Music and Video Iconicity: Theory and Experimental Design

Roger A. Kendall

Music Cognition and Acoustics Laboratory, University of California, Los Angeles, USA

Abstract Experimental studies on the relationship between quasi-musical patterns and visual movement have largely focused on either referential, associative aspects or syntactical, accent-oriented alignments. Both of these are very important, however, between the referential and areferential lays a domain where visual pattern perceptually connects to musical pattern; this is iconicity. The temporal syntax of accent structures in iconicity is hypothesized to be important. Beyond that, a multidimensional visual space connects to musical patterning through mapping of visual time/space to musical time/magnitudes. Experimental visual and musical correlates are presented and comparisons to previous research provided. J Physiol Anthropol Appl Human Sci 24(1): 143–149, 2005 http://www.jstage.jst.go.jp/browse/jpa

Keywords: multimedia, animation, film, perception, cognition, music

Introduction

The connection of musical patterning to visual patterning is an essential and important area of study in the cognitive sciences. Multimedia, DVD video, video games, and television all incorporate interactions of musical and video elements. This study is designed to explore this important aspect of musical communication: How do musical and video elements combine in perceptually and culturally acceptable forms?

Related Literature

The relationship between music and video has been studied in a number of contexts, the most common of which is the assessment of meaning using semantic scales relative to relationships among musical and visual variables. Marshall and Cohen (1988) investigated the relationship between three animated shapes (a large and small triangle and a small circle) and subject ratings on semantic differentials as influenced by two contrasting examples of background music. Independent judgments of the animation and music were made, and then judgments of the combination of music and animation using twelve differentials. Although the activity dimension rating of the two pieces of music was the same, when combined with the animations the activity rating for the geometric ‘characters’ significantly varied. Marshall and Cohen (1988) proposed that there was, therefore, an interaction of visual and musical temporal structures such that visual attention was directed differently depending on the musical example with which it was paired. However it must be noted that technical analysis of the temporal structures of the audio and visual materials, and how specifically these might interact, was not undertaken.

Lipscomb and Kendall (1994) conducted an experiment to evaluate the degree to which the composer’s intent in film music composition was received by the viewer/listener. Five excerpts from Star Trek IV (music by Leonard Rosenman) were extracted. The soundtracks from the five excerpts were interchanged, leaving one match and four distracters. The composites were edited to maximize musical/visual accent alignment by graduate student film composers. The twenty-five combinations were rated on ten verbal attribute scales. Evaluative factor ratings were highest for the combinations that were originally intended. This confirmed data from an experiment where subjects were asked to match one of the five music selections with the film excerpts, a task which confirmed the ability of viewers who had not previously seen the film to determine the intended music. Lipscomb and Kendall (1994) hypothesized that viewers examine the accent relationship between visual and musical elements. If the accent structures align, and associative meaning is culturally viable, then perception and judgment result from the audio/visual composite rather than on each modality in isolation or in lieu of one another.

Further experiments in accent structure alignment between music and visual elements were carried out by Iwamiya (1994). Iwamiya (1994) offset the audio and visual channels of laserdisc video by 500 msec. He found that subjects rated temporally matched versions higher than the mismatched versions. He also added different background music to the originals. Since the audio/visual accent structures were incoherent, subjects rated the different music composites lower in matching than the original versions. Ratings of the audiovisual composites on 22 scales indicated the differential importance of audio visual elements depending on factor complexity: lower order factors influenced ratings across...
conditions, while higher order factors such as ‘uniqueness’ were influential directly when the music and visual materials were well matched. Lipscomb (in press) conducted an extensive investigation into the perception of visual and musical accent structures and their alignment.

