Is expressive timing relational invariant under tempo transformation?

Henkjan Honing

Psychology of Music 2007; 35; 276 originally published online Feb 15, 2007;
DOI: 10.1177/0305735607070380

ARTICLE

Is expressive timing relational invariant under tempo transformation?[1]

Psychology of Music

Psychology of Music
Copyright © 2007
Society for Education, Music
and Psychology Research
vol 35(2): 276–285 [0305-7356
(200704) 35:2; 276–285]
10.1177/0305735607070380
http://pom.sagepub.com

HENKJAN HONING
UNIVERSITY OF AMSTERDAM

ABSTRACT This empirical study is concerned with examining the relation between tempo and expressive timing in music performance. This was investigated by asking listeners ($N = 307$) to distinguish between an original recording and a tempo-transformed version in a musical genre of their preference (jazz or classical). A majority of the participants was able to correctly identify the original recording. This result is taken as support for the tempo-specific timing hypothesis which suggests that the relationship between timing variations and average tempo may function as a cue for identification of a real performance, and counter-evidence for the relationally invariant timing hypothesis that predicts a tempo-transformed performance will sound equally natural.

KEYWORDS: *music perception, music performance, perceptual invariance*

Introduction

An important theoretical issue in cognitive science is how properties of objects or events behave under transformation, and whether these properties remain the same under transformation. The latter phenomenon (often referred to as 'perceptual invariance') is considered an important index to mental representations of the real world and the functioning of our perceptual system (Shepard and Levitin, 2002). Perceptual invariance has been studied and found in several domains, including speech (Perkell and Klatt, 1986), motor behaviour (Heuer, 1991) and object motion (Shepard, 2001). In music perception, too, it has been the topic of several studies (Repp, 1995; Handel, 1992; Hulse et al., 1992). A well-known and relatively uncontroversial example is melody (Dowling and 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 and Hirsch, 1990). Rhythms are timed differently at different tempi (Repp et al., 2002), and listeners often do not recognize proportionally scaled rhythms as being identical (Handel, 1993).

Perceptual invariance under tempo transformation has also been studied in the domain of expressive timing (Clarke, 1999), the existing perceptual studies (Reed, 2003; Repp, 1994; 1995) present rather inconclusive evidence. Repp (1994) asked listeners to distinguish tempo transformed from original MIDI performances and found the responses to be barely above chance level. Repp (1995) found a small but significant effect of tempo in a subjective rating task with the same material. And, lastly, Reed (2003) found no effects of tempo in an identification task, but some in a rating and ranking task.

The music performance literature also presents conflicting evidence regarding the tempo invariance of expressive timing. Some authors found that global tempo does influence expressive timing (i.e. timing being tempo-specific; Desain and Honing, 1994; Friberg and Sundström, 2002): at different tempi different structural levels become salient, and this has an effect on the expressive freedom and variability observed (Clarke, 1999). Other studies found expressive timing to be nearly invariant under tempo transformation (Repp, 1994; 1995), which was interpreted as 'relational invariance' (or proportional duration), a key concept in research on timing control in skilled motor performance (Heuer, 1991). However, the particular properties of the musical material as well as stylistic differences may have been responsible for the different results. Hence, the present study investigates whether expressive timing is perceptually invariant under tempo transformation by considering musical excerpts from both the jazz and classical repertoire.

Experiments

Two experiments were conducted to examine the effect of tempo and musical genre on the identification of an original recording. The participants were asked to listen to 10 sound examples (in their genre of preference) and to indicate which example was an original recording and which was a tempo-transformed version (i.e. a slowed-down or speeded-up version of the original).

Two hypotheses will be considered: the relationally invariant timing 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 to 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 the local deviations from isochrony, as well as more global changes in tempo) is intrinsically related to global tempo. When expressive timing is simply scaled to another tempo (slowing it down or speeding it up proportionally; e.g. using the tempo knob in a MIDI sequencer) 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 in some studies, as mentioned above) it might well be that listeners are sensitive to this as well. When the expressive timing of a tempo-transformed performance is not adapted in a way a musician would normally do, it might sound awkward and makes it easier to distinguish between a real and a tempo-transformed version.

The two experiments discussed below use the same experimental design. One uses recordings from the jazz repertoire (pianists and saxophonists performing original songs), the other fragments from the classical repertoire (pianists performing works by J.S. Bach and L. van Beethoven).

