Is timing in tempo-specific?
An online internet experiment on perceptual invariance of timing in music

Henkjan Honing
University of Amsterdam

[comments are invited]

Running title: Is timing is tempo-specific?

dr Henkjan Honing
Department of Musicology, and
Institute for Logic, Language and Computation (ILLC)
University of Amsterdam
Nieuwe Doelenstraat 16
1012 CP Amsterdam
The Netherlands
E honing@uva.nl
I www.hum.uva.nl/mmm/hh/

Table of Contents

Abstract

This study is concerned with the question whether there is perceptual invariance of expressive timing under tempo-transformation in music performance. This is investigated by asking listeners to compare an original audio recording with a tempo-transformed (or time-stretched) version, using an online experimental design. The results show that a significant proportion of the respondents could identify (and preferred) an original recording. The results are taken as evidence for the timing-is-tempo-specific hypothesis, and counter evidence for the relational invariance hypothesis.

Introduction

An important theoretical issue in cognitive science is the study of 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). Also in music perception it has been a 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, 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 (i.e. being 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 do not generally recognize proportionally scaled rhythms as being identical when scaled to another tempo (Desain, Jansen & Honing, 2000; Handel, 1993).

Another aspect of music that has been studied for perceptual invariance is expressive timing in music performance (Gabrielsson, 1999; Palmer, 1997). The existing studies on perceptual invariance of timing (Repp, 1994; 1995; Reed, 2003) present, however, rather inconclusive evidence. Repp (1994) used a comparison task using tempo-transformed and original MIDI performances and found the responses to be barely above chance level. Repp (1995), however, finds a small, but significant effect of tempo in a subjective rating task. Another, yet preliminary

study (Reed, 2003) found no effects of tempo in an identification task, but some in a rating and ranking task (see General Discussion for more details on these three perceptual studies).

By contrast, the music performance literature presents much stronger, but nevertheless still conflicting evidence. Some authors showed that global tempo does influence the use of expressive timing (Friberg & Sundström, 2002; Desain & Honing, 1994) — at different tempi different structural levels become salient and this has an effect on the expressive freedom and variability observed (see Clarke, 1999). Other studies have shown quite the opposite and found expressive timing to be (roughly) invariant under tempo transformation (Repp, 1994; 1995). This phenomenon was interpreted as *relational invariance* (or proportional duration), a key concept in research on timing control in skilled motor performance (Heuer, 1991; Viviani & Laissard, 1991; Gentner, 1987).

The present study investigates whether expressive timing is perceptually invariant under tempo transformation in a variety of musical repertoires, aiming to resolve this rather undecided issue in music perception.

Two, relatively large-scale experiments (Experiments 1 and 2) were conducted using fragments from commercially available audio recordings from a variety of musical repertoires. Both experiments included original and tempo-transformed versions of these audio recordings and tested whether listeners were able to identify the original recording by focusing on the use of expressive timing.

Experiment 1

The aim of the first experiment was to get an insight of the effect of tempo and musical genre on the identification of an original recording.[1] The participants were asked to compare five pairs of sound examples and 1) to indicate whether it was an original recording or a tempo-transformed version (i.e. a slowed-down or speeded-up version of the original), referred to as *comparison task*, and 2) to indicate which tempo they preferred (i.e. what they considered the most convincing performance), referred to as *preference task*. The sound examples were chosen from a variety of musical genres with different amounts of tempo-transformation (or time-stretching) applied. The experiment used forced-choice responses for the comparison and preference tasks, and open responses for motivation and additional comments.

For the comparison task two hypotheses will be considered: the *relational invariance hypothesis* (Repp, 1994) and the *timing-is-tempo-specific 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 or tempo-transformed fragments: if expressive timing can be scaled proportionally with global tempo, both sound fragments (i.e. the original and tempo-transformed version) will sound equally *natural* — the respondents will consider both versions musically possible and/or convincing performances. On the other hand, if a significant proportion of the respondents is able to identify the original correctly, this will be support for the timing-is-tempo-specific hypothesis. Since expressive timing is the only aspect of the audio recording that is altered — all tempo-transformed fragments are time-stretched with a constant factor, keeping all other aspects of the performance (e.g., pitch, timbre) identical —, it must be the timing that functions as a perceptual cue in deciding whether a performance is an original or tempo-transformed version. If evidence is found for this hypothesis, it would be support for the idea

that expressive timing is intrinsically related to global tempo in music performance — scaling an original recording to another tempo (i.e. simply slowing it down or speeding it up proportionally) makes the expressive timing sound awkward or *unnatural*.

In addition, the preference task was used to see whether there is a significant difference in responses between the identification of an original and actually preferring the tempo of the recording (i.e. considering the excerpt musically convincing). Although this task was not considered central in this study, it offered a subjective rating independent of whether the stimuli were judged to be an original recording or not.[2]

Finally, as a more informal description of the main hypothesis (i.e. timing is tempo-specific) it could be illustrated with a parallel (in a metaphorical sense) from motion perception in film. Think, for instance, of the early b/w films featuring Buster Keaton or Charlie Chaplin. In films of that period, movements, like walking, often look a bit awkward. This is actually caused by a difference in the speed of recording and that of the projection (using a higher frame rate in projection). Interestingly, our perception tells us, immediately but indirectly, that something is wrong with the rate of the projection. Indirectly, because we perceive the *timing* of the movements (e.g., walking) to be strange, and we deduce from that that the *tempo* (or rate of projection) must be wrong. If the timing of walking movements (cf. expressive timing in music performance) would be invariant with rate (cf. global tempo in music performance) we would not have noticed anything peculiar.[3]

METHOD

Participants

The participants (*N* = 174) responded to an invitation that was sent to a variety of professional mailings lists, including members of the "auditory list" (http://www.auditory.org/), the European Society for the Cognitive Sciences of Music (ESCOM), and students of musicology from the University of Amsterdam. Of the respondents 52% reported to be an "expert (musician)", 35% to be "experienced (listen a lot to music)", and 11% to be "average (listen casually to music)". The experiment took on average 17 minutes to complete.

