.

One final examination of timbre's influence on pitch is a study which looked at interval illusions, where two intervals are presented and under certain conditions the smaller of the two intervals is perceived as being larger (Russo & Thompson, 2005). Participants were presented with two intervals, tritone and perfect fifth (P5), and had to rate interval size. Four timbre manipulations were used in the stimuli: dull-static and bright-static, where both tones had a dull timbre or bright timbre shift, a congruent shift, where the pitch and timbre changes moved in the same direction, and an incongruent timbre shift, where pitch and timbre changes moved in different directions. When the tritone had a congruent shift and P5 had an incongruent shift, non-musicians experienced an interval illusion on 60% of trials, whereas musically trained participants experienced the illusion in almost 40% of trials when the interval was descending but not ascending. Although some aspects of this effect can be lessened with musical training, these results indicate that timbre shifts can influence an individual's judgements of pitch.

While all of these studies demonstrate the influence of timbre on pitch, they did not examine timbre's influence on perceptual grouping in complex melodies. For example, it has been shown that grouping cues influence stream segregation and statistical probabilities in melodic sequences (Creel, Newport, & Aslin, 2004*)*. Participants listened to interleaved tone sequences where every odd tone in the sequence together constituted one group, and every even tone constituted a second group. These sequences were presented either with no grouping cue, with a pitch proximity manipulation where the two groups were separated by two octaves, or with a timbral manipulation where the two groups had different timbres where the spectrum of odd tones contained Partials 1 and 5 and even numbered tones contained Partials 2, 3, and 4. It was found that when there was no grouping cue, people used statistical learning in sequences for temporally adjacent tones. When either timbre or pitch proximity was used as a grouping cue, tones were grouped according to that cue and then statistical probabilities were learned. Expectations for a given tone is still dependent on the one immediately before it. When a grouping cue like timbre is added, stream segregation occurs and the two interwoven sequences are perceptually separated. However, within each group, statistical probabilities are learned for temporally adjacent tones.

Although Creel, Newport, and Aslin's study (2004) examines timbre manipulation within a sequence, it does not investigate how perceptual grouping of a tone happens in the context of a sequence. Instead, it looks at timbre as a general cue for stream segregation. Due to this, participants are still making judgements about a tone only in the context of the tone immediately before it. Timbre manipulations here only occur in sequences that are interwoven and timbre is shifting consistently within a given sequence. However, if a sequence were presented with no shifts in timbre and then one target tone was presented that had a change in timbre, the effect may not be the same. In order to fully examine the influence of timbre on melodic grouping, other conditions need to be tested to evaluate how strong timbre's influence is.

Models used in melodic grouping

In order to understand the possible role that timbre may play in perceptual grouping of more complex melodies, a review of existing models is necessary. Several models have been proposed in an attempt to explain melodic grouping. These models include Gestalt principles of perceptual organization (see Balch, 1981), statistical learning theory (see Safran, Newport, & Aslin, 1996), and Narmour's I-R model (1992). Gestalt principles are a set of rules, like pitch proximity and good continuance mentioned above, that attempt to explain how people perceive patterns in stimuli. In previous research, Gestalt principles have been shown to influence melodic expectations (e.g., see Anta, 2013; also see Cenkerova' & Parncutt, 2015). For example, the effects of temporal parameters and frequency on participants' pitch judgements were examined when they were presented two tone pairs where the F_0 differed in one pair and participants had to identify which interval had the different F_0 (Bortchert, Micheyl & Oxenham, 2011). Tones were presented in pairs where one was filtered into a low-spectral region and the other to a high-spectral region. However, tones could either have overlapped temporally, or be completely sequential. Participants were better able to detect F_0 changes when two tones were presented simultaneously compared to being presented sequentially and conditions that increased perceptual segregation impaired a listener's ability to detect differences in F_0. It may be the case that frequency proximity is being used to compare the two complexes. Since the harmonic structure of both complexes will be different, listeners may be comparing where harmonics fall in the spectrum as opposed to normal pitch proximity.

The Gestalt principle of similarity can be seen in a study that examined timbre and loudness as melodic segregation cues in musicians versus non-musicians in order to see which cues required the least amount of perceptual change needed for streaming to occur *(*Marozeau, Innes-Brown & Blamey, 2012*)*. Temporal envelope, spectral envelope, and intensity were all manipulated to see how much change was needed for each acoustic measure for participants to improve the perception of a melody that was interwoven with distractor tones. Musicians required less of a difference compared to non-musicians in loudness, impulsiveness, or temporal envelope, and in the spectral envelope for perceptual streaming. Additionally, musicians required the least amount of change for loudness in order for streaming to occur, especially when the target tone and distractor tones had little pitch difference. For those with less musical training, they required less of a change for loudness and the spectral envelope for streaming to occur compared to changes in the temporal envelope, but did not perform as well as the musicians overall.

Statistical learning theory is more closely related to pattern recognition and learning. Statistical learning can be applied to any kind of information that uses patterns in order to give probabilities of what the next item in a series is (e.g., see Mittag, Takegata, & Winkler, 2016; also see Safran, Newport, & Aslin, 1996). Some fields of study where this can be applied are music, speech, visual presentation tasks, and any number of serial tasks. In music the probability of a target tone is dependent upon the pattern of the tones that were presented before it (e.g., see Mittag, Takegata, & Winkler, 2016). If all tones in a sequence are descending, then there is a much higher probability that the target will continue in the same direction as the sequence. Creel, Newport, and Aslin (2004) investigated how statistical learning operated in interleaved sequences. A grouping cue, either pitch proximity or timbre difference, was present within these sequences. For example, in the pitch condition, one sequence was two octaves above the other sequence. In the timbre condition, complex tones containing five partials were created. Odd tones in the sequence contained partials 1 and 5, while even tones in the sequence contained partials 2, 3, and 4. When there was no grouping cue, statistical probabilities would be created for temporally adjacent tones. However, if there was a grouping cue, timbre and pitch, stream segregation would occur and create two perceptual groups for the listener. This allows for probabilities to be learned within each group. In speech, gap detection between words also uses statistical learning (Safran, Newport, & Aslin, 1996). Every language has common patterns of syllables that are used to create words. For example, in the sentence “Please pass the bottle of soy sauce”, there are many different syllables being uttered. Syllables that have a high probability of occurring together, will be grouped together as a single word. When syllables have a low probability of occurring together, then the start of a new word is perceived.

Much like statistical learning theory, the implication and realization model (I-R, Narmour, 1992) focuses on the expectation for a note given the relationship of the notes that came before it. Bottom-up processes within the I-R model can be broadly broken down into continuations and reversals. As the names might imply, a continuation is when a note proceeds in the same pitch-direction as the previous notes, and a reversal is when registral direction is changed.

Another feature of I-R is intervallic difference, which states that if you have a smaller interval, then it will be expected that the next interval will be of similar size. If, however, you have a large interval, then a smaller interval will be expected. It is important to note that along with registral direction and intervallic difference, aspects of duration, metric, and pitch-specific aspects must also be considered for full melodic implications (see Narmour, 1990). Several other researchers have attempted to find support for these findings and have proposed alternative versions of the model due to inconsistent results (e.g., see Krumhansl, 1995*;* also see Schellenberg, 1996). While the I-R model uses registral direction and intervallic difference as predictors of the model, Krumhansl (1995) proposed registral return, proximity, and closure be added to the model. Using the I-R model to try and predict participants' ratings of how well target tones continued a melodic fragment, it was concluded that the model was over-specified and a simpler model (proximity, registral direction, and registral return) could predict ratings just as well (Schellenberg, 1996).

Expectations of whether a given tone will be a continuation or reversal will be determined by the interval size of the two notes before it. The intervallic scale, as adapted from

Narmour (1990), classifies expectations of continuations and reversals based on what is found in Western music. In the I-R model, a continuation is expected when the interval of the previous tone is small, ranging from a unison to a major third, and a reversal is to be expected when the interval between the two previous tones is large, from a minor sixth (m6) and larger. The I-R model contains a range of pitch-distances, from a Perfect fourth (P4) up to a P5, which is considered a threshold where melodic expectations could be either a continuation or reversal depending on other factors.

It is important to note that components of the I-R model are highly related to pitch proximity, good continuation, and statistical learning. For example, intervallic difference aligns with pitch proximity, statistical learning and good continuation. In intervallic difference when you have a small interval, it will be expected that the next tone presented will be of a similar step size. Good continuation would also expect for the next tone to be of a similar step size, since it would keep step size similar across the sequence. Additionally, having the next tone of the sequence be of similar step size will be a high probability in statistical learning because predictions will be based on the two tones before it that have this small step size. Finally, pitch proximity is stronger the closer two pitches are to each other. Thus, when you have small intervallic differences, you are saying that you have good continuation, a high statistical probability and strong pitch proximity. The opposite will also be true with reversals where there will be expectations for a direction change since the interval before the predicted tone will have weak pitch proximity, will break good continuation and will have a low statistical probability. The components of the I-R model have highly correlated relationships with many of the factors that have been discussed above. In fact, when you talk about concepts like continuations and reversals, you are also indirectly talking about factors like pitch proximity, good continuation, and statistical learning.

Current Investigation

The current investigation evaluated the potential effects of timbre on perceptual grouping in melodic sequences. Previous models of melodic expectation have yet to investigate timbre as a predictor of melodic expectations (e.g., see Narmour, 1992; also see Krumhansl, 1995*;* also see Schellenberg, 1996). Evidence has been found that timbre shifts influence peoples' pitch perception on a note-to-note basis (Russo & Thompson, 2005; Creel, Newport, & Aslin, 2004). When participants were presented tritones with a congruent timbre shift, where timbre changed in the same direction as pitch, and P5 intervals with an incongruent shift, where timbre shifts in the opposite direction as pitch, the tritone is perceived as larger 60% of the time (Russo & Thompson, 2005).

While this effect has been seen in intervals, few studies have looked to see the influence of timbre on perceptual grouping in a melody. Creel, Newport, & Aslin (2004), conducted a study to determine how grouping cues like timbre and pitch proximity effect statistical learning in the presence of stream segregation. When no cue was present all statistical probabilities were learned for each tone based on tones in the sequence that were temporally adjacent. But when timbre or pitch was introduced, tones in the sequence were grouped by this cue and then probabilities were learned within each group. These results reflect the idea that how a sequence is processed changes when different features, like timbre, are introduced as a cue. Also, in the paradigm used by Creel, Newport, and Aslin, timbre was used as a cue within the sequence, so that timbre changed with every tone presented. It may be that if you have a context sequence with no change in timbre followed by a target tone with some timbre manipulation, different results in perceptual grouping could be seen. This could indicate that when examining more complex acoustical structures, like melody, timbre may need to be considered as a factor in order to properly represent how information is being processed.

While it is evident that higher-level processing does influence perceptual grouping decisions (e.g., see Schellenberg & Habashi, 2015; see also Abe & Hoshino, 1990), this does not mean that any influence of low-level factors will be overridden by the high-level factors (e.g., see Balch, 1981; also see Anta, 2013). It may be that lower-level factors, like timbre, are not completely subsumed under high level factors. On the contrary, they may have a unique contribution to bring to perceptual grouping of melodies.

The purpose of this study was to examine the influence of timbre on perceptual grouping in the context of a sequence. In all further instances this sequence will be referred to as “context tones”. In Experiment 1 participants were exposed to context tones consisting of sawtooth-like waves and were asked to make a grouping decision of a target stimulus that was be presented at the end of each sequence. Timbre shifts were imposed on the target tone in order to examine its influence on grouping ratings. Similar to the manipulation used by Russo and Thompson (2005), the target stimulus in the current investigation had static, congruent, and incongruent timbre shifts.

The purpose of Experiment 2 was to attempt to extend the findings of Experiment 1 using natural instruments as sound sources. It was possible that any effect of timbre manipulation in Experiment 1 may not be the same when using instrumental tones that have complex spectral envelopes. In order to examine this, Experiment 2 used tones from an instrument to replace the sawtooth-like waves from Experiment 1. Additionally, low and high-pass filters were used to shift the spectral centroid in target tones to create the desired shifts in timbre.

It was expected, for Experiments 1 and 2, that both congruent and incongruent timbre shifts would influence ratings of perceptual grouping across all participants. When the timbre shift was incongruent it was expected that participants would be more likely to group the target tone with the context tones presented before it, whereas a congruent shift in timbre was expected to be more likely to be perceived as starting a new group. However, based on previous results (Russo & Thompson, 2005), it also was expected that musical training would be negatively correlated with the size of the effect of timbre on grouping performance.

Experiment 1

The purpose of Experiment 1 was to examine the influence of timbre shifts on ratings of perceptual grouping of a target stimulus following a melodic sequence. In order to examine this relationship, tones that were like sawtooth waves, tones that had their peak amplitude at F_0 and dropped at 6-dB per octave, were used as they are complex but have a simple spectral envelope shape. Using sawtooth-like waves allowed for controlled manipulations and for filters to have a systematic effect. The target tone had either a 1st or 5th order low-pass filter imposed upon it in order to shift the spectral centroid of the tone to be a higher or lower in frequency. Spectral centroid is the center of mass for a spectrum, or rather, an amplitude-weighted average frequency for a spectrum, and has been found to correlate with similarity ratings in MDS studies (e.g., see McAdams, 1999). Spectral centroid is perceptually associated with sounding “bright” or “dull” depending on the presence or absence, respectively, of high-frequency energy.

