Love in disguise: Incongruity between text and music in song

Hui-Chieh Hsu [a], Lily I-wen Su [b],*

[a] Department of International Business, National Taiwan University, Taiwan
[b] Graduate Institute of Linguistics, National Taiwan University, 1 Section 4, Roosevelt Road, Taipei, Taiwan

Abstract

Text and music in song are not necessarily congruous; they can be incongruous to each other and still convey meaning. Although the study of the relations between text and music in song has been prevalent throughout the history of Western music (Burkholder et al., 2010), few deal with the problem of text-music incongruity. Via the incorporation of conceptual blending (Fauconnier and Turner, 2002), Zbikowski (1999, 2002, 2009) is able to provide a more integrated account of the identity of song than earlier models, but the problem of text-music incongruity is still left unresolved. In the present research, therefore, through a modified blending model, we demonstrate with two songs as examples how incongruous text and music in song are able to bring about rich emergent meaning, which, though triggered by incongruity, is in fact iconic to the effect "intended" by the composers. With text-music incongruity, the composers are able to not only communicate meaning to the listener, but also bring out the aesthetics of their works.
© 2013 Elsevier B.V. All rights reserved.

Keywords: Conceptual blending; Incongruity; Multimodality; Iconicity; Song

> "Indeed, my poems lead . . . only half a life, a paper-life, black upon white . . . until music breathes the breath of life into them, or at least, when it slumbers within, calls it out and wakens it."
>
> —Wilhelm Müller[1]

1. Zerlina's dilemma

In the first act of Mozart's opera Don Giovanni, the playboy nobleman Don Giovanni encounters and is soon attracted to the newly-engaged peasant girl Zerlina. In order to win her heart, Don Giovanni starts his pursuit with an elegant tune. Concerned about her fiancé and uncertain about Don Giovanni's commitment, Zerlina expresses in words her ambivalence toward the courtship, but sings her hesitation to the exact tune Don Giovanni has just used to court her. In other words, despite the indeterminacy in her words, Zerlina's music already "betrays" her and gives in to Don Giovanni. As the duet unfolds, Zerlina, following her betraying music, indeed grows more and more inclined to accept the pursuit, before finally yielding completely to Don Giovanni in the end.

How is this possible? How are we able to sense Zerlina's "love in disguise" in the first place? How can meaning be emergent from the incongruous text and music?

Due to the similar yet dissimilar nature of language and music,[2] the riveting yet elusive relations between the two modes of communication have long intrigued people. They are similar in that they share many commonalities: both are

---

* Corresponding author. Tel.: +886 2 33664108; fax: +886 2 23635358.
  E-mail address: iwensu@ntu.edu.tw (L. Su).

[1] Quoted in Koch (1902). Though Müller uses in the original German text the word Lieder ("songs") to refer to his poems, the word "poems" is adopted here to avoid confusion.

[2] In the present study, the term "music" refers only to music that is purely instrumental.

unique to human, both make use of sound, both unfold over time, both convert complex acoustic sequences into perceptually discrete elements governed by hierarchical structures (Patel, 2008), and, most importantly, both convey rich meaning (Zbikowski, 2009). Language and music however differ in many aspects. For instance, the musical meaning and the linguistic meaning are not comparable in that music is usually said to have stronger affective power (Patel, 2008). Furthermore, simultaneous events like harmony are possible in music but not in language (Zbikowski, 2009). Language and music, the two similar and yet dissimilar mental capacities that are unique to the human beings, have drawn the attention of philosophers, linguists, musicologists, composers, poets, etc. ever since Plato's time (Patel, 2008). Song, as a combination of language and music, has naturally received much attention as well.

