Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music

Geraint A. Wiggins
Intelligent Sound and Music Systems Group, Centre for Cognition, Computation and Culture
Goldsmiths, University of London, New Cross, London SE14 5SG, UK; g.wiggins@gold.ac.uk

***Abstract*—We argue that it is time to re-evaluate the MIR community's approach to building artificial systems which operate over music. We suggest that it is fundamentally problematic to view music simply as data representing audio signals, and that the notion of the so-called "semantic gap" is misleading. We propose a philosophical framework within which scientific and/or technological study of music can be carried out, free from such artificial constructions. Ultimately, we argue that *Music* (as opposed to sound) can be studied only in a context which explicitly allows for, and is built on, (albeit *de facto*) models of human perception; to do otherwise is not to study *Music* at all.**

I. Introduction

Since ISMIR 2000, the first symposium on music information retrieval (MIR), MIR technology has moved on a long way. Signal processing work now reports a success rate of around 70% in determining the "musical" content of audio signals. Stated thus, there is a positive implication; but the glass is actually 30% empty. These techniques do seem to be limited to success rates around this figure—so clearly so that music information retrieval (MIR) researchers have begun to write about a *glass ceiling* [1] or a *semantic gap* [2].

There can be two responses to this situation, where a research question cannot be fully answered. One is to keep trying. The other is to ask, "Was it the right question?" This latter is the purpose of this paper. In exploring this possibility, we will look at the literature from musicology and psychology, and make argument that *Music*, as opposed to *Sound*, cannot be effectively studied from the standpoint of pure audio engineering—nor, indeed, from that of pure music theory. We argue that the cognitive mechanisms involved in human music perception and cognition *must* be taken into consideration, for a realistic account to be given. This paper is not the first in which this latter point has been made [e.g., 3, 4]; however, we take the argument somewhat further than it has been taken previously, and argue that the *starting point* for all music information retrieval (MIR) research needs to be perception and cognition, and particularly *musical memory*, for it is they that *define* Music. Otherwise, the enterprise is forever doomed to the inadequate success rate of 70%.

In this paper, we explore the philosophical grounding of studies in MIR, with a view to pinning down exactly what it is that MIR research is studying. We will conclude that a change of approach is necessary if we are to understand the very deep problems inherent in what is all too often simplistically assumed to be an easy task. To make this point, however, some other questions must be asked and answered *en route*, and this is the purpose of this paper.

One good heuristic to use, when beginning to ask *what* something is, is first to locate *where* it is. By doing this, we can make sure we are asking our question about the right thing. Before we can do even this, however, we need to say what we mean by the word "Music".

II. What does "Music" mean?

The word "music" is used in many different ways. Some of these are entirely metaphorical, and therefore not relevant here, but nevertheless there are sufficiently many different, directly referential usages for meanings to be confused. For example, a band member may ask her colleague to pass the "music", meaning the physical paper score or instrumental part, from which she is to play; engineers write about "music" processing, meaning the manipulation of audio signals which are generated by musical performance; teenagers are proud of their "music" collection, which is actually a quantity of CDs, tapes and/or MP3 files; and academics discuss "music" analysis, meaning (among other things) the prediction and attempted explication of the relationships between structures notated or (more often) implicit in scores, their composers' motivations, and their effects on listeners.

The composer Milton Babbitt [5] proposed a trinity of psychological music representations divided into categories based on the kind of external representation they are derived from: the *acoustic* (or physical), the *auditory* (or perceived) or the *graphemic* (or notated) *domain* (we must, in the current times, include representations such as the digital encoding used in CDs and mp3 files in the graphemic domain). None of Babbitt's domains is presented as the definitive referent of the word "Music". We follow this view here, taking the philosophical stance that the mysterious thing that is Music is an abstract and intangible cultural construct, which does not have physical existence in itself, but which is *described* by all three of these representational domains; in a sense, something like the notion of the Platonic ideal [6], but without actually existing in the real world. Given this, it is reasonable to think of all of these domain-based representations as being *aspects* of Music, but none of them *is* Music, individually. To emphasise, note that there is no single audio representation of any piece of music performed to multiple listeners: even if a piece is performed only once, listeners in different parts of the auditorium will experience subtly different acoustic effects.

