

THE EFFECT OF MICROTIMING DEVIATIONS ON THE PERCEPTION OF GROOVE IN SHORT RHYTHMS

MATTHEW DAVIES
INESC TEC, Porto, Portugal

GUY MADISON
Umeå University, Umeå, Sweden

PEDRO SILVA & FABIEN GOUYON
INESC TEC, Porto, Portugal

GROOVE IS A SENSATION OF MOVEMENT OR WANTING to move when we listen to certain types of music; it is central to the appreciation of many styles such as Jazz, Funk, Latin, and many more. To better understand the mechanisms that lead to the sensation of groove, we explore the relationship between groove and systematic microtiming deviations. Manifested as small, intentional deviations in timing, systematic microtiming is widely considered within the music community to be a critical component of music performances that groove. To investigate the effect of microtiming on the perception of groove we synthesized typical rhythm patterns for Jazz, Funk, and Samba with idiomatic microtiming deviation patterns for each style. The magnitude of the deviations was parametrically varied from nil to about double the natural level. In two experiments, untrained listeners and experts listened to all combinations of same and different music and microtiming style and magnitude combinations, and rated liking, groove, naturalness, and speed. Contrary to a common and frequently expressed belief in the literature, systematic microtiming led to decreased groove ratings, as well as liking and naturalness, with the exception of the simple short-long shuffle Jazz pattern. A comparison of the ratings between the two listener groups revealed this effect to be stronger for the expert listener group than for the untrained listeners, suggesting that musical expertise plays an important role in the perception and appreciation of microtiming in rhythmic patterns.

Received: January 30, 2012, accepted November 3, 2012.

Key words: groove, microtiming, movement, rhythm, listening experiment

GROOVE IS A CENTRAL ASPECT OF MUSIC PER-ception and appreciation, closely connected to the main functional uses of music; namely, dance, drill, and ritual. When seeking to find a relationship between music and the behavior that groove induces—synchronization and coordination—the temporal properties of the music signal might be crucial to our understanding of groove.

To work towards a formal definition of groove, Janata, Tomic, and Haberman (2012) asked subjects participating in a synchronization experiment to provide their own written descriptions. Based on an analysis of frequently occurring words given by the participants, the authors derived the following definition: “Groove is that aspect of the music that induces a pleasant sense of wanting to move along with the music” (p. 56). This crowd-sourced definition bears strong similarities to the operational definition given by Madison (2006): “wanting to move some part of the body in relation to some aspect of the sound pattern” (p. 201)—a definition that we follow in this paper. Since groove is a pleasant and attractive feature of music, one goal of this research is to understand the physical properties so that we can, in some way, add groove to music. Given that people synchronize movement to musical stimuli (Pressing, 2002), in addition to the utilitarian aspect of putting knowledge of groove to some purpose, we are also interested in the fundamental question of the role of groove in music. A central issue to consider is what physical properties affect the sensation of groove. While these physical properties are presently unknown, the relationship between groove and some global aspects of music signals has been examined; for example, tempo (Janata et al., 2012), musical style (Janata et al., 2012; Madison et al., 2011), rhythmic patterning and timbre (Zeiner-Henriksen, 2010), and beat salience and event density (Madison et al., 2011). In this paper we test a musical property related to performance and expression: microtiming deviations.

A key feature of human performance of music is variability. Expression in music performance can be demonstrated through timing, dynamics, timbre, and pitch, inasmuch as the instrumentation allows for that (for a review, see Gabrielsson, 1999). Indeed, expression in the

form of rubato is considered mandatory in Romantic music (Repp, 1995). Both the modelling of expression, and techniques for adding expression to musical performances, have been widely studied (e.g., Friberg, Bresin, & Sundberg, 2006; Widmer, Dixon, Goebel, Pampalk, & Tobudic, 2003; Windsor & Clarke, 1997).

