EMBODYING THEORETICAL RESEARCH IN MUSIC COGNITION: FOUR PROPOSALS FOR THEORY-DRIVEN EXPERIMENTATION

Andreu Ballús[1], Eric Arnau[1,2], Oriol Nieto[3], Frederic Font[4], Alba G. Torrents[1,5]

[1]Universitat Autònoma de Barcelona, Spain; [2]York University, Canada; [3]New York University, USA; [4]Pompeu Fabra University, Spain; [5]Universidad Nacional De Córdoba, Argentina

ABSTRACT

Research in the field of music cognition typically focuses either on low-level, technically oriented approaches or on highly abstract ontological discussions that lack direct grounding in evidence. To bridge this gap, we propose a revision of the ontology underlying such research, from a perspective restricted to the acoustic and individual aspects of music to an embodied, extended, and anti-individualist approach. We explore the application of these ideas to empirical research in a twofold way: by discussing two experiments conducted by our group and by proposing two ideas for further experimentation. One of the conducted experiments tests whether the ability to play an instrument in any of its dimensions has an influence on how a subject listens to music; the other one explores the impact of visual information on the perception of sound as music. We comment on the results obtained and their theoretical significance. Our work shows that it is possible for abstract theorizing and concrete experimentation to go hand in hand in the field of music studies.

1. INTRODUCTION

Research in the field of music cognition typically consists of low level, technically oriented approaches focused on very specific aspects of musical structure and/or of psychoacoustic processes (e.g.: Crummer et al., 1988; Bigand et al., 2001; Demorest & Morrison, 2003; Janata, 2009), and, on the other end of the spectrum, of highly abstract ontological discussions (e. g.: Hegarty, 2001; Van Nort, 2006). Moreover, due to the relative disconnect between these lines of inquiry, most technically work tends to uncritically assume a particular ontology of music, according to which it would be a fundamentally acoustic phenomenon (Mannes, 2011; Downie et al., 2009).

While acknowledging the value of these approaches and of other narrowly framed contributions (such as the growing body of technical work on music information retrieval), we believe that the field of music cognition benefits greatly, at the present stage, from those studies in which abstract theory and specific data are put to work together, and that it is important to keep checking our implicit high-level theoretical models against empirical reality. In order to do so, our work focuses on reviewing some of the implicit assumptions of existing research in music cognition and elaborating a contrasting set of thesis whose heuristic and explanative capacities could be tested through explorative research. These theses draw heavily on recent developments in what has come to be called the third generation of cognitive science, including philosophical and theoretical positions such as distributed cognition (Hutchins, 1995), the embodied mind thesis (Varela, Thompson & Rosch, 1991), the extended mind thesis (Menary, 2010) and enactivism (Thompson, 2007). Some aspects of these developments have already been applied to music research in recent years (Schellenberg and Trehub, 1999; Cano, 2006; Leman, 2007; Clarke & Clarke, 2011). However, most of these applications focus on specific theses and do not present a general alternative to the core tenets shared by the classical cognitivist stance and its connectionist modification, both of which present musical cognition as a matter of individual symbol processing. Our aim is instead to maintain a broader perspective, using explorative approaches not to reach a definitive conclusion on any technical detail, but to test how does our alternative set of suppositions stand when confronted with empirical data (compared to more classical approaches).

The core of our alternative high-level model is the thesis that music is not a strictly acoustic or psycho-acoustic phenomenon in any restrictive sense (a good initial discussion along these lines can be found in (Wiggins, 2009)). Some of its crucial aspects have nothing to do with either wave propagation or the individual psychological perception of music. This unfolds into three sub-theses: (1) Details of physical embodiment traditionally considered irrelevant to cognitive phenomena are relevant to music cognition (*Embodied Music Thesis*); (2) Interaction with technical instruments and with the environment is relevant to musical phenomena in a way that goes beyond merely enabling the production and transmission of certain sounds (*Embedded / Extended Music Thesis*); (3) Some forms of interpersonal interaction not directly related to the production and propagation of sound are relevant to musical phenomena (*Musical Interactionism / Anti-individualism)*. The experiments presented in this work are designed to explore musical cognition phenomena under this framework.

2. EXPERIMENT A: THE EFFECT OF MUSICAL PRACTICE ON MUSIC PERCEPTION

Our first experiment explores the impact of musical practice to musical perception, with the aim of assessing whether the underlying relevant factors have to do with the bodily engagement with artifacts and its coordination with others.