Method

Participants. Data from 35 participants was collected through the JMU Department of Psychology participant pool and in partial fulfillment of the participant’s course requirements. All participants were at least 18 years of age and reported normal hearing. All participants understood both written and spoken English as all instructions were in English. Due to technological issues with Direct RT scripts and stimuli, partial data sets were collected for 15 participants, resulting in their exclusion from analyses. Another two participants had their data removed due to headphones not working and a self-reported hearing deficit. As a result, data for a total of 18 participants was analyzed.

There were no requirements for participants to have any musical training. Having variability in responses allows for musical training to be analyzed as a continuous manner. Dichotomizing a continuous variable, like musical training, can lead to inaccurate results in analyses (Daly & Hall, 2017). For example, such a dichotomy can, in many instances, lead to inaccurate data parsing/analyses due to similar levels of training being shared between several members of musician and non-musician groups, and in others result in the frequent elimination of potential data from listeners who reflect moderate levels of training.

Materials. The MuTE (Musical Training and Experience) survey was administered in order to determine levels of musical training for each participant and many different aspects of that musical training (Daly & Hall, 2017*)*. This survey of musical training consists of items that assess multiple facets of musical training, including whether or not the participant/respondent has ever taken formal music classes, has performance experience, how much time they spend in practice, how much composition experience they have, and what musical style they most frequently play. Unlike more traditional measures of musical training, the MuTE is designed to measure these facets of musical training in a more continuous manner. The survey used open response items where participants provide values on a continuous scale; as seen in the example below:

How many months/years have you studied/played each instrument (or voice)? Please indicate both duration and the corresponding instrument (ex: piano - 4 months, voice - 3 years, etc.) (8)

If you selected “Yes” above, about how many hours do you spend playing music per week (rehearsal & individual practice)?

(1) _____ hours per week. (13)

These example items above produce a continuous range of possible responses instead of using multiple choice items to group participants together based on some range or pre-determined value.

Stimuli. Context tones were constructed according to rules that are consistent with expectancies from Narmour’s I-R model, Gestalt principles, and statistical learning theory. Specifically, each tone within a given sequence of context tones had strong pitch proximity and had a high probability of occurring. Sequences were constructed in this manner to ensure that the context tones provide strong evidence of being perceived as a single group leading up to the target tone.

The context tones in each sequence consisted of six tones in the A-minor scale and always started at a frequency of 220 Hz, A_3 in musical notation. Context tones were randomly generated, but followed a set of rules so that they all had the same characteristics. Step size from one tone to the next was either one to two scalar-steps along the A-minor scale. This was so that the tones within the sequence maintained strong pitch proximity but moved from the starting pitch position. These qualities made the sequence more likely to be perceptually considered a single stream and also conveyed directional information. Tones in a sequence could have changed direction up to twice within a sequence, and whenever there was a change in direction the next step had to be a single scalar-step change in frequency. This scalar-step limit was so that changes in direction could occur, but that the context tones were still be perceived as a single group. Context tones were not able to change direction until the third tone since direction was not known until at least the second tone of any given sequence. Additionally, if a direction change occurred there could not be a second direction change for the next tone in the sequence. Half of all context sequences were ascending, where the second context tone was a higher pitch than the first context tone, and half were descending, where the second context tone was a lower pitch than the first context tone. All tones, both context and target, were of the same duration, 500 ms, with no silence in-between tones. The purpose of having context tones randomly generated, but following a set of rules, was to keep the context tones novel. Having sequences constantly change helped prevent participants from becoming familiarized with the context tones and made it more likely that they attended to all tones instead of the last couple.

An additional set of 30 sequences, containing only context tones, were randomly generated as trials for an exposure period. All context tones in this exposure block had no target tone attached to them and all tones had a duration of 500 ms with no silence in between. For each possible number of changes in direction (0, 1, or 2) there were five sequences presented for each direction of the contour. Finally, any sequence generated for the exposure period did not match any sequence played in an experimental trial.

The target tone deviated from the last context by either a minor-third (m3), a perfect fourth (P4), or a minor sixth (m6). In the I-R model it is stated that the expectation of a continuance or a reversal depends upon interval size. However, some intervals sizes are ambiguous and could lead to either a continuation or a reversal depending on context (Narmour, 1992). The m3, P4, and m6 intervals were all selected so that the distance of each one to the last context tone would imply a continuation, ambiguity, and a reversal respectively. Also, if there was only one pitch-interval for target tones then every target would have the same mathematical relationship to the tone before it. By varying the interval size and the direction of pitch change from the last context tone to the tone, the listener should not be able to predict the target tone before it is presented.

One third of all target tones were in a baseline static condition where there was no manipulation of timbre. For the rest of target tones, low-pass filters were used to shift the spectral centroid to a lower or higher frequency respectively. One third of remaining target tones had an incongruent timbre shift, where the shift in timbre went in the opposite direction of the pitch-interval between the last context tone and the target tone. Finally, the remaining target tones had a congruent timbre shift, where the timbre shift went in the same direction as the pitch-interval. A series of low-pass filters were used to create each timbre condition. The target remained at -18 dB/octave when there was no shift in timbre, -6 dB/octave for bright timbres, and -30 dB/octave for dull timbres. This kind of timbre manipulation allowed for a systematic manipulation of timbre in all target tones and the cut-off frequency was always at F_0. A total of 108 musical sequences were made with corresponding target tones for experimental trails and the additional 30 sequences were made for the exposure period. A third of all target tones were at each of the target pitch-intervals m3, P4, and m6. All possible timbre shifts had the same probability of being presented across pitch-intervals. Example context tones at each level of directional change in the context tones are shown in Figure 1. The example target tones at each potential interval for both ascending and descending target are also portrayed in Figure 1.

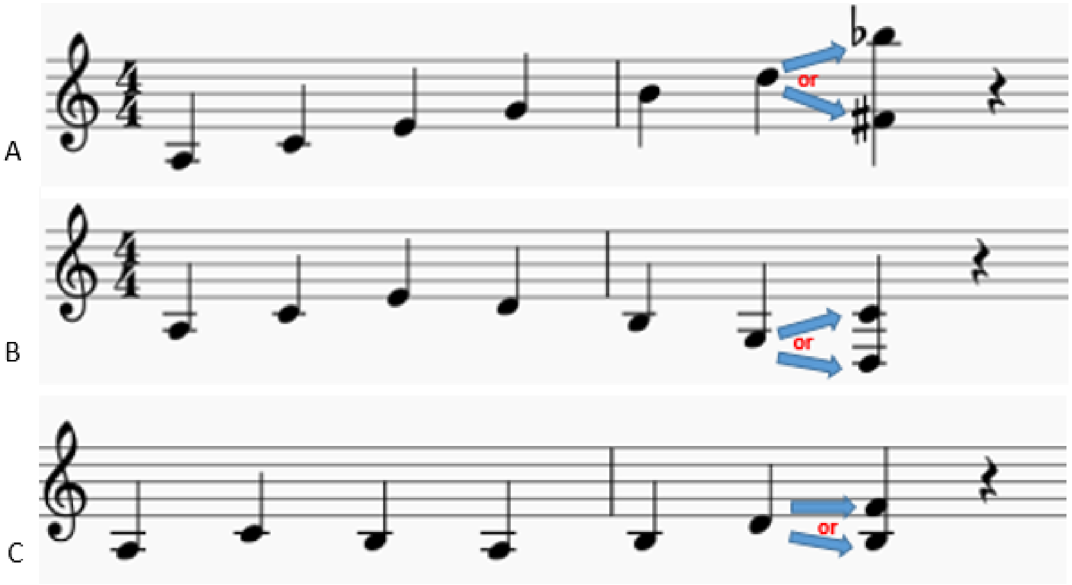



Figure 1: A. Context tones with no direction change paired with potential m6 target pitch-interval. B. Context tones with one direction and paired with potential P4 target pitch-interval. C. Context tones with two direction changes paired with m3 target pitch-interval.

Ableton Live 10, a digital audio workstation (DAW) software package, was used to synthesize and manipulate all tones. Specifically, FormAnt, a formant-based synthesizer, was used to create all tones within the laboratory. Adobe Audition 3 was used to equate tones for root-mean-square (RMS) amplitude and impose filters in order to create the three timbre conditions. Tones were rendered with a 44.1 kHz sampling rate and had a 16-bit bit-depth resolution. A sound-level meter was used in conjunction with manual step attenuators to make sure that peak intensity never exceeded 80 dB[A]. All stimuli were presented to participants through a pair of Sennheiser HD25-SP II headphones and inside a single-walled sound-attenuating chamber.

Procedure. Participants were first provided informed consent document, then they entered the sound-attenuating chamber in order to start experimental trials. The presentation of stimuli and recording of participant responses were controlled on a PC through DirectRT

(Empirisoft Corporation, 2018), a software program that organizes the presentation of stimuli and saves responses out to an Excel spreadsheet**.** For each individual trial, participants were presented a melodic sequence paired with a target tone. The task of participants was to rate if the target tone started a new group/melodic phrase. Ratings were given on a 6-point Likert scale from *1*, indicating that "the tone begins a new group/melodic phrase", to *6*, indicating that "the tone fits well with the phrase before it".

Before participants started the experimental task they listened to the block of 30 exposure sequences in order to be familiarize with all possible kinds of sequences that were going to be played in the experimental blocks. All sequences were played in a random order, and no target tones will be presented with them. Once the exposure block was complete, participants started the experimental task.

Participants completed two blocks of trials for a total of 216 trials per participant. For any given block of trials, participants heard 108 sequences in a randomized order. Of those 108 sequences, each timbre condition, static, congruent, and, incongruent, were presented 36 times; 18 followed ascending context tones and 18 followed descending context tones. Within ascending and descending sequences, each of the 6 target locations, m3, P4, and m6 (in both directions), were presented 3 times. Each unique combination of context tones, pitch-interval, and timbre condition was presented together 3 times per block. Participants were offered a short break in-between the two blocks of trials, in order to minimize fatigue. The second block of trials contained the same sequences as the first block and were presented in a randomized order. However, different target locations and timbre conditions were applied to each group of context tones. Reaction times were recorded for each participant in order to later see if reaction time could predict grouping decisions. It is possible that timbre conditions might have had an effect on perceptual grouping decisions, but the effect would be seen in how long it took participants to respond rather than the value they assigned to a target. Once all ratings were completed, participants were given the MuTE survey to complete in order to assess their musical training. The total duration of the experimental task and the MuTE survey was around one hour.

It was hypothesized that when a target stimulus had an incongruent timbre shift, ratings of the target's perceptual grouping should increase for P4 and m6 pitch-intervals of the target tone, where the higher the rating, the more likely the tone was to be grouped with the previous melody. It has been shown when a timbre shift opposes a pitch change it can cause listeners to perceive a smaller interval between two tones (Russo & Thompson, 2005). This means their sense of pitch proximity should be stronger in this situation for both the P4 and m6 pitch-intervals, thus increasing their ratings of perceptual grouping. However, for m3 pitch-intervals, there was already strong pitch proximity between the target and the context tones, thus grouping ratings should not be affected. Additionally, ratings in the congruent timbre shift condition should decrease ratings of perceptual grouping for P4 and m3 pitch-intervals. Since pitch and timbre changes move in the same direction this promotes a sense of a larger interval distance, weakening pitch proximity. For the m6 pitch-interval, there was already weak proximity between the target and the context tones, thus perceptual grouping should not be affected. If timbre does not affect perceptual grouping, then ratings for the congruent and incongruent conditions should be identical to the static timbre (baseline) condition.

Additionally, it was expected that reaction times for response to P4 pitch-intervals should be slower than that of m3 and m6 pitch-intervals. Expectations for P4 pitch-intervals are ambiguous and depend on context for a grouping decision to be made (Narmour, 1992). For m3 pitch-intervals, it was expected that reaction times would be slower in the congruent timbre condition compared to the incongruent timbre condition. In the congruent condition, m3 pitch-intervals are perceptually being pushed towards ambiguous interval sizes and listeners should take longer to respond. In contrast, m6 pitch-intervals should have a longer response time in the incongruent timbre condition compared to the congruent timbre condition. Here, the m6 pitch-interval is also being perceptually pushed towards ambiguous interval sizes and listeners should be slower to respond to the stimulus. Finally, response times should be faster for the static timbre condition compared to the congruent and incongruent conditions. In the static condition there are not as many acoustical differences between the context tones and the target tone to process, speeding up response times.

Finally, it was expected that musical training would predict task performance. Research has shown that non-musicians tend to mislabel timbre changes as pitch changes, where musicians do not show as many issues in discriminating between changes in these two acoustic dimensions, but do show slower response times (Pitt, 1994). As musical training increases it was expected that there would be a decreasing effect in congruent and incongruent timbre conditions on ratings of perceptual grouping.