Expression arises from musicians' interpretation of what is present in the musical score. For some forms of expression (e.g., significant changes in dynamics) there may be notated guidelines within the score, while other forms of expression may be too small to be formally notated. One such form of expression, microtiming, falls into this second category. Manifested as small temporal deviations from strict notated (or quantized) time (Bilmes, 1993), "unsystematic" microtiming can occur due to motor noise in performance or from variability in the time-keeper, whereas so-called "systematic" microtiming is the result of deliberate manipulation of timing by musicians, as shown in the analysis of professional Jazz drummers who are able to exercise precise, repeated control over timing in performance (Freeman & Lacey, 2002; Honing & de Haas, 2008). The presence and structure of microtiming has been studied across many musical styles including Jazz (Iyer 2002; Waadeland, 2001), Funk (Freeman & Lacey, 2002; McGuinness 2005), Cuban (Alén 1995; Bilmes, 1993), and Samba (Naveda, Gouyon, Leman, & Guedes, 2011; Wright & Berdahl, 2006).

In the wider musical context, microtiming is considered important for musical engagement; without it, performance is thought to sound dull and lifeless (Hellmer, 2006). Hence, in commercial music sequencing software and drum machines, microtiming is a central component of "humanize" functions that add expression by applying (unsystematic) timing deviations to quantized temporal events (Hennig et al., 2011). Beyond adding human-like qualities to sequenced music recordings, microtiming has been studied in terms of a more formal and important role in music, wherein it is an expected feature of certain musical styles and aids listeners in understanding musical structure (Repp, 1998). In groove-based music such as Jazz and Funk, the presence of microtiming is considered key to creating the rhythmic tension necessary to evoke the sensation of groove (Iyer, 2002; Keil, 1995). However, a recent study that measured the correlation between musical descriptors extracted from audio signals and listeners perception of groove (Madison et al., 2011) revealed systematic microtiming to be negatively correlated with groove ratings.

Thus, there apparently is an inconsistency in the literature regarding the role of microtiming for groove.

Most existing research suggests that microtiming is a positive aspect of musical performance and plays an important role in facilitating groove (e.g., Keil, 1995), yet the results of Madison et al. (2011) suggest the opposite. However, this relationship between groove and microtiming was not measured in a systematic way as groove ratings were not recorded across a range of microtiming manipulations in the stimuli. Therefore the principal aim of this study is to experimentally test whether microtiming patterns do indeed facilitate groove. To this end we devised a listening experiment designed to give favorable conditions for a positive relationship between microtiming and groove to emerge while retaining high ecological validity and full experimental control. The basis of the experiment was to have listeners rate their perception of groove in rhythmic sequences that had been manipulated in terms of their microtiming properties.

To explore the domain of microtiming deviations in a comprehensive manner, the stimuli were created to vary across three conditions: music style, microtiming style, and microtiming magnitude. The music styles were selected on the basis of existing research into microtiming: Samba, Funk, and Jazz. For Samba, we built upon the analysis of Naveda et al. (2011), who analyzed global microtiming distributions across a large collection of Samba excerpts. For Funk, we followed research on the timing of the eight-bar drum break of "Funky Drummer" by James Brown (Freeman & Lacey, 2002; Stewart, 2000; McGuinness, 2005, Greenwald, 2002). For Jazz, we leveraged research into the canonical swing pattern (Friberg & Sundström, 2002; Waadeland, 2001). To explore the effect of the amount of microtiming, we varied microtiming magnitude across four conditions: deadpan (zero magnitude), understated microtiming (too small to be noticeable), microtiming at the expected magnitude per music style, and exaggerated microtiming (twice the expected magnitude).

To create the musical stimuli to which the microtiming patterns could be applied, a rhythmic pattern with style-specific percussion instruments was created using MIDI sequencing software. The experimental stimuli were constructed to cover all combinations of music style and microtiming pattern across the range of microtiming magnitudes. The listening experiment was conducted for two different listener groups: nonexpert listeners and musically trained participants.

Since we chose microtiming patterns that occur in real music, applied them in a range of magnitudes that covers those extant in real music, and in all other respects strived to maximize ecological validity, we hypothesized that groove ratings would be higher for

these conditions than for deadpan renditions of each music style. We also hypothesized that the actual magnitude of microtiming patterns in real music constitutes an optimal level, and that exaggerated microtiming patterns would be associated with lower groove ratings. Regarding the application of microtiming patterns typical of one music style to another music style we had no firm hypotheses, but envisaged that it also might increase groove ratings, albeit to a lesser extent than a matched music style and microtiming pattern.

