Evidence for expressive timing in music to be tempo-specific

Henkjan Honing

Music Cognition Group

Universiteit van Amsterdam

[DRAFT 10 July 2005

Please do not cite or distribute– a final version will be made available at

http://www.hum.uva.nl/mmm/abstracts/honing-2005d.html]

Running title: Perceptual invariance of timing in music

Word count: 2956

Henkjan Honing

Music Cognition Group

Institute for Logic, Language and Computation (ILLC)

Universiteit van Amsterdam

Nieuwe Doelenstraat 16

1012 CP Amsterdam

The Netherlands

E honing@uva.nl

I www.hum.uva.nl/mmm/hh/

Abstract

In several domains of cognition perceptual invariance has been studied and found, including the domains of speech, motor behavior, and object motion. In music perception, too, it has been the topic of several studies. However, with regard to the perceptual invariance of expressive timing under tempo transformation in music performance the existing perceptual studies present rather inconclusive evidence. This empirical study addresses the issue using commercially available recordings and an online internet experimental design.. The results show that listeners can decide on what is a real performance when asked two compare two recordings one of which is tempo-transformed to make them similar in overall tempo. This result is taken as support for the *tempo-specific timing hypothesis*: expressive timing can function as a perceptual clue in identifying an original performance, and counter-evidence for the *relational invariance hypothesis* that predicts a tempo-transformed performance to sound equally natural.

Introduction

Invariance and variability have been important topics in the cognitive sciences for several decades now. Perceptual invariance is concerned with whether and how certain objects or event properties remain physically or perceptually invariant under transformation (Shepard & Levitin, 2002). In several domains of cognition perceptual invariance has been studied and found, including speech (Perkell & Klatt, 1986), motor behavior (Heuer, 1991), and object motion (Shepard, 2001). In music perception, too, it has been the topic of several studies (Repp, 1995; Hulse, Takeuchi & Braaten, 1992; Handel, 1992). A well-known and uncontroversial example is melody (Dowling & Harwood, 1986). When a melody is transposed to a different register, it not only maintains its frequency ratios in performance, but it is also perceived as the same melody (i.e., melody remains perceptually invariant under transposition). With respect to other aspects of music, such as rhythm, there is less agreement in the literature. While one might expect rhythm to scale proportionally with tempo in production, and to be perceptually invariant under tempo transformation, several studies have shown that this is not always the case (Handel, 1992; Monahan & Hirsch, 1990). Rhythms are timed differently at different tempi (Repp, Windsor & Desain, 2002), and listeners often do not recognize proportionally scaled rhythms as being identical (Desain, Jansen, & Honing, 2000; Handel, 1993).

Another aspect of music whose perceptual invariance under tempo transformation has been studied is expressive timing: the minute deviations from regularity that contribute to the quality of a musical performance (Clarke, 1999; Palmer, 1997). The existing perceptual studies, however, present rather inconclusive evidence. Repp (1994) asked listeners to distinguish tempo-transformed from original MIDI performances (i.e. recorded and played back on an electronic MIDI keyboard instrument) and found the responses to be barely above chance level. Repp

(1995), however, found a small, but significant effect of tempo in a subjective rating task with the same material. Another, preliminary study (Reed, 2003) found no effects of tempo in an identification task, but some in a rating and ranking task. And finally, Honing (2004b) found a significant effect of tempo in an identification task using stimuli from a variety of musical genres. However, none of these studies controlled for the effect of tempo preference or the effect of artifacts caused by the tempo-transformation method. Both aspects could have biased the results.

These inconclusive results might have been caused by several factors. One could be the particular structural properties of the musical material as well as stylistic differences. Repp (1994) argued that music from the Romantic period might be more susceptible to relational invariance than that from the Baroque period (Desain & Honing, 1994) due to the common use of more global tempo fluctuations (e.g., *tempo rubato*) as compared to Baroque music.