Method

PARTICIPANTS

The participants (N = 307) responded to an invitation that was sent to a variety of professional mailing lists, and to students from the University of Amsterdam and Northwestern University in the USA. Of all respondents, 175 participated in the classical version of the experiment. Of these, 46 percent reported to be an 'expert (musician)', 54 percent 'experienced (listen a lot to music)', and none were of the category 'average (listen casually to music)'. Of the 132 respondents who participated in the jazz version of the experiment 50 percent reported to be an 'expert (musician)', 50 percent 'experienced (listen a lot to music)', and none were of the category 'average (listen casually to music)'.

EQUIPMENT

The responses were collected in an online version of the experiment on the internet using standard web browser technologies (i.e. HTML, CGI and Java scripts). The stimuli were sound excerpts of commercially available recordings (see Tables 1 and 2). These excerpts were converted to 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).

TABLE 1 *Recordings used in classical version of experiment*

Pianist	Composition	Record Label	Duration (s)	Tempo (bpm)
1. Glenn Gould	J.S. Bach, English Suite no. 3 in G minor, Gavotte, BWV 808	Sony SK87765, 2001	10	100
2. Daniel Barenboim	L. van Beethoven, Piano Sonata no. 8 in C minor, Op. 13 (Pathétique), Rondo	EMI 7243 5 57762 0 4, 1995	54	100
3. Sviatoslav Richter	J.S. Bach, English Suite No. 6 in D minor, Gavotte I, BWV 811	Delos GH 5601, 2004	24	83
4. Alfred Brendel	L. van Beethoven, Variation I on 'Nel cor più non mi sento', WoO 70	Philips 432 093–2, 1991	23	80
5. Glenn Gould	J.S. Bach, Two-part Inventions and Three-part Sinfonias, Sinfonia 7 in E minor, BWV 793	Sony SMK 52 596, 1993	22	94

TABLE 2 *Recordings used in jazz version of experiment*

Artist	Composition, Album (Musicians)	Record Label	Duration (s)	Tempo (bpm)
1. Geri Allen	'Invisible', *In the Year of the Dragon* (with Charlie Haden and Paul Motian)	Winter & Winter, 919 027–2, 2002	29	128
2. Yuri Honing	'Hell's Kitchen', *Seven* (with Paul Motian, Gary Peacock and Paul Bley)	Challenge, JIM 75086, 2001	39	140
3. Brad Mehldau	'It might as well be spring', *Progression* (with Larry Grenadier and Jorge Rossy)	Warner Bros 9362–48005–2, 2001	21	80
4. Carla Bley	'Chicken', *Songs with Legs* (with Andy Sheppard and Steve Swallow)	Watt / ECM, 1995	39	110
5. Eric Dolphy	'Miss Ann', *Last Date* (with Misha Mengelberg, Han Bennink and Jacques Schols)	EmAsCy 510 124–1, 1964	39	101

MATERIALS AND STIMULUS PREPARATION

Both versions of the experiment (i.e. jazz and classical) used five original recordings and five tempo-transformed versions of these originals (see Tables 1 and 2). The average tempo of the original and the tempo-transformed versions was Maelzel metronome (M.M.) 103 (*SD* = 20.2),[3] the range of tempi being similar to those used in Repp (1995) and Reed (2003). The tempo-transformed versions were made using standard time-scale modification software (ASD, manufacturer: Roni Music).[4]

Such an algorithm stretches (or compresses) the duration of an audio file in time (hence changing the global tempo proportionally) without altering the pitch. All stimuli were processed using the same equalization and signal processing settings ('Type III', i.e. highest quality).[5] The order (original or tempo-transformed version first) and direction of the transformation (slower or faster) were randomly selected. All sound excerpts were taken from the beginning of a recording (the first *n* seconds) and consisted of one or more musical phrases (see Tables 1 and 2). The resulting 10 stimuli were presented to each participant in random order, except that the two versions of the same music followed immediately upon each other, as such allowing for direct comparison.

