CROSS-MODAL MELODIC CONTOUR SIMILARITY

Jon B. Prince[1], Mark A. Schmuckler[1], and William Forde Thompson[2]

1 - Department of Psychology, University of Toronto, Ontario, Canada

2 - Department of Psychology, Macquarie University, Sydney, Australia

ABSTRACT

In two experiments participants rated the similarity of melodic contours presented as auditory (melodies) and visual (line drawings) stimuli. Longer melodies were assessed in Experiment 1 (M = 35 notes); shorter melodies were assessed in Experiment 2 (M = 17 notes). Ratings for matched auditory and visual contours exceeded ratings for mismatched contours, confirming cross-modal sensitivity to contour. The degree of overlap of the surface structure (the relative position of peaks and troughs), and the strength and timing of the cyclical information (the amplitude and phase spectra produced by a Fourier analysis) in the contours predicted cross-modal similarity ratings. Factors such as the order of stimulus presentation (auditory-visual or visual-auditory), melody length (long versus short), and musical experience also affected the perceived similarity of contours. Results validate the applicability of existing contour models to cross-modal contexts and reveal additional factors that contribute to cross-modal contour similarity.

RESUME

Au cours de deux expériences des participants ont estimé la similarité des contours mélodiques présentés comme stimuli auditifs (des mélodies) et visuels (des dessins au trait). Des mélodies longues (M = 35 notes) ont été évaluées dans la première expérience; des mélodies courtes (M = 17 notes) ont été évaluées dans la deuxième expérience. Les estimations de similarité des contours auditifs et visuels équivalents étaient plus élevées que les estimations de similarité des contours auditifs et visuels différents, ce qui confirme la sensibilité des participants aux contours représentés par des modalités sensorielles différentes. Le degré de chevauchement de la structure superficielle (la position relative des crêtes et des cuvettes), et la force et le rythme de l'information cyclique (les spectres d'amplitude et de phase obtenus par analyse de Fourier) dans les contours ont prédit pour les modalités sensorielles différentes des estimations de similarité élevées. Certains facteurs tels que l'ordre de la présentation des stimuli (auditif-visuel ou visuel-auditif), la durée de la mélodie (longue ou courte), et l'expérience musicale ont aussi affecté la similarité perçue des contours. Ces résultats déclarent valide l'applicabilité des modèles de contours existants aux différents contextes de modalités sensorielles et dévoilent des facteurs additionnels qui contribuent à la similarité des contours dans ces modalités.

1 INTRODUCTION

Contour, or the overall pattern of ups and downs, is a basic attribute of auditory and visual stimuli. In the case of audition, pitch contour plays an important role in two forms of auditory information: language and music. In language, contour is a primary attribute of speech intonation and contributes to the supralinguistic dimensions of speech. Speech intonation provides cues about emphasis, emotional attitude, and syntactic structure, and it may also facilitate the processing of verbal content in tonal and non-tonal languages ('t Hart, Collier, & Cohen, 1990; Lieberman, 1967; Pierrehumbert & Hirschberg, 1990; for a review, see Cutler, Dahan, & van Donselaar, 1997). Contour also plays a crucial role in music cognition, providing one of the most important cues for melody recognition and melodic similarity (Dowling, 1978; Dowling & Harwood, 1986; for a more thorough review see Schmuckler, 1999).

1.1 Contour in music cognition

Listeners can recognize familiar melodies even when the intervals of a melody (the specific pitch distance between successive notes) are severely distorted as long as the contour of the melody, or the relative pattern of rises and falls in pitch, remains intact (Deutsch, 1972; Dowling & Hollombe, 1977; Idson & Massaro, 1978; White, 1960). Moreover, contour is critical for discrimination between (Watkins, 1985) and memory of (Dowling, 1978) novel melodies, especially when there is no tonal framework to aid in constructing a representation of the melody in memory (Dowling, 1991; Dowling & Fujitani, 1971; Francès, 1988; Freedman, 1999). Children and infants also preferentially use contour over more specific, local information when listening for changes in melodies (Chang & Trehub, 1977; Morrongiello, Trehub, Thorpe, & Capodilupo, 1985; Pick, Palmer, Hennessy, & Unze, 1988; Trehub, Bull, & Thorpe, 1984).

Research has elucidated how listeners segment melodies into meaningful units, store this information in memory and subsequently use it for recognition. Pitch accents created by contour reversals (i.e., changes in pitch direction) contribute to the perceptual segmentation of both melodies

and speech (Bregman & Campbell, 1971; Deutsch & Feroe, 1981; Frankish, 1995; Thomassen, 1982), and also direct attention to important notes within a melody (Boltz & Jones, 1986; Boltz, Marshburn, Jones, & Johnson, 1985; Jones & Boltz, 1989; Jones & Ralston, 1991; Monahan, Kendall, & Carterette, 1987). Indeed, alterations to a melody are more obvious when they involve a contour reversal (Dyson & Watkins, 1984; Jones & Ralston, 1991; Monahan et al., 1987; Peretz & Babaï, 1992), and recognizing novel melodies is more challenging as the contour becomes more complex (Boltz et al., 1985; Cuddy, Cohen, & Mewhort, 1981; Morrongiello & Roes, 1990; but see Croonen, 1994). According to Narmour's (1990) implication-realization model, contour reversals represent a crucial feature of melodic structure and listeners expect them to occur after large melodic leaps.

Contour also plays a critical role in melodic similarity. Eiting (1984), for instance, found that similarity judgements of short (3-note) melodic sequences depended primarily on contour. Contour also contributes significantly to similarity judgements of 7-note melodies (Quinn, 1999) and 12-note melodies (Schmuckler, 1999). Categorization of 7-note melodies varying in contour, rhythm, timbre and loudness is almost exclusively determined by the contour (Schwarzer, 1993). More generally, contour is a salient feature in naturalistic passages of music (Halpern, Bartlett, & Dowling, 1998; Lamont & Dibben, 2001).

1.2 Cross-modal melodic contour

Melodic contour can be represented in both auditory and visual modalities. Notated music exemplifies visual depictions of melodic contour. In a musical staff, higher and lower pitches correspond to higher and lower spatial positions on the musical score, allowing a visual analogue of pitch contour. Musical notation in many cultures perpetuates this analogy (and implied relation) by representing pitch in the vertical spatial dimension. Even gross simplifications of Western musical notation preserve this relation – composer and theorist Arnold Schoenberg's (1967) line drawings of Beethoven piano sonatas notated pitch contours in terms of ups and downs based on the frequencies of the notes.

The spatial mapping of pitch height is a pervasive and robust phenomenon. The human auditory system translates the sensation of frequency of vibration (caused by the fluctuations in air pressure from a sound-emitting object) into the psychological construct of pitch. Whether through cultural learning or innate bias, we experience notes of higher and lower pitch according to higher and lower frequencies of vibration, respectively. Pitch is described as having "height" (Bachem, 1950; Ruckmick, 1929; Shepard, 1982), and pitch relations, which form the basis for contours, are described as moving "up" and "down," despite the fact that pitch itself is a function of time (i.e., vibrations per second) not space. In other words, listeners automatically represent pitch height spatially, such that they perceive higher pitches to be above (in a spatial sense) lower pitches. For example, in a pitch height comparison task, congruency between the spatial organization of response keys and the relative pitch

height of isolated tones improves listeners' reaction time (incongruency is detrimental), regardless of the degree of musical expertise (Lidji, Kolinsky, Lochy, & Morais, 2007). Furthermore, both musicians and untrained listeners exhibit activation in visual cortex while attending to pitches within a melody (Démonet, Price, Wise, & Frackowiak, 1994; Perry et al., 1999; Zatorre, Evans, & Meyer, 1994; Zatorre, Perry, Beckett, Westbury, & Evans, 1998). Thus, there is direct physiological evidence that under certain circumstances pitch can be represented spatially. Such spatial representations of pitch are not fully understood, but it is clear that listeners can activate a visual representation of melodic contour. It is possible that this auditory-visual mapping may instantiate a more general and complex process of structure mapping (cf. Gentner, 1983; McDermott, Lehr, & Oxenham, 2008). However, the goal of the present research was not to propose the existence of a unitary mechanism or module by which this transfer occurs. Instead, the primary objective of these studies is to explore the information that listeners use when they consciously compare melodic contours across the auditory and visual modalities.