Experiment 1

METHOD

Participants. Thirty-two participants (13 female, 19 male) took part in the experiment. Ages ranged from 21 and 46 years ($M = 29.3$, $SD = 6.0$) and beyond basic music lessons at school ($M = 1.5$ years musical training), none were considered musicians. Twenty-six of the participants were Portuguese; however, all participants were given instructions in English. The participants were recruited via emails sent to mailing lists at INESC TEC and the Faculty of Engineering at the University of Porto. On completion of the experiment, each participant was paid for their involvement.

Stimuli. Given our aim of investigating groove ratings for a set range of microtiming patterns with specific properties, we created a set of synthetic music examples (MEs) comprised short rhythmic patterns over which we could exercise complete and precise experimental control. With our eventual aim of wanting to explore methods for *adding* groove to music, we chose to work with synthetic music examples rather than manipulate existing musical recordings, in order to approximately replicate the conditions in which a groove-transformation could be applied (e.g., in a similar way to the “humanize” function in music sequencing software).

The MEs used in the experiment varied across three dimensions: music style (M-style), microtiming style (MT-style), and the magnitude of the largest microtiming deviation (MT-magnitude). For each rhythmic style a basic ME without any microtiming was programmed using a MIDI sequencer. To ensure the ecological validity of the MEs, a musician was enlisted to assist in constructing the patterns, setting the loudness of each event and selecting the percussion instruments specific to each music style. Once chosen, these aspects were fixed so that all subsequent MEs used in the experiment differed only in terms of microtiming deviations. The basis for selecting each M-style and how its corresponding MT-pattern was determined is now described.

Music styles and microtiming patterns. 1) Samba was chosen since it is a music style strongly associated with movement and dance, and it has also been studied in terms of its microtiming properties. To determine the microtiming pattern for Samba we follow the work of Naveda et al. (2011) who measured microtiming deviations across a large collection of Samba pieces. In particular the authors illustrated that the third and fourth sixteenth notes of each beat are typically played earlier than in purely quantized time. Statistical tests that compared the measured timing deviations from quantized time were shown to be highly significant, and on this basis we consider these timing deviations to be systematic.

To create a natural sounding ME for Samba, we followed the instrumentation and rhythmic pattern from an existing piece of music “Otau E Eu” by Nicos Jaritz. This piece is characterized by an event on every sixteenth note position of a 4/4 meter and was programmed via MIDI sequencer with appropriate percussion instrument choices for Samba, including: surdo (a large bass drum), cuíca (a Brazilian friction drum), and agogô bells (two metal bells connected by a U-shaped piece of metal). Following Naveda et al. (2011), who report microtiming deviations in Samba for an average tempo of approximately 100 bpm, we created the Samba ME at this tempo. The score notation for the Samba ME and the corresponding microtiming pattern are shown respectively in Figures 1 and 2.

2) As with Samba, Funk was selected due to its strong association with movement (Danielsen, 2006). As we are not aware of the existence of genre-wide microtiming patterns for Funk in the same way they have been shown to exist for Samba, we instead explored the microtiming pattern from a specific piece of Funk music that has been widely studied; the eight-bar drum break of James Brown’s “Funky Drummer” (Freeman & Lacey, 2002; Greenwald, 2002; McGuinness, 2005; Stewart, 2000).

For the Funk ME we worked directly from a drum transcription events that included timing deviations of the “Funky Drummer” drum break (McGuinness, 2005). The tempo of the original recording is approximately 100 bpm and was reprogrammed in MIDI at this tempo, but containing only the principal events that occurred on the sixteenth note metrical grid. All ghost notes (very soft note events) that did not occur on the metrical grid were removed.