2.1 Methodology

The experiment was conducted online, with a sample of 110 subjects of ages ranging 18 to 63 from different nationalities. The first section consisted of a survey to profile the subjects' relation with music, both as performers and as listeners. To avoid falling back into the assumptions of the cognitive/acoustic model, we considered a wide range of dimensions of musical knowledge, production and reception while designing the test, including different levels of bodily implication, degrees of intentionality and individual and collective aspects. To establish their performer profile we had subjects report whether they have musical studies and whether they perform music of any sort, whether they play any instrument, they sing, they perform in public, and they perform with other people. To establish the profile of listening practices, we had subjects report the amount of weekly hours that they listen to music, and the proportions of that listening time that is devoted to listening selected/non-selected music, and to listening music with full/peripheral attention.

In order to analyze how participants perceive music, subjects were asked to listen to several audio samples of diverse musical genres. Subjects had to (i) report the number of voices they could identify in a given sample; (ii) report the number of segments in which a given sample could be segmented; (iii) provide a series of tags describing several attributes of the sample songs; and (iv) provide a number of tags identifying both the similarities and the differences between two given sample songs. These responses were considered in relation with the *know-how* and *know-to* related variables, using supplementary information as educational level and listening practices as a contrast

2.2 Analysis of Results

2.2.1 Identifying Voices and Segments

We first investigated which dimensions of music performance have effects on the discrimination of different voices in a sample with a series of nonparametric tests. Wilcoxon Rank Sum test shows that performing music in any form has a significant effect (assuming $\alpha = .05$) on the number of identified voices (Nyes = 56, Nno = 45, $W = 950$, p = .033), and so does having music studies (Nyes = 59, Nno= 42, W = 824.5, p = .004). However, the effect doesn't seem to be large, considering the differences shown in figure 1.

When decomposing the notion of performing music into further dimensions, we found that playing an instrument ($N_{yes} = 56$, $N_{no} = 45$, $W = 950$, $p = .034$) and singing ($N_{yes} = 43$ $N_{no} = 58$, $W = 860$, $p = .007$) also have a significant yet seemingly small effect on voice discrimination, as illustrated in figure 2. We also found that performing in public has a significant effect on voice

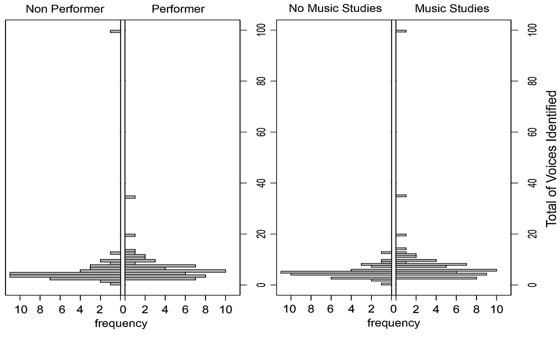

Figure 1: Paired histograms showing the number of voices identified by Performers/Non Performers (left) and subjects with Musical Studies and without (right).

discrimination ($N_{yes} = 31$, $N_{no} = 24$, $W = 221$, $p = .010$), while performing with other people does not ($N_{yes} = 35$, $N_{no} = 18$, $W = 250$, $p = .223$). Figure 3 illustrates these results.

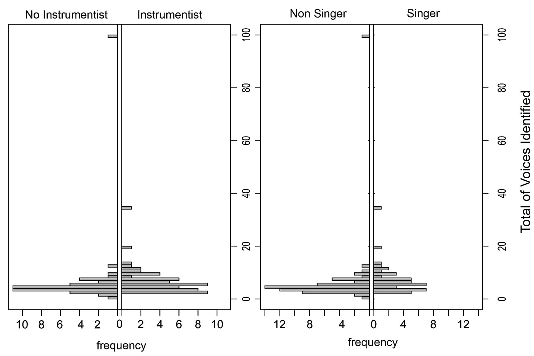

Figure 2: Paired histograms showing the number of voices identified by Instrument player / Non Instrument player (left) and Singers/Non Singers (right).

We performed two-way ANOVAs to further assess the independence of these factors. We found a significant effect of having music studies ($F(1,149) = 5.30$, $p = .023$) as opposed to performing (ns), and a significant effect of performing in public ($F(1,49) = 4.26$, $p = .044$) as opposed to playing with others (ns). No significant interactions between the factors were found. Playing an instrument and singing showed no significant effects. Figure 4 illustrates the pairs where significant effects were found.