1.3 Mechanisms of cross-modal contour perception

Despite the connection between pitch height and spatial height, there is little work specifying how listeners transfer contour information from one modality to the other. What information are listeners using in their mental representation of a melodic contour? In the mapping between auditory pitch height and visuospatial coordinates, what is the nature of the information that listeners use to construct a spatial representation of contour? Is contour represented as a sequence of upward and downward directions between adjacent events, or are relative heights also encoded with respect to nonadjacent events, or even all other events in a sequence? Addressing such questions requires the development of a quantitative model of cross-modal melodic contour perception. Existing models of auditory contour perception may help to account for the cross-modal perception of melodic contour.

Several contour models adopt a reductive approach by condensing contours to a small number of salient events, such as reversal points (changes in the direction of movement) or the location of the highest and lowest (pitch) event. Reductive models have been proposed to account for contour in both speech (e.g., Ladd, 1996; Pierrehumbert & Beckman, 1988; Xu, 2005) and music (Adams, 1976; Dyson & Watkins, 1984; Morris, 1993). Although reductive models provide a parsimonious description of contour, it is questionable whether they provide a complete and accurate characterization of the psychological representation of contour, as they discard important information through their selective focus.

A number of more elaborate models of contour have been developed. These models go beyond simple descriptions such as reversal points and consider (to varying extents and by various statistical means) the relative heights of both adjacent and non-adjacent events. Within the speech domain, several techniques of describing the similarity of two pitch contours have been developed, such as tunnel measures, root

mean square distance, mean absolute difference, and a correlation coefficient (Hermes, 1998b). Hermes (1998a) asked phoneticians to provide similarity ratings for pairs of auditory or visual contours derived from the pitch contour of spoken sentences. Ratings were then compared with the above contour similarity measures (Hermes, 1998b). Of the various measures, the best predictor of rated similarity was obtained by calculating the correlation between piecewise-linear approximations of the pitch contours (reproducing the contour with a concatenation of line segments representing the original shape). As such, a simple correlation measure (hereafter referred to as surface correlation) holds great promise for predicting melodic contour similarity.

1.3.1 Music-specific contour models

There are also contour models developed within the musical domain. One such approach, called CSIM, is based on a combinatorial model of contour (Friedmann, 1985; Marvin & Laprade, 1987; Polansky & Bassein, 1992; Quinn, 1999) in which each pitch event within a melody is coded as either higher or same/lower than every other pitch, resulting in a matrix of pitch relations. Calculating the number of shared elements between the matrices of two melodies quantitatively determines the CSIM contour similarity. In an experimental test of this model, Quinn (1999) found that contour relations between adjacent and non-adjacent notes predicted musicians' similarity ratings of diatonic, 7-note melody pairs. Interestingly, recent work by Shmulevich (2004) suggests that the CSIM measure is algebraically equivalent to surface correlation measures, such as Kendall's tau or Spearman's rho, thus generalizing the surface correlation measure used in speech research (Hermes, 1998b) to music.

An alternative model of contour characterizes melodies through a Fourier analysis of their pitch relations. Fourier analysis represents the cyclic nature of a signal by breaking it down into a set of harmonically related sine waves. Each sine wave is characterized by a frequency of oscillation, an amplitude and phase. The amplitude measure of each frequency represents how strongly that particular sine wave contributes to the original signal, and the phase describes where in its cycle the sine wave starts. This technique efficiently describes the complete contour rather than discarding potentially important cues that a reductive approach might ignore. Using this procedure, Schmuckler (1999, 2004) proposed a model of melodic contour in which a melody is coded into a series of integers; this series is then Fourier analyzed, producing amplitude and phase spectra for the contour. These spectra thus provide a unique description of the contour in terms of its cyclical components. Comparing the amplitude and phase spectra from different melodies gives a quantitative measure of predicted contour similarity. Schmuckler (1999) provided initial support for this model, demonstrating that listeners' perceptions of contour complexity for both atonal and tonal 12-note melodies were consistently predictable based on amplitude (but not phase) spectra similarity. More recently, Schmuckler (2004) described a further test of this model in which similarity judgements of longer, more rhythmically diverse folk melodies were also predictable based on amplitude spectra correspondence. Together, these findings support the idea that the relative strengths of underlying frequency components can characterize the internal representation of a contour.

1.4 Experimental goals

Testing how well these contour models can predict the similarity of auditory and visual contours is a straightforward way of investigating how listeners convert melodic contour between modalities. There is already some work on cross-modal melodic contour perception (Balch, 1984; Balch & Muscatelli, 1986; Davies & Jennings, 1977; Messerli, Pegna, & Sordet, 1995; Mikumo, 1997; Miyazaki & Rakowski, 2002; Morrongiello & Roes, 1990; Waters, Townsend, & Underwood, 1998). Although these studies represent a wide range of research questions, they all address some aspect of how contour contributes to the perception and production of music in both the auditory and visual modalities. Of this work, the most directly relevant for the current purposes are studies by Balch (1984; Balch & Muscatelli, 1986). Balch and Muscatelli (1986), for instance, tested the recognition of six-note melodies using all possible cross-modal combinations of auditory and visual contours, specifically auditory-auditory (AA), auditory-visual (AV), visual-visual (VV) and visual-auditory (VA). In this work, participants experienced pairs of auditory and/or visual contours, and indicated whether the second contour matched the first. Of the four possible cross-modal combinations produced by this design, Balch and Muscatelli (1986) found that overall, performance was best in the VV condition, worst in the AA condition, and intermediate in the cross-modal (AV and VA) conditions. However, speed of presentation influenced recognition; performance in all but the AA condition suffered with increasing speed such that all conditions performed equally at the fastest rate. These findings suggest that it is more difficult to abstract melodic contour information from the auditory than the visual modality, but also generally validate the viability of a direct cross-modal matching procedure.

The goal of current investigation was to examine the cross-modal evaluations of melodic contour similarity. Music is often a multimodal experience and involves frequent transfer of information across modalities. Accordingly, the main theoretical interest is to gain understanding of the transfer across modalities of one of the most salient features in music – melodic contour. Tasks such as reading music, transcribing melodies, and online monitoring of performance accuracy rely on the ability to successfully transfer melodic contours between the visual and auditory modalities.

This research focuses on two primary questions of cross-modal melodic contour. First, can listeners with various levels of musical expertise recognize cross-modal melodic contour similarity? If so, then second, what forms of information can they use? Of particular interest is if listeners use the cyclic nature of pitch height oscillations (as measured by Fourier analysis) and/or more surface-based information (as measured by a correlation coefficient) when comparing melodic

contours cross-modally.

Therefore, the current studies tested if established quantitative models of contour similarity within modalities can predict cross-modal similarity of melodic contours, by directly comparing auditory and visual contours. This procedure should illuminate the features of contour that listeners use to transfer melodic contours across modalities, and shed light on the processes by which melodic and visual contours are mapped onto one another.

2 EXPERIMENT 1

In Experiment 1, participants judged the similarity between melodic and visual contours. On each trial, some listeners heard a melody followed by a visual contour (the auditory-visual, or AV condition); others experienced the opposite order (visual-auditory, or VA condition). Although simultaneous presentation of melodic and visual contours is possible, it is problematic as it allows participants to use a simple element by element matching strategy. In contrast, by presenting only one contour at a time, listeners must extract and represent in memory the information from the first contour and subsequently compare it with the second. Hence, the design highlights the mental representation of contour, and whether theoretical characterizations of a contour are relevant in similarity judgements. Cross-modal presentation of contours also circumvents the impact of an array of potentially confounding auditory factors (e.g., tonal influences, rhythmic and metrical factors) and visual factors (e.g., spatial extent, spatial density, colour) that might arise when using solely melodic or visual stimuli.

If participants can make use of Fourier analysis and surface correlation information then they should judge visual and auditory contours as being similar according to their theoretical degree of similarity, as judged by these models.

2.1 Method

Participants

All participants were undergraduate students in an introductory psychology course at the University of Toronto Scarborough, and received course credit for their participation. There was no prerequisite or exclusion based on participants' level of musical training. There were 19 participants in the AV condition, with an average age of 19.4 years (SD = 1.5), and an average of 5.2 years (SD = 5.9; range = 0 to 13 years) of formal musical instruction. For the VA condition, there were 23 participants, with an average age of 20 years (SD = 1.6), and an average of 4.8 years (SD = 4.4, range = 0 to 15 years) of formal musical instruction.