Equipment, internet support and audio file formats

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 used are sound 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.[4] Alternatively, the stimuli were made available in the (at the time of this study still more widely available) ".wav" format (however, producing files that take considerably longer to load). Of all respondents 87% had software installed that could play MPEG4 files, 13% of the participants used the alternative ".wav" format.

< Insert Table 1 around here >

To test the technology involved (e.g., its workings on a variety of computer platforms and operating systems) a preliminary experiment was run (using the same setup as Experiment 1) among a smaller group of participants (N = 36), mostly colleagues and students from the University of Amsterdam and Northwestern University. Furthermore, it provided an opportunity to get feedback on the clarity of the instructions, effective tempo ranges and the level of difficulty. This test group (i.e. beta-testers) did not participate in Experiment 1, but was invited to take part in Experiment 2.

The experimental setup and stimuli were generated using POCO (Honing, 1990). Standard HTML-Forms[5] were used to collect the responses, custom-made software (CGI and Java scripts) was used to inform respondents of the (intermediate) results in graphical form, and standard e-mail facilities were used to generate automated feedback to the respondents, including their response-form and detailed information on the excerpts used.[6]

Materials and stimulus preparation

The stimuli consisted of five original recordings and five tempo-transformed versions of these originals (see Table 1).[7] The tempo-transformed versions were made using commercial time-stretching software (ASD, manufacturer: Roni Music).[8] All stimuli were processed using the same equalization and signal processing settings ("Type III", i.e. highest quality). The original recordings were zero time-stretched with the same software to minimize differences in sound quality between the original recordings and those tempo-transformed.

The stimuli used in the pilot were selected from a variety of musical styles, including classical, jazz, and popular music. The order (original or tempo-transformed version first), direction of the transformation (slower or faster), and amount of time-stretching (10, 15 or 20%) 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 Table 1).

Procedure

Participants were asked to visit the webpage of the online experiment.[6] There they could select either a Dutch or an English language version of the 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, the participants were instructed 1) to listen –as often as needed– to a pair of sound examples focusing on the use of timing and tempo in each recording, and 2) to answer the questions listed below them. The two questions presented were "Which is the original recording?" and "Which tempo do you prefer (i.e. musically, as a convincing performance)?" The response was forced-choice (either A or B). In addition, the participants were asked to motivate their choice (see Figure 1).

< *Insert Figure 1 around here* >

At the end of the experiment the participants were asked to evaluate the pilot experiment using a short multiple-choice questionnaire: "How did you like the experiment?" ("Challenge", "OK", or "Boring"), "What was the level?" ("Difficult", "Average", or "Easy"), and "What kind of listener are you?" ("Expert (musician)", "Experienced (listen a lot to music)", or "Average (listen casually to music)"). And finally, "Do you have any comments?" allowed for more general remarks.

Analysis

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

RESULTS AND DISCUSSION

Evaluation of the experiment by the respondents

Of all respondents (N = 174) 47% evaluated the experiment to be a "challenge", 50% judged it "ok", and 3% found it "boring". Furthermore, 40% judged the experiment as "difficult", 53% as "average" and 7% as "easy". So, in general, it can be concluded that the respondents liked to participate in the experiment and found it not too difficult a task.

Comparison and preference task

The results of the comparison task ("Which is the original recording?") are shown in Table 2, the results for the preference task ("Which tempo do you prefer musically?") in Table 3. Both are depicted in Figure 2. For the comparison task (see Table 2) it can be seen that, in general, the participants were able to distinguish significantly (one-tailed binomial test) between an original and a tempo-transformed (or time-stretched) recording.

However, the results for the Jazz stimulus pairs were marginally significant, that for the Minimal stimulus pairs significant, but in the unexpected (i.e. opposite) direction.
With regard to the responses to the Jazz stimulus pairs, it might be that the amount of tempo-transformation applied (10% faster) was too small to be able to distinguish between the original and stretched version. By comparison, earlier perceptual studies used a fixed amount tempo-

transformation (20% in either direction in Reed, 2003) or a range of tempi (roughly 23% slower and 44% faster in Repp, 1995). For the Minimal stimulus pairs it might be the lack of expressive timing ("metronomical" timing being typical for the style) that caused participants not to be able to identify the original. Instead they apparently selected the tempo they preferred (i.e. the slower version; cf. preference task).

< Insert Table 2 around here >

< Insert Figure 2 around here >

The results for the comparison task, besides clear with respect to the overall effect of tempo on identification, suggested two main improvements for the follow-up experiment (Experiment 2): to fix the amount of time-stretching to a reasonable scale (i.e. 20%), and to restrict the stimuli to sound examples from musical genres that are known for their use of expressive timing.