Results and Discussion

Average scores were calculated for which combination of condition levels for each participant. For these predictors, response times that represented true outliers were removed from the data. Any trial where the response time was under 150 ms was removed, as stimulus processing is still taking place (i.e., Comerchero & Polich, 1999; Mulert et al., 2008). Additionally, since the response box device did not always register a participant’s first attempt at a response when they thought it had any response time that was over 10 seconds was removed, as it most likely resulted from this technical issue.

For both response times and ratings of the target as a continuation of the melody (perceptual grouping), the individual means indicated above were submitted to a 3-way congruency x target x direction repeated measures analyses of variance (ANOVA). Greenhouse-Geiser adjustments were used to correct degrees of freedom in any instances where sphericity was violated. For any obtained effect that involved a variable with three or more levels, post-hoc pairwise comparisons of means also were conducted using the Sidak adjustment. Finally, to clarify significant three-way interactions, a pair of congruency by target location repeated measures ANOVAs for each level of direction were conducted, and then post-hoc pairwise comparisons were used.

Reaction Times. Response times were not dependent on either congruency or target location. Mean ratings for each level of congruency as a function of target location are depicted in Figure 2. As can be seen in figure, there were no major differences seen in response times across levels of congruency once data was collapsed across target locations, and the same holds true for means of target locations collapsed across congruency. These conclusions were supported by non-significant main effects of congruency and target location, *p's* ≥ .29.

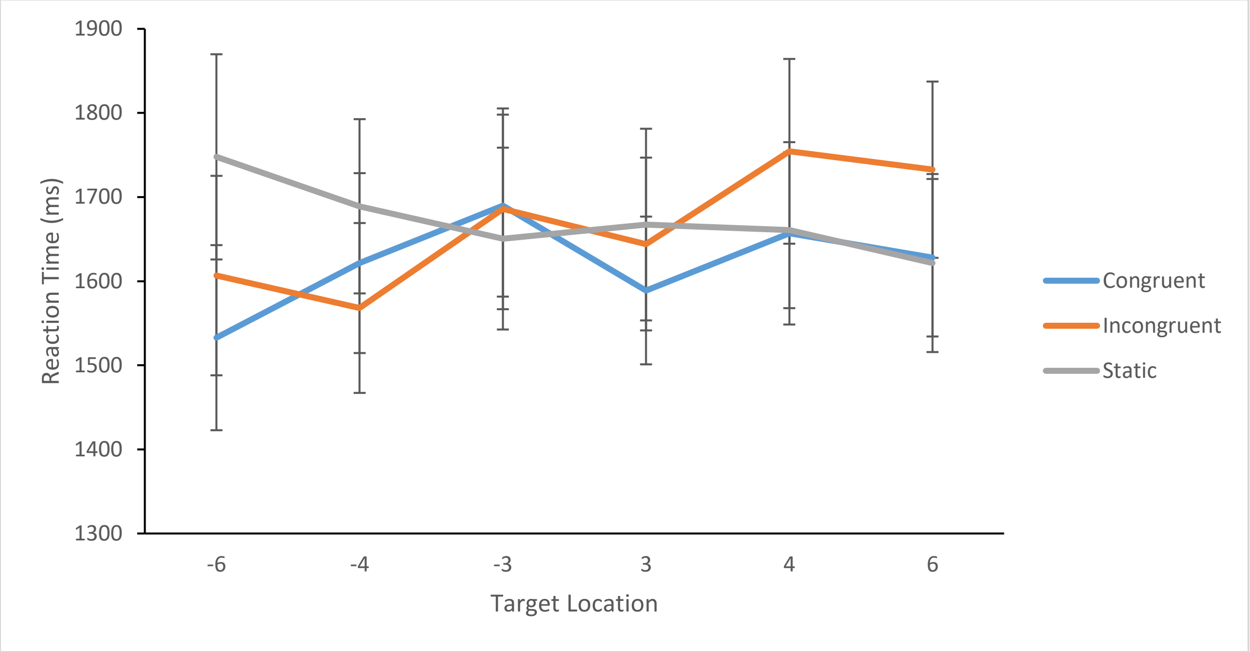


Figure 2: Mean response times each level of congruency as a function of target location. Error bars represent standard error of the mean.

Response times for each target location did depend upon the initial direction of the melody. These mean response times for each direction as a function of target location are summarized in Figure 3. As can be seen in the figure, average response times for ascending melodies decreased when moving from a negative target location (i.e., below the last tone of the preceding melody) to a positive target location (i.e., above the last tone of the preceding melody), whereas the opposing pattern of response times was obtained for descending melodies. This pattern contributed to a significant direction by target location interaction, $F(5, 85) = 4.18$, $p = .002$, $\eta^2 = .015$. Subsequent pairwise comparisons further revealed that when the target location was at “-6” average response times for descending trials were significantly faster than response times for ascending trials, $p = .024$, but were significantly slower when the target location was at “3” and “4” compared to ascending trials, p*’s* $\leq .037$. Contrary to initial hypotheses, target location “-6”, which was considered an unlikely target due to weak pitch proximity, was associated with faster response times instead of slower response times. The larger the interval size, the more unexpected the target tone should be, leading to faster response times as expectations are not being met. For example, target locations “3” and “4” had stronger pitch proximity compared to location “6”, thus more time was needed in order to make a decision of perceptual grouping.

Finally, as can be seen in Figure 3, there was no overall difference in response times between descending and ascending melodies once the data was collapsed across target location. This lack of a difference was confirmed by a non-significant main effect of direction, *F* < 1. Ignoring all other changes in stimuli, initial direction of the melody does not influence response times.

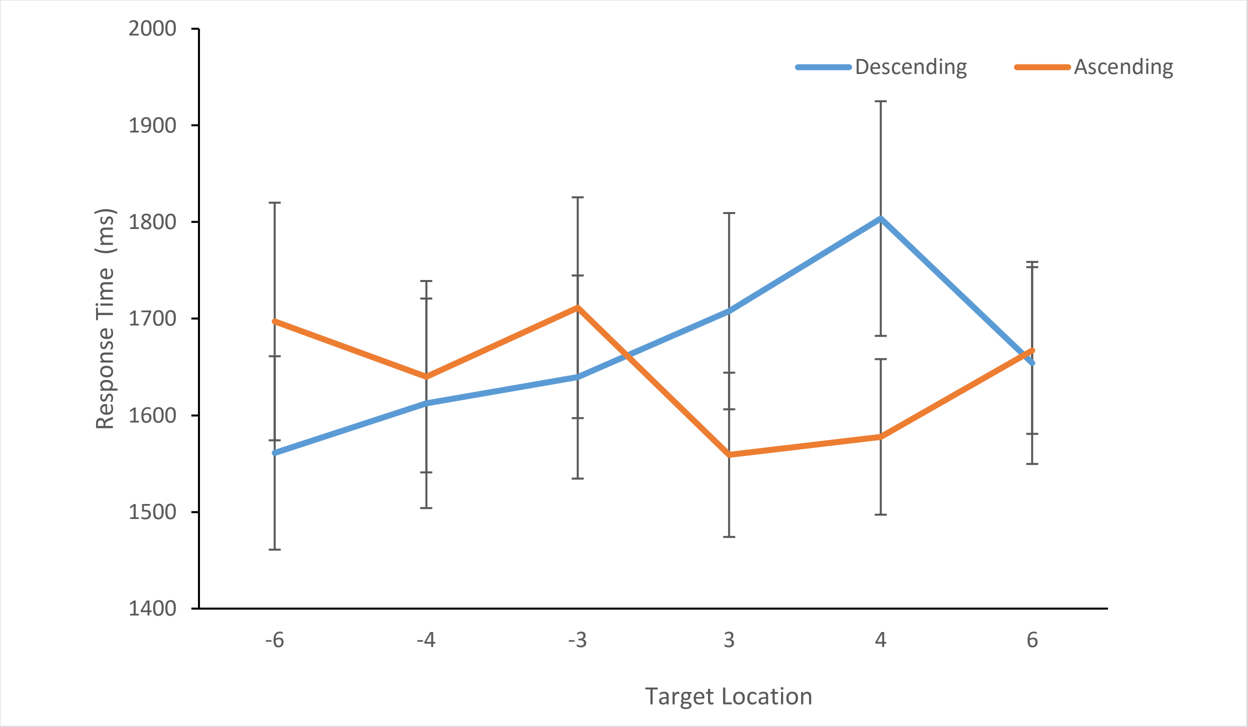


Figure 3: Mean response times for each direction of melody as a function of target location. Error bars represent the standard error of the mean.

Observed effects of congruency on response times depended on direction. Mean response times and corresponding standard errors for each level of congruency as a function of direction is displayed in Figure 4. It can be seen in the figure that response times in the incongruent and static conditions were faster when the beginning of a melody was ascending than when it was descending, whereas response times from the congruent condition were faster when the melody was initially descending. This relationship contributed to a significant congruency by direction interaction, $F(1.50, 24.43) = 5.91$, $p = .006$, $\eta^2 = .013$. Subsequent pairwise comparisons further revealed that response times for the congruent condition were significantly faster for initially descending melodies than initially ascending melodies, $p = .039$, and that response times for the static condition were slower for initially descending compared to initially ascending melodies, $p = .012$. No other effects reached levels of statistical significance.

Contrary to hypotheses, it may be that response times were faster for descending melodies in the congruent condition because the spread of harmonics were closer together compared to that of ascending melodies. As the fundamental frequency of a tone increases, the harmonics that make up that tone spread farther and farther apart because they are integer multiples of each other. In descending melodies, the natural change in harmonic structure was accompanied by a timbre shift that exaggerated this harmonic change, thus making the information more salient in descending melodies. In the static condition, there was no difference in timbre between the target tone and the initial melody, but in ascending melodies there is less timbre information, due to the spread of harmonics (Hall & Rohaly, 2018) to make decisions on, resulting in a faster responses.

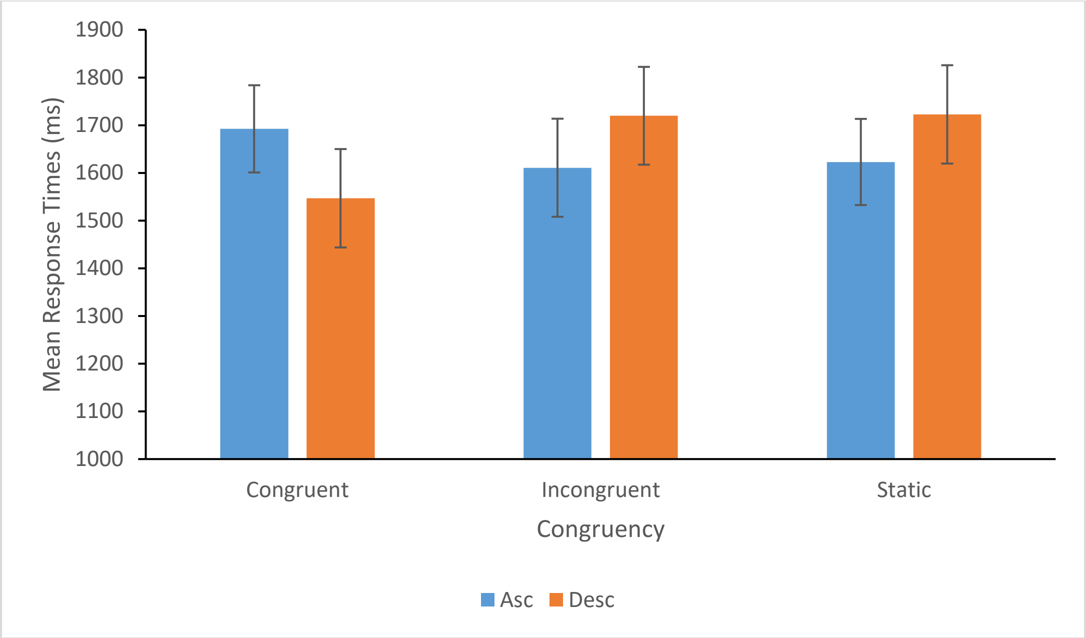


Figure 4: Mean response times for each direction as a function of congruency. Error bars represent the standard error of the mean.

Response times in this study were of secondary concern compared to ratings of perceptual grouping. As a result, the instructions that were read to participants did not include any information about how quickly to respond. If these response times instructions had been included, participants might have responded before the target was presented. Because these instructions for response times were not explicitly stated, it is possible that the other variables of interest did not have a strong effect on response times, as evidenced by the relatively small effect size of the model conducted to examine response times, R^2 = .051. As a result of the small amount of variance in response times explained in Experiment 1, response time data will not be analyzed in Experiment 2.