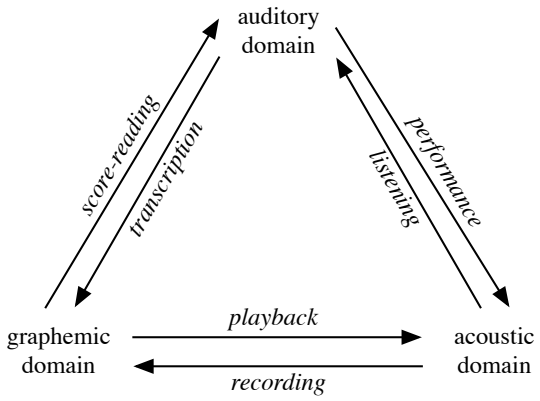

Fig. 1. Babbitt's trinity of representational domains, with our transformations between them.

Therefore, we must consider all performances of a piece as part of its abstract definition, which takes this definition clearly into the theoretical, because we cannot ever do this in practice. However, for our argument here, this presents no difficulties.

Fig. 1 illustrates the relationships between the domains; Music, itself, lies invisibly and intangibly somewhere in the middle, referred to and described by all the domains, but not actually *being* any of them. In quasi-mathematical terms, it is perhaps helpful to think of each of the domains as a projection of the "whole Musical object", which is itself not directly available. To experience (or study) music, then, is to experience (or study) one or more *aspects* of this "whole Musical object", and perhaps the interactions between those aspects, as they become perceptible in the world. It quickly becomes clear that the particular lens through which an individual chooses to view Music focuses on particular qualities: for example, the musicology of the Romantic period [e.g., 7] is very heavily based on the Western score, and this evidently informs and affects its precepts, with knock-on effects—not always positive—in its conclusions.

Our starting position, then, is that Music has no definitive existence in the physical world. This position is probably uncontroversial; however, we strengthen it and say that Music has no existence in the physical world *at all*. Rather, it is fundamentally an invisible, intangible entity, which leaves *traces* in the real world (audio signals and notation), much as light cannot itself be seen, but leaves traces everywhere around us as it rebounds from objects. This position, in turn, raises questions about the study of music. In particular, is it adequate to propose a study of music which is based mostly, if not exclusively, on the study of a restricted subset of the traces left by music in the world, rather like studying the dynamics of ice floes by looking only above the water-line? We examine this question in the next section.

III. Where and What is Music?

While it may not be possible to specify exactly what "Music" means, since we can only pin it down by saying "it is what all these other things describe", we can perhaps identify a preferred viewpoint from which to study it, with the help of Babbitt's trinity of domains. One criterion by which to make such an identification is the amount of information we can gain from the various available viewpoints. This is essentially the same reasoning that identifies *expressive completeness* and *structural generality* as desirable features of music representation systems [8].

Recently, the audio engineering community has coined the term "the semantic gap" [e.g., 2], to refer to perceived musical information which, though its existence is agreed by listeners, stubbornly refuses to be extracted from audio signals in isolation; the same community has identified what it calls a "glass ceiling" at about 70% accuracy in the results it can achieve by audio processing alone [1, 9]. One reason for this could be that the techniques applied are just not good enough. The other (which, along with Aucouturier [4], we suggest is actually the case) is that the information sought *is not in the audio signal*. We suggest, therefore, that a reasonable explanation of the experience of the signal processors is that the audio signal is *less than the whole Music*, and we will return below to the reasons why this might be the case. As such, the perceived existence of the "semantic gap" is evidence that the acoustic domain is not an entirely adequate description of Music—perhaps contrary to naïve expectations.