For the preference task (see Table 3) it can be seen that, in general, the participants had a clear preference for either one of the tempi presented (two-tailed binomial test). To see whether the results for the preference task are different from the comparison task, a significance test on two proportions was performed. Contrary to what was expected, only the responses for the Jazz stimulus pairs showed a significant difference (one-tailed binomial test; $p < 0.05$).
These results indicate that there were few interpretable differences between the two tasks. The relatively small individual differences and, overall, only one significant difference, suggests that the respondents preferred the tempo of what they thought was an original recording.

< Insert Table 3 around here >

Finally, the actual correctness of the responses was investigated. On average three out of five correct identifications were made (M = 2.90, SD = 1.12; for a detailed overview see Table 4).

< *Insert Table 4 around here* >

Qualitative responses (motivation)

The motivation given by the respondents was only informally analyzed. Some examples of the motivation given to describe the comparison task were: "X had a more natural feeling", "X has more energy and vibrancy" "X seems too slow and sluggish", "X sounds too fast, uncomfortable pacing for the music." "X just sounds better", "X invites dancing", or "X sounds like tripping over itself". These qualitative responses confirm that the participants indeed focused on the quality of the expressive timing in the sound examples used.

The motivation was also analyzed for remarks on possible artifacts of the time-stretching method used in Experiment 1. Of all respondents (N = 174) four participants (2.3%) made a remark on audible artifacts, especially on the excerpt containing snippets of voice (i.e. "3. Soul"). However, only one of them identified all excerpts correctly (the others three). As a consequence, in Experiment 2 no fragments with voice were included in the stimuli, and the tempo-scale was fixed and chosen to guarantee optimal sound quality.

Furthermore, two respondents made a remark about being uncomfortable with forced-choice. In Experiment 2 therefore a confidence rating was added.

The final question ("Do you have any comments?") showed that, in general, the respondents found it a challenge to identify the original recording. Some examples of responses are: *"My first reaction was 'how can you tell, if you don't know the song?' but after listening a second time, I did form an opinion about each of them"* (identified excerpts 1, 2, and 5 correctly), *"Experiment is fun, but somewhat frustrating. The differences seem subtle in most cases."* (identified excerpts 1 and 2 correctly), and *"I've offered no motivation because I have none, except that one tempo sounds 'right' "* (identified excerpts 1, 2, and 3 correctly).

No further systematic analysis was done on this qualitative data, except using the textual responses as a way of filtering the occasional unserious participant from the invitation list for Experiment 2.

Experiment 2

The aim of the second experiment was to systematically study the effect of tempo on the identification of an original recording in two musical genres: "Jazz" and "Classical". As in Experiment 1, the participants were asked to listen to a number of sound examples and to indicate whether it was an original recording or a tempo-transformed version (i.e. a slowed-down or speeded-up version of the original). However, with regard to Experiment 1 a number of aspects were changed and/or improved: all tempo-transformed sound excerpts were time-stretched by the same amount (either 20% faster or slower), a larger set of sound examples was used (ten in each musical genre), all responses were forced-choice (no open responses) and a confidence scale was added. And finally, all excerpts were individually judged (i.e. not explicitly compared, as in Experiment 1).

The experiment came in two versions: one used recordings from the Jazz repertoire, the other fragments from the Classical repertoire. Except for the stimuli used, the design of both versions was identical.

METHOD

Participants

From the 174 listeners that participated in Experiment 1, 124 responded to an invitation to take part in Experiment 2. Three Gift Certificates were raffled among all who submitted their responses within four weeks of being invited.

Of all respondents, 76 participated in the Classical version of the experiment (N = 76). Of these 49% reported to be a "expert (musician)", 51% "experienced (listen a lot to music)", and none to be of the category "average (listen casually to music)". In addition, of all respondents, 48 participated in the Jazz version of the experiment (N = 48). Of these 52% reported to be a "expert (musician)", 48% "experienced (listen a lot to music)", and none to be of the category "average (listen casually to music)". The experiment took on average 11 minutes to complete.

Equipment

Same as Experiment 1.

Materials and stimulus preparation

The experiment came in two versions, Jazz and Classical, using different stimuli but an identical design. The stimuli consisted of five original recordings and five tempo-transformed versions of these originals (see Tables 5 and 6). The tempo-transformed versions were made using commercial time-stretching software (ASD, manufacturer: Roni Music).[8] All stimuli were processed using the same equalization and signal processing settings ("Type III", i.e. highest quality). The original recordings were zero time stretched with the same software to minimize differences in sound quality between the original recordings and those tempo-transformed. 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 5 and 6). The resulting ten stimuli were presented in random order and blocked per artist.

< Insert Table 5 around here >

< *Insert Table 6 around here* >

Procedure

Participants were asked to visit the webpage of the online experiment.[9] 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 with (or like to listen too). Finally, the participants were instructed 1) to listen –as often as needed– to a single sound example, focusing on the use of timing and tempo — as if they were a judge in a music performance master class, and 2) to answer the questions listed below them. The questions presented were "Is this an original recording?" (response categories "Yes" or "No") and "Are you sure?" (response categories "Yes", "Somewhat" or "No")(see Figure 3). Furthermore, they could review their judgments, before sending the response form. Finally, the participants were asked to evaluate the experiment (same as Experiment 1).

< *Insert Figure 3 around here* >

Analysis

Same as Experiment 1.

RESULTS

Evaluation of the experiment by the respondents

From the all respondents 76 decided to participate in the Classical version of the experiment. Of these 53% evaluated the experiment to be a "challenge", 43% judged it "ok", and 4% found it "boring". With respect to the level of the experiment 61% judged the experiment as "difficult", 28% as "average" and 11% as "easy". And finally, 95% of the participants indicated that they liked to be invited for a possible follow-up study.