None of the factors involving music production showed significant effects on the discrimination of segments on a different sample. Finally, no factors involving the listening practices of the subjects (total amount of listening hours and relative amounts with different degrees of control or attention), showed any significant effects on the discrimination of voices or segments.

These results show that, while people's habits and practices of listening to music have no significant relation with their capacity to discriminate voices, the latter is related to whether they have musical studies and to whether they perform music. While the extent of such effects remains uncertain, our results seem to suggest that the know-what aspect that comes with music studies is a more relevant factor than the embodied and know-how aspect that comes with musical practice.

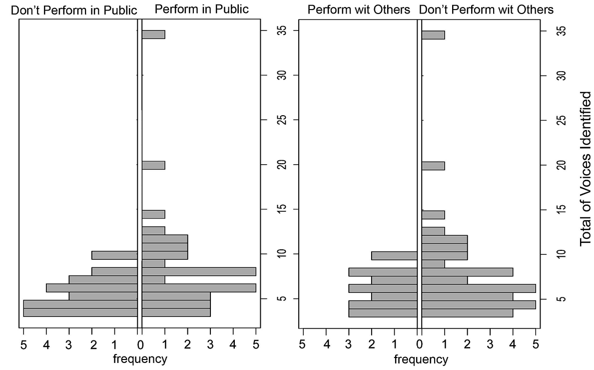

Figure 3: Paired histograms showing the number of voices identified by subjects who perform / do not (left) and subjects who play with others / do not (right).

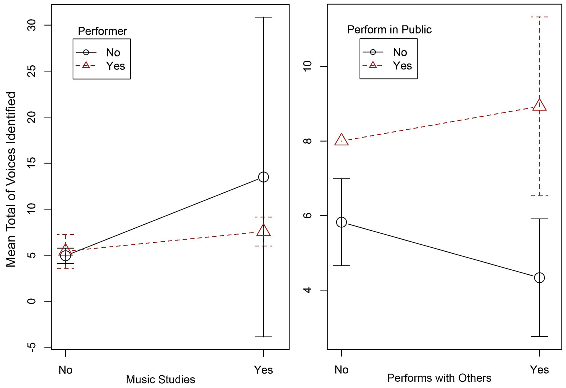

Figure 4: Mean number of voices identified by subjects decomposed according to the 'Performer' and the 'Musical Studies' partitions of the sample (left) and to the 'Performance with Others' and the 'Public Performance' (right).

2.2.2 Describing and Comparing Tracks Using Tags

In order to assess what features subjects perceive as more salient when listening to music, the responses obtained on the open descriptive and comparative questions were categorized into different semantic fields as depicted in figure 5.

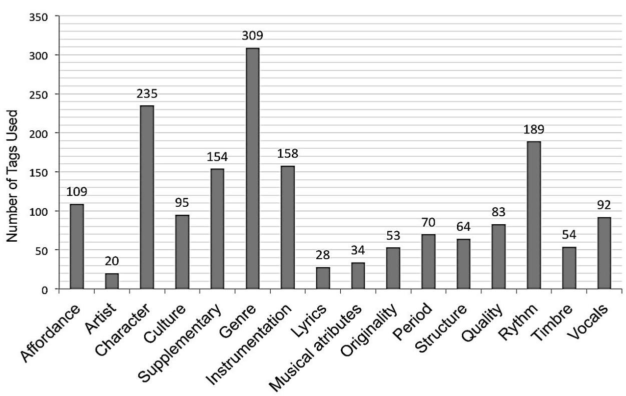

Figure 5: Categorization of the tags used by subjects to describe and compare music samples, with their frequencies. We take the categories to be self-explanatory, except for 'Supplementary', which stands for supplementary, non-conventional symbolic or metaphoric projections.

We also partitioned all tags into two second order categories: 'Acoustic' encompasses tags referring to features that can allegedly be reduced to acoustic or psychoacoustic properties – those in falling within the categories of Rhythm, Timbre, Structure and Musical Attributes (melody, pitch, etc.)– , and 'Non-Acoustic' includes the remaining tags that cannot refer to features heavily relying on extra-acoustic features. Overall, as figure 6 shows, most judgments involve non-acoustic features.