The study of the relations between language and music in song[3] has been prevalent throughout the history of Western music: Some argued for the dominance of music, while others spoke for the critical role of language (Arnold, 1957; Burkholder et al., 2010). In the last century, several theoretical frameworks for the analysis of song (Langer, 1953; Kramer, 1984; Agawu, 1992; Cook, 1998) have been proposed, but none is able to provide a satisfactory account of the identity of song,[4] let alone text-music incongruity. A more integrated account, via the incorporation of conceptual blending theory (Fauconnier and Turner, 2002), is advanced by Zbikowski (1999, 2002, 2009) to explain how language and music can be put together to create artistic works with rich emergent structure. Nevertheless, besides undermining the critical roles of cross-space vital relations and frame (Fillmore, 1985, 2006; Fillmore and Atkins, 1992), Zbikowski's analyses deal only with songs with congruous text[5] and music. Text-music incongruity is still left unaccounted for.

Similar to form-meaning mismatches in language, blending incongruous text and music is in fact feasible; were it not, our appreciation of songs like the Mozart duet above would not be possible. As Fauconnier and Turner (2002: 135–36) put it, "Some networks have what appear at first sight to be extremely tenuous cross-space mappings between the two frames organizing the inputs, but they lead ultimately to extraordinarily rich emergent meaning and global insight." Indeed, the blending of inputs that seem hardly related might be exactly where our imagination brings about the most fascinating results.

In their analysis of a Toblerone ad, Fauconnier and Turner (2002) illustrate how, with the help of emergent vital relations, it is cognitively possible to construct meaning by blending inputs with only tenuous or even contradicting relations. Toblerone is a chocolate brand selling the famous triangular chocolate candies that come in various sizes. In a Toblerone ad, the Egyptian pyramids of Giza—one large, one medium, and two small—are shown, together with a group of four Toblerone chocolates in the same configuration, though much smaller in size. Also shown are the captions "Ancient Tobleronism?" and "Toblerone: Inspires the World," the slogan of the brand. With an eureka in mind, we understand that the relations between the chocolate candies and the ancient pyramids are hardly accidental but are in fact causal: Toblerone have been there in the history all along, inspiring great human achievements like the pyramids of Giza, which were built in order to honor the triangular chocolate candies. In the conceptual integration network (Fauconnier and Turner, 2002; henceforth "CIN") for this ad, the primary emergent structure is the cross-space vital relations that are intuitively in the wrong direction. By constructing these seemingly mistaken vital relations of Causality, Intentionality, and Time, strong and essential connections between the two hardly related inputs are built, giving rise to extremely rich emergent structure and highly creative interpretations of the ad.

Another study, conducted by Teng and Sun, also shows how seemingly incompatible elements can be put together to bring forth rich emergent meaning. In the study, Teng and Sun (2002) propose a working hypothesis called the Image Grouping Hypothesis to systematically account for the underlying cognitive operations of pictorial grouping, pictorial simile, and pictorial oxymoron. According to the Hypothesis, when images are symmetrically aligned, the viewer of the juxtaposed images is invited to view them as belonging to the same category. If the aligned images are originally from the same category, a pictorial grouping is formed. If originally they belong to different categories, the alignment invites the viewer to view them as coming from the same category, and a pictorial simile established. If the juxtaposed images not only come from different categories but are incompatible or even contradictory to each other, a pictorial oxymoron is constructed and further invites the viewer to come up with novel and open-ended interpretations.

Though proposed for pictorial elements, the Hypothesis does provide an analogy for how text and music can be combined and give rise to song. According to the Hypothesis, images aligned in space invite the viewer to view them as belonging to the same category. Analogously, in a song, the simultaneously perceivable—or "temporally aligned"—text

---

[3] Although there are quite a few "songs" without music (e.g. Heine's Das Buch der Lider ["The Book of Songs"]) and "songs" without words (e. g. Felix Mendelssohn's Lider ohne Worte ["Songs Without Words"]), the present study deals only with songs where both text and music are present.

[4] By the term "the identity of song," we view song as an integrated and inseparable whole.

[5] The "text" is used in a global sense, referring to the entire global discourse structure. Different from the text, the "word" is limited to individual linguistic forms. According to Barthes (1964/1977), the text is not something one can hold in hands but something beyond the concrete form: It is filiation-independent, open, interactional, process-oriented, and plural. The term "text" should also be distinguished from the term "language," which denotes a mode of communication where the text is given rise to by words.