From the all respondents 48 decided to participate in the Jazz version of the experiment. Of these 58% evaluated the experiment to be a "challenge", 42% judged it "ok", and none found it "boring". With respect to the level 58% judged the experiment as "difficult", 35% as "average" and 6% as "easy". And finally, 98% of the participants indicated that they liked to be invited for a possible follow-up study.

From these results it can be concluded that the participants, in general, enjoyed doing the experiment and found it a challenging task.

Classical results

The results of the identification task ("Is this an original recording?") are shown in Table 7 and in Figure 4. It can be seen that, as in Experiment 1, listeners can correctly identify the original. All responses are moderately to highly significant (one-tailed binomial test). There are however two exceptions: an original recording by Richter and Gould. The difference in responses is in the right direction but they are non-significant. This could well be caused by the rather idiosyncratic styles of both pianists that makes it hard to judge the "naturalness" of the performance.

< *Insert Table 7 around here* >

< *Insert Figure 4 around here* >

With respect to correctly identifying an original, the respondents were most confident in the case of Barenboim (on average .76), and least confident in the case of Gould (.50). As can be seen in Table 7, confidence rates show higher values for originals as compared to tempo-transformed versions. This supports the idea that listeners might perceive an original to be more convincing than a tempo-transformed version, in the latter case introducing more doubt as to whether it could be intentionally timed as such.

With respect to the correctness of the responses the participant is this study did slightly better than in Experiment 1 (M = 3.21, SD = 1.11; for details see Table 8). This is probably due to the fact that in this experiment they were judging a style they reported to be familiar with.

< *Insert Table 8 around here* >

Jazz results

The results of the identification task ("Is this an original recording?") are shown in Table 9 and in Figure 5. Here as well, listeners seemed to be able to correctly identify the original; All responses are highly significant (one-tailed binomial test). In comparison to the Classical version of the experiment, the results in the Jazz version are more pronounced. Suggesting that, indeed, in jazz, expressive timing plays an even more important role: expressive timing cannot just be scaled to another tempo without sounding awkward.

There is, however, one intriguing exception: the fragment performed by the Mehldau trio was identified in the unexpected (i.e. opposite) direction. This means that the tempo-transformed version was judged by a significant majority to be an original, and vice versa.[10] The fragment was taken from a live recording of a composition that was originally recorded at another tempo. This suggests a number of interpretations of this peculiar mix-up. The fragment used could have been perceived as relatively loose and using a non-typical type of timing.[9] On closer inspection, there is a considerable amount of asynchrony between piano, bass and drums. In a slower version these timing variations are lessened and the slower tempo could therefore be preferred. Another interpretation could be that the participants were familiar with the piece and/or the studio recording and, in addition, were not able to base their judgment on the timing alone, and therefore used a tempo preference instead. However, the experimental design does not allow to distinguish between these effects.

< Insert Table 9 around here >

< Insert Figure 5 around here >

With respect to the overall correctness of the responses the participants is this study did better than in Experiment 1 (M = 3.27, SD = 1.01; see for details Table 8). Like in the Classical version, this is probably due to the fact that in this experiment participants were judging a style they reported to be familiar with.

General Discussion

The two experiments reported in this article were concerned with the question whether there is perceptual invariance of expressive timing under tempo-transformation in music performance.

This was investigated by asking listeners to compare an original audio recording with a tempo-transformed (time-stretched) version (Experiment 1), and to indicate whether a recording was an original or tempo-transformed version (Experiment 2). The results showed that listeners can, on the basis of timing alone, decide on whether a recording is an original performance or not. By judging the "naturalness" of the expressive timing used, listeners were able to identify the original audio recording. Since expressive timing was the only musical parameter manipulated, the participants must have used expressive timing as a perceptual clue for whether something was a real or artificially scaled recording.

Interestingly, since the expressive timing in the tempo-transformed stimuli was in fact relational invariant with the original (timing was scaled proportionally with tempo, using a time-stretching algorithm) the *relational invariance hypothesis* (i.e. timing is perceptually invariant under tempo transformation; see Repp, 1994) predicts no preference for the original over the tempo-transformed version. As said, this contradicts the experimental results of the present study: listeners were, in most cases, able to identify (Experiments 1 and 2) and generally prefer (Experiment 1) the original over the tempo-transformed version. This was taken as evidence for the *timing-is-tempo-specific hypothesis* in large variety of musical repertoires (most notably Jazz and Classical music). Furthermore, this confirms what has been found in several music performance studies (Clarke, 1982; Desain & Honing, 1994; Palmer, 1997; Clarke, 1999).

These results might come as no surprise to musicians. In the wider music literature there is often spoken of how to select the appropriate tempo and how and when to apply the appropriate timing (Rink, 1995). Musicians tend to adapt their timing to the tempo used, bringing out other structural levels of the music at different tempi (see Clarke, 1999). Besides changing the depth of the expressive timing (relative modulation depth or amount of *rubato*) — which still could be

proportional to the timing at a slower tempo (cf. Repp, 1995) — also the timing patterns themselves change significantly (Clarke, 1982, 1999; Honing, 2005).

As a concrete example, Friberg & Sundström (2002) showed that the swing-ratio in jazz performance (the typical timing pattern of consecutive eight notes) does not stay the same —as the relational invariance hypothesis would predict—, but found that this ratio changes with tempo. Apparently, to produce the same sense of swing at different tempi, the ratio between consecutive notes has to be adapted. However, whether a swing-ratio has to be changed with tempo to give the same sense of swing in perception has not been systematically studied as yet.

