MUSIC, LANGUAGE AND

Introduction

Connections between music and language have been a perennial concern of scholars, poets, music, and literary theorists and musicians going back to antiquity. The basis of this interest lies in certain commonalities which are intuitively understood to lie at the heart of the two capacities but which become complex and problematic when one attempts to elucidate their precise nature. Confusion on the question of the interaction of music and language is not surprising since the underlying basis of music and language as independent objects has been poorly understood until recently. Nor is it surprising, given traditional disciplinary divisions, that scholars have tended to focus on developing descriptive and explanatory frameworks (see *DESCRIPTIVE, OBSERVATIONAL, AND EXPLANATORY ADEQUACY*) for each which take for granted their status as independent rather than common faculties. With the exception of Rousseau's (1997 [1763]) "Essai sur l'origine des langues," which claims an ancestral proto-language from which language and music both derive, most attempts at engaging the question have stressed that not only do music and language serve distinct ends, one mainly aesthetic the other mainly communicative, they access distinct underlying psychological means.

The umbrella of cognitive science has provided a context for a renewed discussion of some of these points of comparison. Perhaps most striking is the re-emergence of arguments for an evolutionary pre-cursor in the form of what Brown (2000) refers to as "musi-language" whose essential characteristics are identified in Mithen (2005) by the acronym "hmmmm" ("holistic, manipulative, multimodal, musical, and mimetic"). These approaches are somewhat controversial not only in their endorsement of what might be called a neo-Rousseauvian perspective but in their assumption of a significant overlap in some of the cognitive structures

which underlie both music and language. The best known explorations of the common ground, Leonard Bernstein's *The Unanswered Question*, Deryk Cooke's *The Language of Music*, and Joseph Swain's *Languages of Music* have not suggested any specific shared mechanisms. Rather they, and others, have applied certain aspects of the descriptive apparatus and general methodologies of linguistic theory to yield, as Brown notes, analogies between music and language, helpful and suggestive analogies to be sure, but which do not constitute arguments for a shared cognitive basis.

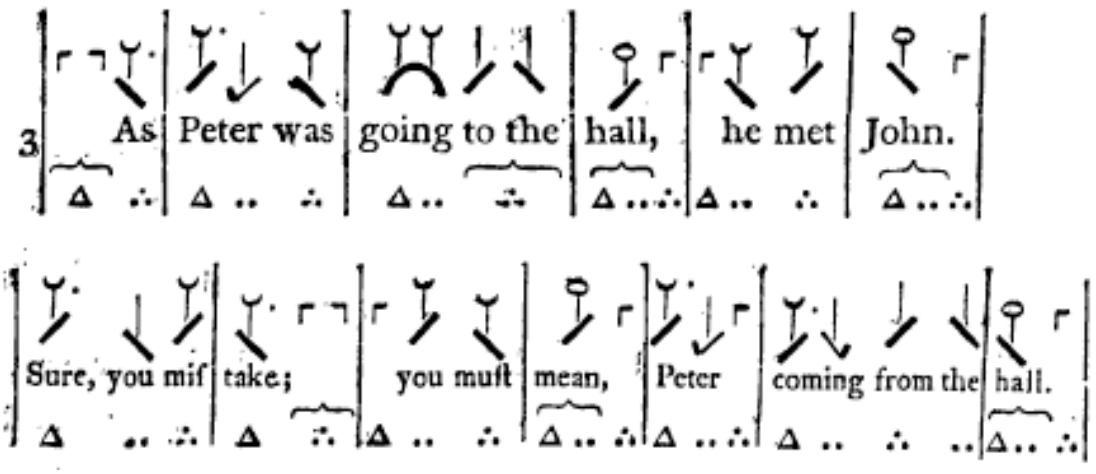
Musicalist Representation of Linguistic Structure

The studies just alluded to are representative of recent scholarship in that they attempt to ground subjective judgments with respect to musical structure on the hard foundation provided by linguistic science. While this has been the prevailing direction of influence, it has at times extended in the opposite direction. Most notably, musical notation for several centuries constituted the only effective means for visually representing the structure of audible sound. Among the acoustical phenomena rendered visible and thereby amenable to a structural analysis were speech sounds of English carefully transcribed into a modified form of musical notation by Joshua Steele in his *Essay towards Establishing the Melody and Measure of Speech* (Steele 1775).

In a recent review of the work, Jamie Kassler (2005) credits Steele as being among the first to identify linguistic suprasegmentals-the tier of linguistic structure computed independently of and mapped onto phonemic segments. *TONE*, the hierarchically related sequence of *PITCH* locations assigned to voiced segments is one such suprasegmental and is relatively naturally represented in musical notation. Steele also recognized that unlike musical pitch which tends to

be discrete, the target pitches of speech are consistently connected by continuous glissandi or slides, represented in Steele's scores by diagonal line segments of various types attached to note stems shown in the following:

Example 1.