PROCEDURE

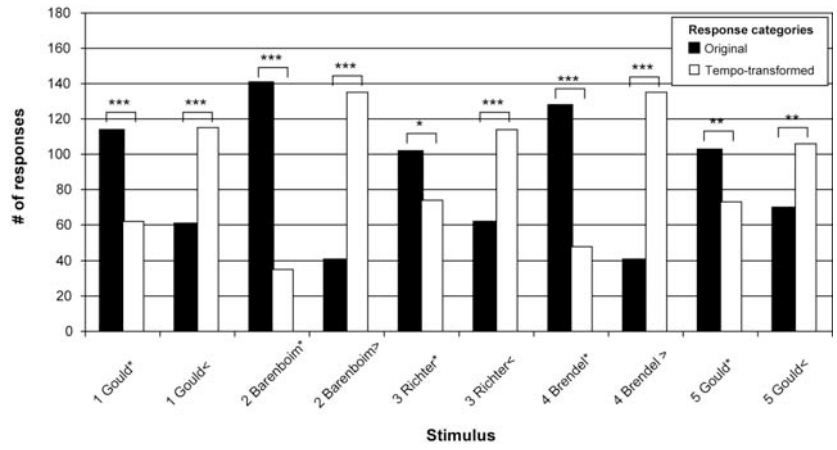
Participants were asked to visit a temporary webpage of the online experiment. First, they were asked to test their computer and audio system with a short sound excerpt, and adjust the volume to a comfortable level. Next, they were asked to select the musical genre ('jazz' or 'classical') with which they considered themselves most familiar. Finally, the participants were instructed (1) to listen – as often as needed – to a single sound example, focusing on the use of expressive timing (as if they were a judge in a music performance master class) and to ignore possible timbral artefacts; and (2) to answer the questions listed on the screen. The questions presented were 'Is this an original recording?' (response categories 'Yes' or 'No') and 'Are you sure?' (response categories 'Yes', 'Somewhat' or 'No'). The participants could review all their judgments. 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 14 minutes to complete.

Analysis

The response forms were automatically sent by email to the author and converted to a tabulated file for further analysis using POCO (Honing, 1990). JMP (version 5.0, SAS) was used for the statistical analyses.

RESULTS

The results of the identification task ('Is this an original recording?') for the classical version of the experiment are shown in Figure 1. It can be seen that



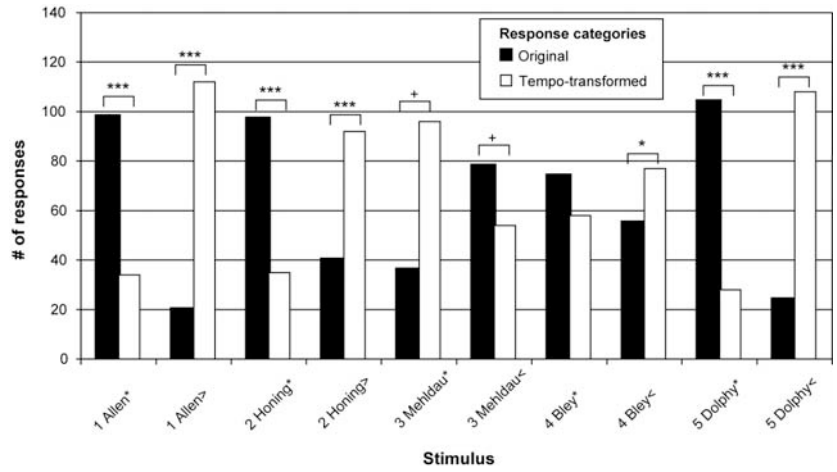
Note: An * in the stimulus-label refers to an original recording, a < and a > respectively to a 20% slower and faster tempo-transformed version. Significance levels are indicated with asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

FIGURE 1 *Results of the classical version of the experiment ($n = 175$).*

listeners ($n = 175$) could correctly identify the original. All 10 stimuli pairs differ significantly from chance (one-tailed binomial test). Furthermore, the confidence ratings correlated positively with the 'Yes' responses ($r = .83$), but less so with the 'No' responses ($r = .39$).

The results of the identification task for the jazz version of the experiment are shown in Figure 2. Here as well, the participants ($n = 132$) seemed to be able to correctly identify the original. Seven of the 10 stimuli pairs differ significantly from chance (one-tailed binomial test), one stimulus ('3 Bley*') being non-significant, and one pair ('3 Mehldau') an intriguing case of misinterpretation (see discussion below). Furthermore, the confidence ratings correlated positively with both the 'Yes' and 'No' responses (respectively, $r = .88$ and $r = .89$).

The overall results of the classical version of the experiment were significantly different from chance (two-tailed *t*-test; $p < .0001$). In the Jazz version this was the case as well (two-tailed *t*-test; $p < .05$). While the classical responses were significantly different from the jazz responses (two-tailed *t*-test; $p < .01$), the set of stimuli is too small to make substantial observations on a possible effect of music genre. But clearly both repertoires allow for expressive timing to be a cue in recognizing whether a performance was performed at the original tempo. To make sure the participants did not simply base their judgments (against the instructions) on 'preferred tempo' (M.M. 100; cf. Clarke, 1999), the correlation between the response proportions and the distance between preferred tempo and the tempi of the stimuli was examined. No positive correlation was found (classical, $r = 0.05$; jazz, $r = -.23$). However, since this is a relatively crude measure, we cannot be sure that a preference for a certain tempo had at least some effect (see discussion below).