For the Samba ME we could rely upon the outcome of statistical tests to infer whether a timing deviation was systematic or not. However, with only eight transcribed bars in the available Funky Drummer data, there was an

The figure displays a percussion score for three rhythmic styles: Samba, Funk, and Jazz. The score is organized into seven main staves, each representing a different instrument. The top two staves (Snare and Whistle) are for Samba. The next three staves (Agogos, Cuíca, and Surdo) are for Funk. The bottom two staves (Shaker and Tambourine) are for Jazz. Below these are three stacked staves for Hi-hat, Snare, and Kick, and another three stacked staves for Ride cym., Snare / Stick, Kick, and Pedal Hat. The notation includes various rhythmic patterns, such as eighth notes, sixteenth notes, and rests, with accents and dynamic markings.

FIGURE 1. Percussion score notation for each rhythmic style (top to bottom: Samba, Funk, Jazz).

insufficient number of data points to determine which deviations were systematic using the same approach. Instead, we adopted a simpler method, where for each metrical position on the sixteenth note grid, we retained only the mean deviations greater than a fixed threshold of ± 10 ms and set all timing deviations below this threshold to zero. The score notation for the Funk ME is shown in Figure 1, along with the microtiming pattern and ± 10 ms threshold in Figure 2. Note, the mean timing deviations for all metrical positions are shown, but only those above the threshold are used in the experiment.

3) Jazz has been widely studied in terms of its microtiming properties, particularly in relation to the timing

of canonical swing rhythm that characterized by an alternating “long-short” pattern of eighth notes (Freeman & Lacey, 2002). The ratio between the long and short events—the so-called “swing-ratio”—was shown to vary with tempo, with slower tempi having a higher ratio than faster tempi where the ratio approaches unity (Friberg & Sundström, 2002).

As both the Samba and Funk patterns were characterized by a sixteenth note on every position of a 4/4 metrical grid, this posed no problems for the later application of the Funk microtiming pattern onto Samba and vice versa. However, for Jazz, this constraint was problematic since a typical eighth note swung Jazz pattern is not characterized by an event on every sixteenth note.

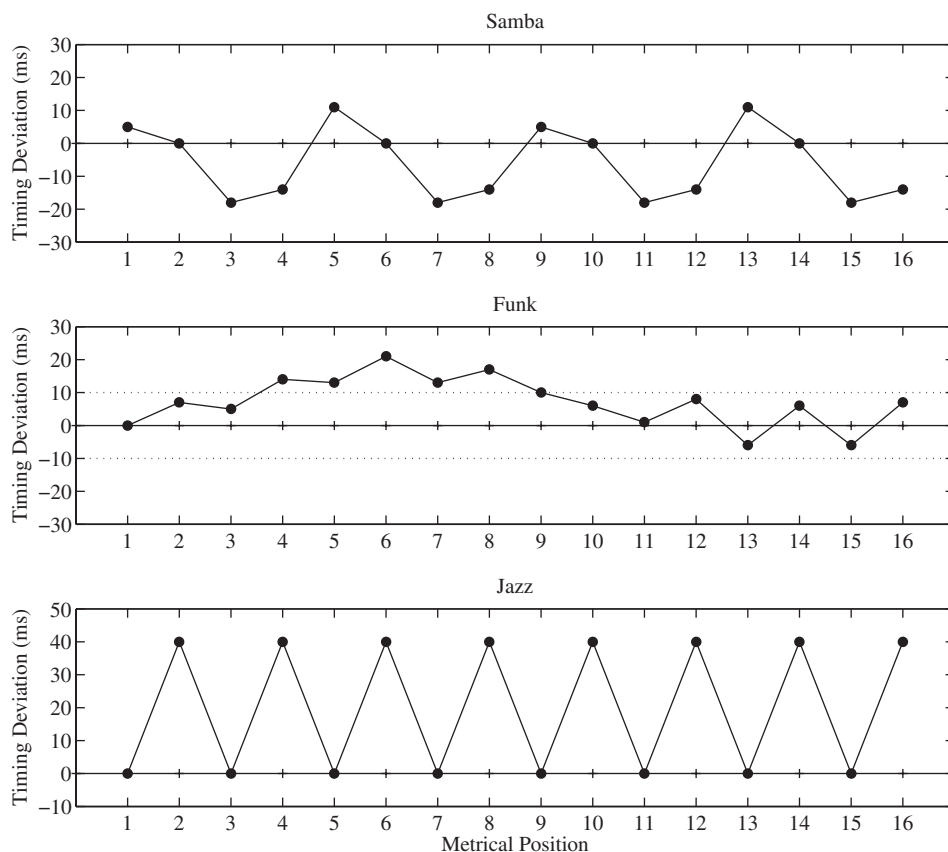


FIGURE 2. Visualization of microtiming patterns (top to bottom: Samba, Funk, Jazz).