Still, with respect to the music performance literature, we are left with some support for the relational invariance hypothesis. One explanation could be the influence of musical genre or repertoire on the contradicting results. 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 (Friberg & Sundström, 2002; Desain & Honing, 1994).

An acknowledged problem in music performance studies is that all kinds of stylistic and idiosyncratic issues can interfere with the phenomenon studied. As Repp (1995) noted, it is unrealistic to expect a performer to perform a piece of music identically at different tempi — as such questioning the idea of studying invariance of timing in performance methodologically. Therefore, perception is a far more direct way of testing relational invariance in music performance.

As was briefly discussed in the introduction, a number of studies have systematically studied perceptual invariance in music performance (Repp, 1994; 1995; Reed, 2003). Repp (1995) found

a small, but significant effect of tempo in subjective judgments using an experimental design in which ten pianists listened to manipulated MIDI performances played back on an electronic keyboard. In a more recent, but preliminary study with the same group size but using audio fragments (Reed, 2003), rather inconclusive evidence was found. No effect of tempo was found in an identification task but some effects in a subjective rating and rating task. However, this exploratory study didn't report any statistical tests to confirm these interpretations.

These rather inconclusive results in the music perception literature might be caused by the relatively small number of participants taking part, as well as some problems in the methodology applied.

With regard to the latter, Repp (1995) used a MIDI performance of a single pianist that was tempo-transformed and rated by a panel. The tempo-transformation method used (see Repp, 1995 for details) included several "regularizations" applied to, for example, onset asynchronies and articulation. All these could well interfere with the perceived quality of the performance, and, arguably, caused the responses to be less receptive for an "natural/unnatural" judgment. In that sense, audio recordings (as used in Reed, 2003 and the current study) can be considered more ecologically-valid stimuli.[11]

With respect to the number of participants typically used in these type of perceptual experiments, it should be noted that is not uncommon to use just ten subjects (as was the case in the studies mentioned). However, the recent advances of internet technology and the possibility to playback high quality audio on a wide-variety of computer platforms allows for collecting a considerably higher number of responses than usual, and especially with categorical frequency data can give far more significant results. The experimental design might serve as an example of

how to use standard technologies in music perception and performance using ecologically-valid stimuli.

In addition, the present study can also be seen as an evaluation of the state-of-the-art time-stretching technology. It suggests that time-stretching algorithms might need additional information in order to keep the quality of the original timing similar under tempo transformation. Recent sound signal processing research is indeed focusing on such enhancements (Gomez et al., 2003), trying to incorporate structural and stylistic knowledge to make the tempo-transformation sound more natural.

Finally, the present study showed that relational invariance is, in general, too simplistic a model of the interaction between expressive timing and global in music performance. It suggests the need of richer models of expressive timing and tempo than might be currently considered (cf. Honing, 2002; 2004; 2005).

Acknowledgments

I thank all participants (and especially the beta-testers) for their enthusiasm in doing the experiment, and for the suggestions they provided. Marijke Engels (Department of Psychology, University of Amsterdam) is thanked for her advice on the statistical analyses.

References

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.

Clarke, E.F. (1999). Timing in the performance of Erik Satie's 'Vexatations.' *Acta Psychologica*, 50, 1-19.

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* Keele, UK: Keele University, Department of Psychology.

Desain, P., & Honing, H. (1994). Does Expressive Timing in Music Performance Scale Proportionally with Tempo? *Psychological Research*, 56(4), 285-292.

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

Friberg, A., & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19(3), 333-349.

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

Gentner, D. R (1987) Timing of Skilled Motor Performance: Tests of the Proportional Duration Model. *Psychological Review*. Vol. 94 (2), pp. 255-276.

Gomez, E., Grachten, M., et al. (2003) Melodic characterization of monophonic recordings for expressive tempo transformations. In *Proceedings of Stockholm Music Acoustics Conference.*

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.* 364-368. San Francisco: Computer Music Association.

Honing, H. (2002) Structure and interpretation of rhythm and timing *Tijdschrift voor Muziektheorie.* 7(3), 227-232.

Honing, H. (2004) Computational modeling of music cognition: a case study on model selection. ILLC Prepublication PP-2004-14.

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.

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 (ESCOM).* 558-561.

Repp, B. H., Windsor, W. L., & Desain, P. (2002) Effects of tempo on the timing of simple musical rhythms. *Music Perception*, 19(40), 565-593.

Repp, B.H. (1994) Relational invariance of expressive microstructure across global tempo changes in music performance: An explorative study. *Psychological Research*, 56(4), 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.

Rink, J. (Ed.) (1995) *The Practice of Performance: Studies in Musical Interpretation*. Cambridge: Cambridge University Press

Shepard, R. & Levitin, D. (2002) Cognitive psychology and music. In Levitin, D. (Ed.) *Foundations of Cognitive Psychology: Core Readings*. Cambridge, MA: MIT Press.

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

Viviani. P. & Laissard, G. (1991) Timing control in motor sequences. In J. Fagard & P.H. Wolff (Eds.) *The development of timing control and temporal organisation in coordinated action* (pp. 1-36). Amsterdam: Elsevier.