Babbitt [5] helps us here, by giving two alternative places to look: the graphemic domain, and the auditory. Since most people cannot hear a score, but can nevertheless listen to and imagine music, and since Music existed long before notation was invented, it follows that Music does not primarily reside in notation. What is more, it is evident that a musical score conveys less information about a piece of music (it is much less expressively complete) than does (is) an audio signal, despite the fact that it contains more information about the structure of the music (and is therefore more structurally general). We can therefore label the graphemic domain as an unlikely primary residence of the essence of Music.

Therefore, we can only argue that the place where Babbit's three domains come together is really in the auditory domain: in the human mind. This is borne out by the transformations between the domains (added by the current author) in Fig. 1, where the processes converting directly between the graphemic and the acoustic domains are the only ones *not* involving brains (at least, not beyond pushing the "start" button). Our suggestion, therefore, is to take Babbitt's taxonomy quite literally: as Babbitt proposed, what happens on the outside of an ear should strictly be called "acoustic" (audio signals); what happens on paper, or as notation carried by other media, should strictly be described as "graphemic" (symbolic representations); and what is, necessarily, everything else should be described as being firmly in Babbit's "auditory" domain—that is, the *psychophysiological effect* of a stimulus, audio or notated, which is the human response to and/or memory of Music. It is to be noted that this domain is often missing altogether from MIR research papers, except perhaps for suggestions for future work.

For completeness, one must also ask whether there are other domains, outside Babbitt's scheme, where the centre of Music might reside. One proposal might be a "compositional" domain, which describes the intent of composers and songwriters as they construct pieces of music. However, since these people are a minority of listeners to music, it is difficult to support the claim that this domain is primary to what is ultimately a shared

cultural experience (albeit increasingly a mediated one). This special example aside, it is hard to argue that what is missing from audio signals can be found anywhere other than in the auditory, because there is simply nowhere else for it to be found. The very credible argument that music is an embodied aspect of cognition, supported by evidence from linguistics and neuroscience, does not undermine this, because, where intention is involved, brains are at the centre of embodiment.

Evidence to support our view, and to show what can be learned from it, is easily found. For example, the difficult problem of F0 estimation can be made easier if a filter model based on human pitch perception is used [10]; Large [11] proposes a more detailed and exact auditory model that has the capacity to make even better judgements, because it is based not on filters, but on resonating oscillators—*like the mammalian ear*. But why should a biologically-inspired method work best? Because *fusion* in the mammalian auditory system has evolved to group harmonic oscillations and to associate them together into a single cognitive object, the selection pressure to do so being that in this way acoustic phenomena can be efficiently associated with the physical objects making the noise (which may be either prey or predator), and because Music (as opposed to sound) has co-evolved to match the human ear. Indeed, musical instruments have been progressively redesigned to optimise (with respect to particular aesthetic objective functions) their acoustic/auditory relationship with this system over the past few hundred years. Interestingly, with respect to this particular example, musical instrument builders have for centuries understood that the perceptually important harmonics in locating F0 are in fact F1, F2 and F4: this fact is used to give the impression that impossibly large organs had been built into small churches, implying very low frequency bass F0s, by using appropriately tuned, paired pipes pitched at F1 and F2 and sometimes including higher harmonics. Rameau [12] applied the same principles to harmony, accounting for the function of the major triad in terms of its component notes as F3, F4 and F5 of an implied bass, 1 or 2 octaves lower.

Large's hearing model is particularly effective because phase matters: mammalian ears have evolved to transmit information about phase synchrony between harmonics, as well as pitch, to the brain [13]; and Cariani [14] gives a plausible mechanism of harmonic fusion at the neuronal level, again not based on passive filter banks. It is unlikely that such complexity would evolve unless it were important. However, many (if not most) computational attempts at F0 estimation discard phase information entirely.

Similarly, there is clear evidence that patterns (and therefore styles) in music are *learned*, that the learned model is used to generate expectations, which are implicated in the experience of listening to music, and that all this can be modelled [15, 16] and used for music-analytical purposes [17]. However, most F0 estimation methods work locally, with no prior estimation of what note(s) to expect next.