Rating Data. Mean ratings of perceptual grouping were dependent upon levels of timbre congruency. Average ratings and corresponding standard errors for each level of congruency as a function of target location are summarized in Figure 5. As can be seen in the figure, participants rated target tones within the static condition as better continuations of the melodies than target tones within either congruent or incongruent conditions. These differences contributed to a significant main effect of congruency, $F(1.42, 24.1) = 168.52$, $p < .001$, $\eta^2 = .28$. Pairwise comparisons further confirmed that this effect was due to the fact that mean ratings of the target as a continuation of the melody were significantly higher for the static condition ($M = 4.25$, $SD = .46$) than mean ratings for both the congruent ($M = 2.73$, $SD = .42$) and incongruent ($M = 2.75$, $SD = .40$) conditions, *p's* < .001, which did not significantly differ from each other, $p = .961$.

Contrary to hypotheses that target tones within the incongruent condition would be judged better continuations of the melody than in the remaining timbre conditions, and likewise that targets in the static condition would be judged as better continuations than those in the congruent condition. the obtained pattern of results indicates that mean ratings decreased when there was a timbre change of any kind. Two possible explanations are that participants were either unable to separate pitch and timbre information in order to make a rating decision, or they may not have completely understood task instructions.

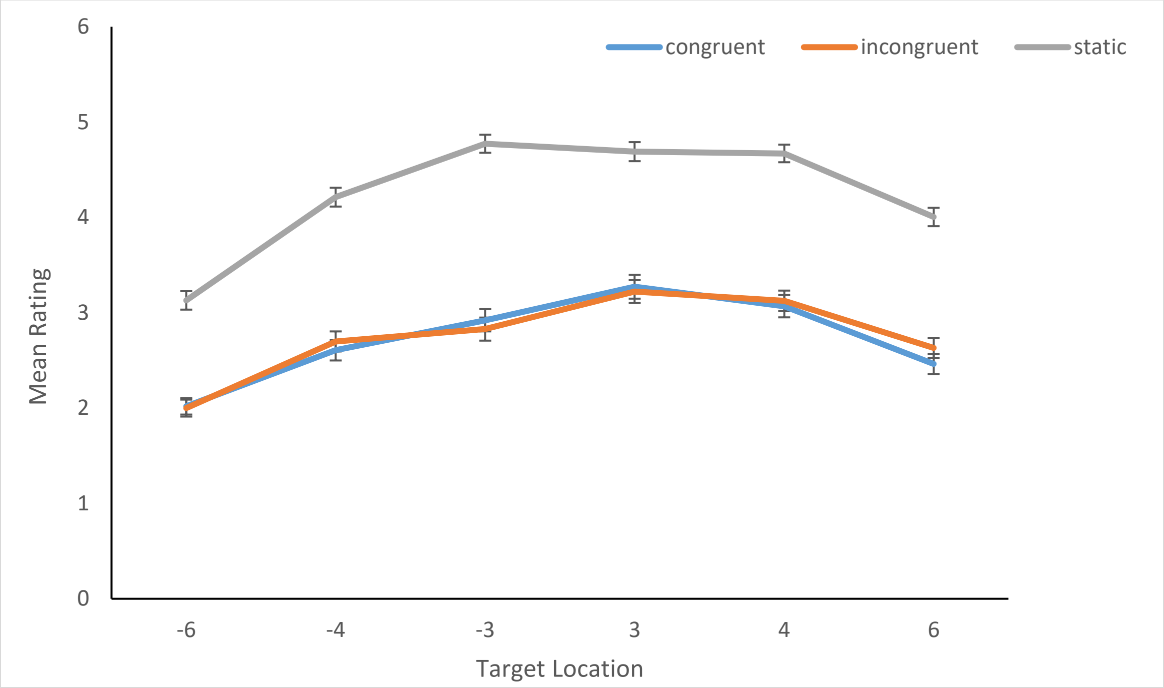


Figure 5. Mean ratings for each level of congruency as a function of target location. Error bars represent the standard error of the mean.

It was hypothesized that mean ratings for each level of congruency would depend upon target location such that observed differences between timbre conditions would be restricted to target locations that reflected smaller interval sizes. As can be seen in Figure 5, mean ratings across timbre conditions did indeed change as a function of target location. This pattern contributed to a significant congruency by target location interaction, $F(5.36, 91.18) = 5.06$, $p < .001$, $\eta^2 = .007$. Subsequent pairwise comparisons further revealed that ratings of the target as a continuation of the melody were significantly higher for static ratings compared to congruent and incongruent ratings at every target location, *p's* < .001. Additionally, ratings of perceptual grouping at each target location did not differ significantly from each other when comparing congruent and incongruent ratings, *p's* ≥ .079. Contrary to the hypothesized pattern of results, timbre manipulations decreased ratings of perceptual grouping regardless of the target's location. It may be that in these stimuli, timbre information was prioritized over pitch information. While ratings did show a quadratic shape that would be expected if pitch proximity were influencing perceptions, participants may have relied on timbre more.

The effect of congruency on mean ratings also depended upon the initial direction of the melody (See Figure 6). Mean ratings for each level of congruency as a function of target location are displayed in panels A and B of Figure 6 for initially descending and ascending melodies, respectively. This pattern of responses contributed to a significant congruency by direction interaction, $F(1.19, 20.90) = 224.96$, $p < .001$, $\eta^2 = .332$. Subsequent pairwise comparisons further revealed that ratings of the target as a continuation of the melody for the congruent ($M = 3.74$, $SD = .48$) and static ($M = 4.36$, $SD = .50$) conditions were significantly higher for initially ascending melodies compared to congruent ($M = 1.71$, $SD = .50$) and static ($M = 4.14$, $SD = .47$) conditions for initially descending melodies, $p\text{'s} \leq .006$. In contrast, mean ratings of perceptual grouping for the incongruent condition ($M = 1.86$, $SD = .48$) were significantly lower for initially ascending melodies compared to the incongruent condition ($M = 3.64$, $SD = .51$) for initially descending melodies, $p < .001$.

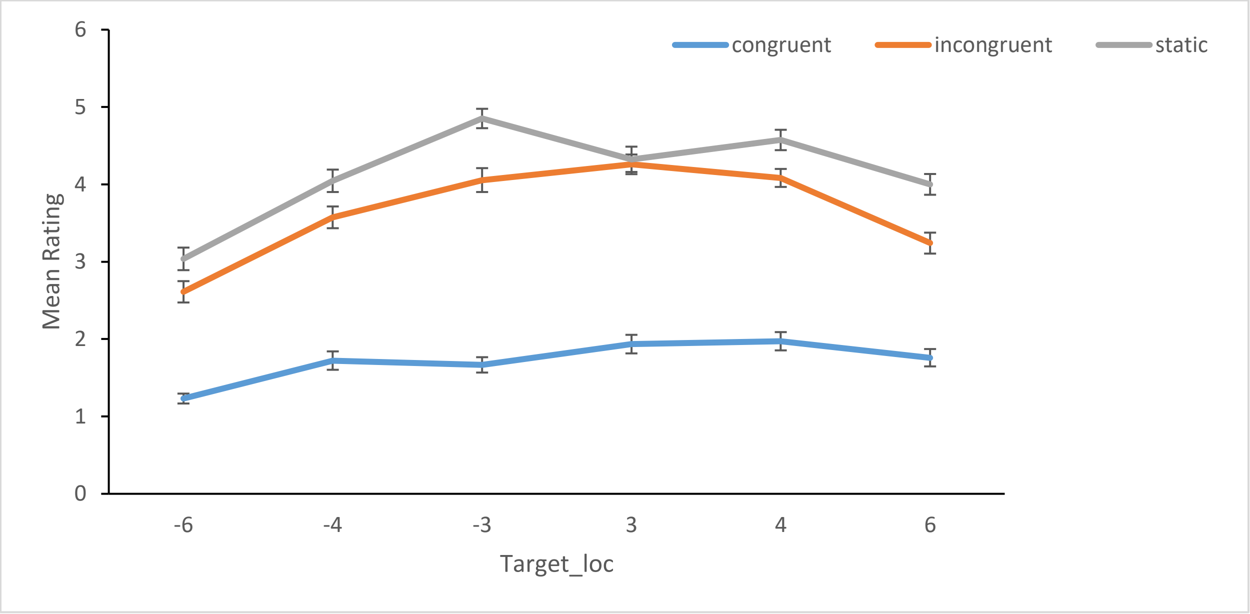
One possible explanation for this interaction is that timbre might influence peoples' expectations differently depending on the direction of the melody. Based on the pattern of observed results, listeners may prioritize grouping by interval size more in descending melodies compared to ascending melodies, as can be seen by participants rating incongruent target tones higher than congruent target tones in initially descending melodies. Other grouping cues, like harmonicity, may be prioritized over interval size for ascending melodies, since ratings for the congruent condition were higher than those of the incongruent condition in initially ascending melodies. However, making such a determination is beyond the scope of this and this possible explanation warrants further investigation.

The relationship between congruency and direction also depended upon target location (see Figure 6). This trend contributed to a significant congruency by target location by direction interaction, $F(10, 170) = 10.80$, $p < .001$, $\eta^2 = .019$. When levels of direction were analyzed separately there was still a significant congruency by target location interaction for both descending, $F(10, 170) = 7.60$, $p < .001$, and ascending melodies, $F(10, 170) = 8.10$, $p < .001$. Subsequent pairwise comparisons further revealed that ratings of the target as a continuation of the melody for the incongruent condition of were significantly higher than for the congruent condition at all levels of the target location when the melody was initially descending $p's < .001$. Additionally, ratings of the target for the congruent condition were significantly higher than ratings for the incongruent condition at every level of the target location when the initial melody was ascending, $p's < .001$. Finally, ratings of perceptual grouping for the static condition were significantly higher than for the congruent condition across all levels of the target location for initially descending melodies $p's \leq .025$, except for the "3" location, $p = .78$, and ratings of perceptual grouping for the static condition were significantly higher than ratings for the incongruent condition at every level of the target location for initially ascending melodies, $p's \leq .029$.

Listeners use multiple strategies in order to group pitches in a melody (i.e. pitch proximity, step inertia, and harmonicity). It is possible that while listeners in this study still used these different strategies to group target tones to the melody, timbre could have affected these strategies in different ways for ascending and descending melodies. Previous research has indicated that listeners create expectations based on the style of music they are familiar with and that the strategies they use to complete melodies are dependent upon these familiar styles (Abe & Hoshino, 1990). Because of this, familiarity with common musical styles in Western culture could influence how these different grouping strategies are used by listeners.

These strategies could also be influenced by the timbre manipulations used. For example, if pitch proximity was prioritized in descending melodies, as can be seen in Figure 6 (panel A), the incongruent shift should be rated higher because the interval size would likely be perceived as small than it actually was, strengthening pitch proximity. In contrast, if harmonicity were a more salient cue in ascending melodies, it could be that interval sizes the congruent shift were likely perceived as being larger in such a way that its harmonic relationship to the melody was one that was more commonly seen in familiar music. However, making such a determination is beyond the scope of this and this possible explanation warrants further investigation.

A



B

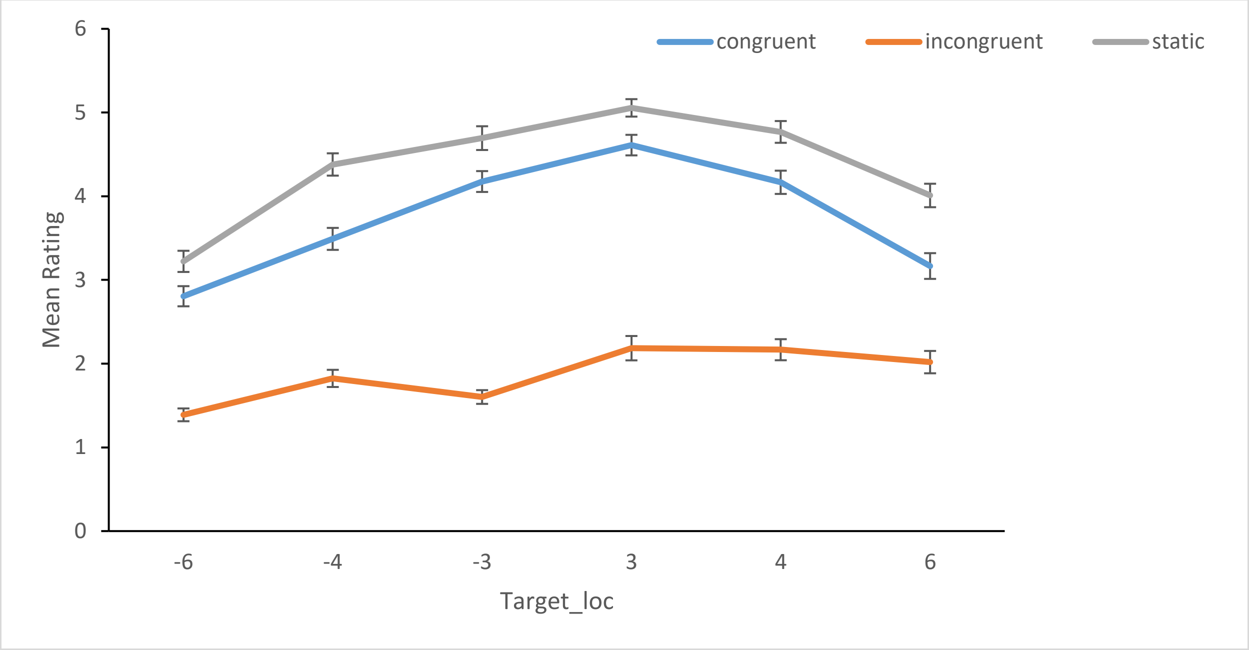


Figure 6: Mean ratings for each level of congruency as a function of target location for initially descending and ascending melodies (in panels A and B, respectively). Error bars represent the standard error of the mean.