Tables

Table 1. Recordings used in Experiment 1

Genre	Artist: Album, Composition	Record Label	Duration (s)
1 Baroque	Glenn Gould: J.S. Bach, Two-part Inventions and three-part Sinfonias, Sinfonia 9 in F minor, BWV 795	Sony SMK 52 596, 1993	37
2 Jazz	Bradford Marsalis Quartet: Requiem, Bullworth	Columbia 069655 2, 1999	39
3 Soul	James Brown: Out of Sight, Funky Drummer	Polydor 589297-2, 2002	23
4 Minimal	Nurit Tilles and Edmund Niemann: Steve Reich, Pianophase	Nonesuch 979 169-2, 1987	59
5 Samba	Paulinho da Viola: Brasil: A Century Of Song, Cenarios	Blue Jacket 5002-2, 1995	10

Table 2. Results comparison task (*N* = 174)

Excerpt	Original/Stretched	Number	(%)
1 Baroque	Original	102	(58.6)*
	Stretched (-15%)	72	(41.4)*
2 Jazz	Original	92	(52.9)
	Stretched (-10%)	82	(47.1)
3 Soul	Original	113	(64.9)***
	Stretched (-20%)	61	(35.1)***
4 Minimal	Original	71	(59.2)+
	Stretched (20%)	103	(40.8)+
5 Samba	Original	127	(73.0)***
	Stretched (-15%)	47	(27.0)***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; + significant, but in unexpected direction (see text)

Table 3. Results preference task (N = 174)

Excerpt	Original/Stretched	Number	(%)
1 Baroque	Original	88	(50.6)
	Stretched (-15%)	86	(49.4)
2 Jazz	Original	75	(43.1)*
	Stretched (-10%)	99	(56.9)*
3 Soul	Original	105	(60.3)**
	Stretched (-20%)	69	(39.7)**
4 Minimal	Original	79	(45.4)*
	Stretched (20%)	85	(54.6)*
5 Samba	Original	132	(75.9)***
	Stretched (-15%)	42	(24.1)***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4. Number of correct identifications

# Correct	Count
5	10
4	45
3	60
2	38
1	19
0	2

Table 5. Recordings used in the Classical version of Experiment 2

Pianist	Composition	Record Label	Duration (s)
1 Glenn Gould	J.S. Bach, English Suite nr.3 in G minor, Gavotte, BWV 808	Sony SK87765, 2001	10
2 Daniel Barenboim	L. van Beethoven, Piano Sonata no. 8 in C minor, Op. 13 (Pathéthique), Rondo	EMI 7243 5 57762 0 4, 1995	54
3 Sviatoslav Richter	J.S. Bach, English Suite Nr. 6 in D minor, Gavotte I, BWV 811	Delos GH 5601, 2004	24
4 Alfred Brendel	L. van Beethoven, Variation I over Nel cor più non mi sento, WoO 70	Philips 432 093-2, 1991	23
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

Table 6. Recordings used in the Jazz version of Experiment 2

Artist	Composition, Album (Musicians)	Record Label	Duration (s)
1 Geri Allen	Invisible, In the year of the Dragon (with Charlie Haden and Paul Motian)	Polygram. 1989	29
2 Yuri Honing	Seven (with Paul Motian, Gary Peacock and Paul Bley)	JIM 75086, 2001	39
3 Brad Mehldau	It might as well be spring, Progression (with Larry Grenadier and Jorge Rossy)	Warner Bros 9362-48005-2, 2001	21
4 Carla Bley	Chicken, Songs With Legs (with Andy Sheppard and Steve Swallow)	Watt / ECM, 1995	39
5 Eric Dolphy	Miss Ann, Last Date (with Misha Mengelberg, Han Bennink and Jacques Schols)	Verve / Limelight, 1964	39

Table 7. Results identification task (*N* = 76)

Pianist	Original/Stretched	Original?	Number	(%)		Confidence
1 Glenn Gould	Original	Yes	47	(61.8)	*	0.50
		No	29	(38.2)	*	0.50
	Stretched (20%)	Yes	27	(35.5)	**	0.59
		No	49	(64.5)	**	0.57
2 Daniel Barenboim	Original	Yes	57	(75.0)	***	0.76
		No	19	(25.0)	***	0.58
	Stretched (-20%)	Yes	20	(26.3)	***	0.65
		No	56	(73.7)	***	0.71
3 Sviatoslav Richter	Original	Yes	40	(52.6)		0.55
		No	36	(47.4)		0.53
	Stretched (20%)	Yes	31	(40.8)	*	0.45
		No	45	(59.2)	*	0.52
4 Alfred Brendel	Original	Yes	56	(73.7)	***	0.60
		No	20	(26.3)	***	0.37
	Stretched (-20%)	Yes	16	(21.1)	***	0.44
		No	60	(78.9)	***	0.53
5 Glenn Gould	Original	Yes	44	(57.9)		0.57
		No	32	(42.1)		0.52
	Stretched (20%)	Yes	31	(40.8)	*	0.53
		No	45	(59.2)	*	0.58

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 8. Number of correct responses in both versions of Experiment 2

# Correct	Classical	Jazz
5	11	5
4	19	15
3	28	18
2	13	8
1	5	2
0	0	0

Table 9. Results identification task (*N* = 48)