The fact that several key tenets of music: the *sensation* of pitch, the *perceived* function of harmony, and the *expectation* of musical structure, are tied together in this intimate way, lends some force to the argument that it is perception and not sound *per se* that contains the essence of Music. Specifically, it explains *why* notes are harmonic series; it explains *why* tonal harmony is an attractor in the search space of possible musics; it explains *why* there is musical tension, and so on—and without a scientific explanation of *why*, any scientific account of any phenomenon is incomplete.

Thus, we challenge the common, but naïve, supposition, implicit in philosophies based exclusively in the acoustic domain, that Music is *merely* organised sound. It is certainly true that Music *is* organised sound: there is a journal of that name to support the assertion. However, the epithet hides a massive potential presumption: that *any organisation will do*. The work of the 20th Century experimental modernists—for example, Schoenberg's *Tonfarbenmelodie* [18]—demonstrate that there are some sonic things that simply *do not work* in a way that listeners of the time or, indeed, composers, in that particular example—expect from what they call Music. It follows that only certain kinds of organisation are adequate to constitute Music (though it is fundamentally important to note that *which* kind is determined by cultural milieu and date, as well as by perceptual primitives, not by some Romantic absolute notion of greatness, like that proposed by [7]—this point follows naturally from our argument here). John Cage went to great lengths to demonstrate that music can be *found* in appropriately presented randomness [19]; his work demonstrates very clearly what some of the requirements are, but also underlines that the music experienced by a listener *can only* be a perceptual or cognitive construct of their own making, for the notes chosen by the composer were not *deliberately* chosen by the composer, nor by anyone else.

What, in turn, decides what are the appropriate kinds of organisation? There are only two positions that can answer this question: either music is indeed a Platonic structure, whose nature is mysteriously externally determined simply to be the way it is; or it is entirely a sociocultural construct, and as such, entirely determined by the people who make it and listen to it. It is very hard indeed to defend the former of these alternatives without recourse to superstition or at best pseudoscience, especially since the latter alternative constitues a complete and quite straightforward (though complex) explanation, thus satisfying Ockham's Razor.

It is still worth recalling that there is a larger meaning of "Music", arising from all Babbitt's domains (and more) combined in a diachronic, historical context; but the fundamental *source* of Music—that without which Music cannot exist—is *in the mind*: if no mind is involved, there is no Music, but only sound. In this view of the world, both audio signal and notation are stimulus and/or result, applied to or derived from a cognitive process, depending on the current activity. The final, clinching argument in this discourse is the fact that Music can exist *without* sound, in the imagination of a musical human.

In the final analysis, then, Music, as opposed to sound, can exist without either or both of the acoustic or the graphemic domains, but it cannot exist without the auditory, and so that is where its primary residence must be. To study the acoustic domain, or the graphemic, in isolation is therefore to ignore the root of the problem; these domains are the tips of the metaphorical icebergs, and ignoring the mass of auditory ice under the water can only lead to shipwreck.

Having located Music firmly in the mind, we can now move to the edge, and stare into the "semantic gap".

IV. Where, if anywhere, and what, if anything, is the "semantic gap"?

Our title polemically questions the term and notion of "semantic gap". This should not be taken to imply that there is nothing to be identified in the metaphorical place to which Celma and Serra [2] and others refer by this term. And there can be no doubt that some of what is there is related to what is experienced as musical meaning. However, whether music can really be said to have "semantics" is an on-going debate [e.g., 20]. There are many different interpretations of the word, and that makes the concept slippery. One convincing and specific definition, grounding semantics in perception and then extending it to the more familiar referential, truth-functional semantics of language, is given by Gärdenfors [21], who argues that other common technical usages (e.g., as in "the Semantic Web") are substantively incorrect [22]. Like Gärdenfors, we take the position that meaning is generated only by and in minds, and therefore anything representing that meaning in a non-mental inference system (e.g., the propositional calculus) must by definition be metaphorical; only the mental meaning is actually semantics.