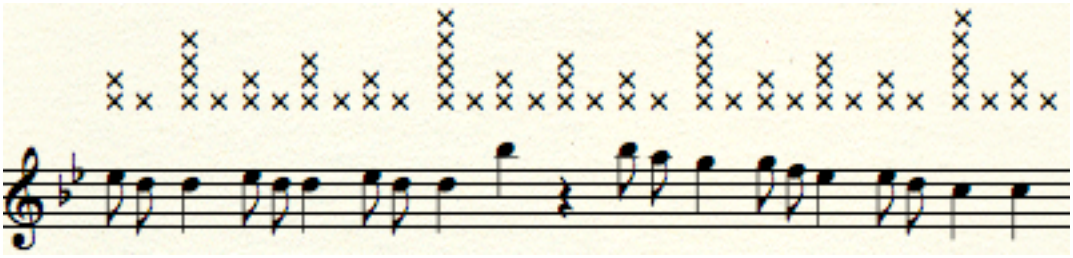
The other linguistic suprasegmental identified by Steele, *STRESS*, emerges somewhat obliquely from Steele's transcriptions. One of Steele's important insights was to have recognized that a particular type of musical accent, the metrical accent, is associated with linguistic stress. Thus, for example, the initial beat of the musical measure is metrically strong and the most stressed *SYLLABLES* of a text assigned to a tune, (generally the stressed syllables of polysyllabic words “Peter,” “going,” “mistake,” and “coming” in example 1) are assigned to what Steele refers to as the ictus position. Finally, and most significantly, Steele recognized that metrical accent is not an objective feature of the musical event but is a psychological attribute inherited from its temporal location. A strong position will be perceived as such regardless of whether the event occupying the position is objectively accented in the form of higher pitch, amplitude, or length. Indeed, it may be heard as strong even when it is vacant-occupied by a rest. Metrical accent is therefore, in Steele’s words, a subjective "mental sensation" deriving from "a sense of

pulsation . . . giv[ing] the mind an idea of emphasis . . . independent of any actual increment of sound or even of any sound at all" (p. 117). In recognizing the abstract character of *METER*, Steele anticipated twentieth century cognitivist approaches which view linguistic stress, along with most other salient characteristics of language, as mental constructs, *PHONOLOGICAL* rather than *PHONETIC*, psychologically real but only obliquely related to the acoustical or physiological surface form.

The Grid Representation

The "measure of speech" referred to in Steele's title—the patterned occurrence of strong and weak metrical positions- is represented in Steele's transcriptions by a three level hierarchy appearing below the staff in example 1: heavy, light and lightest locations within each measure are assigned a triangle, three and two dots respectively. This would be the first, and for many years one of the few attempts to make explicit the underlying form corresponding to how meter is mentally constructed by listeners. When this objective would be reinitiated in the 1970s, most notably within the generative theory of Ray Jackendoff and Fred Lerdahl (1983), the representation would take the form of the metrical grid shown in example 2.

Example 2: Mozart Symphony 40



It will be noticed that example 2 omits the conventional notational means for indicating the metrical hierarchy-barlines, beaming of eighth notes and the time signature. It can do so since

these are indicated with greater precision by the grid which identifies the relative prominence of particular locations by their inclusion at successive horizontal tiers referred to as higher levels of the grid. Relatively strong positions at the measure, half note, and quarter notes are represented by columns appearing above these locations while weak positions at the eighth note level appear only at the lowest level of the grid.

That metrical structure is a fundamental component of music—or, to put it informally, that music frequently "has a beat"—is, of course, self-evident to most listeners. That normal linguistic utterances are rhythmic in anything like the same sense is less apparent and remains a subject of some controversy within linguistics. For this reason, it might appear surprising that as phonologists confronted a range of data provided by a cross section of the world's languages, it became apparent that the same representation, namely the grid, would emerge the optimal means to represent linguistics stress. Indeed, metrical stress theory, the dominant explanatory framework within the generativist paradigm would be defined by the grid representation, one variant of which is shown in the following:

Example 3 (from Halle and Vergnaud 1987)

```
          *       level 3
          *       level 2
 * *      *       level 1
 * *  * * *       level 0
Ticonderoga
```

A comparison of the grids in examples 2 and 3 reveals two essential differences between linguistic and musical structure. First, while the stress grid projects syllables onto higher metrical levels, the bottom level of the musical grid indicates not actual musical events (i.e., notes) but rather temporal locations. As a consequence, empty metrical locations such as those in example 2 which are a necessary component of any reasonable description of musical meter are excluded from the stress grid. Secondly, as Jackendoff and Lerdahl (1983) show, musical structure

imposes strict requirements on the geometric form which grids may assume, limited to what they refer to as a small class of well formed structures. In contrast, there are no a priori constraints on the form taken by the stress grid. The successive positions projected onto line 1 of example **[3]** would be ruled out as a potential metrical structure in music where strong positions need to be to be separated by at least one position on level 0. This violation of musicalist well-formedness does not, however, prevent example 3 from accurately characterizing the pattern of secondary and primary stress for the word in question.