Note: An * in the stimulus-label refers to an original recording, a < and a > respectively to a 20% slower and faster tempo-transformed version. Significance levels are indicated with asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; + significant misidentification; see text for details).

FIGURE 2 *Results of the jazz version of the experiment (n = 132).*

Summary and discussion

The two experiments reported in this article were concerned with the question of whether listeners can identify an original recording when asked to focus on the expressive timing used. This was investigated by asking listeners to distinguish between an original audio recording and a tempo-transformed version. The results showed that a significant majority of the participants could identify an original performance.

Since the expressive timing in the tempo-transformed stimuli was in fact relationally invariant with the original stimuli, the relationally invariant timing hypothesis predicts no significant difference in responses to the original and the tempo-transformed version. This contradicts the empirical results of the present study: listeners were, in most cases, able to identify the original and the tempo-transformed version. This is taken as evidence for the tempo-specific timing hypothesis: the relationship between timing variations and global tempo can function as a cue for identification of a real performance. This is consistent with what has been found in several music performance studies (Clarke, 1982; Desain and Honing, 1994; Friberg and Sundström, 2002; Palmer, 1997), which showed that performers adapt their expressive timing with global tempo. However, some alternative explanations of the reported results have to be considered. One could be the possible artefacts caused by the signal processing method that may have helped the identification of the tempo-transformed stimuli. While the parameter settings and tempo range used were carefully decided on to minimize artefacts (using the results of a pilot study; Honing, 2005a), and listeners were instructed to

focus on the expressive timing (not on possible timbral artefacts), we cannot be sure this had any effect.[6] However, arguing the other way around, if artefacts played a role in deciding on what is an original recording, one would expect much higher identification rates. Furthermore, it cannot explain the misidentification of the Mehldau example.

The jazz excerpt is an intriguing counter-example. The tempo-transformed version was judged by a significant majority to be an original, and vice versa. What might have caused this peculiar mix-up is a topic for further study.[7]

Another factor that could have influenced the results is that listeners may have based their judgments on tempo preference, instead of the instruction to judge whether expressive timing was used in a musically convincing way. It may have been the case that some listeners, when in doubt, simply selected the tempo they preferred. However, a different experimental design is needed to be able to distinguish between these factors. One way of addressing this issue is to perform a comparison task in which listeners are asked to compare two audio recordings of the same composition in two interpretations, one of which was tempo-transformed to become similar in tempo, as such controlling for a possible global tempo preference. This is a topic of current research (Honing, 2006b).

These results might come as no surprise to musicians. In the music literature one often finds discussions of how to select the appropriate tempo, and how and when to apply the appropriate timing (e.g. Rink, 1995). Musicians tend to adapt their timing to the tempo used, bringing out different structural levels of the music at different tempi (Clarke, 1982), that is, when the tempo change is large enough (Repp, 1995). Besides changing the depth of the expressive timing (relative modulation depth or amount of rubato) the timing patterns themselves also change significantly (Clarke, 1999; Honing, 2005b).

Still, the music performance literature provides some support for the relationally invariant timing 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 Classical period (Desain and Honing, 1994) or jazz (Friberg and Sundström, 2002). The different results might also be explained by the number of participants (only 10 in the perceptual studies mentioned), familiarity with the material, and differences resulting from the methodology used (MIDI versus audio excerpts, rating versus identification task, etc.). Some of these aspects are susceptible to further investigation (see Honing, 2006a). In conclusion, the present preliminary study showed that relational invariance is, in general, too simplistic a model to describe the interaction between expressive timing and global tempo in music perception, and it suggests the need of richer models of timing and tempo in music. How precisely rhythm, timing and tempo interact is a topic of current research using both computational and empirical methods (see Honing, 2006).

ACKNOWLEDGMENTS

Marijke Engels, Bruno Repp, Renee Timmers, and two anonymous referees are gratefully thanked for their comments on an earlier draft.