Another factor that could have influenced the results is the kind of stimuli used. Repp (1994; 1995) presented MIDI performances at different tempi played back on an electronic keyboard. These performances included several 'regularizations' applied to, for example, onset asynchronies and articulation. These regularizations could well interfere with the perceived quality of the performances, and, arguably, made it more difficult to make judgments on the 'naturalness' of the performances. In that sense, audio recordings, as used in Reed (2003) and Honing (2004b), can be considered more ecologically valid stimuli. However, as said, these studies did not control for the effect of possible artifacts of the tempo transformation method used.

Hence, this study applies a different experimental design than that used in the perceptual studies mentioned above. To minimize the influence of tempo preference, in the current design two recordings of the same composition are used. These are made similar in tempo by tempo-transforming one of them. Listeners were are asked (using a comparison task) to indicate which of the two is an original recording (Experiment 1), focusing on the expressive timing of the performances. Furthermore, to control for the effect of artifacts, in a second experiment audio experts were asked to focus on the sound quality of the recordings and to indicate whether they contained an artifact that could be attributed to the signal processing method (Experiment 2).[1]

Two hypotheses will be considered: the *relational invariance hypothesis* and the *tempo-specific timing hypothesis*. In the experimental design used, the first hypothesis is in fact the null hypothesis. It predicts no significant difference in responses between the original and tempo-transformed excerpts: both excerpts will sound equally *natural*, so that the respondents will consider both versions musically plausible performances, and, consequently, just guess what is an original recording.

On the other hand, if a significant proportion of the respondents is able to identify the original correctly, this will support the tempo-specific timing hypothesis. This hypothesis is based on the idea that expressive timing in music performance (defined as both the local deviations from isochrony as more global changes in tempo) is intrinsically related to global tempo. When expressive timing is simply scaled to another tempo (i.e., slowing it down or speeding it up proportionally) this may make the performance sound awkward or *unnatural*, and hence easier to identify as a tempo-transformed version. In addition, one could argue that if performers adapt their timing to the global tempo in a non-proportional way (as was shown at least for some musical styles; Friberg & Sundström, 2002; Desain & Honing, 1994) it might well be that listeners are sensitive to this as well. A performance that is tempo-transformed may sound awkward since the expressive timing is not adapted in a way a musician would normally do.

Experiment 1

The participants were asked to listen to pairs of audio fragments from commercially available recordings by well-known pianists (see Table 1): one being an original recording, the other a manipulated, tempo-transformed recording. The tempo-transformed recording was originally performed at a different tempo, but has been time-stretched (or time-compressed) to become close in tempo to the other performance of the pair. The experiment used fourteen original and fourteen tempo-transformed recordings. The two stimulus pairs derived from each performance pair were presented to different groups of listeners: this to avoid the possibility of remembering characteristics of one pair and use them to make the response to the other pair. Furthermore, all stimuli were randomly presented between and within stimulus pairs to control for order effects.

< *Insert Table 1 around here* >

METHOD

Participants

The participants (N = 143) responded to an invitation that was sent to a variety of professional mailing lists in a previous study (Honing, 2004b) in which they indicated they liked to be invited for in a follow-up study. In addition, the participants were music students from the Universiteit van Amsterdam and Northwestern University. Three gift certificates were raffled among those that responded. Of all participants 43% reported to be an "expert (musician)", 41%

"experienced (listen a lot to music)", and 16% to be of the category "average (listen casually to music)".

Equipment

The responses were collected in an online internet version of the experiment using standard web browser technologies (i.e., HTML, CGI, and Java scripts). The stimuli were excerpts of commercially available recordings (see Table 1). These excerpts were converted to the MPEG4 file format to guarantee optimal sound quality on different computer platforms, at different data transmission rates.[2] The experimental setup and stimuli were generated using POCO (Honing, 1990).

Materials and stimulus presentation

The experiment used fourteen original and fourteen tempo-transformed recordings. The two stimulus pairs derived from each performance pair (A/B) were presented to different groups of listeners. Group I (*N* = 74) was presented seven pairs A/B' (Prime indicating a transformed recording), Group II (*N* = 69) were presented seven pairs A'/B.