The asymmetries in the two forms of representation are, it would seem, necessary for a description of the output of each system: as mentioned, the projection of musical meter requires an underlying temporal periodicity which is neither intuitively obvious nor empirically demonstrable in language except as a statistical regularity (see Patel and Danielle 2003 for discussion). In addition, the asymmetry reflects essential differences in the character of the basic elements of musical versus linguistic structure. The assignment of stress is a formal computation effected on syllables from the rich *PHONEMIC* inventory of particular languages. In contrast, the computation of musical meter, what is known in the music perception literature as "beat induction," can be effected on a highly impoverished musical input. As has been shown repeatedly, a listener will unproblematically assign a metrical structure even when the events to which it is assigned appear as series of pitchless claps, clicks, or drum beats. The sorts of subtle variations in timbre and pitch characteristic of the phonemic repertoire may tip the balance between competing metrical interpretations when these appear in a musical context, however, they are in themselves insufficient for the inference of meter.

Rhythmic Structure in Language and Music

It is worth noting that the uncoupling of a musicalist interpretation from metrical grids in their application in most variants of metrical stress theory is in some respects inconsistent with the grid notation as it was proposed in work by Mark Liberman (1975) undertaken concurrently with Jackendoff and Lerdahl (1983). Here, the intention was explicitly musicalist, namely to relate the metrical structure of simple children's songs and chants to the syntactic and phonological structure of the words and phrases assigned to them. In contrast to most approaches, cognitivist and traditional, which view musical and linguistic computations as an independent and self contained, Liberman's objective was to establish an "equivalence" between the underlying representation of song and speech, a connection which was understood by Liberman to be "in some ways a very deep one" (p. 81). This hypothesis, whatever its ultimate heuristic or conceptual value, has not been influential within the field of linguistics or in music theory.

Two partial exceptions should, however, be mentioned. While word level stress provides evidence for the disassociation of metrical structure in language and music, higher levels of linguistic structure provides some evidence for a musicalist interpretation of linguistic performance. In particular, phrasal stress, unlike word stress, is not phonologically austere and requires for its computation in addition to *MORPHOLOGICAL, SYNTACTIC* and *PRAGMATIC* factors, the quasi-musical considerations of what is referred to as phrasal "euphony." Most conspicuous among these is the "stress clash" resulting from stressed syllables from two words appearing adjacent to each other within the same phrase:

Example 4

```
                x          L(2)
          x     x          L(1)
      x   x     x          L(0)
a)* Thir teen men
```

```
                x           L(2)
        x       x           L(1)
        x   x   x           L(0)
b)  Thirteen men
```

The unacceptable form in a) triggers the application of the *RHYTHM* rule which achieves euphony by retracting leftward the first of the two syllables involved in the stress clash to produce the acceptable form b). While not a validation of the musicalist view, the terminology which is adopted by linguists as well as the mechanisms by which this particular phenomenon is explained is suggestive of a shared basis underlying the computation of phrasal stress and the assignment of musical meter.

A second point of contact, as noted by Kassler (2005), is the musicalist interpretation of metrical structure incorporated into certain approaches to formal *PROSODY* and in its transformational variant generative metrics. The stated goal of these approaches is to define the abstract structure of lines of texts composed in a poetic meter. But even here the adoption of a musicalist view as representing a significant aspect of the relevant empirical domain remains controversial. While poets routinely invoke the "music of poetry" and the "rhythms of verse" and appeal to explicitly musical terms such as "phrasing," "staccato," "harmony," etc., it remains an open question whether poetic rhythm has any relationship to musical rhythm as musicians understand as term. The reaction to Derek Attridge's *Rhythms of English Poetry* (1982) sheds some light on these questions. Attridge posits a scansion which assigns syllables to alternating beat and offbeat positions (indicated by b and o, respectively) some of which can remain vacant, most notably at the end of lines (see *VERSELINES*). The general approach and the representation of "empty positions" in particular is criticized by Tarlinskaya (2002) for failing to recognize that "though music and language used to be intertwined, they parted ways long ago.

Musical meter and meter of verse texts cannot be equated; musical theories of meter need no resurrection" (p. 39).