Artist	Original/Stretched	Original?	Number	(%)		Confidence
1 Geri Allen	Original	Yes	35	(72.9)	**	0.69
		No	13	(27.1)	**	0.50
	Stretched (-20%)	Yes	6	(12.5)	***	0.42
		No	42	(87.5)	***	0.85
2 Yuri Honing	Original	Yes	39	(81.2)	***	0.71
		No	9	(18.7)	***	0.67
	Stretched (-20%)	Yes	9	(18.7)	***	0.67
		No	39	(81.2)	***	0.64
3 Brad Mehldau	Original	Yes	11	(22.9)	+	0.55
		No	37	(77.1)	+	0.50
	Stretched (20%)	Yes	37	(77.1)	+	0.57
		No	11	(22.9)	+	0.55
4 Carla Bley	Original	Yes	28	(58.3)		0.70
		No	20	(41.7)		0.57
	Stretched (20%)	Yes	19	(39.6)		0.53
		No	29	(60.4)		0.64
5 Eric Dolphy	Original	Yes	44	(91.7)	***	0.72
		No	4	(8.3)	***	0.25
	Stretched (20%)	Yes	4	(8.3)	***	0.50
		No	44	(91.7)	***	0.77

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; + significant, but in unexpected direction (see text)

Figure Captions

Figure 1. Fragment of the online interface of Experiment 1.

Figure 2: Results of Experiment 1 (*N* = 174). The left panel shows the responses on the comparison task, the right panel those on the preference task.

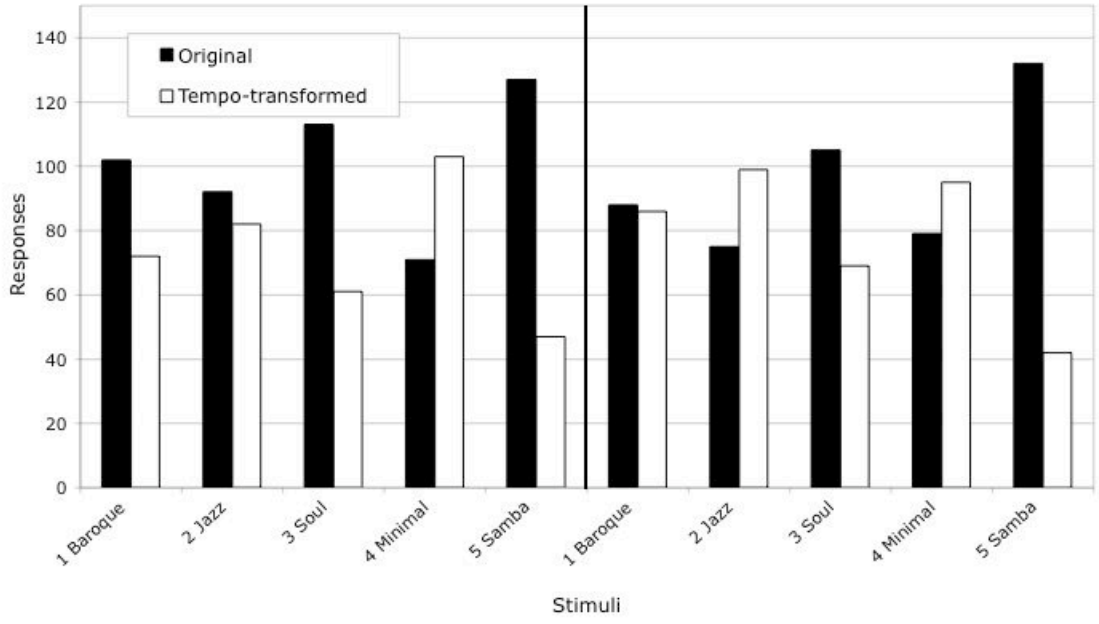
Figure 3. Fragment of the online interface of Experiment 2.

Figure 4. Results of the Classical version of Experiment 2 (*N* = 76). An * in the stimulus-label refers to an original recording, a < and a > respectively to a slower and faster tempo-transformed version.

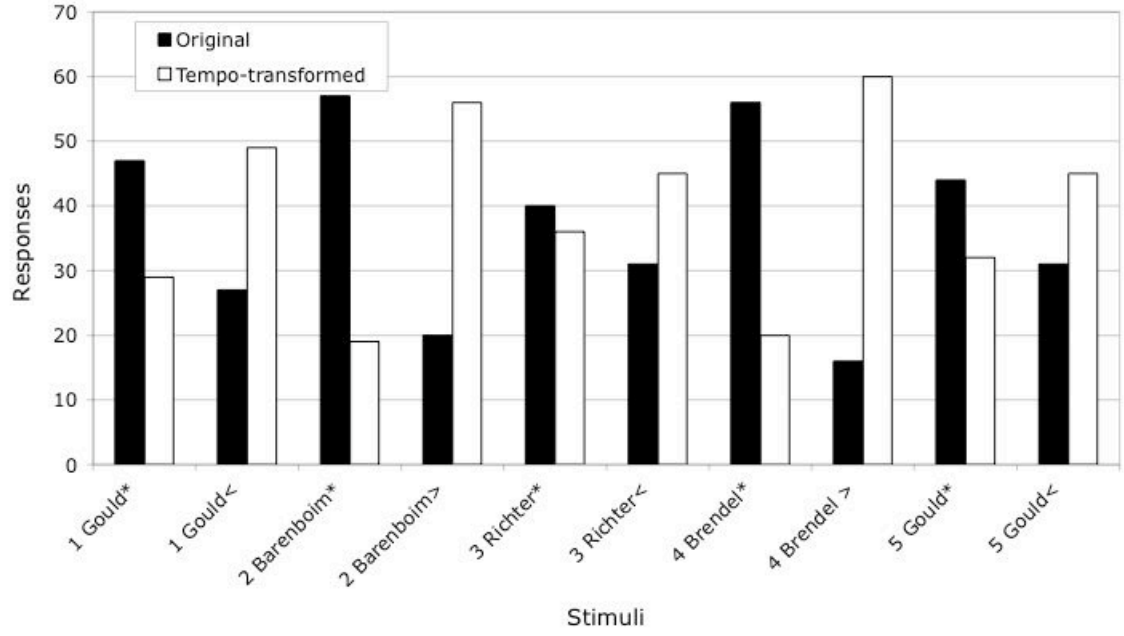
Figure 5. Results of the Jazz version of Experiment 2 (*N* = 48). An * in the stimulus-label refers to an original recording, a < and a > respectively to a slower and faster tempo-transformed version.