Psychologists are interested in the magnitude and order of musical and visual variables in the ratings scale judgments mentioned above. However, associative meanings such as those produced by semantic differential ratings are not the only types of meanings induced by the film and animation experience. Dowling and Harwood (1986) juxtapose the theories of Meyer (1956) and Peirce (1931–1935) in their discussion of various types of musical meaning. The concept of ‘referentialism’ in Meyer (1956) largely is synonymous with Peirce’s ‘index.’ These meanings are the type of associations between musical and visual elements that are uncovered by semantic differential ratings methodologies such as those mentioned previously. An interesting, and central, point in Meyer’s (1956) theory of meaning is ‘expressionism.’ This type of meaning is embodied (contained within) the music itself. Kendall (1999, 2000) proposed that, instead of these categories, this implies a continuum of referentiality. At one end of the continuum are those meanings that are associative in a completely arbitrary way; the sign and signified are linked only by convention and not by any aspect of form or structure. At the other end of the continuum is the referential; the meaning embodied within the structure and patterning of the music or visual itself. In the middle between the arbitrary denotation of the referential and the meaning derived from pure syntactical patterning, is the connotative meaning of icon. Icons are connotative meanings in which the patterning or form in one domain or modality suggests or implies a connection to another domain. The common example in English is the phrase ‘weeping willow.’ This tree is named such because the branches droop, suggesting weeping (the tree obviously does not physically cry). Iconic meanings are between association/index and syntax on the continuum of referentiality. The continuum as presented here is unidimensional. It is quite clear that dimensional analysis of verbal responses indicates that associative meanings alone are themselves multidimensional. See Kendall (in press).

Iconic meanings in visual motion and music have as their basis syntax. An example in film and music is the Galley Rowing scene in Ben Hur (1959). The conveyed meaning is not just in the fact that the music pulse and tempo matches the rowing pace of the slaves, but that the melodic pitch ascends when the oars are brought up and in, and descends when the oars are brought down and outward. The combination of temporal accent alignment in the visual and musical structure with the connotation (suggestiveness) of musical variable magnitude change through time and visual change in space melds into an iconic composite.

Kendall (in press) conducted an experiment to see if subjects could rate audiovisual examples from commercial movies and broadcasts on a scale of referentiality. Six music and video combinations drawn from the commercial media were selected and hypothesized to span a continuum from referential (syntactical) through iconic to referential. The six examples were Shanghai Lil from Footlight Parade (movie, 1933), Cool from West Side Story (movie, 1961), Galley Rowing from Ben Hur (movie, 1959) Inside the Ship from Close Encounters of the Third Kind (1977), Paul Wyle Ice Skating (Olympic Broadcast, 1992), and Fantasia (animated film, 1940/1990). Six graduate students rated the randomly presented excerpts on a scale from referential (left side) through icon (middle) to index/referential (right side) Repeated measures ANOVA of the resulting data showed a significant main effect ($p<.0001$) across the six means. Post hoc analysis of paired mean differences showed that means were statistically different ($p<.05$) except for the two syntax/icon combinations, as hypothesized. The data thus demonstrates that trained musicians can use the continuum in a non-categorical manner.

Relevant to the present discussion are two important recent studies that approach elements of iconicity in visuals with music. Iwamiya and Ozaki (2004) studied the animation of three visual stimuli with combinations of seven musical pitch patterns, three of which involved two-line contrary motion. The visual stimuli periodically expanded and contracted at the center of the screen. They note, “In the present experiments, the shape of the image was monotonously enlarging or reducing. This type of visual monotonous transformation matched unidirectional and continuous descending or ascending pitch patterns.” The formal congruency they speak of is very close to the concept of iconicity outlined in Kendall (1999, 2000, in press). However, it is not at all clear whether the temporal periodicity of the pitch patterns and of the visual stimuli was always coordinated or whether there were out of phase conditions.

Lipscomb and Kim (2004) conducted experiments with a carefully nested set of visual and musical variables. Three pitch patterns with an ascending (3 notes) then descending (2 notes) pitch contour were created with three interval sizes, from small to large with a C3 tonic. Timbre patterns, attached to each of the unique pitch chroma, were drawn from perceptual scalings in Kendall et al. (1999) to produce three magnitudes of timbre change. Similarly, loudness contours and note length contours were produced. Visual magnitude contours of color, size, shape, and location along the y-axis of 2-D space were paired with the musical contour patterns. Results indicated that pitch contour was matched best with position along the y axis, a result that corresponds with Iwamiya and Ozaki (2004) and connects to the concept of iconicity outlined below. Loudness was matched with size, and timbre with shape (although Fig. 1, p. 74 shows location and loudness at relatively high match levels, and shape nearly equally matched with pitch, loudness, and timbre. No post hoc analysis was provided). These results largely match those found in Walker (1987).