Stimuli

Twenty-five tonal melodies composed by Bach, Mozart, Beethoven, Schubert and Brahms were selected from a sight singing textbook (Ottman, 1986) for this study. All of these melodies remained in a single key. The average length of the melodies was 35 notes (SD = 8) and the average duration was

14 s (SD = 3). In these melodies the tempo (the level of the metric pulse) was 120 beats per minute (.5 s per beat), and the timbre was an acoustic grand piano MIDI patch. A series of integers represented the fundamental frequency of each pitch in the melodies, where the lowest note had a value of 0 and the highest note a value of n-1, with n equal to the number of unique notes in the melody (see Schmuckler, 1999). The integer series was graphed as a stair plot, whereby each step of the stair represents a discrete pitch in the melody. Stair plots were then saved as a graphics file (jpeg) to serve as the "matching" visual contour. Figure 1 displays a sample melody from this study (in musical staff notation) and its matching visual contour.

Along with the matching visual contour, a family of mismatching visual contours was created for each melody by randomly reordering the values in the original series. There were some restrictions on these mismatched series. First, the initial two and final three numbers of the original series were the same for all related mismatches so as to prevent participants from relying exclusively on beginning (i.e., primacy) or ending (i.e., recency) information in their similarity judgements. Second, the number of intervals in the mismatches that were bigger than three steps could not vary more than 5% from the number of such intervals in the original. Lastly, no interval in the mismatched series could be larger than the largest interval in the original. This final restriction ensured that the mismatched series did not contain any distinctive features that obviously differentiated them from the original.

For each original sequence, there were initially nine mismatched sequences, with these mismatches varying in their theoretical similarity relation to the original series. Specifically, both Fourier analysis and surface correlation techniques assessed the theoretical similarity between contours. For the Fourier analysis measure, the amplitude and phase spectra of each integer coding were calculated. The amplitude spectra for these contours were then converted to percent variance (technically, the energy spectra), which normalizes the relative strengths of the various sine wave components. For simplicity, this measure will be referred to as amplitude spectra (as they are essentially a normalized derivative of this information). As phase spectra are, by definition, already normalized, there is no need to modify these values. Correlating the amplitude spectra between the original series and the mismatch series determined the amplitude similarity; phase spectra were not considered given the earlier results suggesting that amplitude, not phase information is critical for auditory contours (Schmuckler, 1999, 2004). There were nine mismatched sequences because there was one sequence for each tenth of amplitude similarity between mismatch and original from 0 and 1. In other words, there was one mismatch with an amplitude spectra correlation with the original between 0 and .1, another between .1 and .2, and so on up to between .8 and .9.

For the nine mismatches, the surface correlation similarity was derived by calculating the correlation coefficient of the original (the integer code representing the coded pitch height of the notes in the original melody) with each mis-
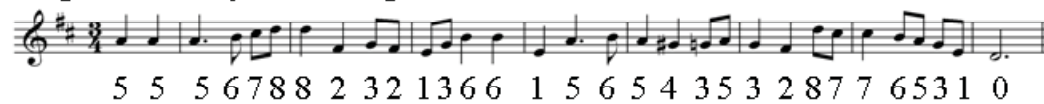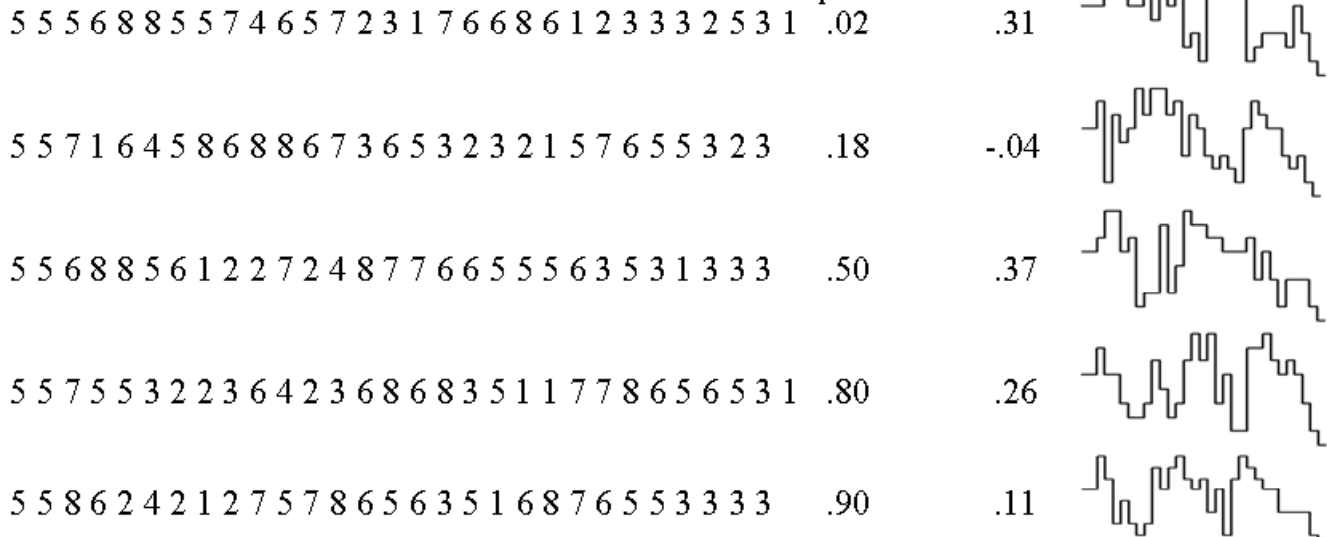
Figure 1. Sample stimulus melody (in musical notation), its integer coding, and line drawing. Below, the integer codes for the final (chosen) five mismatches as well as their line drawings are shown. Measures of similarity between each mismatch and the original are also listed, specifically the correlation of the amplitude spectra and the surface correlation.

match. Ultimately, five of these mismatches were chosen for presentation to participants, selected by choosing the five series with the lowest surface correlation with the original, in an effort to empirically separate (as much as possible) the potential effect of surface correlation and amplitude spectra similarity. Although this attempt to disentangle amplitude spectra and surface correlation did so by minimizing surface correlations, both measures nonetheless produced a fairly wide (and equivalent) range of correlation coefficients with the original series (Fourier analysis: .01 to .9; surface correlation: -.49 to .52). The final five mismatched series were graphed as line drawings and saved as graphics files in the same manner as the matching visual contour. Figure 1 also displays the five mismatched integer series for the corresponding sample melody, along with the amplitude spectra correlations and surface correlations with the original, and the line drawing resulting from these series. Combined with the matching stimulus, this procedure yielded six possible visual contours for comparison with each auditory melody.

Apparatus

Generation of random mismatched series, and analyses of all (original and mismatch) sequences were performed using code written in MATLAB, in conjunction with the midi toolbox (Eerola & Toiviainen, 2004). Presentation of the stimuli and the experimental interface were programmed with MATLAB 7.0 using Cogent 2000 (developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department

of Imaging Neuroscience). Two Pentium(R) 4 computers (3.0 and 1.7 GHz) were used for running the experiment. Auditory stimuli for this study were generated using Audigy Platinum Soundblaster sound cards, and were presented to listeners over Audio Technica ATH-M40fs or Fostex T20 RP Stereo Headphones, set to a comfortable volume for all participants. Visual stimuli appeared on either a Samsung 713V or LG Flatron L1710S 15" monitor.

Procedure

Participants in the auditory-visual (AV) condition heard a melody, followed by a picture that represented the shape of a melody, and then rated the similarity between them. Each trial for the AV participants began with the phrase "Listen carefully to the melody" displayed on the computer monitor while the melody played. After the melody finished, the computer loaded and displayed the graphics file as quickly as possible (due to hardware limitations, this was not immediate, however the delay was always less than one second). This contour remained present until listeners entered a response, at which point the monitor was blank for 250 ms, until the beginning of the next trial. Participants in the visual-auditory (VA) condition experienced the same stimuli but in the reverse order. For the VA participants, the line drawing was displayed for 2.5 seconds before being replaced by the phrase "Listen carefully to the melody" (placed at the same location in order to mask residual visual input). Concomitantly, the melody began playing.

All participants (AV and VA) rated the similarity of the

contour between the picture and the melody on a scale of 1 to 7 (1 being not at all similar, 7 being very similar). Trials were presented in random order, with the restriction that no individual melody was heard twice in a row. Twenty-five possible (original) melodies combined with six possible visual displays (the match plus five mismatches) resulted in 150 trials in total. To clarify, because only the original melodies were presented, there were no additionally generated melodies. Instead, pairing generated visual sequences with the original melody constituted a mismatch. Participants were either run individually or in pairs (on different computers, separated by a divider). The entire experimental session lasted about one hour for both AV and VA conditions.