As a result of deciding that Music resides in the mind, a whole new literature of cognitive science becomes relevant to our question. One significant set of results, from neuroscience [23], suggests that some brain processes associated with the processing of meaning in language are active in the same way when visually presented words are primed with sentences, as when they are primed with musical phrases—musical structures seem to be communicating real-world meanings, and, notwithstanding the fact that perceptual metaphor or analogy is evidently involved (for example, a staircase seems to be associated with a rising musical figure), when one studies the stimuli, the connections are persuasive; and they match with Gärdenfors' notion of perceptual semantics. Based on these ideas, Aucouturier [4] presents an elegant set of empirical studies around a computational model of word grounding, in the context of musical genre, and makes some of the same arguments that are made here, from a practical perspective.

Both of these pieces of work support the idea that there are agreed meanings of words and (linguistic) phrases which are consensually associated with listeners' experience of music, in ways that may be analogous to symbol-grounding on real-world objects. However, in Koelsch's work, the music is isolated motifs, and the primes are perceptually objective; in Aucouturier's, the stimuli are entire popular songs and the terminology is abstract and socially determined; this makes it is difficult to combine the two results in conclusions. This dichotomy shows how the apparent analogy with symbol-grounding may be deceptive. When we say "table" and point to a table, thus grounding the term, we have substantial, specific and stable information about what is perceived by our collocutor as the object of that grounding process; the same is true of a short, isolated musical motif or short chord sequence. But the same is not true of a whole song from which we cannot point out particular sub-structures, without recourse to specialist terminology unavailable to most listeners. So to motivate the attempt to apply linguistic labels to entire pieces by reference to the neuroscientific evidence is suspect: either the labels must be summary terms, and therefore restricted in precision, or the music must be extremely uniform. This is the difficulty with genre classification, which renders that task pointless in the majority of contexts, for reasons rehearsed elsewhere [e.g., 24]. What is more, these terms cannot be said to be semantics in any realistic sense: they are a very long way indeed from a precise description of the state of meaning induced by the music in the listener. What is relevant here is the effect experienced by listeners (that is, the semantics of the music, in Gärdenfors' terms; e.g., harmonic function for Steedman [30]), and not the words used to describe it, which are separated from the effects by at least one level of indexing.

Another reason that "*semantic* gap" is a misleading term is that there is plenty of musical *syntax* that is perceived and fundamentally relevant to the experience of listening, but not explicit in an audio signal (though it is sometimes explicit in the corresponding score). The most obvious example is that of repetitive structure. While actual repetitions are objectively evident in an audio signal (even if reliable automatic detection is elusive), the relationship between them in time (mediated by memory, to which we return below) and the effect this has on cognitive response to the music (for example, recognition of the return of a theme after an extended period of musical development, which engenders a hard-to-explain feeling of satisfaction) are only implicit, and must be *induced* by the listener from his/her own remembered experience. The more general (and harder) case, where the repetitions are not exact, has spawned an extensive literature [e.g., 25, 26, 27], which is mostly focused on detection of the structures involved, rather than explication of the listener's resulting experience; this is appropriate, methodologically, since the objective aspects need to be understood before the subjective ones [28].

Another example, which makes us look at the music from a different angle, is harmony. Harmonic function is often reduced, in MIR, to identification and comparison of chords. This is a gross oversimplification: harmonic function is a perceptual construct which arises from a conjunction of stylistic expectation, extended local context, and, only finally, the current chord. It is possible—and entirely reasonable, in an engineering context—to finesse this kind of issue, by techniques which equate to approximate matching, such as the chord distribution comparisons of [29]; but in doing so, we lose what may be crucial information, such as the canon of stylistically acceptable Jazz chord substitutions [30].