The tempo-transformed versions were made using state-of-the-art time-scale modification software (Bonada, 2000).[3] Of all recordings the tempo of the first four bars was measured with a metronome, and checked perceptually by synchronizing it with the music. The resulting tempo estimate was used to calculate the tempo-scaling factor to make the A/B' and A'/B pairs similar in tempo. All sound excerpts were taken from the beginning of a recording (i.e. the first sixty seconds). The presentation of the stimuli was randomized between and within pairs for each participant, as was group assignment.

Procedure

Participants were asked to visit a non-public webpage of the online experiment. First, they were asked to test their computer and audio system with a short sound excerpt, and to adjust the volume to a comfortable level. Next, they were referred to a webpage containing the actual experiment (see Figure 1) containing the following instructions: *"You will be presented seven pairs of audio fragments: one being an original recording (by one pianist), the other a manipulated, tempo transformed recording (by another pianist). The tempo transformed recording originally had a different tempo, but has been time-stretched (or time-compressed) to become close in tempo to the other performance of the pair. For you to decide which is which. 1) Listen to a pair of audio fragments once for its full length. 2) Focus on the use of expressive timing by the performer, such as note asynchrony, tempo rubato and articulation. (Please* ignore *any audio recording quality related phenomena like noise, ticks, and/or miking technique. The sound quality of the recordings is* not *relevant here.) 3) Then answer the two questions listed next to them: which is a real, original recording (i.e. the most natural performance), the top or bottom excerpt, and, how confident are you? 4) Please do this for all seven pairs of audio fragments."* At the end of the experiment they were asked to fill in a short multiple-choice questionnaire to obtain information on, e.g., musical experience. The experiment took on average sixteen minutes to complete.

< Insert Figure 1 around here >

Analysis

The response forms were automatically sent by e-mail to the author and converted to a tabulated file for further analysis using POCO (Honing, 1990). The responses to the "What is the original?" question were converted to percent correct, the responses to the "Are your sure?" question were converted to a factor (1= "Yes", .5 = "Somewhat", and 0 = "No"). JMP (version 5.0, manufacturer: SAS) was used for the statistical analyses.

RESULTS

The results of the comparison task are shown in Figure 2. A majority of the participants could correctly identify the original recoding (*M* = 70%, *SD* = 9.1%). For each stimulus pair the majority of responses is highly significant (one-tailed binomial test; marked with asterisks in Figure 2). A *t*-test showed the effect of direction of presentation (either A/B' or A'/B) was not significant. While some pairs were apparently more difficult to judge than others, overall the confidence ratings correlated positively with percentage correct responses (*r* = .34).

< *Insert Figure 2 around here* >

Experiment 2

To make sure that possible artifacts of the signal processing method (Bonada, 2000) did not bias the responses, a control experiment was performed. In this experiment the same stimuli as used in Experiment 1 were judged individually for artifacts by a control group that consisted of mainly of audio experts.

Participants

The participants (*N* = 43) responded to an invitation that was sent to the AUDITORY mailing list. Three gift certificates were raffled among those that responded. Of all participants 56% reported to be an "audio expert", 26% "experienced (listen a lot to music)", and 18% to be of the category "average (listen casually to music)". The experiment took on average thirty-two minutes to complete.

Equipment

Same as Experiment 1.

Materials and stimulus presentation

The same stimuli were used as in Experiment 1 presented individually and in random order.

Procedure

Participants were asked to visit a non-public webpage of the online experiment. First, they were asked to test their computer and audio system with a short sound excerpt, and to adjust the volume to a comfortable level. Next, they were referred to a webpage containing the actual experiment containing the following instructions: *"This listening experiment investigates whether an advanced time-stretching method used in a related experiment causes any audible artifacts. Please do the following: 1) Listen to each excerpt once for its full length, using headphones. 2) Focus on possible timbral artifacts (unnatural transients, phasiness, loss of attack sharpness, etc.) in the audio recording. (Please* ignore *any performance related phenomena. The musical quality, e.g., the timing or tempo used, is* not *relevant here - this is the topic of a parallel study.) 3) Then answer the two questions listed next to them: is the recording manipulated in some way (or is it an original recording), and, are you sure? 4) Please do this for all twenty-eight sound excerpts."* Furthermore, for each sound except the recording date was mentioned. At the end of the experiment the participants were asked to fill in a short multiple-choice questionnaire to obtain information on, e.g., their expertise.