2.2 Results

To provide a baseline measure of maximal similarity, participants' ratings for the matching auditory-visual stimuli were first compared with the ratings for the mismatched stimuli by means of a one-way repeated-measures Analysis of Variance (ANOVA). The within-subjects factor was match (matching versus mismatching). In the initial analysis the different levels of auditory-visual mismatch were thus collapsed. For the AV condition, this analysis revealed that ratings of similarity were significantly higher for matches (M = 4.83, SD = .55) than for mismatches (M = 4.27, SD = .65), $F(1,18) = 15.69$, MSE = .19, $p < .001$, $\eta_p^2 = .47$. Interestingly, two participants failed to show this trend; this result indicates that they were not attending to the task, and therefore their data were removed from further analyses. Similar results were observed for the VA condition; matches (M = 4.86, SD = .72) were rated as being more similar to the melody than mismatches (M = 4.18, SD = .54), $F(1,22) = 72.58$, MSE = .07, $p < .001$, $\eta_p^2 = .77$. In this case, one participant did not show this pattern; the data of this participant were removed from further analyses. Overall, therefore, the average ratings of perceived similarity of melodies and matching sequences exceeded those of melodies and mismatching sequences.

The preceding analysis demonstrates that participants were sensitive to the similarity between auditory and visual melodic contours. However, the analysis does not determine whether listeners differentiated between visual contours having varying degrees of similarity with the auditory contour. To explore this issue, subsequent analyses focused on examining whether or not the various models of contour similarity described earlier could predict listeners' perceived contour similarity. Because this question is one of predicting perceived levels of mismatch between auditory and visual stimuli, these analyses focused on the mismatched sequences only and excluded the match trials.

Based on the various contour models described earlier, a host of contour similarity predictors were generated, including models based on those outlined by Schmuckler (1999). The Fourier analysis model produces two possible predictors (as already discussed): amplitude spectra and phase spectra similarity. As described in the stimulus section, amplitude and phase spectra information for all integer series were calculated, and absolute difference scores standardized to the length

of the melody were computed between the auditory (original) and visual (mismatch) sequences[1]. Along with these Fourier analysis measures, Schmuckler (1999) also described an oscillation model, in which the interval information between consecutive pitches is quantified to produce both a summed and a mean interval measure (see Schmuckler, 1999, for detailed discussion of these measures). Accordingly, four measures were derived from this earlier work – amplitude and phase spectra difference scores, and summed and mean interval difference scores.

Along with these measures, three additional theoretical predictors were calculated. The first is based on the combinatorial model (Friedmann, 1985; Marvin & Laprade, 1987; Polansky & Bassein, 1992; Quinn, 1999) and involves the CSIM measure described earlier, which characterizes each contour as a matrix in terms of whether a subsequent tone is higher (coded as 1) or equal to/lower (coded as 0) than each of the other tones in the melody. Then, the mean number of shared elements between the matrices of each mismatch and its corresponding match was calculated and used as the CSIM predictor. Second, a surface correlation measure was calculated by simply correlating the integer codes for each melody. Third, a measure based on comparing the number of reversals in the match and mismatch was calculated. Dividing the number of reversals in the match by the number of reversals in the mismatch gave a ratio of reversals. This ratio was subtracted from 1 so that the absolute value of this difference indicated the percent difference in number of reversals between match and mismatch (a higher number would indicate greater difference, thus presumably less similarity).

Preliminary analyses revealed that the length of the melody was a strong predictor of perceived similarity, perhaps because two of the 25 melodies were longer than the rest (56 notes; beyond two standard deviations of the mean of 35 notes). Given that remembering the first contour and comparing it to the second was a challenging task, and only these two melodies were much longer than the others, listeners may have systematically rated longer melodies (and line drawings) as more similar than shorter stimuli. Therefore, the data of these two melodies were excluded, leaving 23 melodies (each with five mismatches); in addition melody length was included as a potential predictor of similarity.

Table 1 provides an intercorrelation matrix for these eight measures across all the mismatching stimuli in this study. This table reveals a few significant intercorrelations between variables. As expected, CSIM and surface correlation measures were essentially equivalent ($r = .96$, $p < .001$), corroborating Shmulevich's (2004) calculations. Melody length correlated significantly with amplitude spectra, summed interval and mean interval. These correlations are not surprising given that these three variables were all standardized to the length of the melody. Mean interval was significantly correlated with amplitude spectra and reversal ratio; reversal ratio was also related to summed interval. The interrelation of these variables most likely indicates the extent to which these measures mutually indicate some aspect of the cyclical ups and downs of contour.

Table 1: Intercorrelations of Theoretical Predictors of Contour Similarity for Experiment 1

Predictor	Phase Spectra	Summed Interval	Mean Interval	Surface Correlation	CSIM	Reversal Ratio	Melody Length
Amplitude Spectra	-.07	-.01	.27**	.14	.15	.07	-.62***
Phase Spectra		.03	-.01	-.18	-.12	-.03	.10
Summed Interval			-.18	-.09	-.11	.35***	.28**
Mean Interval				.12	.11	.30**	-.30**
Surface Correlation					.96***	.05	-.14
CSIM						.02	-.17
Reversal Ratio							.07

** $p < .01$. *** $p < .001$.

All eight of these predictors were correlated with the averaged similarity ratings for the AV and VA conditions. The results of these analyses appear in Table 2 and demonstrate that surface correlation, CSIM, and melody length all significantly correlated with listeners' cross-modal similarity ratings in both conditions. Amplitude spectra difference scores correlated negatively and significantly with the VA similarity ratings but not for the AV condition. The AV and VA ratings themselves were significantly related, ($r = .39$, $p < .001$).

As a follow-up to these analyses, two multiple regression analyses were performed to determine the unique contribution of these models to predicting perceived similarity, for the AV and VA conditions separately. Given the high correlation between the surface correlation and CSIM variables (leading to an unacceptably low tolerance value of .087 in the regression equation), and the fact that surface correlation had the larger unique contribution of explanatory variance in both AV and VA conditions, only surface correlation was retained in the final regression equations. Both AV and VA similarity ratings were thus predicted from the three variables of amplitude spectra differences, surface correlation, and melody length. For the AV condition these three variables significantly predicted similarity ratings, $R (3,111) = .41$, $p < .001$, with significant contributions by surface correlation, $B = .61$, $\beta = $

.28, $p < .01$, and melody length, $B = .03$, $\beta = .36$, $p < .01$. In contrast, amplitude spectra failed to contribute significantly, $B = 1.49$, $\beta = .03$, ns. For the VA condition these three variables also significantly predicted similarity ratings, $R (3,121) = .49$, $p < .001$, with significant contributions from amplitude spectra, $B = -3.06$, $\beta = -.24$, $p < .05$, and surface correlation, $B = .18$, $\beta = .37$, $p < .001$. In this case, melody length failed to contribute significantly, $B = .004$, $\beta = .19$, ns.

Finally, a set of analyses looked at the impact of musical experience on contour similarity. For this analysis each participant's ratings were averaged across the 23 matching stimuli and compared with the average ratings from four different sets of mismatches. The first set consisted of the averaged ratings for the complete set of mismatches ($N = 115$); the second set consisted of averaged the ratings for the 23 mismatches with the largest amplitude spectra difference score; the third set consisted of averaged the ratings for the 23 mismatches with the largest phase spectra difference score; the fourth set consisted of averaged the ratings for the 23 mismatches with the lowest surface correlation with each melody. Each participant's data were transformed into z-scores (each participant as a separate population), and the differences between the z-scores of the matches and the four mismatched sets were calculated. Thus, each participant had four scores: an over-

Table 2: Correlations of Theoretical Predictors with Auditory-Visual (AV) Similarity Ratings and Visual-Auditory (VA) Similarity Ratings of Experiments 1 and 2

Predictors	Experiment 1		Experiment 2	
	AV similarity rating	VA similarity rating	AV similarity rating	VA similarity rating
Amplitude Spectra	-.15	-.30**	-.10	-.22*
Phase Spectra	-.12	.06	-.37***	-.30**
Summed Interval	.06	.04	-.08	-.04
Mean Interval	-.05	-.07	.10	-.01
Surface Correlation	.24*	.31***	.46***	.45***
CSIM	.19*	.27**	.41***	.39***
Reversal Ratio	.07	.02	.19*	.13
Melody Length	.30**	.28**	-.15	-.02

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3: Correlations Between the Years of Musical Training and Difference Score Measures in Similarity Ratings for Experiment 1 and 2

Difference score	Experiment 1		Experiment 2	
	AV Condition	VA Condition	AV Condition	VA Condition
Overall	.54*	-.15	.57*	.05
Amplitude Spectra	.33	-.12	.59*	-.05
Phase Spectra	.49*	-.34	.49*	.08
Surface Correlation	.53*	.03	.63**	.01

* $p < .05$. ** $p < .01$.

all difference score, an amplitude spectra difference score, a phase spectra difference score and a surface correlation difference score. These difference scores were then correlated with participants' degree of musical training (for AV and VA conditions separately), as indexed by the number of years of formal instruction on an instrument or voice. Table 3 shows the results of these analyses.