Both these examples are problematic for purely audio-based approaches because they explicitly require knowledge which is not explicit in the signal, some of which is only available in each particular listener's mind at a level considerably more discrete than the acoustic; they are sometimes referred to as using "mid-level representations". Pampalk et al. [3] suggest that experiments involving human subjects are needed in the evaluation of MIR research, and they are right. But this is not the whole story, and from the viewpoint of the auditory domain, it is easy to see why: to be 100% successful, an MIR system would need to directly encode and use knowledge not just about music, but about the listener him/herself. For a generic MIR system, then, the goal of 100% correctness (we return to the question of what that means, below), is in most cases quixotic, and possibly even undesirable [4].

V. Memory and Musical Similarity: The Truth is not on the Ground

This brings us to the central key issue in MIR: that of musical similarity; any activity in MIR which is based on musical content (as opposed to meta-data) must address it. Tversky [31] notes that judgements of similarity between sequentially-presented stimuli cannot be symmetrical with respect to presentation order, because of the effect of *priming*, in which the act of perceiving the first stimulus directly but non-consciously affects the perception of the second. In conjunction with our argument that musical culture is learned and not innate, it follows that any judgement of similarity is affected by the musical culture of the judge—right down to personal musical choice. This and related arguments have been rehearsed elsewhere [e.g., 32, 33, 4]; ultimately, it follows that MIR systems based on absolute encoding of standardised musical properties will need the ability to allow for listener variation, either by direct familiarisation, or by statistical approximation. This leads us to the question of evaluation, and the misleading notion of *ground truth*, which is borrowed from the geographical sciences. In that context, it refers to the existence and inter-relation of substantive, real-world structures: the "ground" is thought of as immovable. In geography, this is probably reasonable, but in MIR there is no ground. Because music, as a cultural construct, is defined by the collective action of minds over time, it is not constant. Therefore, we cannot think of music, or MIR, as having a "ground truth". Every question is context-dependent, and every musical process is derived from cognitive structures, which are simply not accessible to scrutiny. This, in turn, means not only that MIR researchers need to look to psychology for methods of evaluating the quality of their systems [4], but also that *the cognitive nature of the phenomena those systems aim to capture needs to be taken into account*.

VI. Representation of Musical Syntax and Semantics

This last point is not just a naïve restatement of the argument that we might as well look to biological systems for inspiration, because they already work; it runs much deeper. Music is not just *processed* by these biological systems: it is *defined* by them. Therefore, any system that deals with Music effectively is *de facto* a cognitive model (even if a "black box"), because Music is *fundamentally cognitive*; and by the same token, *only cognitive models are likely to succeed* in processing Music in a human-like way. To treat Music in a way which is not human-like is *meaningless*, because Music is *defined by humans*.

Therefore, we do not need a new term to describe the metaphorical place where the illusory "semantic gap" is not: musicologists have been exploring it for decades [7], so have "symbolic" AI researchers [8, 34] and music-cognitive scientists [23]. Working from scores and other discrete representations of music to elucidate the structure experienced in it is an example; we must explore the concepts carefully determined and labelled in music theory over the past 200 years: the syntactic elements of music which, perhaps compositionally, but at least combined, give rise to musical experience. It is an indication of the prevalent poverty of understanding of Music in the MIR community that symbolic representations based on stable perceptual notions (e.g., "note" or "chord") are often dismissed as unrealistic simplifications of the "real" (audio) problem. This is a damaging and factually incorrect view, that should be questioned at every encounter.

What is more, cognitively- and perceptually-motivated intermediate (mid-level) representations allow us to boot-strap the process of automating music understanding. Human listeners (the definers of music) experience notes and chords, and a very substantial part of the process of music listening is predicated on their perception [16]. Thus, research on methods starting at, for example, the *musical surface* [35] of notes is a useful way of proceeding in parallel, cutting to the musical chase, while the audio engineering gets to the level required to begin to match human competence at hearing.