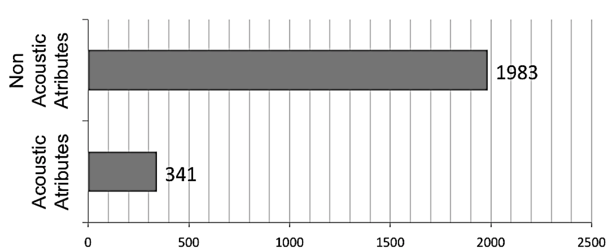

Figure 6: Second order categorization of the tags used by subjects, with their frequencies.

With regard to the amount and variety of the responses, our first finding is that the total number of tags is significantly related to performing music ($N_{yes} = 43$, $N_{no} = 58$, $W = 936$, $p = .032$), but not to having musical studies. More specifically, nonparametric tests indicate that playing and instrument does not show significant effects, but singing does ($N_{yes} = 54$, $N_{no} = 47$, $W = 938$, $p = .024$). Moreover, performing a two-way ANOVA for the factors Instrument and Singer with regard to the total of tags, shows that Instrument is a significant factor ($F(1, 97) = 4.73$, $p = .032$), and being a singer is not ($F(1, 97) = 0.00$, $p = .959$), but there is a significant interaction effect ($F(1, 97) = 4.27$, p = .041).

If we analyze the amount of different categories used by subjects instead of the number of tags, the results are less clear. It appears that the only factor close to statistical significance is playing an instrument ($N_{yes} = 56$, $N_{no} = 45$, $W = 973.5$, $p = .049$). However, the results of the ANOVA take out its significance ($F(1, 97) = 3.41$, $p = .068$). Neither performing in public or with others, or any of the variables related to listening practices showed any effect on the overall prolificacy and semantic diversity of subjects' responses.

So, there is not strong evidence about that people who make music have a more prolific and diverse impression of music. Nevertheless, more interesting results come up when we target specific categories of musical judgment. We found effects on the usage of tags related to 'Genre', 'Instrumentation', 'Musical Attributes' and 'Quality'. Music performers significantly more often choose these qualities.

Using Fisher's Exact Test (two-sided) we found that musical practice has a significant effect on the frequency of genre ascriptions when describing and comparing samples ($p = .046$), while having music studies does not show such effect. On a finer level, playing an instrument has a significant effect ($p = .024$), while singing does not. Such effects are confirmed by the ANOVA, which shows 'Performer' to be a significant factor ($F(1, 97) = 8.87$, $p = .003$) as opposed to 'Music Studies' (ns) , and 'Instrument' to be a significant factor ($F(1, 97) = 8.39$, $p = .005$) as opposed to 'Singer' (ns). No interaction effects were found in either case (figure 7).

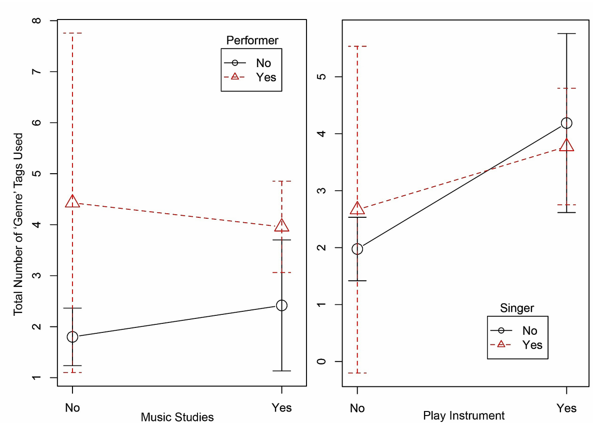

Figure 7: Mean number of 'Genre' tags used by subjects decomposed according to the 'Performer' / 'Musical Studies' partitions of the sample (left), and to the 'Instrument' / 'Singer' (right).

The same pattern appears if we consider the amount of judgments referring to musical attributes (melody, pitch etc.). Performing music has a significant effect (p = .016), while having music studies does not. On a finer level, playing an instrument has a significant effect (p = .002), while singing does not. Such effects are again confirmed by an ANOVA, which shows 'Performer' to be a significant factor ($F(1, 97) = 9.58$, p = .003) as opposed to 'Music Studies' (ns), and playing an instrument to be a significant factor ($F(1, 97) = 9.55$, p = .003) in front of 'Singing' (ns). No interaction effects were found in either case (figure 8).