Participants in the AV condition with more formal training differentiated more between matches and mismatches, and relied more on both phase spectra and surface correlation differences to form their ratings of perceived similarity. For the VA condition, however, musical training did not affect participants' difference scores. There were no overall differences between the AV and VA condition in the absolute size of the overall difference score, $F(1,37) < 1$, MSE = .08, ns, the amplitude spectra difference score, $F(1,37) = 2.91$, MSE = .11, ns, the phase spectra difference score, $F(1,37) < 1$, MSE = .12, ns, or the surface correlation difference score, $F(1,37) < 1$, MSE = .11, ns.

2.3 Discussion

There are three main findings of Experiment 1. First, listeners matched contours of long melodies cross-modally, as demonstrated by higher similarity ratings between the auditory melodies and matching visual representations of their contour, relative to ratings of similarity between melodies and mismatched visual representations. Second, established theoretical models of contour similarity can partly explain the perceived similarity of cross-modal melodic contours, however there were differences between the AV and VA conditions. Third, only in the AV condition did musical expertise aid listeners in rating the difference between match and mismatch; it also enabled them to be more sensitive to phase spectra and surface correlation in forming their ratings.

Our observation that listeners were able to recognize the similarity between contours presented cross-modally replicates previous findings on cross-modal contour perception (Balch, 1984; Balch & Muscatelli, 1986; Davies & Jennings, 1977; Messerli et al., 1995; Mikumo, 1997; Miyazaki & Rakowski, 2002; Morrongiello & Roes, 1990; Waters et al., 1998). Because only one contour (either auditory or visual) was presented at a time, listeners could not simply compare the auditory and visual contours element by element and check for differences. Accordingly, this task required listeners to extract and subsequently remember contour informa-

tion for use in a later comparison.

What attributes of the contours contributed to listeners' perceived similarity? Both conditions showed strong effects of surface correlation, a finding that extends previous research on within-modal auditory contour similarity (Hermes, 1998a, 1998b; Quinn, 1999) to cross-modal applications. To the extent that surface correlation conveys both the local, note-to-note characteristics and overall global shape of a contour, this finding implies that listeners can use a combination of both local and global cues when converting contours between the auditory and visual domains, regardless of the modality in which the contour is initially presented.

The effect of Fourier components on perceived similarity was mixed, and varied for the AV and VA conditions. Phase did not contribute to the AV or VA condition regressions, a result that replicates and extends findings of the unreliable nature of phase in modeling melodic contour perception (Schmuckler, 1999, 2004; although see Schmuckler, 2008). In contrast, amplitude spectra differences were significant, but only in the VA condition. These results suggest that listeners can use the global cues of cyclic oscillation that Fourier analysis captures for evaluations of cross-modal melodic contour similarity, but only when comparing a visual contour to a subsequently occurring auditory contour. However, discussing this finding in detail requires reference to the results of the second experiment, therefore the general discussion considers the implications of this finding.

Largely because Experiment 2 replicates the findings of the variable role of musical expertise for AV and VA conditions, this result also is explored in greater detail in the general discussion. However, the fact that this finding emerges only in the AV condition suggests that converting contour information from the auditory to the visual domain exploits the skills that musical training confers. It is likely that the AV condition is more challenging than the VA condition due to differential memory demand. Specifically, because the melodies were presented in a gated (note-by-note) fashion, participants had to remember the melody in its entirety in the AV condition and subsequently compare it to a visual contour. Conversely, in the VA condition participants could compare their memory of the visual contour to the gated presentation of the melody as it progressed note-by-note rather than waiting until the melody finished. Thus the relatively higher memory demand of the AV condition may differentiate across levels of musical training more so than the VA condition.

Indeed, one potential concern with this study concerns the high memory demand of the task. In particular, this study employed melodies of considerable length, which may have strained listeners' memory capacities and made the evaluation of cross-modal contour similarity difficult. Accordingly, it is of interest to replicate the principal findings of this work with melodies that make lesser memory demands. Specifically, can listeners recognize cross-modal melodic contour similarity, and can current models of contour information such as surface correlation and Fourier analysis components explain perceived similarity when memory demands are less? Experiment 2 provided such a replication by testing the cross-modal similarity of shorter melodies than those employed here, thus also extending this work.

3 EXPERIMENT 2

The results of Experiment 1 suggest that surface correlation and Fourier analysis components both contribute to the perceived similarity of long melodies compared across auditory and visual modalities. However, listeners often hear shorter melodies, and furthermore most of the previous work on melodic contour (within and across modalities) uses much shorter melodies. It is also possible that the length of the stimulus melodies and the concomitant memory demands might have influenced the nature of listeners' cross-modal comparisons, along with how well these different approaches characterized cross-modal contour similarity.

Therefore it is of interest to replicate these results with shorter melodies, for two main reasons. First, these models of melodic contour may perform differently under conditions more similar to existing melodic contour research. Thus, repeating these tests with shorter melodies can investigate this possibility and potentially extend the validity of these models to melodies of various lengths. Second, listeners may or may not use similar contour information for short as well as long melodies. Consequently, testing shorter melodies provides the opportunity to ascertain if listeners use the same information to evaluate cross-modal melodic similarity regardless of contour length.

To test these possibilities, Experiment 2 employed the same task as the earlier study but used new, shorter melodies for cross-modal comparisons.

3.1 Method

Participants

Participants were undergraduate students in an introductory psychology course at the University of Toronto at Mississauga, and received course credit for their participation. There was no prerequisite or exclusion based on participants' level of musical training.

There were 17 participants in the AV condition, with an average age of 18.5 years (SD = .86), and an average of 1.5 years (SD = 2.6; range = 0 to 10 years) of formal musical instruction. For the VA condition, there were 17 participants, with an average age of 19.1 years (SD = 2.19), and an aver-

age of 1.3 years (SD = 2.9, range = 0 to 10 years) of formal musical instruction.

Stimuli, Apparatus, and Procedure

Twenty-five tonal melodies from a compilation of sight singing melodies (Smey, 2007) were used for this study. All of these melodies were between 14 and 18 notes long, and did not modulate to a new key. The average length of the melodies was 16.7 notes (SD = 1.3) and the average duration was 7.5 s (SD = .4). As in Experiment 1, the tempo of these melodies was 120 beats per minute, and the timbre was an acoustic grand piano MIDI patch. The melodies were coded as integer series in the same manner as Experiment 1, to form the "matching" visual contours. The non-matching mismatched visual contours were created in the same fashion as Experiment 1, using the same rules and theoretical similarity measures.

The apparatus and procedures were the same as in Experiment 1. There were 150 trials in total, and the experimental session lasted about 45 minutes.

3.2 Results

As in Experiment 1, an initial step in the data analysis was designed to establish the average similarity rating for conditions of maximal similarity (match). A one-way repeated measures ANOVA compared participants' ratings for the matching auditory-visual stimuli with the ratings for the mismatched stimuli, with the within-subjects factor of match (matching versus mismatching). For the AV condition, this analysis revealed that ratings of similarity were significantly higher for matches (M = 4.93, SD = .72) than for mismatches (M = 4.13, SD = .55), F(1,16) = 18.86, MSE = .29, p < .001, η_p^2 = .54. For the VA condition the results were similar, with matches (M = 5.18, SD = .61) rated as more similar to the melody than mismatches (M = 4.22, SD = .47), F(1,16) = 40.93, MSE = .19, p < .001, η_p^2 = .72. Again, therefore, listeners recognized the greater similarity of contours that matched the melodies to those that were mismatched.