Analysis

Same as Experiment 1.

RESULTS

The results of the identification task are shown in Figure 3. Overall the participants did not do better than chance (*M* = 52%, *SD* = 17%). Furthermore, the number of correct responses in the case of a tempo-transformed recording was not significantly different from chance, and not significantly different from the correct responses in the case of an original recording.
There were, however, some individual exceptions. Stimuli 02 Ga', 05 RT' and 13 RT' attracted a significantly higher amount of correct responses. Apparently, these did contain artifacts and, consequently, the responses to the pairs containing these stimuli in Experiment 1 (02 Gb/Ga', 05 GG/RT', and 13 GG/RT') could have been biased because of this.

< *Insert Figure 3 around here* >

Summary and Conclusion

The two experiments reported here were concerned with the question of whether listeners can identify an original recording when asked to focus on the expressive timing. This was investigated by asking listeners to distinguish between an original audio recording by one pianist and a tempo-transformed recordings of the same composition by another pianist. Experiment 1 instructed listeners to focus on the expressive timing of the performance (ignoring the sound quality) and to indicate which was an original recording and which a tempo-transformed recording. Experiment 2 (Control) instructed listeners to focus of the sound quality (ignoring the musical aspects) and to indicate whether they heard an artifact that could be attributed to the signal processing method.

The results of Experiment 1 were highly significant. Apparently, listeners can decide on what is an original recording by focusing on the expressive timing of a performance. For Experiment 2 only a few stimuli differed significantly from chance. However, overall the results were not significantly different from chance. As such we can be sure that artifacts of the tempo transformation method did not bias the responses, and we can take the results of Experiment 1 (N.B. without the three pairs mentioned) as support for the *tempo-specific timing hypothesis*: expressive timing, when applied at the appropriate tempo, can function as a clue in identifying an original performance. The results are counter-evidence for the *relational invariance hypothesis* that predicts no preference for the original over the tempo-transformed version: both are predicted to sound equally natural.

Still, the music performance literature provides some support for the relational invariance hypothesis. Relational invariance might be a good approximation for the use of expressive timing in piano music from the Romantic period (Repp, 1994), but less so with music from other repertoires, such as music from the Baroque period (Desain & Honing, 1994) or Jazz (Friberg & Sundström, 2002). Next to a possible effect of the musical material used, the different results might also be explained by differences resulting from the methodology applied (using fragments in MIDI versus audio format, using rating versus identification or comparison tasks, etc.).

Furthermore, the current study did not control for a possible effect of familiarity. If listeners were familiar with a particular recording, and thought they recognized the performer, they could have based there judgment (against instructions) on tempo instead of the expressive timing used.[4] For expert pianists it has been shown that they, at least to some extent, can recognize their own performances (Repp & Knoblich, 2004). However, in how far listeners are capable of remembering and/or recognizing the timing details of performances of others than themselves is less clear. With respect to the memory for tempo (Levitin & Cook, 1996), it was shown that the phenomenon of absolute tempo is apparent in pop and rock music, but less clear for music from the classical repertoire. The difference was attributed to a larger variety of tempi used for one composition in classical music, having a less strong effect on an iconic memory for tempo (Levitin & Cook, 1996). Hence, one could argue that the current experimental design is less sensitive to such an effect.

The results presented are important for models of rhythm perception and production in music. If relational invariance was observed this would have been an indicator of the existence of a generalized motor program having a variable rate parameter (Heuer, 1991). Several models of expressive timing in music performance indeed suggest this (Honing, 2005); They predict timing to be relational invariant with global tempo (or rate). But clearly, the relation between timing and tempo in music perception and music production is far more intimately coupled: one can not be changed without affecting the other (Honing, 2004a). What could be the nature of this relation and the effect of previous exposure and expectancy is a topic of current research (EmCAP, 2005).

Acknowledgments

Special thanks to Jordi Bonada (MTG/UPF) for providing his time-scale modification algorithm and making the tempo-transformed examples, Bruno Repp (Haskins Labs) for advice on the experimental design, and all beta-testers from the Universiteit van Amsterdam and Northwestern

University for their time and suggestions improving the online internet version of the experiment.