Mean ratings also changed as a function of target location. Consistent with hypotheses, mean ratings of the target tone as a continuation of the melody decreased as the target location moved farther away from the initial melody (see Figure 6). A significant main effect of target location was found, $F(1.93, 32.84) = 47.98$, $p < .001$, $\eta^2 = .114$. Subsequent pairwise comparisons of the effect further revealed that mean ratings of perceptual grouping for target locations “3” ($M = 3.73$, $SD = .48$) and “4” ($M = 3.62$, $SD = .48$) were significantly greater than locations “6” ($M = 3.03$, $SD = .59$), “-4” ($M = 3.17$, $SD = .41$), and “-6” ($M = 2.38$, $SD = .40$), *p’s* $\leq .016$. Ratings of perceptual grouping for target location “-3” ($M = 3.51$, $SD = .37$) were higher than locations “-4” and “-6”, *p’s* $< .001$, and ratings of perceptual grouping for location “-4” were higher than location “-6”, $p < .001$. This pattern of responses can be explained by pitch proximity. As the interval size between the target tone and the last tone of the melody increased, pitch proximity decreased, lowering ratings of perceptual grouping.

The effect of target location on mean ratings of the target as a continuation of the melody was dependent on direction (see Figure 6). This dependency contributed to a significant target location by direction interaction, F(5, 85) = 2.66, $p = .028$, $\eta^2 = .003$. Subsequent pairwise comparisons further revealed that when the target was at location “3” mean ratings of perceptual grouping were significantly higher when the melody was ascending compared to descending, $p = .001$. No other statistically significant differences were found, $p \geq .064$. This obtained difference supports Narmour’s I-R model (1992), where small intervals would generate an expectation for the next tone to go in the same direction and be of the same interval size as the tones that came before it. In ascending melodies it would be expected that the next tone in a sequence be another ascending tone that has a small interval size.

Mean ratings also were generally dependent on the initial direction of the melody. Average ratings of the target as a continuation of the melody for ascending trials ($M = 3.32$, $SD = .38$) were higher than for descending trials ($M = 3.16$, $SD = .37$). This difference contributed to a significant main effect of direction, $F(1, 17) = 10.80$, $p = .004$, $\eta^2 = .003$. This main effect was based on a specific definition of direction which was originally defined as the direction of the first two tones of the melody. As such, melodies were measured to find out the proportion of ascending and descending melodies that maintained that direction across the whole melody. For example, melodies where the first two tones reflected an ascending pitch interval were measured to see what percentage of those melodies had an overall direction that was ascending. A total of 83% of initially ascending melodies had a final tone that was higher than the first tone of the melody, and a total of 78% of initially descending melodies had a final tone that was lower than the first tone of the melody. With that in mind, listeners seemed to have higher expectations for ascending melodies compared to descending melodies, but due to how the direction variable was designed, interpretations of this effect are beyond the scope of this study. Further investigations into the influence of direction on melody perception is suggested.

Conclusions that can be reached from this experiment are necessarily limited. This limitation is due to nature of the instructions read to participants. Participants were not explicitly asked to give ratings based on pitch and as result they could have used any acoustic information in order to make judgements. A possible alternative explanation is that participants, based on these instructions, chose to focus on timbre rather than pitch. Further investigation is needed in order to either support or eliminate this possibility.

Experiment 2

The purpose of Experiment 2 was to extend the results of the Experiment 1 using recordings from naturally produced instrumental tones. While the design of Experiment 1 afforded considerable stimulus-related control, the results from the experiment cannot easily be generalized to many real-world settings insofar as the stimuli, unlike natural events, were synthesized as static timbres that did not vary over time. It is certainly possible that more natural, dynamic changes within tones could influence how melody is perceived, and might even reduce any observable effects of timbre. Thus, an evaluation of melodic grouping with dynamic timbre stimuli was critical to permit any broader conclusions to be reached regarding how those grouping decisions might occur under more typical listening conditions. This was the focus of Experiment 2. If this experiment were to show corresponding effects of timbre influencing perceptual grouping in melodies, as observed in Experiment 1, then it could more readily be argued that potential timbre effects should be considered more seriously when designing melody studies.

Similar to Experiment 1, it was expected that timbre shifts that were incongruent, shifting in the opposite direction as pitch, would be more likely to group with the context tones before it. In contrast, a congruent shift, where timbre shifts in the same direction as pitch, would be less likely to be grouped with the context tones before it. Like Experiment 1, the duration of Experiment 2 was around an hour.

Method

Participants. Twenty-six listeners were recruited through the JMU Department of Psychology participant pool and participated for partial fulfillment of course requirements. All participants were at least 18 years of age, self-reported no hearing problems, and understood both written and spoken English. Due to technological difficulties data could not be analyzed for two participants. There were no requirements for participants to have any musical training. This was to ensure as much variability in musical training in the sample as possible. Having variability in responses allowed for musical training to be analyzed in a continuous manner.

Stimuli. Instead of using sawtooth waves, as in Experiment 1, natural instrument tones were used for stimuli. Instruments have more complex spectral envelope shapes that also change over time. This kind of spectral is far more common in real world sound sources. Using instruments instead of sawtooth allowed for any potential results to more comparable to real world situations.

Because few instruments have a 3-4 octave range that matches each other, Cello samples, from the Ableton Live 10 sample database, were used with both low and high-pass filters to create the timbre conditions. Since every sequence started at 220 Hz, the spectral centroid of that tone was measured and used as the cut-off value for both of the filters. The spectral centroid was measured to be right around 1,100 Hz, so that was the cut off value used. Both the low-pass and high-pass filters were first-order Butterworth filters applied using Adobe Audition 3. Because these filters caused a shift in the spectral centroid of the tone, such changes, due to the nature of the spectral centroid, should be perceived as a change in the size of the instrument body. Additionally, due to using both high- and low-pass filters, the original samples constituted the static timbre condition and the filtered tones constituted bright and dull conditions, respectively.

Another change was that the number of sequences were reduced from Experiment 1 to Experiment 2. Due to the Cello's limited natural range of fundamental frequencies, sequences that had no changes in direction were not able to be played. As a result, all context tones in Experiment 1 with no changes in direction were removed from Experiment 2. A total of 144 trials were created. Of those 144 sequences, each timbre condition, static, congruent, and, incongruent, were presented 48 times; 24 followed ascending context tones and 24 followed descending context tones. Within ascending and descending sequences, each of the 3 target locations, m3, P4, and m6, were presented 4 times.

Procedure. Participants first provided informed consent, and then entered the sound-attenuating chamber in order to start experimental trials. The presentation of stimuli and recording of participant responses were controlled on a PC through DirectRT (Empirisoft Corporation, 2018), a software program that organizes the presentation of stimuli and saves responses out to an Excel spreadsheet**.** For each individual trial, participants were presented a melodic sequence paired with a target tone. The task of participants was to rate if the target tone started a new group/melodic phrase. Ratings were given on a 6-point Likert scale from *1*, indicating that "the target pitch was a bad continuation of the initial melody/began a new group/melodic phrase”, to *6,* indicating that “the target pitch was a very good continuation of the initial melody.”

Participants completed 2 blocks of 72 trials for a total of 144 trials per participant. Presentation of stimuli within each block was randomized. Between the two blocks of trials, participants were offered a short break in order to minimize fatigue. Once all ratings were completed, participants were given the MuTE survey to complete in order to assess their musical training. The total duration of each experiment session was around one hour.

Similar results were anticipated for Experiment 2 as they were in Experiment 1. It was hypothesized that when a target stimulus has an incongruent timbre shift, ratings of the target’s perceptual grouping should increase for P4 and m6 pitch-intervals of the target tone. It has been shown when a timbre shift opposes a pitch change it can cause listeners to perceive a smaller interval between two tones (Russo & Thompson, 2005). This means their sense of pitch proximity would be stronger in this situation for both the P4 and m6 pitch-intervals, thereby increasing their ratings of perceptual grouping. However, for m3 pitch-intervals, there should already be strong pitch proximity between the target and the context tones. Thus, ratings would not be affected. Additionally, congruent timbre shifts were expected to decrease ratings of perceptual grouping for P4 and m6 pitch-intervals. Since pitch and timbre changes move in the same direction this promotes a sense of a larger interval distance, weakening pitch proximity. For the m6 pitch-interval, there was already weak proximity between the target and the context tones. As a result perceptual grouping should not be affected. If, alternatively, timbre did not affect perceptual grouping, then ratings for the congruent and incongruent conditions should be identical to the static timbre (baseline) condition.

Finally, it was expected that musical training would predict task performance. Research has shown that non-musicians tend to mislabel timbre changes as pitch changes, where musicians do not show as many issues in discriminating between changes in these two acoustic dimensions, but do show slower response times (Pitt, 1994). As musical training increases it was expected that there would be a decreasing effect in congruent and incongruent timbre conditions when the direction of the sequence is ascending. It has been found that interval illusions were not present for musicians when the interval was ascending, but were present when the interval was descending (Russo & Thompson, 2005). Since this effect was seen previously under similar timbre conditions, it was expected that the same trend would be seen here.

Results and Discussion

Data were analyzed using the same types of statistical analyses, corrections, and post-hocs as Experiment 1. Specifically, mean ratings of perceptual grouping were submitted to a 3-way congruency x target x direction repeated measures ANOVA. Greenhouse-Geiser adjustments were used to correct degrees of freedom in any instances where sphericity was violated. For any obtained effect that involved a variable with three or more levels, post-hoc pairwise comparisons of means also were conducted using the Sidak adjustment. Finally, to clarify significant three-way interactions, a pair of congruency by target location repeated measures ANOVAs for each level of direction were conducted, and then post-hoc pairwise comparisons were used. Additionally, simple regression analyses were conducted to determine if level of musical training predicted the difference between mean scores obtained from each of the timbre conditions.

Mean ratings of perceptual grouping (along with corresponding standard errors) for each level of timbre congruency as a function of target location are summarized in Figure 7. As can be seen in the figure, ratings of the target as a continuation of the melody were influenced by congruency manipulations. Mean ratings of perceptual grouping for static trials were higher than ratings for either congruent or incongruent trials. This pattern of responses contributed to a significant main effect of congruency, $F(1.1, 24.1) = 121.34$, $p < .001$, $\eta^2 = .414$. Subsequent pairwise comparisons for this effect confirmed that mean ratings of perceptual grouping for the static condition ($M = 4.51$, $SD = .64$) were significantly higher than mean ratings of perceptual grouping for both congruent ($M = 2.61$, $SD = .65$) and incongruent ($M = 2.61$, $SD = .72$) conditions, p's $< .001$, which did not significantly different from each other, $p = 1$. The finding that ratings of perceptual grouping for the incongruent condition were also worse than the static condition was not hypothesized. One possible explanation of this finding is that a majority participants chose to ignore instructions to judge pitch, and instead used all possible acoustic differences to make rating judgements. If this were to happen, participants would rate any acoustical difference as not being a good continuation of the melody, as a target with a different timbre would not match the initial melody. However, it does not seem likely that every participant would choose to ignore directions in the same manner. An alternative, more likely possibility, is that participants were unable to separate pitch and timbre information when making rating decisions. It is possible that participants grouped timbre and pitch information together and made decisions based on both, even if they thought they were only giving responses based on pitch information.