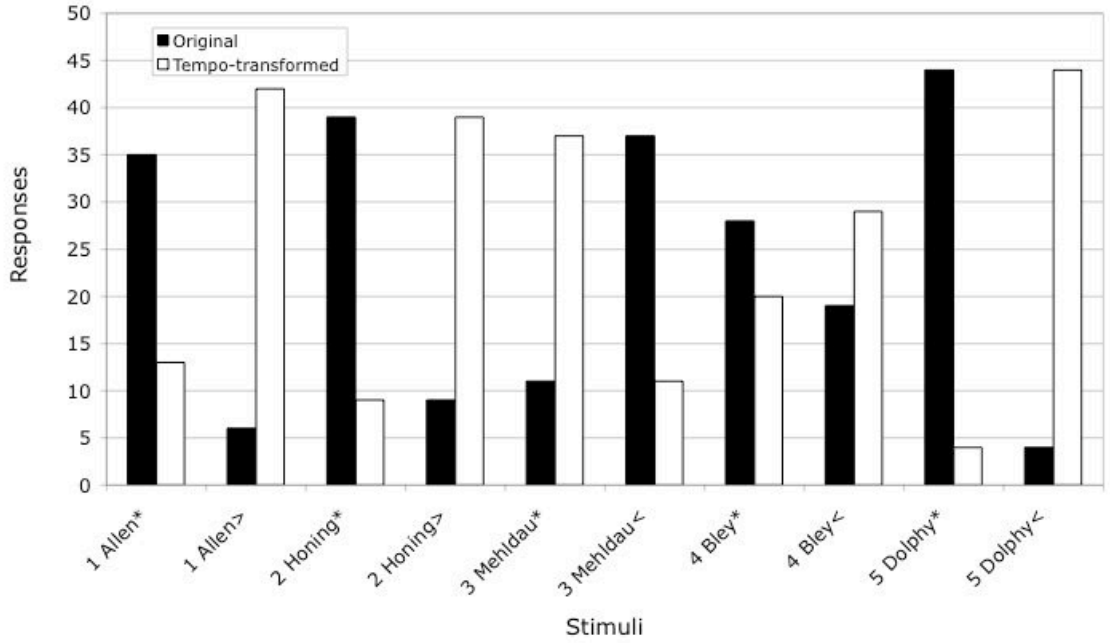
Figures


Question 2 (of 5):
A:
(alternative format 2a.wav)
B:
(alternative format 2b.wav)
2a. Which is the original recording?
A
B
Check one that applies
2b. Motivation
Motivation for selection
2c. Which tempo do you prefer
A
B
Check one that applies
2d. Motivation
Motivation for selection


Original
Tempo-transformed
Responses
0
20
40
60
80
100
120
140
1 Baroque
2 Jazz
3 Soul
4 Minimal
5 Samba
1 Baroque
2 Jazz
3 Soul
4 Minimal
5 Samba
Stimuli


Excerpt 2 (of 10):
Alternative format: 2.wav
2a. Is this an original recording?
Yes
No
2b. Are you sure?
Yes
Somewhat
No


70
60
50
40
30
20
10
0
Responses
Original
Tempo-transformed
1 Gould*
1 Gould<
2 Barenboim*
2 Barenboim>
3 Richter*
3 Richter<
4 Brendel*
4 Brendel >
5 Gould*
5 Gould<
Stimuli


Original
Tempo-transformed
Responses
50
45
40
35
30
25
20
15
10
5
0
1 Allen*
1 Allen>
2 Honing*
2 Honing>
3 Mehldau*
3 Mehldau<
4 Bley*
4 Bley<
5 Dolphy*
5 Dolphy<
Stimuli

Footnotes

[1] Experiment 1 was presented to the participants as a pilot experiment, and it was, in fact, also used as such: next to giving an indication of the effect of genre, it provided a number of improvements that were implemented in a follow-up study (Experiment 2).

[2] We cannot relate the responses to the notion of *preferred tempo* (Fraisse, 1957), since there are too few systematically tempo-transformed stimuli used in the experiment as to be able to test this.

[3] Interestingly, this informal example might actually be seen as counter-evidence, at least in perception, for the relational invariance hypothesis in motor behavior (see Heuer, 1991).

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

[5] See http://www.w3.org/.

[6] Experiment 1 can be found at http://www.hum.uva.nl/mmm/exp/.

[7] The stimuli are available at http://www.hum.uva.nl/mmm/exp1/.

[8] See http://www.ronimusic.com.

[9] The stimuli are available at http://www.hum.uva.nl/mmm/exp2/.

[10] This was not, simply, due to mix-up of stimuli, neither that the recording itself was manipulated.

[11] It has to be noted that current quality of time-stretching techniques were not readily available at the time of the perceptual studies mentioned.