References

Bonada, J. (2000). Automatic technique in frequency domain for near-lossless time-scale modification of audio. *Proceedings of International Computer Music Conference.* San Francisco: Computer Music Association.

Clarke, E.F. (1999). Rhythm and Timing in Music. In D. Deutsch (Ed.), *Psychology of music, 2nd edition* (pp. 473-500). New York: Academic Press.

Desain, P., & Honing, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research*, 56(4), 285-292.

Desain, P., Jansen, C., & Honing, H. (2000). How identification of rhythmic categories depends on tempo and meter. In *Proceedings of the Sixth International Conference on Music Perception and Cognition* (pp. 51). Keele, UK: Keele University, Department of Psychology.

Dowling, W.J., & Harwood, D. (1986). *Music cognition.* New York: Academic Press.

EmCAP Consortium (2005). *Emergent Cognition Through Active Perception.* Unpublished Research Proposal for the European Sixth Framework Research Programme (FP6-IST).

Friberg, A., & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19, 333-349.

Gabrielsson, A. (1999). Music Performance. In Deutsch, D. (ed.), *Psychology of music, 2nd edition.* (pp. 506-602). New York: Academic Press.

Handel, S. (1992). The differentiation of rhythmic structure. *Perception & Psychophysics*, 52, 497-507.

Handel, S. (1993). The effect of tempo and tone duration on rhythmic discrimination. *Perception & Psychophysics*, 54, 370-382.

Heuer, H. (1991). Invariant relative timing in motor-program theory. In J. Fagard & P.H. Wolff (Eds.), *The development of timing control and temporal organisation in coordinated action* (pp. 37-68). Amsterdam: Elsevier.

Honing, H. (1990). POCO: an environment for analysing, modifying, and generating expression in music. In *Proceedings of the 1990 International Computer Music Conference* (pp. 364-368). San Francisco: Computer Music Association.

Honing, H. (2004a). Computational modeling of music cognition: a case study on model selection. ILLC Prepublication PP-2004-14. [under review *Music Perception*]

Honing, H. (2004b). Is timing tempo-specific? An online internet experiment on perceptual invariance of timing in music. *ILLC Prepublication* PP-2004-34.

Honing, H. (2005, in press). Is there a perception-based alternative to kinematic models of tempo rubato? *Music Perception.*

Hulse, S., Takeuchi, A. H., & Braaten, R. F. (1992). Perceptual invariances in the comparative psychology of music. *Music Perception*, 10, 151-184.

Levitin, D., & Cook, P. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 56 (4): 414-423.

Monahan, C. B., & Hirsch, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, 47, 227-242.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48, 115-138.

Perkell, J.S., & Klatt, D. H. (1986). *Invariance and variability in speech processes.* Hillsdale, NJ: Erlbaum.

Reed, R. (2003). Tempo change and interpretation preference. *Proceedings of the European Society for the Cognitive Sciences of Music* (pp. 558-561). Hannover: University of Hannover.

Repp, B.H. (1994). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research*, 56, 269-284.

Repp, B.H. (1995). Quantitative effects of global tempo on expressive timing in music performance: Some perceptual evidence. *Music Perception*, 13, 39-57.

Repp, B.H., & Knoblich, G. (2004). Perceiving action identity. How pianists recognize their own performances. *Psychological Science*, 15 (9), 604-609.

Repp, B. H., Windsor, W. L., & Desain, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception*, 19, 565-593.

Rink, J. (Ed.) (1995). *The practice of performance: Studies in musical interpretation*. Cambridge: Cambridge University Press.

Shepard, R. (2001). Perceptual-cognitive universals as reflections of the world. *Behavioral and Brain Sciences*, 24, 581-601.

Shepard, R., & Levitin, D. (2002). Cognitive psychology and music. In Levitin, D. (Ed.) *Foundations of cognitive psychology: Core readings* (pp. 503-514). Cambridge, MA: MIT Press.