In contrast with this pattern, different factors appear to be relevant for the production of judgments involving instrumentation. Here we find that both being a performer (p = .004), and having music studies (p = .003) have significant effects. However, the ANOVA between these two factors indicates that only the former effect is significant ($F(1, 49) = 11.13$, p = .001). Unlike the previous cases, when we decompose 'Musical Performing' into playing an instrument and singing, the categorical variable (i.e. whether subjects make judgments about 'Instrumentation') shows a significant relation with singing (p = .008), but not with playing. Yet the ANOVA points to the opposite direction, indicating that playing an instrument is the statistically significant factor ($F(1, 97) = 4.86$, p = .030). No interaction effects were found in either case. Moreover, no effects were found on performing in public, performing with others, or having different listening habits. Similar results are found with regard to the production of quality judgments. Both being a performer (p = .026), and having music studies (p = .004) have significant effects.

As a result, what we notice first is that making music is strongly coordinated with the type of features a subject focuses on. More interestingly, we see that the features that attract a special attention to instrument players –genre, melody– are different from those features that attract singers –instrumentation, quality. Not only playing an instrument and singing have different effects, but also the effects of singing seem to align with those of having musical studies. This might indicate that the effect of singing has more to do with knowledge and familiarity, and less with the embodied engagement with musical artifacts.

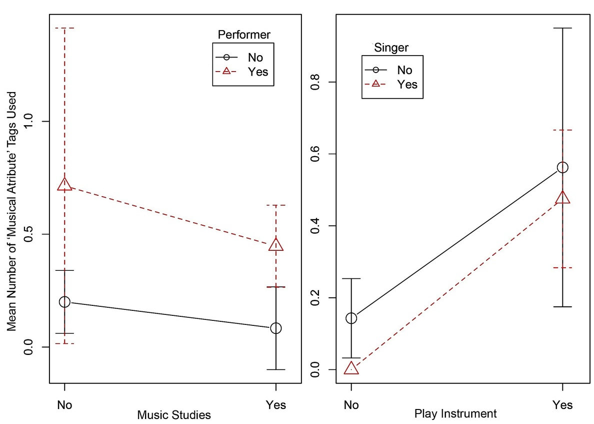

Figure 8: Mean number of 'Musical Attribute' tags used by subjects decomposed according to the 'Performer' / 'Musical Studies' partitions of the sample (left), and to the 'Instrument' / 'Singer' (right).

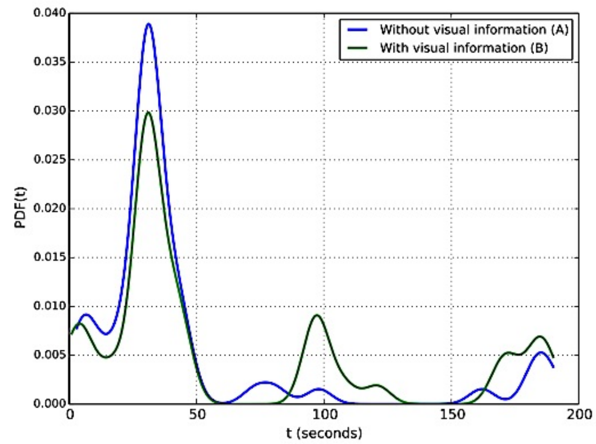

Figure 9: Probability density function of instant t for subjects listening to the sample audio with and without visual information.

3. EXPERIMENT B: EFFECT OF VISUAL CUES ON MUSIC IDENTIFICATION

Experiment B focuses on the conscious identification of an acoustic stream as music, and tries to determine whether the perception of visual images with music-related content has any significant effect on it. Our aim is to heuristically test the consideration of non-acoustic, extended aspects of music as proper elements of it.

3.1 Methodology

A four minute custom piece of audio was created for experiment B, in which different musical elements slowly and subtly appear against a backdrop of noise, so that it sounds like noise at the beginning, and at some point it starts sounding musical. The same subjects (n=110) that performed experiment A were randomly assigned to one of two conditions. Subjects in condition A were asked to listen to the piece of audio and report the exact moment in which they started to perceive music. Subjects in condition B were asked to watch a video clip with the same audio paired with footage from an experimental music performance, and to report, equally, the moment in which they started to perceive music[1].