Tarlinskaya probably represents a consensus position among scholars of metrics in doubting there is significant evidence for the influence of musicalist rhythm on the structure of literary verse. An offshoot of generative metrics has been able to avoid this problem by taking as its primary empirical domain texts which are unambiguously intended as functioning with a musical context, namely lyrics of familiar strophic song forms. The basis of this work is the recognition that average listeners encountering unfamiliar texts for a familiar strophic song will effect sometimes considerable modifications in the structure of the original to accommodate the text. Thus, as noted in Halle and Lerdahl (1993), those minimally competent in the relevant linguistic and musical idiom will delete three of the original ten notes of the song "The Drunken Sailor" when they encounter the seven syllable text "Keel haul him 'til he's sober" while augmenting the original melody with three additional notes when confronted with the thirteen syllables, "Scrape the hail off his chest with a hoop iron razor." These strikingly uniform intuitions constitute the core data of what Hayes (in press) designates as "the textsetting problem" for which Hayes proposes an *OPTIMALITY THEORETIC* solution. It remains to be seen whether this work will validate Lieberman's initial insight of the deep connection between linguistic and musical structure or whether the relevant intuitions will apply solely to the narrow artistic domain which with which these analyses are concerned.

Conclusion: An Internalist Perspective on Language and Music

One possible explanation for the discrepancies in the forms taken by linguistic and musical representations, even when they are superficially similar, is that this discontinuity simply

reflects the fact of the matter. That is, no significant overlap in the empirical domains of music and language exists beyond that both are, in an important sense, products of our minds and which make use of our psychological capacities for structuring the external world. But it should be recognized that however close or distant the ultimate relationship, granting a significant psychological basis to musical structure is itself testimony to the influence of linguistics, namely the recognition by modern linguists of the status of language as grounded not in the external reality of speech, its acoustic and physiological structure (e-language), but in the underlying psychological mechanisms which give rise to linguistic behavior (*I-LANGUAGE*).

In contrast, musical scholarship has remained largely a *STRUCTURALIST* enterprise, primarily devoted to describing the external tokens of music, most commonly musical scores. Approaches which take as primary the unconscious knowledge which listeners (and composers in their capacity as listeners) access in making sense of what they hear and compose are decidedly peripheral within the field. Consequently, confusions as to what a theory of music is a theory of more or less routinely arise. Insofar as traditional structuralist theories are seen as offering the only empirically sound and intellectually satisfying accounts of musical form then linguistic and musical scholarship, aside from occasional points of convergence are likely to continue on their separate paths. If, on the other hand, the posing of interesting questions and a viable theoretical framework relies crucially on viewing musical works as psychologically-based "natural objects" in the Chomskyan sense, in this respect what we understand about language has a great deal to offer our understanding of music.

--John Halle

Works Cited and Suggestions for Further Reading

Attridge, Derek. 1982. *The Rhythms of English Poetry*. New York: Longman.

Bernstein, Leonard. 1976. *The Unanswered Question: Six Talks at Harvard*. Cambridge, MA: Harvard University Press.

Brown, Stephen. 2000. "The 'musilanguage' model of music evolution." In *The Origins of Music,* ed. N. L. Wallin, B. Merker, and S. Brown, 271-300. Cambridge, MA: MIT Press.

Cooke, Deryk. 1959. *The Language of Music*. Oxford: Clarendon Press.

Halle, John, and Fred Lerdahl. 1993. "A Generative Textsetting Model." *Current Musicology* 55: 3–26.

Halle, Morris, and Jean-Roget Vergnaud. 1987. *An Essay on Stress*. Cambridge, MA: MIT Press.

Hayes, Bruce. In press. "Textsetting as Constraint Conflict," Proceedings of the April 2005 Paris conference on the Typology of Poetic Forms, ed. Jean-Louis Aroui.

Jackendoff, Ray, and Fred Lerdahl. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Kassler, Jamie. 2005. "Representing Speech Through Musical Notation." *Journal of Musicological Research* 24: 227-39.

Liberman, Mark. 1975. *The Intonational System of English*. PhD diss, MIT. Reprinted 1979 by Garland, New York and London.

Mithen, Stephen. 2005. *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld and Nicolson.

Patel, A. D., and J. R. Daniele. 2003. "An empirical comparison of rhythm in language and music." *Cognition* 87: B35-B45

Rousseau, Jean-Jacques. 1997 [1763]. *Essai sur l'origine des langues*. Reproduction of the Neuchâtel manuscript. Paris: H. Champion.

Steele, Joshua. 1775. *An Essay Towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols.* London: J. Almon.

Swain, Joseph. 1997. *Musical Languages*. New York: Norton.

Tarlinskaya, **Marina**. 2002. "Verse Text: Its Meter and Its Oral Rendition." In *Meter, Rhythm, and Performance*, ed. C. Kueper, 39-55. Frankfurt am Main: Peter Lang.