TIMING IS TEMPO-SPECIFIC*

Henkjan Honing
Music Cognition Group
ILLC / University of Amsterdam
www.hum.uva.nl/mmm

ABSTRACT

This study is concerned with the question whether there is perceptual invariance of expressive timing under tempo-transformation in audio recordings. This was investigated by asking listeners to distinguish between an original recording and a time-stretched (i.e. tempo-transformed) version. The original recordings were identified by a significant proportion of the participants. The results suggest that expressive timing can function as a clue in identifying a real performance. This is taken as evidence for the *tempo is timing-specific hypothesis*, and counter evidence for the *relational invariance hypothesis* that predicts proportionally scaled expressive timing to be perceived natural as well. The results are discussed in the context of whether there is perceptual invariance of expressive timing under tempo transformation. These findings suggest the need for improvements to state-of-the-art time-stretching algorithms.

1. INTRODUCTION

Perceptual invariance is an important theoretical issue in cognitive science. It concerns the study of whether, and if so, how certain event properties remain perceptually invariant under transformation (e.g., [20]). Also for computer music software it is a relevant topic since it will influence the ease with which perceptually convincing transformations of musical data can be supported.

A well-known and uncontroversial example of perceptual invariance in music is melody. 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. As such, melody remains perceptually invariant under transposition. Sequencer and notation software take advantage of this characteristic, and hence the transposition transformation is easily supported.

With respect to other aspects of music, such as rhythm, supporting transformations is less trivial. While one might expect rhythm to scale proportionally with tempo (i.e. being perceptually invariant under tempo transformation), several empirical studies have shown that this is not always the case (e.g., [8]). Rhythms are timed differently at different tempi ([17]), and listeners do not generally recognize proportionally scaled rhythms as being identical when scaled to another tempo ([3], [9]).

However, the relation between timing and tempo has long be assumed perceptual invariant, both in computer music and music cognition research. Most sequencers have a tempo controller, suggesting timing to be scalable with tempo. It is a result of representing timing and tempo-change in computer music systems as a continuous function of score position (a so-called *tempo curve*; [4], [11]). While such a representation captures the tempo deviations as measured in a recording, it in fact also suggests that the shape of a tempo curve is independent of the number of events (or note density), the rhythmic structure (i.e. differentiated durations), and the overall tempo of the performance. However, a simple test, like changing the tempo of a recorded track of a drummer playing a certain *groove*, will reveal to a listener that timing cannot be simply represented like that: a tempo-transformation will sound awkward ([11]).

2. THIS STUDY

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.* A relatively large-scale experiment ($N = 307$) was 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 in those performances.

3. EXPERIMENT

3.1. Aim

The aim of the experiment was to systematically study the effect of tempo on the identification of an original recording in two musical genres: "Jazz" and "Classical". The participants were asked to listen to a number of sound examples and to indicate whether it was an original recording or a time-stretched version (i.e. a slowed-down or speeded-up version of the original), referred to as *identification task*. All tempo-transformed sound excerpts were time-stretched by the same amount (either 20% faster or slower), ten sound examples were used for

*This is research in progress (March 2005). Related and more elaborate studies are available as [13] and Honing (in preparation), see www.hum.uva.nl/mmm under 'Publications'.

each musical genre, all responses were forced-choice (no open responses), and a confidence scale was used.

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.

3.2. Hypotheses

For the identification task 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 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 (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 an 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; [5], [6]) 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.

3.3. A parallel: motion in early film

The main hypothesis (i.e. tempo-specific timing) could be informally illustrated with an example from motion perception in film. Think, for instance, of early films featuring, .e.g., Buster Keaton. 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.

3.4. Method[1]

3.4.1. 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.

3.4.2. Internet support

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 experimental setup and stimuli were generated using POCO ([10]).

3.4.3. 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 (in MPEG4 format). The tempo-transformed versions were constructed using commercial time-stretching software (ASD, manufacturer: Roni Music).[2] All stimuli were processed using the same equalization and signal processing settings. 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. The resulting ten stimuli were presented in random order and blocked per artist.