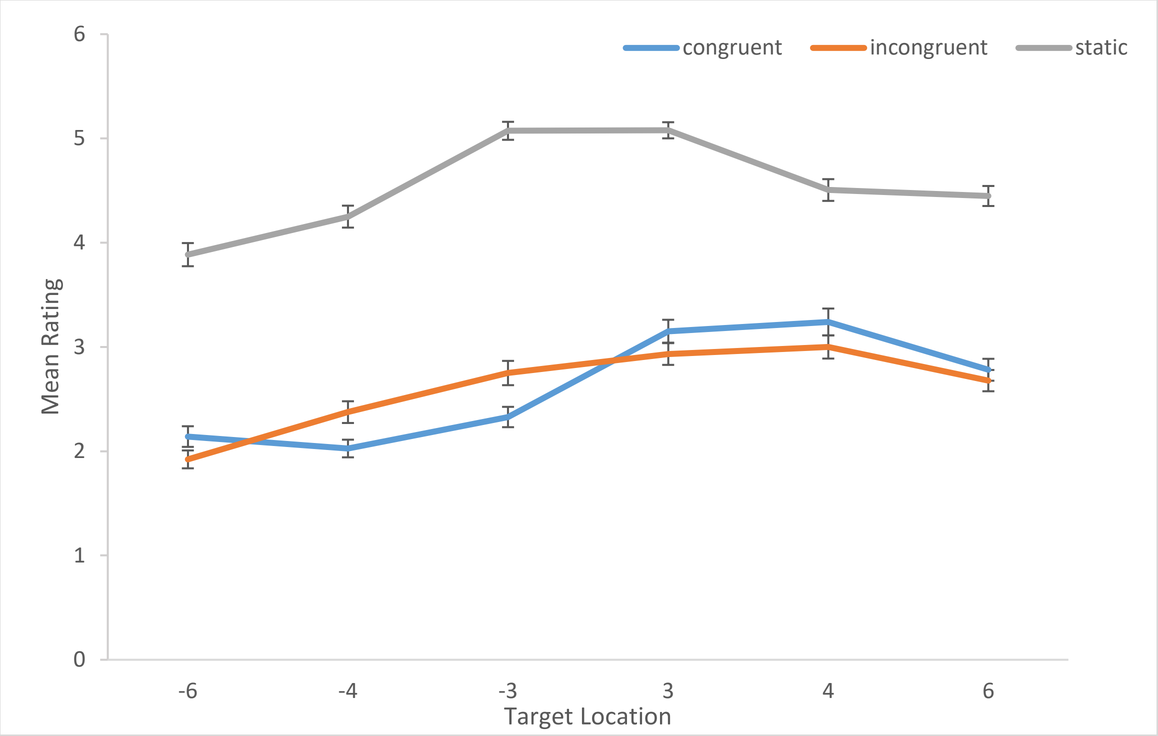


Figure 7. Mean ratings for each congruency condition at every level of target location. Error bars represent the standard error of the mean.

Additionally, the observed effects of congruency depended upon target location. Specifically, incongruent and congruent ratings of the target as a continuation of the melody were higher when the target's pitch location was above the last tone of the preceding melody relative to when it was below that last tone (see Figure 7). This pattern contributed to a significant congruency by target location interaction, $F(10, 230) = 7.59$, $p < .001$, $\eta^2 = .017$. Subsequent pairwise comparisons revealed that the interaction was primarily due to the fact that mean ratings of perceptual grouping were significantly higher for the congruent condition than the incongruent condition at locations "4" and "-6" (p's $\leq .045$), but were significantly lower at locations "-3" and "-4", p's $\leq .003$. Finally, mean ratings of perceptual grouping for the static condition were higher than both incongruent and congruent conditions at every level of target location, p's $< .001$.

One possible explanation for this interaction is listeners had a preference for the bright timbre over the dull timbre. With the exception of target location “-6”, ratings of perceptual grouping were higher for the bright timbre than the dull timbre. The high-pass filter that was used to make the bright timbre was perceptually more similar to the static condition compared to the low-pass filter that was used to make the dull timbre. All timbre manipulations were examined with an F_0 of 220 Hz, and it was very apparent to the researchers that the spectral layout of the bright timbre was much more similar to the static timbre than was the dull timbre.

The effect of congruency on ratings of perceptual grouping did not match hypothesized results. Based on the pattern of obtained results, stream segregation occurred for both the timbre cue and pitch proximity cue across all conditions. Stream segregation for congruency conditions could be due to the fact that very strong timbre manipulations were used and were thus prioritized when using grouping cues.

The observed effects of congruency also were dependent upon direction. Mean ratings for each level of congruency as a function of target location, where panels A and B represent initially descending and initially ascending melodies, respectively, are depicted in Figure 8. When the melody was descending, ratings of the target as a continuation of the melody were highest for the static condition and lowest for the congruent condition, whereas when the melody was ascending, ratings of perceptual grouping were lowest for the incongruent condition. These differences contributed to a significant congruency by direction interaction, $F(1.28, 29.39) = 16.54$, $p < .001$, $\eta^2 = .024$. Subsequent pairwise comparisons revealed that mean ratings of perceptual grouping for the congruent condition were significantly higher for ascending trials ($M = 2.89$, $SD = .77$) relative to descending trials ($M = 2.34$, $SD = .66$), $p < .001$, while mean ratings of perceptual grouping for the incongruent condition were significantly higher for descending trials ($M = 2.86$, $SD = .87$)

relative to ascending trials (*M* = 2.36, *SD* = .70), *p* = .001. No other significant differences were found *p* = .184. A possible explanation for this interaction is that listeners may create different expectations based upon the initial direction of the melody. If a melody is initially ascending, a listener should expect the next note of the melody to also be ascending. However, if the overall direction of the melody stays level, or was descending, expectations would change. For example, this can be seen in ratings for the congruent condition being higher for ascending melodies compared to descending. Since there would already be expectations for the tone to be ascending, target tones that were ascending would not be as influenced by a congruent timbre shift since the expectation was still met.

As can be seen in Figure 8, the congruency by target location interaction also depended on direction. Mean ratings of the target as a continuation of the melody for the congruent condition were higher than for the incongruent condition in initially ascending melodies, but were lower than for the incongruent condition in initially descending melodies. A significant congruency by target location interaction was found in both initially descending trials, *F*(10, 230) = 14.31, *p* < .001 and initially ascending melodies, *F*(5.63, 129.49) = 9.45, *p* < .001. Subsequent pairwise comparisons further revealed that ratings of perceptual grouping for the incongruent condition were significantly higher compared to ratings for the congruent condition at target locations “-3”, “-4”, and “-6” when the melody was initially descending, *p’s* < .016. When the melody was initially ascending, ratings of perceptual grouping for the congruent condition were significantly higher than ratings for the incongruent condition at target locations “-6”, “-4”, “-3”, and “4”, *p’s* < .033. Ratings of perceptual grouping for the static condition were higher than the other conditions for both levels of direction, *p’s* ≤ .001. One potential factor that may explain these results could be how the timbre manipulation was created. The first order Butterworth high-pass filter and low-pass filter removed energy above and below 11,000 Hz, respectively. These two filters change the spectral shape of the tone in different ways, and thus could influence listeners differently.

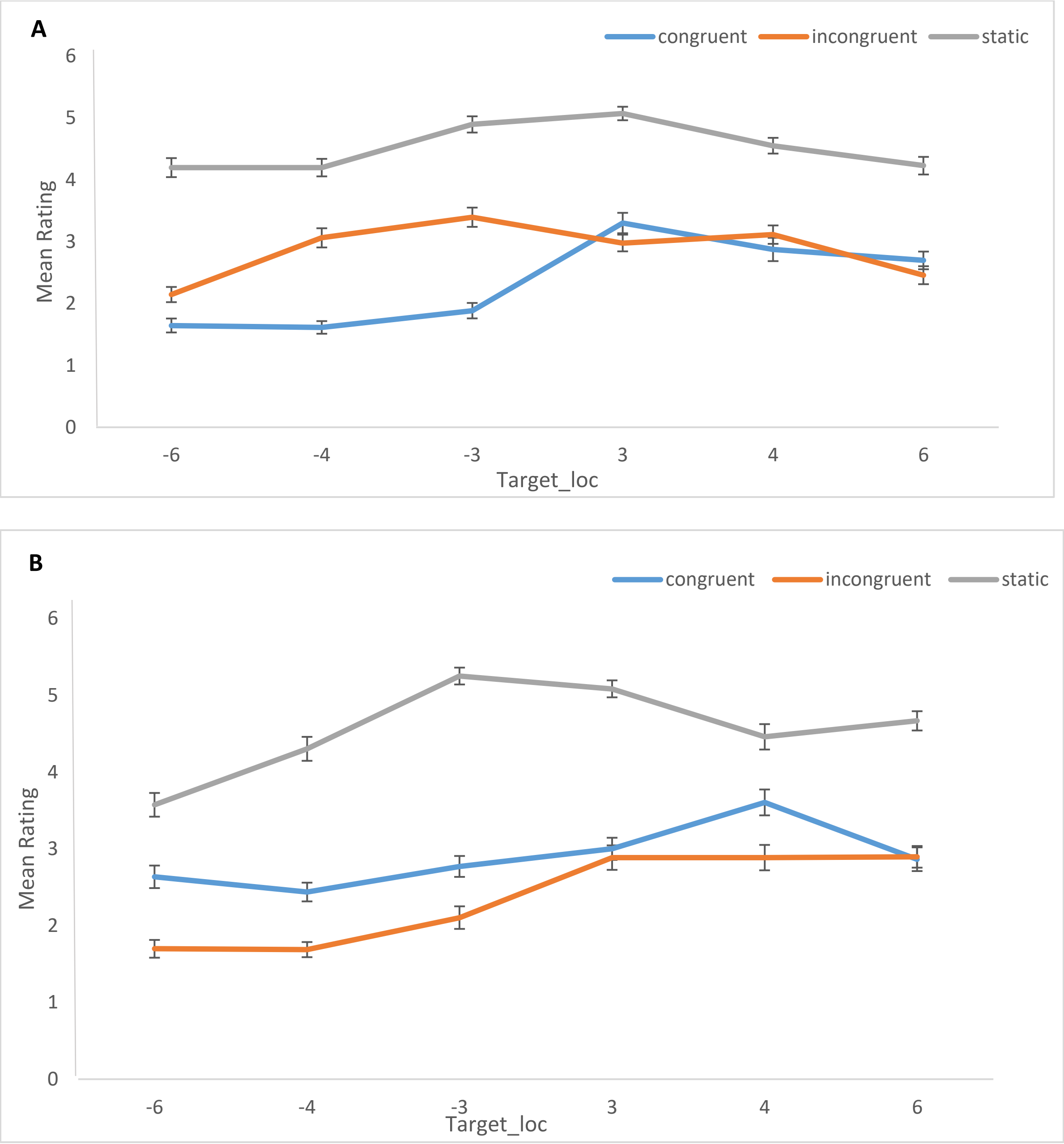


Figure 8: Mean ratings of congruency as a function of target location when the direction of the melody is descending and ascending (panels A and B, respectively). Error bars represent the standard error of the mean.

Consistent with hypotheses, mean ratings of the target as a continuation of the melody also decreased with increases in the size of the target location variable, and thus, with increases in the interval between the last tone of the melody and the target tone (see Figure 8). This pattern of results contributed to a significant main effect of target location, $F(2.71, 62.41) = 28.29$, $p < .001$, $\eta^2 = .072$. Subsequent post-hoc analyses further revealed that that target location “3” ($M = 3.72$, $SD = .60$) was judged a better continuation than every location except “4” ($M = 3.58$, $SD = .60$), $p's \leq .005$. Likewise, location “4” was judged better than every other location except “6” ($M = 3.3$, $SD = .62$), p’s ≤ .028, and location “6” was judged better than “-6” ($M = 2.74$, $SD = .63$), $p = .006$. Finally, location “-3” ($M = 3.31$, $SD = .60$) was judged better than locations “-4” ($M = 2.88$, $SD = .65$) and “-6”, $p's < .001$. As interval size increased ratings of perceptual grouping decreased, indicating that target tones with weak pitch proximity were rated as not belonging to the melody compared to tones with strong pitch proximity.

As can be seen in Figure 8, the influence of target location on mean ratings of the target as a continuation of the melody were influenced by the initial direction of the melody. Both ascending and descending ratings of perceptual grouping followed a quadratic shape across levels of target location. This pattern of responses contributed to a significant direction by target location interaction, $F(5, 115) = 4.00$, $p = .002$, $\eta^2 = .004$. Subsequent post hoc analyses revealed that this interaction stemmed from ratings of perceptual grouping for the target at location “6” being significantly higher when the melody was ascending compared to descending, $p = .01$. No other statistically significant differences were found, $p \geq .07$. One likely reason for this effect was that participants were more likely to group a target tone as being part of the melody if the target went in the same direction as the melody. Thus, if a melody is ascending, there should be an expectation for the next tone to also be ascending. Lastly, no significant main effect of direction was found, $F(1, 23) = 1.66$, $p = .21$, $\eta^2 < .001$.

Some conclusions that can be reached from this experiment are necessarily limited by how variables were defined. One of these is the definition of direction. As in Experiment 1, stimuli were analyzed in order to get a proportion of trials where the initial direction of a melody matched the overall direction. These proportions reveal how well the original definition of direction can be generalized to a definition that describes the overall direction of the melody. When a melody was initially ascending, the direction of the melody as a whole was ascending 75% of the time. When melodies were initially descending, the overall direction stayed descending 67% of time. Because the variable, as it was defined, predicted the direction of the overall melody fairly well, it is likely that the variable of direction is actually reflecting listeners following the shape of the overall melody instead of just the first two tones. It is possible that the influence of direction on perceptual grouping would be stronger if direction indicated the overall direction of the melody 100% of the time.

Effects Related to Musical Training. In order to examine the relationship between musical training and the effect of congruency, difference scores between the congruent and static conditions, as well as between the incongruent and static conditions, were determined. For each participant these difference scores were calculated using their average score for each level of congruency. These difference scores were then used as the dependent variables for the two regression analyses with musical training as a predictor.