Leaving gaps in the metrical grid for the Jazz example would mean certain metrical positions to be microtimed would have no associated event and therefore be unaltered. To address this issue, we constructed a basic Jazz pattern based around the characteristic swing pattern, with additional grace notes using kick drum, snare, and rim shot sounds to complete the quantized sixteenth note metrical grid. Consistent with the Funk and Samba examples, the Jazz example was constructed without any microtiming (i.e., in straight feel) in such a way that the later application of the Jazz microtiming pattern would create the representative swing pattern.

To generate the Jazz pattern we followed the standard “long-short” swing pattern and set an appropriate swing ratio for a sixteenth note pattern at 100 bpm; that is, equivalent to the swing ratio for an eighth note pattern at 200 bpm (Friberg & Sundström, 2002). The score notation for the Jazz ME is shown in Figure 1 along with the Jazz microtiming pattern in Figure 2.

Unsystematic microtiming deviations. To allow a comparison between the style-specific systematic

microtiming patterns described above and microtiming that could represent unintentional deviations, we also created a baseline condition using a set of unsystematic microtiming patterns. For the unsystematic case we generated a random deviation at each sixteenth note position. For these deviations to be consistent with the style-specific patterns described above, the random deviations were calculated from a uniform distribution in the range $[-1, 1]$ and then all values subsequently were normalized so that the maximum absolute deviation was 1. This allowed the deviations to be subsequently scaled to cover the required range of microtiming magnitudes, as explained in the following subsection. For each of the required conditions and for each bar of the MEs a different unsystematic microtiming pattern was created.

Magnitude of deviations. To explore the effect of the magnitude of the microtiming on groove ratings, we selected three levels to cover a reasonable range of magnitudes. We wished to have one magnitude below the perceptual threshold that we expected would be

indistinguishable from the quantized condition, one magnitude consistent with the expected largest deviation, and a magnitude that was too large and would likely produce a negative effect. To this end we selected the three maximum magnitudes as 5 ms (too small), 20 ms (approximately the mean of the maximum deviation of the Funk and Samba patterns), and 40 ms (too large). To implement the microtiming patterns at these different magnitudes, each was first normalized so that the maximum absolute deviation was 1, and then scaled linearly by each of the three magnitude levels.

While these three levels of magnitudes satisfied our criteria for Funk and Samba, the largest magnitude of 40 ms was closer to the ideal swing ratio for Jazz. To incorporate a condition of too much microtiming for Jazz, one further ME was included with the Jazz microtiming pattern with a magnitude of 80 ms. This was only used for Jazz MT applied to the Jazz MEs, with the 80 ms magnitude not used to scale any of the Funk, Samba, or unsystematic microtiming patterns. The raw microtiming patterns for each musical style along with the normalized versions are shown in Table 1 in the Appendix.

Implementation of MEs. To generate all of the permutations of the MEs required for the experiment, a script was used to apply a given microtiming pattern and magnitude to the MIDI file corresponding to a given music style. Any simultaneous events in the pattern (e.g., a hi-hat and kick drum event on the same metrical position) were identically shifted according to the microtiming pattern.