3.4.4. 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 *a)* 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 (ignoring possible timbral artifacts), and *b)* to answer the questions listed on the screen (see Figure 1).

Figure 1. Fragment of the online interface.

[1] Due to space restrictions, for a details see [13].

[2] The manufacturer could not provide any information on the signal processing method used; Future experiments will use an even more advanced time-stretching algorithm ([2]).

4. RESULTS

4.1. Results for the Classical version

The results of the identification task ("Is this an original recording?") are shown in Figure 2. It can be seen that the participants ($N = 175$) could correctly identify the original. All responses are highly significant.

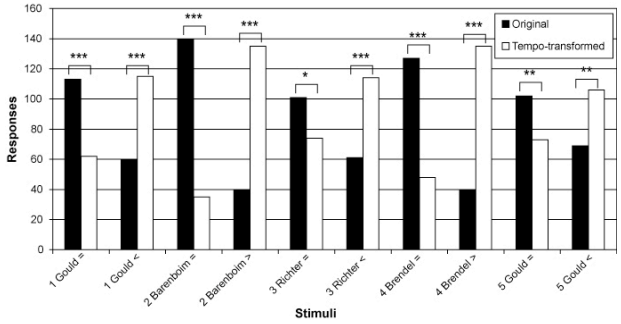

Figure 2. Results of the Classical version of the Experiment ($N = 175$). An = in the stimulus-label refers to an original recording, a < and a > respectively to a slower and faster tempo-transformed version (* $p<0.05$; ** $p<0.01$; *** $p<0.001$).

4.2. Results for the Jazz version

The results of the identification task ("Is this an original recording?") are shown in Figure 3. Here as well, the participants ($N = 132$) seemed to be able to correctly identify the original; All responses significantly differ from chance (one-tailed binomial test), except for one recording (discussed below). 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 (e.g., *swing* or *groove*) cannot just be scaled to another tempo without sounding awkward.

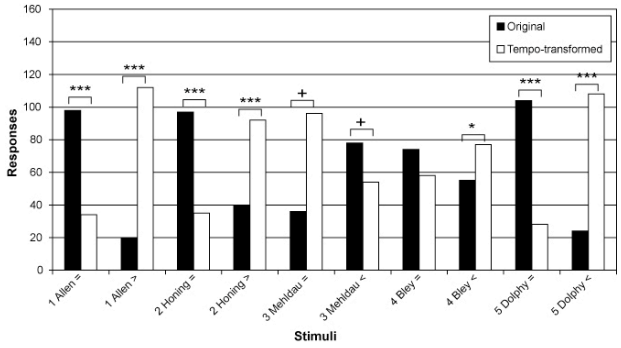

Figure 3. Results of the Jazz version of the Experiment ($N = 132$). An = in the stimulus-label refers to an original recording, a < and a > respectively to a slower and faster tempo-transformed version (* $p<0.05$; ** $p<0.01$; *** $p<0.001$; + significant misidentification).

As mentioned, there was one intriguing exception: the jazz fragment performed by the Mehldau trio. This was significantly misidentified. This means that the tempo-transformed version was judged by a significant majority to be an original, and vice versa. However, it is unclear what caused this peculiar mix-up.[3]

Finally, the overall results of the Classical version of the experiment were significantly different from the null hypothesis ($\chi^2[9] = 122.50$, $p<.0001$). In the Jazz version of the experiment this is even more pronounced ($\chi^2[9] = 133.12$, $p<.0001$). This suggests a stronger sensitivity to the use of expressive timing as a cue in recognizing an original recording for the jazz repertoire.

5. SUMMARY AND DISCUSSION

The experiment reported in this short paper was 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 show that a significant majority of the participants could identify an original performance, and more clearly so for sound excerpts from the Jazz than the Classical repertoire.

Since the expressive timing in the tempo-transformed stimuli was in fact relationally invariant with the original stimuli, the *relational invariance hypothesis* predicts no preference for the original over the tempo-transformed version. This contradicts the experimental results of the present study: listeners were, in most cases, able to identify the original versus the tempo-transformed version. This is taken as evidence for the *tempo-specific timing hypothesis* in jazz and classical music: expressive timing, when applied at the appropriate tempo, can function as a clue in identifying an original performance.