Tables

Table 1. Recordings Used in Experiment 1 and 2

Code	Pianist	Composition	Recording
01 Ga	Glenn Gould	J.S. Bach, Goldberg Variations (1981), BWV 988, Variation I	Sony, SMK 64126, 1999
02 Gb	Glenn Gould	J.S. Bach, Goldberg Variations (1955), BWV 988, Variation I	Sony, SK 52594, 1992
03 GG	Glenn Gould	J.S. Bach, English Suite No. 4, BWV 809, Allemande	Sony, SK 87766, 2001
04 SR	Sviatoslav Richter	J.S. Bach, English Suite No. 4, BWV 809, Allemande	Delos, GH 5601, 2004
05 GG	Glenn Gould	J.S. Bach, WTC II, BWV 890, Prelude 21	Sony, SX4K 60150, 1997
06 RT	Rosalyn Tureck	J.S. Bach, WTC II, BWV 890, Prelude 21	BBC, BBCL 4116-2, 2002
07 AR	Arthur Rubinstein	L. v. Beethoven, Piano Sonata No. 14, Op. 17, no. 2. Allegretto	RCA, 09026-63056-2, 1999
08 VA	Vladimir Ashkenazy	L. v. Beethoven, Piano Sonata No. 14, Op. 17, no. 2. Allegretto	Decca, 452 982-2, 1997
09 CA	Claudio Arrau	F. Chopin, Grande Valse Brillante, op.18	Philips, 468 391-2, 2001
10 VA	Vladimir Ashkenazy	F. Chopin, Grande Valse Brillante, op.18	Decca, 417 798-2, 1990
11 VH	Vladimir Horowitz	R. Schumann, Kinderszenen, Träumerei	DGG, 474 370-2, 1991
12 CA	Claudio Arrau	R. Schumann, Kinderszenen, Träumerei	Philips, 468 391-2, 2001
13 GG	Glenn Gould	J.S. Bach, WTC II, BWV 880, Fugue 11	Sony, SX4K 60150, 1997
14 RT	Rosalyn Tureck	J.S. Bach, WTC II, BWV 880, Fugue 11	BBC, BBCL 4116-2, 2002

Figure Captions

Figure 1. Fragment of the online internet user interface showing the presentation of seven pairs of sound excerpts.

Figure 2. Results of Experiment 1 (Group I, N= 74; Group II, N= 69). Significance levels are indicated with asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Figure 3. Results of Experiment 2 ($N = 43$). Significance levels are indicated with asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Figures