NOTES

1. Portions of this work were presented at the 10th Rhythm Perception and Production Workshop (RPPW10), University of Ghent, Belgium (July 2–6, 2005) and the International Computer Music Conference (ICMC), Universitat Pompeu Fabra, Barcelona, Spain (September 5–9, 2005).
2. See http://www.apple.com/mpeg4/ for technical details.
3. The tempi were estimated from the first four measures using a metronome (see Tables 1 and 2).
4. See Bonada (2000) for details, and http://www.hum.uva.nl/mmm/exp1/ for the excerpts used.
5. In a pilot study (Honing, 2005a) it was examined whether the signal processing method caused audible artefacts. Piano sounds performed best and fragments containing voice worst within a 25 percent tempo range. Hence, to minimize audible artefacts, in this study only instrumental music was used and a tempo scaling of +/– 20 percent.
6. In more recent experiments we try to control for this in a control experiment that uses the same material but instructs listeners to *ignore* the expressive timing and instead focus on possible timbral artefacts (Honing, 2006b).
7. The stimuli can be found at http://www.hum.uva.nl/mmm/exp1/.

REFERENCES

Bonada, J. (2000) 'Automatic Technique in Frequency Domain for Near-lossless Time-Scale modification of Audio', in T. Noll et al. (eds) *Proceedings of International Computer Music Conference*. San Francisco: Computer Music Association.

Clarke, E.F. (1982) Timing in the performance of Erik Satie's "Vexations"', *Acta Psychologica* 50: 1–19.

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. and Honing, H. (1994) 'Does Expressive Timing in Music Performance Scale Proportionally with Tempo? *Psychological Research* 56: 285–92.

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

Friberg, A. and Sundström, A. (2002) 'Swing Ratios and Ensemble Timing in Jazz Performance: Evidence for a Common Rhythmic Pattern', *Music Perception* 19: 333–49.

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: 370–82.

Heuer, H. (1991) 'Invariant Relative Timing in Motor-program Theory', in J. Fagard and 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 S. Arnold and G. Hair (eds) *Proceedings of the 1990 International Computer Music Conference*, pp. 364–8. San Francisco: Computer Music Association.
Honing, H. (2005a) 'Timing is Tempo-specific', in X. Serra (ed.) *Proceedings of the International Computer Music Conference*, pp. 359–62. San Francisco: Computer Music Association.
Honing, H. (2005b) 'Is there a Perception-based Alternative to Kinematic Models of Tempo Rubato? *Music Perception* 23: 79–85.
Honing, H. (2006a) 'Evidence For Tempo-specific Timing in Music Using a Web-based Experimental Setup', *Journal of Experimental Psychology: Human Perception and Performance* 32(3): 780–6.
Honing, H. (2006b) 'Computational Modelling of Music Cognition: A Case Study on Model Selection', *Music Perception* 23(5): 365–76.
Hulse, S., Takeuchi, A.H., and Braaten, R.F. (1992) 'Perceptual Invariances in the Comparative Psychology of Music', *Music Perception* 10: 151–84.
Monahan, C.B., and Hirsch, I.J. (1990) 'Studies in Auditory Timing: 2. Rhythm Patterns', *Perception & Psychophysics* 47: 227–42.
Palmer, C. (1997) 'Music Performance', *Annual Review of Psychology* 48: 115–38.
Perkell, J.S., and Klatt, D.H. (1986) *Invariance and Variability in Speech Processes*. Hillsdale, NJ: Erlbaum.
Reed, R. (2003) 'Tempo Change and Interpretation Preference', in R. Kopiez, A.C. Lehmann, I. Wolther and C. Wolf (eds) *Proceedings of the European Society for the Cognitive Sciences of Music*, pp. 558–61. 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: 269–84.
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., Windsor, W. L., and Desain, P. (2002) 'Effects of Tempo on the Timing of Simple Musical Rhythms', *Music Perception* 19: 565–93.
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. and Levitin, D. (2002) 'Cognitive Psychology and Music', in D. Levitin (ed.) *Foundations of Cognitive Psychology: Core Readings*, pp. 503–14. Cambridge, MA: MIT Press.

HENKJAN HONING is affiliated with the Department of Musicology and Institute for Logic, Language and Computation (ILLC) at the University of Amsterdam. He heads the Music Cognition Group and conducts research in the areas of music cognition and systematic, computational, and empirical musicology. The aim is to arrive at a cognitive science of music, with a special focus on its temporal aspects, such as rhythm, tempo and timing. For additional information, see http://www.hum.uva.nl/mmm/hh.
Address: Music Cognition Group, Department of Musicology and Institute for Logic, Language and Computation (ILLC), University of Amsterdam, Nieuwe Doelenstraat 16, 1012 CP Amsterdam, The Netherlands. [email: honing@uva.nl]