However, some alternative explanations of the reported results have to be considered. One could be the possible artifacts of the signal processing method that may have helped the identification of the tempo-transformed stimuli (cf. footnote 2). While the parameter settings and tempo range used were carefully decided on to minimize artifacts (using the results of a pilot study; [13]), and listeners were instructed to focus on the expressive timing (not on possible timbral artifacts), we cannot be sure this had at least some effect. However, arguing the other way around, if artifacts would have 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.

Another factor that could have influenced the results is that listeners may have based their judgments on tempo preference, instead of whether expressive timing was used in a musically convincing way. While the latter was instructed, it may be that some listeners, when in doubt, simply selected the tempo they preferred. However, a further experiments are needed to separate between these important factors.

Nevertheless, these results might come as no surprise to musicians. In the music literature one often

[3] The stimuli used can be found at http://www.hum.uva.nl/mmm/exp1

finds discussions of how to select the appropriate tempo, and how and when to apply the appropriate timing. Musicians tend to adapt their timing to the tempo used, bringing out other structural levels of the music at different tempi (see [1]). 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. [19]) — also the timing patterns themselves change significantly (cf. [15]). Furthermore, the results confirm what has been found in several in music performance studies ([1], [5], [6]).

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 (e.g., [7], [20]), trying to incorporate structural and stylistic knowledge to make a tempo-transformation sound more natural. Computational models of rhythm perception might be of good use in such research ([11], [14]).

In conclusion, the present study shows that relational invariance (cf. a tempo controller on a MIDI sequencer) is, in general, too simplistic a model of the interaction between expressive timing and global tempo in music performance. It suggests the need for richer models of expressive timing and tempo than might be currently considered in computer music systems ([11], [14], [15]).

6. ACKNOWLEDGEMENTS

I would like to thank the participants —and especially all beta-testers— for their enthusiasm and for the suggestions they provided. Thanks to Marijke Engels of the Department of Psychology, University of Amsterdam for her advice.

7. REFERENCES

[1] 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.

[2] Bonada, J. (2000). Automatic Technique in Frequency Domain for Near-Lossless Time-Scale Modification of Audio. *Proceedings of International Compute Music Conference*. San Francisco: Computer Music Association.

[3] Desain, P. & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3), 341-365.

[4] Desain, P., & Honing, H. (1991). Tempo curves considered harmful. A critical review of the representation of timing in computer music. In Proceedings of the *International Computer Music Conference*. San Francisco: Computer Music Association.

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

[6] 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.

[7] Gomez, E., Grachten, M., Amatriain, X., & Arcos, J. L. (2003) Melodic characterization of monophonic recordings for expressive tempo transformations. In *Proceedings of Stockholm Music Acoustics Conference*. CD-ROM.

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

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

[10] 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.

[11] Honing, H. (2001) From time to time: The representation of timing and tempo. *Computer Music Journal*, 35(3), 50-61.

[12] Honing, H. (2002) Structure and interpretation of rhythm and timing *Dutch Journal of Music Theory*, 7(3), 227-232.

[13] Honing, H. (2004a) Is timing tempo-specific? An online internet experiment on perceptual invariance of timing in music. *ILLC Prepublication* PP-2004-34. [http://dare.uva.nl/en/record/140322]

[14] Honing, H. (2005a, under review) Computational modeling of music cognition: a case study on model selection. *ILLC Prepublication* PP-2004-14. [http://dare.uva.nl/en/record/120423]

[15] Honing, H. (2005b, under review) Is expressive timing relational invariant under tempo transformation? *Music Perception*.

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

[17] 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.

[18] 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.

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

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

[21] Grachten, M., Arcos, J., & López de Mántaras, R. (2004) TempoExpress, a CBR Approach to Musical Tempo Transformations. In Advances in Case-Based Reasoning. Proceedings of the 7th European Conference, *Lecture Notes in Computer Science*. Springer.