The magnitude of observed effects for the timbre manipulation also differed with changes in the duration of musical training, which is defined as the number of years an individual has played an instrument. Consistent with hypotheses, the influence of timbre on ratings of the target as a continuation of the melody decreased as musical training increased. Musical training regressed on both the difference in mean ratings of perceptual grouping between the static and congruent conditions (congruent-static), as well as ratings of perceptual grouping between the static and incongruent conditions (incongruent-static); these relationships are summarized in panels A and B of Figure 9, respectively. As can be seen in the figure, there was a negative relationship between incongruent-static ratings and musical training where small difference scores were associated with high levels of musical training. The nature of this relationship was confirmed by musical training being a significant predictor of both the incongruent-static ratings, $F(1, 23) = 8.05$, $p = .01$, $R^2 = .27$, $b = -.09$, and congruent-static ratings, $F(1, 23) = 7.18$, $p = .01$, $R^2 = .25$, $b = -.08$.

A potential explanation for these results is that as musical training increases, a listener is better able to separate timbre and pitch information from each other. The influence of timbre have less of an effect on musically trained listeners has been seen before in previous literature (Pitt, 1994). This may be due to the fact that those who have played music for a long time have received formal training on not only how to play their respective instrument, but to decompose music they listen to, so that they can tell what sound source is playing a particular part/theme within a piece of music. In contrast, those with little to no training may not have a working concept of what timbre is.

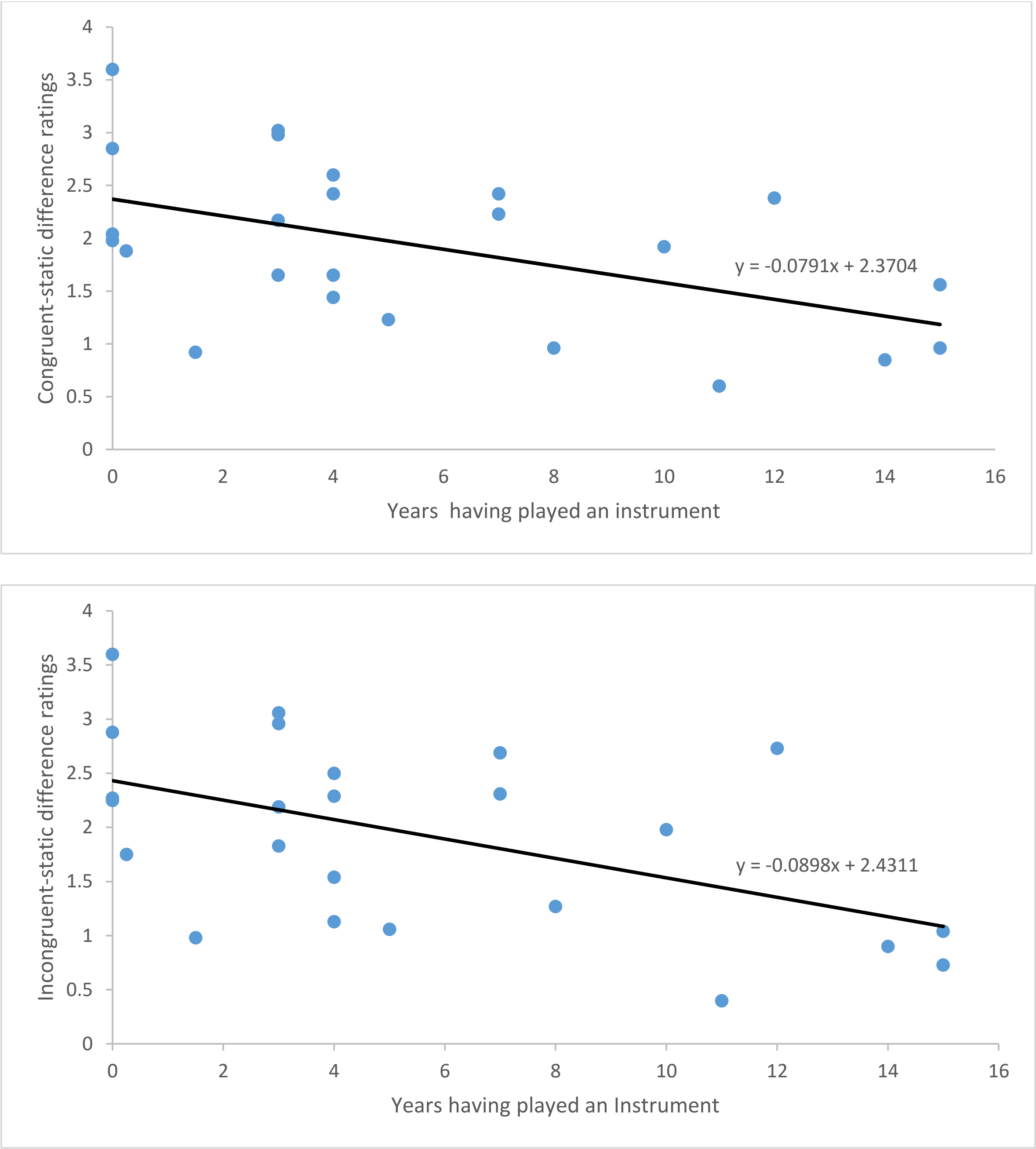


Figure 9: Mean rating differences between the congruent and static ratings and difference between static and incongruent ratings as a function of years having played an instrument for panels A and B respectively.

Conclusions. When considering the entire sample, listeners were unable to separate pitch and timbre information. Results from this study have indicated than any change of timbre in the target tone will result in a lower rating of perceptual grouping. Regardless, participants used stream segregation to group auditory streams based on timbre and pitch proximity in order to make judgements of perceptual grouping. This can be seen by looking at each individual level of congruency across target location where each level of congruency tends to have a quadratic shape, indicating that pitch proximity is being used. Finally, response patterns are correlated with musical training, where higher levels of musical training are related to lower difference scores of timbre conditions.

General Discussion

Observed effects of target location, or interval size, were observed across experiments. These effects were hypothesized to be consistent with Narmour's I-R model (1992), which states that listeners create expectations for a tone based on the interval size and direction of the tones that came before it, and pitch proximity. These expected results supported the I-R model and pitch proximity in both Experiment 1 and Experiment 2. Specifically, as the interval size between the last tone of the melody and the target tone increased, ratings of the target tone decreased. Ratings for the target tone decreased at larger intervals because it did not match the interval sizes of the tones in the initial melody and pitch proximity was weaker, making the target less likely to be grouped with the tones that came before it.

While effects of timbre were obtained across both experiments, they did not match the initial hypotheses. Specifically, when the timbre of the target tone shifted in any way, ratings of perceptual grouping decreased. There are several possible explanations as to why this happened. For example the lack of an explicit mention of pitch in the task instructions in Experiment 1 makes it possible that participants in that experiment were actually grouping by timbre instead of by pitch. There are a few indicators that make this account unlikely. First, similar patterns of responding were obtained regardless of whether or not the listener was aware of additional information about the purpose of the experiment. Three lab members ran through the experiment, as a subset of the data, and their mean ratings for each level of congruency as a function of target location can be seen in Figure 10 (panel A). Despite the fact that these participants had a greater awareness of the purpose of the study, the pattern of obtained results for these listeners were similar to those seen in the entire sample. Ratings for the congruent condition were higher when the target was located below the initial melody while ratings for the incongruent condition were rated higher when the target was located above the initial melody. This pattern is showing that the dull timbre was preferred to the bright timbre. When the target tone is below the melody, the congruent condition is a dull timbre because it moves in the same direction as the target tone. Likewise, the timbre was dull for the incongruent condition when the target was above the initial melody because it moved in the opposite direction as the target.

Additionally, task instructions for Experiment 2 were changed slightly so that pitch was explicitly mentioned as what participants were making judgements about. Despite this change in task instructions, a similar pattern of responses was seen in Experiment 2 for a subset of lab personnel (*n* = 5) who were told to completely ignore timbre when completing the task (See Figure 10 B). As can be seen in Figure 10 (panel B), these participants , despite the fact that they had a greater awareness of the purpose of the study, showed a pattern of obtained results that were similar to those seen in the entire sample. As with Experiment 1, these participants showed the same pattern of results as the overall results in Experiment 2, namely that timbre shifts led to a decrease in ratings of grouping. It therefore appears that listeners were unable rely on pitch information to the exclusion of timbre regardless of the degree to which they may have understood task instructions.

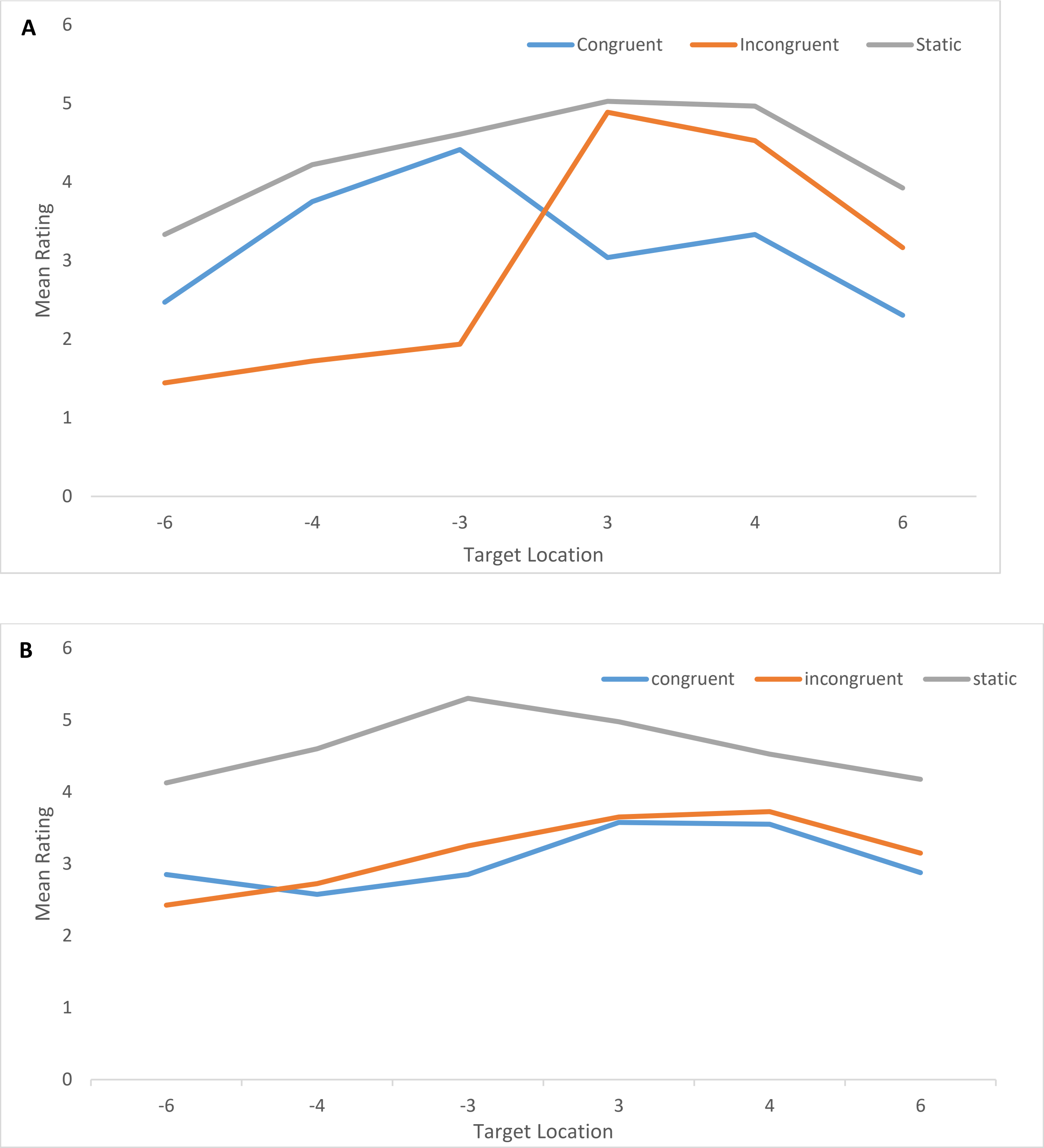


Figure 10: mean ratings for each level of congruency as a function of target location for a subset of participants who knew the general purpose of the study for Experiment 1 (Panel A) and Experiment 2 (Panel B).

A second possible explanation is that participants understood the instructions to use pitch across experiments, but chose to ignore those instructions. This also seems unlikely. Lab personnel showed the same pattern of results even though they had no known motivation to ignore. Additionally, it is unlikely that a vast majority participant would choose to ignore instructions in the same way.

A third possible explanation for the obtained effect of congruency is that participants were unable to focus on pitch information to the exclusion of timbre. This alternative explanation seems reasonable insofar as each experiment used a salient timbre manipulation. For example, slopes of the low-pass filters used differed by ± 12 dB per octave, which past research as indicated could be a discriminable difference (Li & Pastore, 1995). Furthermore, in Experiment 2 centroids were calculated for the 220 Hz tone in each timbre condition and found that centroid shifted by a couple hundred Hz for each timbre shift. Finally, stimuli were listened to by the author and the faculty supervisor and timbre shifts were deemed to be easily recognizable. Such a strong manipulation could have made it a stronger cue for perceptual grouping than pitch proximity. If participants were unable to separate these two dimensions of sound it could explain why the same pattern of results were found in Experiment 2 after instruction changes were made. It would also explain why the pattern of the overall data resembled that of the subset of data from lab personnel. Compared to a majority of participants choosing to ignore instructions, it is more likely that they were unable to separate pitch and timbre information from each other.