3.2 Analysis of Results

Probability density functions (PDF) were calculated for the music detection time reported by both groups. PDF was estimated using Kernel Density Estimation. Figure 9 depicts PDF(t) for the two groups. It shows three main t instants where participants tend to indicate that the audio track transitions from non-musical noise to music (t_1~30, t_2 ~100 and t_3~180). Those instants correspond to well-defined changes in the audio track, but no particular events occur in the video. In t_1 a slightly dissonant harmonic element with a low volume is introduced in the track. In t_2 a musical pad

[1] URL to the video clip: **https://www.youtube.com/watch?v=dSirlyqLR40**

appears with a rather strong presence in the mix. Finally, at t_3 the background noise disappears from the audio track and only a melody is left. Both PDFs have their maximum values around t_1, but PDF of group B has stronger peaks in t_2 and t_3 than PDF of group A. On average, participants listening to the audio track without visual information tend to indicate that transition from non-musical noise to music occurs 18 seconds before than other participants, but the comparison is not statistically significant (F= 320, p = .650, RankSum Test). Overall these results suggest that the visual information may be acting as a distracting factor and thus some participants of group B might not be noticing the change in t_1.

To get more insight on this analysis, we further divided groups A and B according to their musical studies (yes/no). Figure 10 shows the estimated PDF(t) for the four resulting groups. As it can be seen, participants with musical studies (regardless of the group) also have a tendency to indicate that transition to music happens earlier than participants without musical studies. In fact, the comparison happens to be stronger than when comparing groups A and B. On average, participants with musical studies indicate that transition from non-musical noise to music occurs 36 seconds earlier than participants without musical studies (F = 160.5, p = .006). This result indicates that having musical studies has a bigger impact than being exposed to intentional visual information paired with the audio. One possible explanation is that the slightly dissonant harmonic element introduced at instant t_1 on the audio track is probably not being considered as music by participants without musical studies.

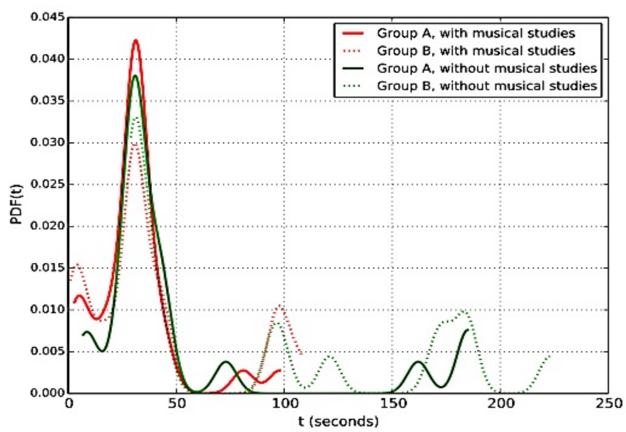

Figure 10: Probability density function of instant t for condition groups A and B decomposed into those with music studies and those without.

4. EXPERIMENTS DISCUSSION

We take our findings to support our contention that research in musical cognition can be illuminated by adopting a framework that distances from the common assumption that music is a fundamentally an acoustic phenomenon. First, they show that music practice has several effects on musical perception. People who make music seem to have a different capacity to discriminate voices, and they perceive some features as more salient. This is still more relevant against the fact that none of these effects were found for different listening practices. It seems that listening to more or less music, or in different conditions, does not affect these results. We also found that having music studies has several significant effects, but they do not align with the effects of music performance. This discrepancy between the effects of the domain of the abstract know-what and the domain of the embodied know-how can be accounted for in terms of our *embodied music thesis* –details of physical embodiment traditionally considered irrelevant to cognitive phenomena are relevant to music cognition.

However, the dimension of making music is still too broad. By considering further dimensions of musical performance, we see that across our different tests, the effects exhibited by playing an instrument differ quite systematically from those observed for being a singer, which align much more with the effects found by having musical studies. We argue that this pattern points to the fundamental distinction between the bodily engagement with an external artifact that comes with playing an instrument, and the practice of singing, in which there is no integration of such external scaffolds. This goes along the lines of our *embedded/extended music thesis*: interaction with artifacts and environmental scaffolds are relevant to musical phenomena in a way that goes beyond merely enabling the production and transmission of certain sounds.

In order to assess our *musical anti-individualist thesis* –some forms of interpersonal interaction not directly related to the production and propagation of sound are relevant to musical phenomena–, we considered two further variables: whether subjects perform in public and whether they play with others. With that regard, only a barely significant effect was found that might relate public performance with a tendency to focus more on acoustic features of music. No effects were observed from playing with others. But our impression is that, unlike the embodied and extended dimensions of musical practice, which have a more longstanding transformative effect, the collective dimension of music might well be a more fleeting phenomenon, which, ideally, should be observed on the fly. We next sketch some ways to circumvent this and other limitations of our current experiments.