Subsequent analyses determined the extent to which the various contour similarity models correlated with participants' perceived similarity ratings, again focusing only on the ratings of the mismatch trials. This analysis tested the same contour similarity predictors as Experiment 1, including the difference score measures of amplitude spectra, phase spectra, summed interval and mean interval, as well as the CSIM/surface correlation measure, reversal ratio and melody length measures. Table 4 shows the intercorrelations between the predictors for Experiment 2. The correlations between these predictors and the perceived similarity ratings for the AV and VA conditions appear in Table 2. For both the AV and VA conditions, phase spectra, surface correlation and CSIM measures were significantly related to participants' ratings. Counterintuitively, the reversal measure was significantly positively correlated with AV similarity ratings, a finding suggesting that a greater difference in reversals between a melody and its mismatch produced higher perceived similarity. As in Experiment 1 the amplitude spectra significantly

Table 4: Intercorrelations of Theoretical Predictors of Contour Similarity for Experiment 2

Predictor	Phase Spectra	Summed Interval	Mean Interval	Surface Correlation	CSIM	Reversal Ratio	Melody Length
Amplitude Spectra	.08	.24**	.37***	-.14	-.17	.07	-.23**
Phase Spectra		.12	.07	-.42***	-.37***	-.14	.01
Summed Interval			.20*	.08	.07	.01	-.27**
Mean Interval				.02	.00	.45***	-.42***
Surface Correlation					.97***	.11	-.22*
CSIM						.09	-.23**
Reversal Ratio							-.22*

*$p < .05$. ** $p < .01$. *** $p < .001$.

correlated with perceived similarity in the VA condition only. Finally, the AV and VA condition similarity ratings correlated significantly with each other ($r = .58$, $p < .001$).

Two multiple regression analyses examined the strength and unique contribution of each of the potential predictors to perceived similarity. As in Experiment 1, surface correlation was included instead of CSIM in both AV and VA conditions because of its stronger relation with similarity ratings. For both sets conditions, similarity ratings were predicted from amplitude spectra differences, phase spectra differences, surface correlations, and reversal scores.

For the AV condition, these variables significantly predicted similarity ratings, $R(4,120) = .51$, $p < .001$, with significant contributions of phase spectra differences, $B = -.22$, $\beta = -.19$, $p < .05$, and surface correlation, $B = .44$, $\beta = .36$, $p < .001$. In contrast, there was no significant effect of either amplitude spectra differences, $B = -.79$, $\beta = -.04$, ns, or of reversals, $B = .23$, $\beta = .13$, ns. For the VA condition, these variables also significantly predicted similarity ratings, $R(4,120) = .49$, $p < .001$, with significant contributions by amplitude spectra differences, $B = -3.68$, $\beta = -.16$, $p < .05$, and surface correlations, $B = .47$, $\beta = .37$, $p < .001$. In contrast, there was no significant effect of either phase spectra differences, $B = -.15$, $\beta = -.11$, ns, or of reversals, $B = .18$, $\beta = .1$, ns.

The last set of analyses tested the effect of musical experience on contour similarity ratings. Each participant's ratings for the 25 matching stimuli were averaged and compared with the same four sets of mismatches described in Experiment 1. As in this previous study, the differences between the z-scores of the matches and the four mismatch sets were calculated, and correlated with participants' degree of musical training, as indexed by the number of years of formal musical instruction. Table 3 presents these analyses, and indicates the same general pattern as Experiment 1. Participants in the AV condition with more formal training differentiated matches and mismatches more, were better able to use amplitude and phase spectra and surface correlation differences between matches and mismatches in forming a perceived similarity rating. But in the VA condition, musical training did not correlate with participants' difference scores. Also similar to Experiment 1, there were no differences in absolute size of the difference scores between the AV and VA condition. Neither

the overall difference score $F(1,32) < 1$, MSE = .14, ns, nor the amplitude spectra difference score, $F(1,32) < 1$, MSE = .19, ns, nor the phase spectra difference score, $F(1,32) = 1.2$, MSE = .16, ns, nor the surface correlation difference score, $F(1,32) = 1.8$, MSE = .17, ns showed any difference in absolute size between the AV and VA condition.

3.3 Discussion

In Experiment 2, listeners again succeeded at recognizing matching cross-modal melodic contours. Furthermore, surface correlation and Fourier components predicted their ratings of perceived similarity between non-matching contours. Lastly, musical expertise allowed listeners to make better use of the available cues in evaluating contour similarity in the AV condition. Therefore the results of Experiment 2 are quite similar to those of Experiment 1, while ruling out the potentially confounding effects of melody length from Experiment 1.

The surface correlation measure was a good predictor of cross-modal contour similarity ratings for both the AV and VA condition, again demonstrating the importance of correlation coefficients in modeling contour similarity and generalizing its validity to cross-modal perception. The Fourier components, on the other hand, varied in their predictive value depending on the order of presentation of the contours. Specifically, listeners' ratings were related to phase spectra for the AV condition, and amplitude spectra in the VA condition. Neither Fourier component significantly predicted perceived similarity in both conditions. Other than the significant contribution of phase in the AV condition of Experiment 2 (that did not occur in Experiment 1), these results echo Experiment 1.

4   GENERAL DISCUSSION

4.1 Summary

Together, Experiments 1 and 2 provide a number of insights into contour processing. First, and most fundamentally, these studies demonstrate that listeners can recognize the similarity of melodic contours when presented cross-modally, regardless of melody length. Both studies revealed higher similarity ratings for matching auditory and visual contours

relative to mismatching contours. Although this finding may seem relatively intuitive, this result is noteworthy in the sense that the majority of research on cross-modal melodic contour (Balch, 1984; Balch & Muscatelli, 1986; Cupchik, Phillips, & Hill, 2001; Lidji et al., 2007; Mikumo, 1994; Miyazaki & Rakowski, 2002; Morrongiello & Roes, 1990; Waters et al., 1998) has used relatively short melodies (five to seven notes) that were within the capacity of working memory. Because both studies in this work employed melodies well beyond the limitations of short term processes, recognition of cross-modal similarity in this case is not a foregone conclusion, particularly given that the sequential presentation of the contours exacerbated the difficulty of the task. Nevertheless, listeners were able to recognize the similarity of cross-modal melodic contours.

Second, these results provided an additional validation of the applicability of current models of contour structure and similarity to a previously untested domain. Specifically, theoretical similarity between cross-modal contours was predictable based on the combinatorial CSIM (or surface correlation) model proposed by Quinn (1999), as well the Fourier analysis model of Schmuckler (1999, 2004, 2008). In both experiments, these two models significantly predicted cross-modal contour similarity. This result suggests that at least some of the information that listeners use when constructing a mental representation of an auditory or visual contour is embodied by these quantitative contour descriptions.

4.2 Differences between experiments

One important distinction between these two models in these studies that merits deeper consideration is the variable success of the Fourier components (amplitude and phase) across modality presentation order and melody length. Whereas the surface correlation model was predictive across both presentation orders and melody lengths, amplitude and phase were not. Specifically, amplitude spectra differences were predictive of contour similarity for both short and long melodies, but only when the visual contour preceded the auditory contour (the VA condition), but not for the opposite order (the AV condition). In contrast, phase spectra differences were predictive only for the AV presentations with the short melodies.

Why might a VA, but not an AV, ordering of contours allow for the use of amplitude spectra information, whereas an AV ordering with short melodies enable the use of phase spectra information? One possibility is that listeners mentally convert what they remember of the contour presented first into the modality of the contour that occurs second to facilitate a direct comparison between the two. That is, listeners might attempt to create an auditory analogue of a visually presented contour for a VA ordering, or vice versa for an AV ordering. Such a recoding would make similarity judgements predictable based on the optimum way of characterizing the latter contour. Research on the applicability of Fourier analysis to visual scenes has revealed that in general, phase information is more important than amplitude information for visual perception (Bennett & Banks, 1987, 1991; Kleiner, 1987; Klein-

er & Banks, 1987). Conversely, amplitude spectra information is more important than phase spectra information when perceiving auditory contours (Schmuckler, 1999, 2004).

There is good reason for the variable importance of amplitude and phase for audition and vision, respectively. In vision, variation in amplitude corresponds to stimulus energy (essentially degrees of light and dark), whereas phase corresponds to stimulus structure, or roughly the presence and placement of edge information. Clearly, of the two, edges and their locations are more fundamental for visual object recognition. For auditory contours, however, stimulus energy indexes the relative strength of the cyclic components (i.e., whether the signal repeats once, twice, and so on, over its length), whereas phase indexes the relative timing within the contour of ascending and descending patterns. Although both forms of information are potentially important in understanding the general shape and structure of a melody, the former intuitively seems to have a greater perceptual priority. In support of this idea, Schmuckler (1999) found that listeners can make use of phase information for perceived contour similarity when the melodies were constructed specifically to contain important phase relations. More recently, Schmuckler (2008) found a consistent correlation between phase spectra differences and perceived contour similarity when phase information was calculated based on a rhythmically weighted contour code (see Schmuckler, 1999, 2004, for discussions of this form of coding). However, in a multiple regression context phase spectra differences failed to add significantly to predictions of contour similarity.