Taken together, Experiments 1 and 2, provide strong evidence supporting the idea that timbre information was not completely subsumed/overcome by higher-level processes, like expectations of melody. Regardless of interval size, any shift in timbre resulted in listeners being less likely to rate the target tone as being a good continuation of the melody. In Experiment 1, the main effect of congruency accounted for more variance in ratings than every other effect in the model except for the congruency by direction interaction. In contrast, in Experiment 2, the congruency by direction interaction accounted for very little variance in participant ratings, but the main effect of congruency accounted for the most. Since the presence (or lack of) a timbre shift explained large amounts of variance in participants' response patterns compared to other factors (like interval size), timbre should be considered an independent factor in the perception of melody. However, it is important to note that this experimental task held many high-level processes of melody, like tempo, rhythm, and tone duration, constant across melodies. In order to examine the influence of timbre in completely natural music, these factors would need to be manipulated along with timbre to see if the effects found in this study are consistent across these high-level cues of melody perception.

Cross-Literature Comparisons

The influence of timbre on pitch perception concurs with the findings and conclusions made by Russo and Thompson (2005). Specifically, congruent and incongruent timbre shifts influenced ratings of perceptual grouping, albeit in an unexpected way. For example, in the Russo and Thompson study ratings of interval size increased with a congruent shift and decreased with an incongruent shift. In contrast, any timbre shift at all, in this present study, lead to a lower rating of the target tone being a good continuation of the initial melody.

There are several possibilities as two why the direction of effect changes across studies. The first is that in the Russo and Thompson study there was no expectation of timbre staying the same within each interval. One-third of intervals were static, one-third had a congruent timbre shift, and one-third had an incongruent timbre shift. Since all conditions were equally likely to occur, there should have been no expectations regarding the type of timbre that would follow the initial tone. For this present study, the distribution of timbre conditions in target tones were the same as the Russo and Thompson study. However, because timbre never changed within the initial melody, expectations that timbre of the target tone would stay the same likely increased. It may be that the congruent and incongruent timbre conditions did shift listener's perceptions of pitch in different ways, but because neither timbre shift met the expectations for the target to have the same timbre as the initial melody, ratings of grouping decreased regardless of the kind of shift used.

A second factor between studies to consider is differences in task instructions. In the Russo and Thompson study, participants were asked to rate the interval size of each stimulus. The rating of interval size task could not be used for this current study because it would allow participants to completely ignore the melody up until the last two tones. Instead, participants in Experiment 2 were asked to rate how well the target pitch continued the initial melody.

Results of the current investigation are also consistent with the findings of Creel and colleagues (Creel, Newport, & Aslin, 2004). In their study, they used a change in timbre as a cue for stream segregation in interleaved sequences. Results found that listeners would group based on this cue and then use statistical learning within each group. In this present study, target tones that had timbre shifts were segregated into different streams, but only for the target tone. The stimuli in the Creel study are similar to music that would be composed in the counterpoint style, whereas the stimuli from the current investigation would follow more traditional forms of melody, which is more generalizable to popular music. In the Creel study there was an expectation for every other tone in the sequence to be the same timbre. However, for this current investigation, there was an expectation for every tone in the melody to be the same. Regardless of this difference in expectations, stream segregation still occurred in the same manner.

Due to the results seen in both the current investigation and Creel et al. (2004), it may be possible to create stimuli where participants create expectations for different kinds of timbre shifts. This might change the pattern of responses that were seen in this present study, due to the fact that only sudden changes in timbre were used. If expectations for a certain timbre were generated, ratings of grouping for a target could then be predicted based on the strength of the timbre shift and the kind of timbre shift used in the melody. This suggested methodology would be a good extension of the current investigation because it would allow researchers to examine how the context of timbre in the initial melody influences grouping decisions of the target. In this current investigation only sudden timbre shifts were used within the target tone when listeners were expecting the target to have the same timbre as that of the initial melody. As such, conclusions about how timbre changes in the initial melody effect grouping of a target tone cannot be made. This suggested extension of the current investigation would allow for this effect to be examined.

The influence of timbre on pitch should not be unexpected. There are many examples of this influence in previous research across many different research questions. Some specific examples include the Garner speeded classification task (Pitt, 1994), missing fundamental phenomenon (Seither-Preisler, et al., 2007), and the tritone paradox (Deutsch, 1986a; 1986b; 1991), the influence of pitch on timbre judgements (Hall & Rohaly, 2018), and harmonic tuning (Becker & Hall, 2014). Furthermore, other studies have shown that timbre is one of the more memorable dimensions of music, (Schellenberg & Habashi, 2015). Due to the number of demonstrations of the timbre and pitch interaction at lower perceptual levels, it should not be surprising that timbre influences melody as well. As a results, it is important that timbre be considered as a manipulation in melody research. Many studies of melody do control timbre in order to examine different dimensions of music (e.g., Abe & Hoshino, 1990; Anta, 2013), but do not necessarily conduct follow-up studies to see if results from the original study are influenced by timbre.

This follow-up study would not be important to look at just because timbre may influence perception, but even if you control timbre by using a single sound source, timbre perception has been shown to be influenced changes in pitch (Hall & Rohaly, 2018). As the fundamental frequency of a sound source increases, the spread of harmonics get father apart since they are integer multiples of each other. As the spread of harmonics get father apart, less spectral information about what the sound source is will be conveyed, creating a change in timbre information. These changes are different for each sound source, since the filter for each source creates different patterns of resonances. Thus, one model of how timbre changes across pitch cannot be used since this change in timbre will be dependent upon what the source itself is. If experiments are conducted where an instrument is played across a fairly wide pitch range, timbre will already by changing slightly as a function how this spread of harmonics change. Pitch and timbre information are highly relation to each other and cannot be completely independent.

While the effects in the current experiments were strong, there were necessary limitations. One important consideration was the strength of timbre manipulation. Even though it is clear the timbre is able to influences target tones in a melody, it does not address any question that involves timbre and pitch cues being in direct competition with each other. If this direct competition were the main research question of interest, it might be best to start with cues that were empirically matched for the strength of each cue. This design would help determine relative salience in each cue and their contributions to melody perception. However, because listeners are very sensitive to changes in frequency, timbre manipulations would need to be designed to be able to match sensitivity of these changes. In order to see if this might be possible, the researchers took sample instruments, clarinet and bassoon, and created a highly familiar tune, Mary had a Little Lamb. While the researchers could group all tones based on pitch, it was very difficult to do so and it required forcing oneself to completely disregard all timbre information. This anecdotal example, like the primary findings from the current investigation, suggests that higher-order processes involved in melody perception are influenced by timbre. Given what was found in the Mary had a Little Lamb example, it is expected that timbre cues could be more salient a cue than pitch.

A third limitation that was mentioned was the variable direction. Because it was not designed to reflect the overall direction of the melody it is not quite as meaningful as it could be. This change in direction should be accompanied by a follow-up study to look at the congruency by target location by direction interaction. This would use stimuli where the direction of the melody was designed to describe the entire melody instead of just the first two tones. Doing this would eliminate confounding explanations as to why these results were found.

In conclusion, timbre has a large impact on the grouping of pitches within a melody under the right conditions. Because the bright and dull timbres were different enough from the static timbre, listeners used timbre, in addition to pitch proximity, as a cue for stream segregation when comparing the target tone to the melody . Minimally, these results indicate that effects of timbre should be evaluated when considering how listeners perceive melody, given that timbre represents a salient grouping cue to help listeners distinguish streams of incoming auditory information, including melodies, at least occasionally above and beyond the effects of other available cues.

References

Abe, J. & Hoshino E. (1990). Schema driven properties in melody cognition: experiments on final tone extrapolation by musical experts. *Psychomusicology, 9(2), 161-172.*

Anta, J., F. (2013). Exploring the influence of pitch proximity on listener's melodic expectations. *Psychomusicology: Music, Mind, and Brain, 23(3), 151-167.*

Balch, W.R. (1981). The role of symmetry in the good continuation ratings of two-part tonal melodies. *Perception & Psychophysics 29*(*47*). doi: 10.3758/BF03198839

Becker, C. J., & Hall, M. D. (2014). Effects of fundamental frequency removal and low-pass filtering on pitch comparisons. *Psychomusicology: Music, Mind, and Brain, 24(3), 195-206.*

Bergeson, R., T., & Trehub, E., S. (2006). Infants perception of rhythmic patterns. *Music Perception: An Interdisciplinary Journal 23*(*4*)*, 345-360*. doi: 10.1525/mp.2006.23.4.345

Borchert, O.M.E., Micheyl, C., & Oxenham, J.A. (2011). Perceptual grouping affects pitch judgments across time and frequency. *Journal of Experimental Psychology: Human Perception and Performance, 37(1), 257-269.*

Cenkerova', Z., & Parncutt, R. (2015). style-dependency of melodic expectations: changing the rules in real time. *Music Perception*, *33*(1), 110-128.

Creel, C.S., Newport, L.E., & Aslin. N.R. (2004). distant melodies: statistical leaning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30(5), 1119-1130.*

Comerchero, M.D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110(1), 24-30.*

Daly, H. R., & Hall, M. D. (2018). Not all musicians are created equal: Statistical concerns regarding the categorization of participants. *Psychomusicology: Music, Mind, and Brain, 28(2), 117-126.*http://dx.doi.org/10.1037/pmu0000213

Deutsch, D. (1986). A musical paradox. *Music Perception*, *3*, *275-280*.

Deutsch, D. (1986). An auditory paradox. *Journal of the Acoustical Society of America*, 80(93).

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception*, 8, 335-347.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America, 54*(6), 1496-1516. http://dx.doi.org/10.1121/1.1914448

Hall, M. D., & Rohaly, T. (2018, November). *Predicting pitch's influence on instrument timbres and vowels.* Poster session presented at the Auditory Perception, Cognition, and Action Meeting, New Orleans, LA.

Halpern, A. R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *Quarterly Journal of Experimental Psychology*, *61*(9), 1371–1384. https://doi.org/10.1080/17470210701508038

Hamaoui, K & Deutsch, D. (2007). *The perceptual grouping of musical sequences: Pitch and timing as competing cues*. Presented at the 11th International Conference on Music Perception and Cognition (ICMPC11), Seattle, WA.

Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, 17, 53-90.

Li, X., & Pastore, R. E. (1995). Perceptual constancy of a global spectral property: Spectral slope discrimination. *Journal of the Acoustical Society of America*, 98(4), 1956-1968.

http://dx.doi.org/10.1121/1.413315

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia, 7,* 128-134. http://dx.doi.org/10.1007/BF02156143

Marozeau, J., Innes-Brown, H., & Blamey, P.J., (2012). The effect of timbre and loudness on melody segregation. *Music Perception 30(3), 259-274.*

McAdams, S. (1999). Perspectives on the Contribution of Timbre to Musical Structure. *Computer Music Journal, 23*(3), 85-102.

Mittag, M., Takegata, R. & Winkler I. (2016). Transitional probabilities are prioritized over stimulus/pattern probabilities in auditory deviance detection: memory basis for predictive sound processing. *The Journal of Neuroscience, 36(37), 9571-9579.*

Mulert. C., Seifert. C., Leicht. G., Kirsch. V., Ertl. M., Karch. S., Moosmann. M., Lutz. J., Möller. H., Hegerl. U., Pogarell. O., Jäger. L. (2008). Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *NeuroImage, 42(1), 158-168*.

Narmour, E, (1990). *The analysis and cognition of basic melodic structures: the Implication-realization model*. Chicago: University of Chicago Press.

Narmour, E., (1992). The influence of embodied registral motion on the perception of higher-level melodic implication. In Jones, R. M. & Holleran, S., *Cognitive Basis of Musical Communication* (69-90). Washington D.C: American Psychology Association.

Pitt, A., M. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance 20(5), 976-986.*

Prinzmetal. W., & Banks, W. P. Good continuation affects visual detection. (1977). *Perception & Psychophysics, 21, 389-395.*

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception, 26,* 645–665. doi:10.1068/p260645

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics, 67*(4), 559-568.

Safran, R. J., Newport, L. E., & Aslin, N. R. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language 35, 606-621*.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, *58, 75-125*.

Schellenberg, E.G. & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Mem Cogn* 43: 1021. DOI: 10.3758/s13421-015-0519-1.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(*3*), *743-751*. doi:10.1037/0096-1523.33.3.743

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America, 36*, 2346–2353. doi:10.1121/1.1919362

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science, 7*(*3*), 289-296. DOI: 10.1111/j.1467-7687.2004.00348.x

Warren, J. D., Jennings, A. R., Griffiths, T. D. (2005). Analysis of the spectral envelope shape in the human brain. *NeuroImage, 24(4),* 1052-1057. doi.org/10.1016/j.neuroimage.2004.10.031