5. CONCLUSIONS AND FURTHER RESEARCH

The experiments we performed seem to confirm the heuristic fruitfulness of our general theoretical model (i.e. that music is not a strictly acoustic or psycho-acoustic phenomenon), but some of the results are certainly not informative enough. To correct this, and as an indication of possible directions for further research, we propose two ideas for follow-up experiments.

Proposed experiment C focuses on the relationship between music discrimination and the subject's listening practices while making use of the work already done on the analysis of music descriptions and categorizations: in a context and setting similar to that of the performed experiments, subjects are provided with pairs of musical segments of different styles and with different relationships (same musical style, performances of the same piece, etc.). They are asked to provide detailed information on their listening practices and to answer short open questions on the paired segments. In one group the pieces are accompanied by some "exemplifying" descriptions; the influence of this elements in the responses and its interaction with listening practices is analyzed in contrast with the control group.

Proposed experiment D would be an extension of the experiment designed to explore the relationship between the perception of sound as music and the concurrent activity of the listeners. Different groups of listeners are asked to perform (i) a simple puzzle-solving task, (ii) a puzzle-solving-task involving sound elements and (iii), a puzzle solving-task in coordination with another subject, and the influence of these practices in the perception of sound as music is compared to that of the control group.

These, of course, are just some potential directions in which further research could be developed on the basis of the exploratory approach we have presented. We encourage researchers to follow any of these potential paths and to contribute their testing of the heuristic value of our basic alternative model.

6. ACKNOWLEDGMENTS

The authors would like to thank Isam Alegre and Obsidian Kingdom for the music track created specifically for our experiment.

7. REFERENCES

Aucouturier, J. (2007). Sounds like teen spirit: Computational insights into the grounding of everyday musical terms. In J. Minett & W. Wang (eds.). *Language, Evolution and the Brain*, Honh Kong, City University of HK Press, 35–64.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, 81(1), B11-B20.

Crummer, G. C., Hantz, E., Chuang, S. W., Walton, J., & Frisina, R. D. (1988). Neural basis for music cognition: Initial experimental findings. *Psychomusicology: A Journal of Research in Music Cognition*, 7(2), 117.

Clarke, D., & Clarke, E. (Eds.). (2011). *Music and Consciousness: Philosophical, Psychological, and Cultural Perspectives*. Oxford University Press.

Demorest, S. M., & Morrison, S. J. (2003). Exploring the influence of cultural familiarity and expertise on neurological responses to music. *Annals of the New York Academy of Sciences*, 999, 112-117.

Downie, J. S., Byrd, D., & Crawford, T. (2009). Ten years of ISMIR: Reflections on challenges and opportunities. In *Proceedings of the 10th International Conference on Music Information Retrieval* (pp. 13–18).

Hegarty, P. (2001). Noise threshold: Merzbow and the end of natural sound. *Organised Sound*, 6(3), 193-200.

Hutchins, E. (1995). Cognition In The Wild. Cambridge, MA: MIT Press.

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, 19, 2579-2594.

Leman, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.

Mannes, E. (2011). The power of music: Pioneering discoveries in the new science of song. Bloomsbury Publishing USA.

Menary, R. (Ed.). (2010). *The extended mind*. Cambridge, MA: MIT Press.

Cano, R. L. (2006). 'What kind of affordances are musical affordances? A semiotic approach'. In: *Simposio Internazionale sulle Scienze del Linguaggio Musicale. Bologna*, 23, 25.

Schellenberg, E. G., & Trehub, S. E. (1999). Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology*, 74, 107-127.

Tillmann, B., & Bigand, E. (1998). Influence of global structure on musical target detection and recognition. *International Journal of Psychology*, 33, 107-122.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cabridge, MA: Harvard University Press.

Van Nort, D. (2006). Noise/music and representation systems. *Organised Sound*, 11(2), 173.

Varela, F. J., Thompson, E. T., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Wiggins, G. A. (2009). *Semantic Gap?? Schemantic Schmap!! Methodological Considerations in the Scientific Study of Music*. In Proc of the 11th IEEE International Symposium on Multimedia, 477–482.