Although the data are valuable, several issues may impede connection of this research to that outlined below. Firstly,
physical magnitudes of color, for example, were used based on wavelength. Color, like pitch, is multidimensional, and I will suggest that only perceptual magnitudes, expressed in musical terms, be used iconic studies. It is likely that the use of spectral wavelength resulted in a brightness contour that decreased as pitch increased, thus negating the iconic connection. Similarly, with timbre, the instruments employed cannot play at the tessitura indicated if indeed the tonic is C3. Thus the Kendall et al. (1999) data may not provide adequate information for timbre magnitudes, since the pitch chroma employed in that study was Bb4. Finally, the pitch pattern ascends and descends from the tonic, and the corresponding animation on the y-axis moves up and down around the origin. These minor details in an important study can be explored and clarified in future research.

**Experimental Design**

Musical icons emerge in multimedia from visual magnitude and music magnitude and direction change through time. In the projected experiments, the visual field is divided into temporal units on the x, y, and z space axes. Musical time is expressed by conventional notation that matches the temporal structure of the visual graphs.

A few iconic archetypes are illustrated in Figure 1.

The top row presents ascending and descending ramp. In visual space, this could be motion up or down on any two axes of 3-d space. In addition, a ramp can be increasing visual magnitude such as a zoom with increasing size along the z-axis. Musical ramps are common and exist in ascending and descending scalar pitch patterns and crescendi and decrescendi.

The burst, called in movies a stinger, is a sudden, temporally brief contrast of increasing magnitude in the visual (e.g. brightness, size, color) and musical (e.g. sforzando, cymbal crash, consonant-dissonant shift, key change) domains.

The second row presents the arch, which can be diagramed either in the semicircular form as illustrated above, or as the connection of two ramps in temporal contiguity.

The third row simply indicates that combinations of icons can be layered. In music, a perfect example is the start of Suite No. 2 in Ravel’s (1912) Daphnis et Chole. The woodwinds perform continuous arch patterns suggesting a brook while the strings and brass perform a series of loudness and pitch ramps, suggesting a gradual sunrise.

An important syntactical feature in musical/visual iconicity is congruence of contours in time. Changes of pitch direction in time, in the manner described in Dowling and Harwood...
(1986), are hypothesized to connect to changes of object direction in time. Below we consider some examples for use in a perceptual experiment.

**Stimuli**

The present experiment diagrams a number of visual animations using a circle and, in the future, a texture-mapped externally-lighted sphere with camera angle perspective (since camera perspective can create visual/temporal patterns in panning and zooming).

Figures 2–9 below are hypothesized to illustrate visual ramps and are hypothesized to be best matched with musical pattern changes that decrease in perceptual value at the same rate. Below the visual figures are musical patterns which are hypothesized to correlate with the visual animations. In the experiments, one variable is temporal rate, so the frame represents 2400 msec (M. M. = 100 in common time) or 2800 msec (M. M. = 70). Not shown here is the non-motion stopping point (at the half note musically) of 1200 or 1400 msec.

The first experimental procedure follows the related literature. All combinations of musical patterns and animations, at two tempos, will be combined. The audio-visual combinations will be rated by subjects for the degree of fit.

In the second experiment, the animations and music patterns that are best matches in experiment 1 will be rated for similarity in all possible pairs. The similarity matrix will be
subjected to multidimensional scaling, thus revealing the cognitive space for iconic combinations. Additional analysis should reveal the perceptual salience of different variables resulting in the composites.

References


Fig. 7  This arch starts at the center of the visual field and alternates above and below the center point. Similarly the pitch pattern arches around the tonic.

Fig. 8  This is an example of arches and ramp superposition. The animation rotates on the x-y axis and enlarges on the z-axis (not shown).

Fig. 9  These music patterns are included because they do not correspond syntactically and precisely to any of the visual patterns.
Received: October 30, 2004
Accepted: November 8, 2004
Correspondence to: Roger A. Kendall, Department of Ethnomusicology, Program in Systematic Musicology, University of California at Los Angeles, (UCLA), Box 951657, Los Angeles, CA 90095–1657, USA
e-mail: kendall@ucla.edu