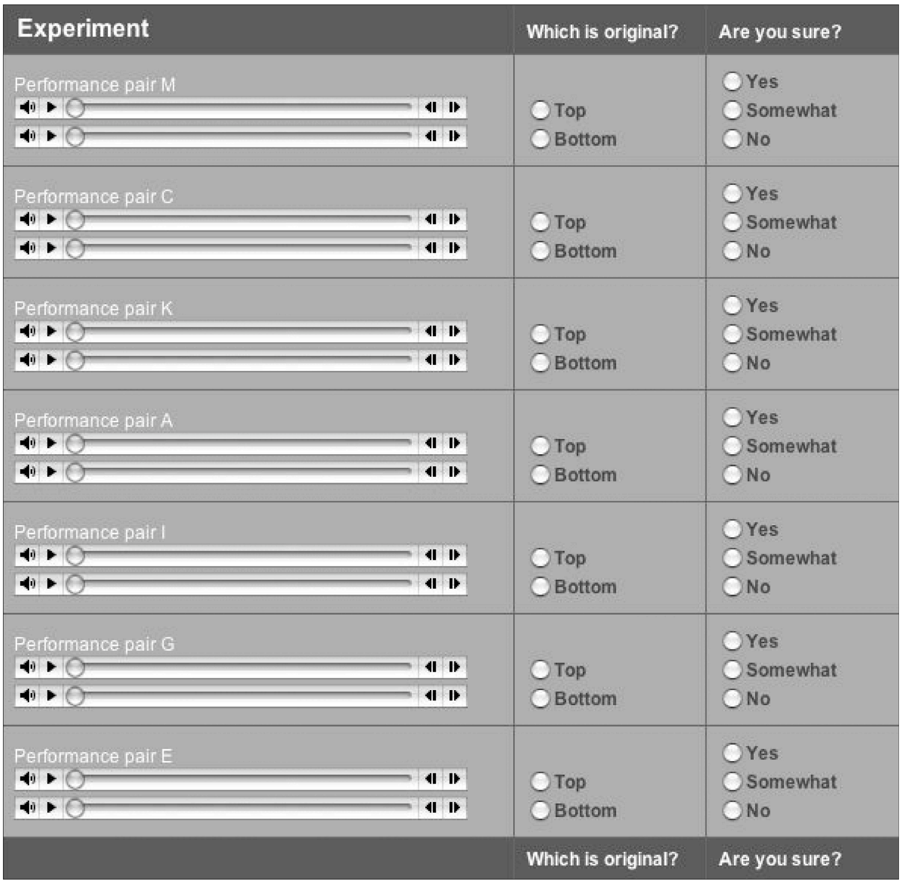


Figure 1.

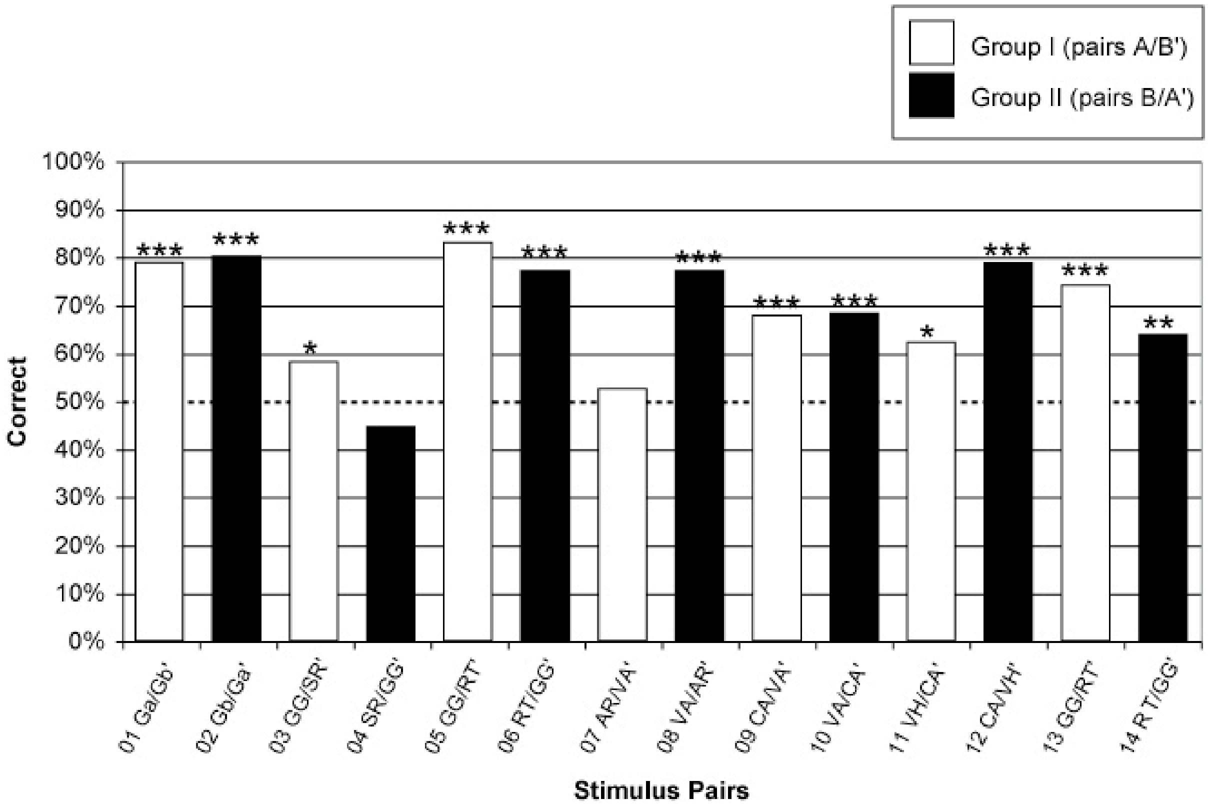
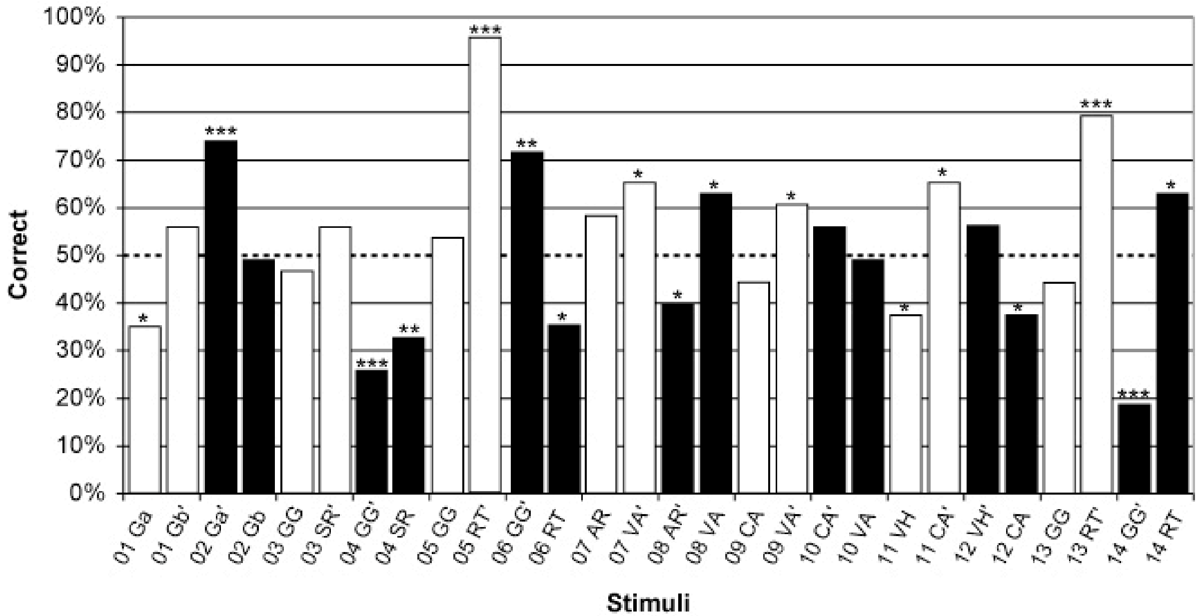