The idea that listeners convert what they remember of the first contour into the modality of the second contour predicts well the observed pattern of results for the amplitude spectra differences. Specifically, because the VA condition would encourage listeners to recode the visual contour into an auditory one, amplitude spectra information would thus become maximally important for contour comparisons; this was what was observed in this study. This hypothesis, however, also predicts the opposite pattern for the AV condition. In this case, listeners would mentally convert the initial auditory contour into a visual analogue, with similarity judgements primarily predictable based on phase spectra differences. In partial support of this idea, similarity judgements in the AV condition were predictable based on phase information, at least for the shorter melodies of Experiment 2. However, phase played no role in the AV condition for the longer melodies of Experiment 1, implying a melody length effect on the use of phase information.

In short, the predictive value of phase changed across melody length, indicating that cross-modal contour similarity may be evaluated differently under varying musical conditions. But why should melody length have such an impact on listeners' use of phase, but not amplitude? Simply put, because phase information indexes the relative timing of the ups and downs in an auditory signal, shorter melodies enable the use of local ups and downs, and thus foster listeners' mental recoding of the melodic contour as a visual analogue. However, longer melodies (on average 35 notes in Experiment 1)

vitiate the usefulness of local information, such as the timing and/or position of rises and falls in the contour, as measured by phase spectra. Accordingly, phase information will be of less use with such melodies. In contrast, because amplitude information captures global contour shape, such information is equally accessible in short and long melodies; in fact, global contour information is likely the most accessible information in longer melodies. Consequently, melody length should have less influence on the use of amplitude spectra information, provided that the melodies are long enough to contain sufficiently differentiated amplitude spectra (see Schmuckler, 2004, for a discussion of this point).

A final point about the differences between Experiments 1 and 2 concerns the relationship between AV and VA similarity ratings. In Experiment 1, the correlation between similarity ratings for the AV and VA conditions was relatively low (r = .39) compared to Experiment 2 (r = .58). This difference is likely a result of greater task difficulty of the first experiment due to longer melodies (and thus increased memory demand), thereby introducing more variability into the similarity ratings. However, the level of task difficulty did not only differ across experiments, but also within conditions; the latter variance reveals some interesting findings with regard to musical expertise, discussed below.

4.3 Role of musical experience

The third principal finding from these studies involves the role of musical experience in cross-modal melodic contour similarity. In both experiments, musical training aided participants' ability to differentiate between matches and mismatches, but only in the AV conditions. Further, in these conditions, musical training enabled listeners to make better use of amplitude spectra, phase spectra, and surface correlation information. These results give rise to two questions – how can musical training confer an advantage on listeners' cross-modal melodic contour perception generally, and why is this facilitation specific to the AV condition?

Musical training involves extensive practice with cross-modal contours. Specifically, musicians receive extensive practice in translating between written musical notation (essentially a system of horizontal lines with vertically arranged dots) and auditory sequences, experience that intuitively seems quite comparable to the tasks used in these studies. Accordingly, simple practice effects with comparably structured stimuli may account for the overall advantage conferred by musical training. In keeping with this argument, there are reports in the literature of processing advantages in cross-modal musical stimuli due to musical training. Brochard et al. (2004) similarly found that musicians possess a spatial advantage for processing dots placed above and below horizontal lines (similar to musical notation). Further, these authors also observed that musicians processed dots placed to the left and right of vertical lines faster than nonmusicians. Lidji et al. (2007) had similar findings, in that pitch height automatically activated congruent left-right spatial mappings for musicians but not nonmusicians. Specifically related to contour perception, Balch and Muscatelli (1986) found that musi-

cians outperformed nonmusicians in all contour comparison tasks, including within-modal (AA and VV) and cross-modal (AV and VA) conditions. Furthermore, accuracy at recognizing transformations to melodic contours predicts the ability to judge spatial transformations of three-dimensional figures (Cupchik et al., 2001). Thus musical training may improve the perception and processing of cross-modal contour more generally.

However, musical experience was not helpful in the current studies for all conditions, but only the AV condition. The relative difficulty of the AV versus VA condition may explain why the facilitation effect of musical training only occurred in the AV condition, as evidenced by the difference score measures (Table 3). In both experiments, the similarity ratings between melodies and their matching visual contours was higher than for non-matching, mismatched contours, but the effect was always larger for the VA condition than for the AV condition. Inspecting the partial eta-squared values reveals that the effect size of differentiating between match and mismatch was higher for VA than AV conditions in Experiment 1 (AV $\eta_p^2 = .47$; VA $\eta_p^2 = .77$) and Experiment 2 (AV $\eta_p^2 = .54$; VA $\eta_p^2 = .72$). This difference makes sense intuitively, because the AV condition was more taxing on memory demands than the VA condition. Accordingly, the more difficult task of the AV condition accentuated the difference in abilities to compare melodic contours cross-modally as a result of musical training. Conversely, the VA condition was less difficult for participants, and so the contour processing advantage of musically-trained listeners was not as apparent. Thus if listeners encounter a situation that resembles a music-specific task, then musicians' experience will give them an advantage. However if the task changes (in this case, even just the order of presentation of stimuli), the domain-specific skills that musicians have developed may not confer the same benefits. Additionally, presenting the visual line drawing in a gated fashion may make the VA condition more difficult and consequently differentiate more between musically trained and untrained listeners.

4.4 Limitations

Along with the positive findings of these studies, there are a number of important limitations to this work that require consideration. Probably the most critical such concern involves the fact that although the various theoretical models of contour structure were predictive of cross-modal similarity, ultimately these models only explained part of the variance in such predictions. Such a finding raises the question of exactly how important such information is in participants' perception and processing of contour. As a partial answer to this concern, it is worth noting that the level of predictiveness of these variables is generally equivalent to what has been previously reported in the literature (Eerola, Järvinen, Louhivuori, & Toiviainen, 2001; Quinn, 1999; Schmuckler, 1999). Accordingly, although there are clearly many other factors that also enter into contour perception, the information captured by these predictors seems to be a consistently influential. Both Eerola and colleagues (Eerola & Bregman, 2007;

Eerola, Himberg, Toiviainen, & Louhivuori, 2006; Eerola et al., 2001) and Schmuckler (1999, 2004) have posited and investigated a variety of other factors, ranging from rhythmic components to structural factors (such as tonality and meter) to individual contour features, with varying degrees of success.

A second form of limitation with this work involves issues with the theoretic predictors themselves. Specifically, both Fourier analysis and surface correlations have inherent constraints that raise concerns when applying such procedures to models of contour structure and perceived similarity. For instance, both correlation techniques and Fourier analysis procedures are constrained by factors related to the length of the series being analyzed. Correlation measures are adversely affected by sequence length, such that the shorter the sequence the more susceptible the measure is to outlying values of the individual elements. Accordingly, shorter melodies limit the utility of correlation measures. Correlations are also limited in that they can only be applied to sequences containing the same number of elements. Given that contour comparisons rarely involve contours of the same length, this poses a methodological problem for applying surface correlations to models of melodic contour.

Fourier analysis techniques also present important methodological concerns. For one, as a mathematical procedure Fourier analysis makes a variety of assumptions about the signal that are generally not met in an application to melodic contour. Perhaps the most obvious is that Fourier analysis assumes that the signal is continuous and periodic (i.e., it has been on forever and will continue indefinitely). Needless to say, other than the occasional annoying tune that perversely gets stuck in one's head, melodies do not repeat ad infinitum. Yet in order to achieve continuity and function as a cohesive piece of music, repetition of some of the musical structure must occur; contour is one of the most important forms of pitch structure and as such could function as one of components that help to achieve this continuity. Another assumption of Fourier analysis concerns the length of the signal. When the signal is too short, Fourier analysis spectra are prone to distortions such as edge effects. The length of the melodies used in this research help to insulate the Fourier analysis from this phenomenon, but this is an issue in any application of this tool. Ultimately, the success of this approach in predicting contour similarity in this and other contexts provides support for the applicability of these procedures for the quantification of contour perception and processing, despite these potentially problematic issues.

5 Conclusion

In conclusion, this investigation into the cross-modal similarity of melodic contour has enabled insights into how listeners accomplish the transfer of contour information between the visual and auditory modalities. The multimodal nature of music highlights the importance of understanding how listeners convert musical information between modalities, and melodic contour is a prime example. For example, there are numer-ous musical skills that depend on the accurate conversion of melodic contour between the visual and auditory modalities, such as the ability to read music, record melodies in written form and monitor the accuracy of musical performances in real-time.