VII. Conclusion

In summary, the "semantic gap", viewed from the perspective of the audio domain, is a chasm which must be bridged, and for which tools do not exist. Viewed from the auditory domain, however, no "gap" is visible: there is instead a discrete spectrum of structure, theoretically explicable in psychological and musicological terms, which is realised in or stimulated by an audio signal. There is evidence that *some* of the information in that spectrum is cognitively processed in a way related to (language) semantics, but a significant amount of it is essentially syntactic: so "semantic" only covers part of the story. Therefore, viewed from the proper perspective, there is no "gap", except, perhaps, in the bibliography of papers on the subject. And, even if there were a gap, "semantic" would not be the right word to describe it, because the transition from acoustics to semantics relies on *syntactic* mechanisms considerably more subtle than the broad judgements involved in the straw man of genre classification, and these constitute a significant part of the necessary mechanism.

Because music is first and foremost a psychological construct, there can be no externally defined truth, and systems which aim to encode musical similarity must, by definition, do so in a human-like way. Therefore, technology and evaluation methodology must be imported from cognitive science to allow MIR to proceed in a meaningful scientific direction [24].

It is time for a change. No matter how unsinkable the Titanic of audio-only MIR becomes, the auditory iceberg will be still waiting, 30% below the waterline.

References

[1] J.-J. Aucouturier and F. Pachet, "Improving timbre similarity: How high's the sky?" *Journal of Negative Results in Speech and Audio Sciences*, vol. 1, no. 1, 2004.

[2] Ò. Celma and X. Serra, "FOAFing the music: Bridging the semantic gap in music recommendation," *Journal of Web Semantics*, vol. 6, no. 4, 2008. [Online]. Available: doi:10.1016/j.websem.2008.09.004

[3] E. Pampalk, A. Flexer, and W. G., "Improvements of audio-based music similarity and genre classifcation," in *Proceedings of the 6th International Symposium on Music Information Retrieval (ISMIR 2005)*, J. D. Reiss and G. A. Wiggins, Eds., 2005. [Online]. Available: www.ismir.net

[4] J.-J. Aucouturier, "Sounds like teen spirit: Computational insights into the grounding of everyday musical terms," in *Language, Evolution and the Brain*, ser. Frontiers in Linguistics, J. Minett and W. Wang, Eds. Taipei: Academia Sinica Press, 2009.

[5] M. Babbitt, "The use of computers in musicological research," *Perspectives of New Music*, vol. 3, no. 2, pp. 74–83, 1965, available at `http://www.jstor.org/`.

[6] Plato, *The Republic* [Original 385BC, translated by J. L. Davies and D. J. Vaughan]. Ware, Hertfordshire, UK: Wordsworth Editions Ltd., 1997.

[7] H. Schenker, *Beethoven's ninth symphony : a portrayal of its musical content, with running commentary on performance and literature as well*. New Haven: Yale University Press, 1992, edited by John Rothgeb.

[8] G. A. Wiggins, E. Miranda, A. Smaill, and M. Harris, "A framework for the evaluation of music representation systems," *Computer Music Journal*, vol. 17, no. 3, pp. 31–42, 1993, machine Tongues series, number XVII.

[9] P. Herrera, J. Bello, G. Widmer, M. Sandler, Ò. Celma, F. Vignoli, E. Pampalk, P. Cano, S. Pauws, and X. Serra, "SIMAC: Semantic interaction with music audio contents," in *Proceedings of 2nd European Workshop on Integration of Knowledge, Semantic and Digital Media Technologies*, 2005.

[10] A. Klapuri, "A perceptually motivated multiple-f0 estimation method," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2005. [Online]. Available: http://www.cs.tut.fi/sgn/arg/klap/waspaa2005.pdf

[11] E. W. Large, "A generic nonlinear model for auditory perception," in *Auditory Mechanisms: Processes and Models*, A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer, Eds. Singapore: World Scientific, 2006, pp. 516–517.

[12] J.-P. Rameau, *Traité de lharmonie réduite à ses principes naturels*. Paris: Ballard, 1722.

[13] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 2nd ed. London: Academic Press, 1982.

[14] P. Cariani, "Temporal coding of periodicity pitch in the auditory system: An overview," *Neural Plasticity*, vol. 6, no. 4, 1999.