Once the modifications had been made to a MIDI file, it was then rendered to a 16-bit stereo wave file at a sampling frequency of 44,100 Hz using Timidity¹ with the appropriate sound library specific to each music style. All MEs used in the experiment and the scripts to generate them are available online.²

Rating scales. The main purpose of the experiment was to test whether microtiming patterns typical for a particular M-style increased the sensation of groove. To minimize response bias, this particular purpose and the nature of the manipulations were not disclosed to the participants, who were instead told that the study was about the more general topic of rhythm perception. This was partly the reason for including other rating scales, so that groove would not stand out as a central property. Three additional rating scales (speed, preference, and naturalness) were included to assess whether

the manipulations were valid as examples of music. Speed, in particular, was included to assess the psychometric properties of the scales and of the participants' ratings behavior, since speed has previously been found to be quite sensitively and accurately rated (Madison & Paulin, 2010).

In the experiment instructions, groove was defined as, "the sensation of wanting to move some part of your body in relation to some aspect of the music" and Naturalness was defined as, "how much the music example sounds like a typical musical performance." In response to the global question, "How well do the following words describe your experience of the music?" participants entered ratings ranging from "0" ("not at all") to "10" ("entirely"). The ratings were entered on a separate horizontal slider for each term, with the default position of the sliders set to 0. The order of the terms presented to the participants remained constant for each ME.

Design. The dependent variables were the four rating scales and the independent variables were M-style (Samba, Funk, Jazz), MT-style (Samba, Funk, Jazz, unsystematic), and MT-magnitude (0 ms, 5 ms, 20 ms, 40 ms, 80 ms). The core conditions for the experiment were constructed by all combinations of the independent variables with the MT-magnitude conditions restricted to 5 ms, 20 ms, and 40 ms. Since MT-style has no effect for an MT-magnitude of 0 ms, only a single ME was required to represent the quantized condition per M-style, and the 80 ms condition was a special case only for Jazz with Jazz MT.

We adopted a repeated-measures within participant design where each participant rated all conditions once. Conditions were presented in a different random order for each participant. During the experiment each participant was allowed to hear each ME only once, and they were required to listen to the complete ME before entering any rating information. These design constraints were incorporated into a stand-alone application developed for use in the listening test.

Procedure. Participants were tested individually; they were seated at the computer in a quiet room and listened to the MEs on headphones. Prior to starting the experiment, each participant read the experimental instructions, which included the definitions of each rating scale. They were asked if they understood the task and if they required clarification on any terminology. They were then given a consent form to read and sign in line with the ethical requirements of INESC TEC. The participants undertook a short training session in which they rated three MEs, set the volume to a comfortable level, and familiarized themselves with