There are several potential implications of the current results. First, they validate the applicability of theoretical models of contour structure to cross-modal investigations. Second, these findings have the potential to inform models of music expertise and cross-modal music cognition. Third, this research may have relevance to practical applications such as remedial speech perception training and pedagogical approaches to musical instruction.

REFERENCES

't Hart, J., Collier, R., & Cohen, A. J. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge, UK: Cambridge University Press.

Adams, C. R. (1976). Melodic contour typology. *Ethnomusicology, 20(2)*, 179-215.

Bachem, A. (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychologica, 7*, 80-88.

Balch, W. R. (1984). The effects of auditory and visual interference on the immediate recall of melody. *Memory and Cognition, 12(6)*, 581-589.

Balch, W. R., & Muscatelli, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception and Psychophysics, 40(5)*, 351-358.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6(3)*, 501-515.

Bennett, P. J., & Banks, M. S. (1987). Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision. *Nature, 326(6116)*, 873-876.

Bennett, P. J., & Banks, M. S. (1991). The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vision Research, 31(10)*, 1759-1786.

Boltz, M. G., & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology, 18(4)*, 389-431.

Boltz, M. G., Marshburn, E., Jones, M. R., & Johnson, W. W. (1985). Serial-pattern structure and temporal-order recognition. *Perception and Psychophysics, 37(3)*, 209-217.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89(2)*, 244-249.

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54(2)*, 103-109.

Chang, H.-W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. J*ournal of Experimental Child Psychology, 24(2)*, 324-331.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7(4)*, 869-883.

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cogni-*

tion, 46(3), 373-382.

Davies, J. B., & Jennings, J. (1977). Reproduction of familiar melodies and perception of tonal sequences. *Journal of the Acoustical Society of America, 61(2)*, 534-541.

Démonet, J. F., Price, C. J., Wise, R., & Frackowiak, R. S. J. (1994). A pet study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. **Brain, 117(4)**, 671-682.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception and Psychophysics, 11(6)*, 411-412.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88(6)*, 503-522.

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

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

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

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego: Academic Press.

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics, 21(1)*, 60-64.

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

Eerola, T., & Bregman, M. (2007). Melodic and contextual similarity of folk song phrases. *Musicae Scientiae, Discussion Forum 4A-2007*, 211-233.

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and african folk melodies by western and african listeners. *Psychology of Music, 34*, 337-371.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception, 18*, 275-296.

Eerola, T., & Toiviainen, P. (2004). *Midi toolbox: Matlab tools for music research.* University of Jyväskylä: Kopijyvä, Jyväskylä, Finland. Available at http://www.jyu.fi/musica/miditoolbox/

Eiting, M. H. (1984). Perceptual similarities between musical motifs. *Music Perception, 2(1)*, 78-94.

Francès, R. (1988). *The perception of music*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology, 9*, S5-S22.

Freedman, E. G. (1999). The role of diatonicism in the abstraction and representation of contour and interval information. *Music Perception, 16(3)*, 365-387.

Friedmann, M. L. (1985). A methodology for the discussion of contour, its application to schoenberg music. J*ournal of Music Theory, 29(2)*, 223-248.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7(2)*, 155-170.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1998). Perception of mode, rhythm and contour in unfamiliar melodies: Effects of age and experience. *Music Perception, 15(4)*, 335-355.

Hermes, D. J. (1998a). Auditory and visual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research, 41(1)*, 63-72.

Hermes, D. J. (1998b). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research, 41(1)*, 73-82.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, 24(6), 551-565.

Jones, M. R., & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review, 96(3)*, 459-491.

Jones, M. R., & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory and Cognition, 19(1)*, 8-20.

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior & Development, 10(1)*, 49-59.

Kleiner, K. A., & Banks, M. S. (1987). Stimulus energy does not account for 2-month-olds' face preferences. *Journal of Experimental Psychology: Human Perception and Performance, 13(4)*, 594-600.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, England: Cambridge University Press.

Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception, 18(3)*, 245-274.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33(5)*, 1189-1207.

Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: M.I.T. Press.

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours - extensions of a theory for contour. *Journal of Music Theory, 31(2)*, 225-267.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science, 19(12)*, 1263-1271.

Messerli, P., Pegna, A., & Sordet, N. (1995). Hemispheric dominance for melody recognition in musicians and non-musicians. *Neuropsychologia, 33(4)*, 395-405.

Mikumo, M. (1994). Motor encoding strategy for pitches of melodies. *Music Perception, 12(2)*, 175-197.

Mikumo, M. (1997). Multi-encoding for pitch information of tone sequences. *Japanese Psychological Research*, 39(4), 300-311.

Miyazaki, K., & Rakowski, A. (2002). Recognition of notated melodies by possessors and nonpossessors of absolute pitch. *Perception and Psychophysics, 64(8)*, 1337-1345.

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

Morris, R. D. (1993). New directions in the theory and analysis of musical contour. *Music Theory Spectrum, 15(2)*, 205-228.

Morrongiello, B. A., & Roes, C. L. (1990). Developmental changes in children's perception of musical sequences: Effects of musical training. *Developmental Psychology, 26(5)*, 814-820.

Morrongiello, B. A., Trehub, S. E., Thorpe, L. A., & Capodilupo, S. (1985). Children's perception of melodies: The role of contour, frequency, and rate of presentation. *Journal of Experimental Child Psychology, 40(2)*, 279-292.

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

Ottman, R. W. (1986). *Music for sight-singing* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia, 30(3)*, 277-292.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport, 10(18)*, 3979-3984.

Pick, A. D., Palmer, C. F., Hennessy, B. L., & Unze, M. G. (1988). Children's perception of certain musical properties: Scale and contour. *Journal of Experimental Child Psychology, 45(1)*, 28.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge, MA: The MIT Press.

Pierrehumbert, J., & Hirschberg, J. (1990). *The meaning of intonation contours in interpretation of discourse*. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), Intentions in communication (pp. 271-311). Cambridge, MA: M.I.T. Press.

Polansky, L., & Bassein, R. (1992). Possible and impossible melody - some formal aspects of contour. *Journal of Music Theory, 36(2)*, 259-284.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception, 16(4)*, 439-456.

Ruckmick, C. C. (1929). A new classification of tonal qualities. *Psychological Review, 36*, 172-180.

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. Music Perception, 16(3), 295-326.

Schmuckler, M. A. (2004). *Pitch and pitch structures*. In J. Neuhoff (Ed.), Ecological psychoacoustics (pp. 271-315). San Diego, CA: Elsevier Science.

Schmuckler, M. A. (2008). Melodic contour similarity using folk melodies. Submitted.

Schoenberg, A. (1967). *Fundamentals of musical composition.* New York: St. Martins.

Schwarzer, G. (1993). Development of analytical and holistic processes in the categorization of melodies/entwicklung analytischer und holistischer prozesse bei der kategorisierung von melodien. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25(2)*, 89-103.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89(4)*, 305-333.

Shmulevich, I. (2004). A note on the pitch contour similarity index. *Journal of New Music Research, 33(1)*, 17-18.

Smey, D. (2007). *Sight-singing bonanza*. from http://davesmey.com/eartraining/sightsing.pdf

Thomassen, J. M. (1982). Melodic accent - experiments and a tentative model. *Journal of the Acoustical Society of America, 71(6)*, 1596-1605.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55(3)*, 821-830.

Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in musical sight reading: A study of pianists. *British Journal of Psychology*, 89(1), 123-149.

Watkins, A. J. (1985). Scale, key, and contour in the discrimination of tuned and mistuned approximations to melody. *Perception and Psychophysics*, 37(4), 275-285.

White, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology, 73*, 100-107.

Xu, Y. (2005). Speech melody as articulartorily implemented communicative functions. *Speech Communication, 46*, 220-251.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14(4)*, 1908-1919.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences of the United States of America, 95(6)*, 3172-3177.

Author Notes

Notes

Schmuckler (1999) used both difference scores and correlational measures for computing perceived similarity. Interestingly, in that work as well as subsequent research (Schmuckler, 2004), difference scores have proven to be somewhat more sensitive than correlations to perceived contour similarity. One possible reason for this finding is that outliers can greatly influence correlation values. Such extreme values occasionally occur with Fourier analysis information, in terms of the relative strengths of high frequency information, which typically tends to be quite low. Such outliers, then, would have a more dramatic effect on correlations than average difference scores, and could thus lead to somewhat distorted similarity predictions.