[15] J. R. Saffran, E. K. Johnson, R. N. Aslin, and E. L. Newport, "Statistical learning of tone sequences by human infants and adults," *Cognition*, vol. 70, pp. 27–52, 1990.

[16] M. T. Pearce and G. A. Wiggins, "Expectation in melody: The influence of context and learning," *Music Perception*, vol. 23, no. 5, pp. 377–406, 2006.

[17] K. Potter, G. A. Wiggins, and M. T. Pearce, "Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music," *Musicae Scientiae*, vol. 11, no. 2, pp. 295–324, 2007.

[18] A. Schoenberg, *Letters*. London: Faber, 1974, edited by Erwin Stein. Translated from the original German by Eithne Wilkins and Ernst Kaiser.

[19] D. Revill, *The Roaring Silence: John Cage: A Life*. Arcade Publishing, 1993.

[20] G. A. Wiggins, "Music, syntax, and the meaning of "meaning"," in *Proceedings of the First Symposium on Music and Computers*, Corfu, Greece, 1998.

[21] P. Gärdenfors, *Conceptual Spaces: the geometry of thought*. Cambridge, MA: MIT Press, 2000.

[22] ——, "How to make the semantic web more semantic," in *Formal Ontology in Information Systems*, A. Variz and L. Lieu, Eds. Amsterdam, NL: IOS Press, 2004, pp. 17–34.

[23] S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. Gunter, and A. d. Friederici, "Music, language and meaning: brain signatures of semantic processing," *Nature Neuroscience*, vol. 7, no. 3, pp. 302–307, 2004.

[24] A. Craft, G. A. Wiggins, and T. Crawford, "How many beans make five? the consensus problem in music-genre classification and a new evaluation method for single-genre categorisation sysytems," in *Proceedings of ISMIR*, Vienna, Austria, 2007.

[25] T. Crawford, C. S. Iliopoulos, and R. Raman, "String-matching techniques for musical similarity and melodic recognition," *Computing in Musicology*, vol. 11, pp. 73–100, 1998.

[26] D. Meredith, K. Lemström, and G. A. Wiggins, "Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music," *Journal of New Music Research*, vol. 31, no. 4, pp. 321–345, 2002.

[27] S. Abdallah, M. Sandler, C. Rhodes, and M. Casey, "Using duration models to reduce fragmentation in audio segmentation," *Machine Learning, special issue on Machine Learning for Music*, vol. 65, no. 2–3, pp. 485–515, 2006, dOI: 10.1007/s10994-006-0586-4.

[28] J. C. Forth and G. A. Wiggins, "An approach for identifying salient repetition in multidimensional representations of polyphonic music," in *London Algorithmics 2008: Theory and Practice*, ser. Texts in Algorithmics, J. Chan, J. Daykin, and M. S. Rahman, Eds. College Publications, 2009.

[29] J. Pickens, J. P. Bello, G. Monti, M. B. Sandler, T. Crawford, M. Dovey, and D. Byrd, "Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling approach," *Journal of New Music Research*, vol. 32, no. 2, pp. 223–236, 2003.

[30] M. J. Steedman, "The blues and the abstract truth: Music and mental models," in *Mental Models In Cognitive Science*. Mahwah, NJ: Erlbaum, 1996, pp. 305–318.

[31] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, pp. 327–52, 1977.

[32] R. H. Allan, G. A. Wiggins, and D. Müllensiefen, "Methodological considerations in studies of musical similarity," in *Proceedings of ISMIR*, Vienna, Austria, 2007.

[33] G. A. Wiggins, "Models of musical similarity," *Musicae Scientiae, Discussion Forum 4a*, pp. 315–337, 2007.

[34] G. Widmer and A. Tobudic, "Playing Mozart by analogy: Learning multi-level timing and dynamics strategies," *Journal of New Music Research*, vol. 32, no. 3, pp. 259–268, 2003.

[35] R. Jackendoff, *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press, 1987.