Group I (pairs A/B')
Group II (pairs B/A')
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Correct
***
*
**
01 Ga/Gb'
02 Gb/Ga'
03 GG/SR'
04 SR/GG'
05 GG/RT'
06 RT/GG'
07 AR/VA'
08 VA/AR'
09 CA/VA'
10 VA/CA'
11 VH/CA'
12 CA/VH'
13 GG/RT'
14 RT/GG'
Stimulus Pairs

Figure 2.



Correct
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Stimuli
01 Ga
01 Gb'
02 Ga'
02 Gb
03 GG
03 SR'
04 GG'
04 SR
05 GG
05 RT'
06 GG'
06 RT
07 AR
07 VA'
08 AR'
08 VA
09 CA
09 VA'
10 CA'
10 VA
11 VH
11 CA'
12 VH'
12 CA
13 GG
13 RT'
14 GG'
14 RT

Figure 3.

Footnotes

[1] Alternatively, since piano music was used in this study, a design using, e.g., a MIDI grand piano (i.e. modern pianola) would avoid the problem of artifacts in the manipulating audio data. However, the current setup was preferred to take advantage of the wide variety of audio recordings currently available, ranging an enormous variety of musics by expert musicians.

[2] See http://www.apple.com/mpeg4/ for technical details.

[3] See Bonada (2000) for details on the signal processing method, and http://www.hum.uva.nl/mmm/exp2/ for the stimuli used.

[4] The two famous recordings of the *Goldberg Variations* by Glenn Gould (stimulus pairs 01 and 02) could well be susceptible of such an effect.