¹ <http://timidity.sourceforge.net/>

² <http://smc.inescporto.pt/shakeit/data>

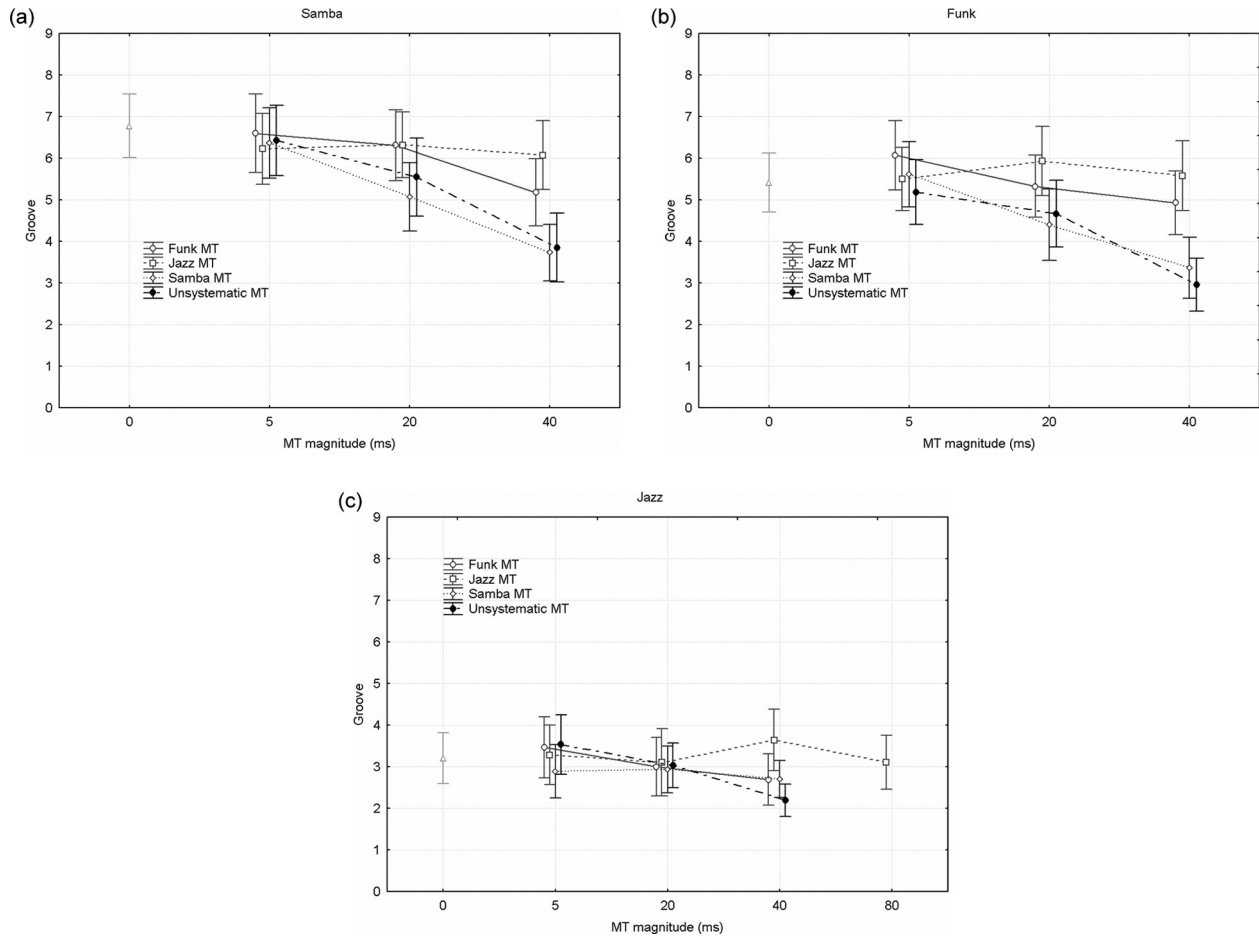


FIGURE 3. Mean groove ratings for nonexperts as a function of MT style and microtiming magnitude for (a) Samba, (b) Funk and (c) Jazz. The quantized condition has an MT magnitude of 0 ms.

how to enter their ratings using the sliders. On completion of the experiment, participants were asked about their experience with the experiment and whether they found it difficult, if the length of the experiment was appropriate, and whether they felt tired. The entire session took around 30 min to complete.

RESULTS AND DISCUSSION

The interviews indicated that all participants understood the instructions and did not find the test difficult or tiring. However, many commented that the MEs were quite repetitive, and some were concerned their opinions and ratings changed during the experiment, a factor that was mitigated against by presenting the MEs in a different random order for each participant. A summary of the groove ratings for Funk, Samba, and Jazz are shown in Figure 3a-c.

The most immediate result shown in these figures is that microtiming causes a decrease in groove for all cases, except when the MT-style is Jazz, in which case the groove ratings are largely unaffected. So, disregarding for the moment Jazz, groove ratings decrease as a function of MT magnitude, and this occurred regardless of whether the MT-style was matched to the M-style. There was no difference in this respect between the systematic and unsystematic MT patterns. Confidence intervals indicate that for Jazz, only 40 ms unsystematic MT brought about a significant decrease (Cohen’s $d' = 0.71$).

While this general trend was similar for Samba and Funk, the Jazz ratings were quite low, meaning that any negative tendency for groove to decrease with MT-magnitude may have been compressed. The globally lower ratings for the Jazz MEs created a significant, although trivial, style interaction. For both Funk and

Samba, the decrease was significant for Samba MT and Unsystematic MT at both 20 and 40 ms ($d' = 0.51, 0.37, 1.02, \text{ and } 1.22$, respectively), with the addition of Funk MT at 40 ms for Samba ($d' = 0.72$).

Thus far all results contradict our hypothesis that MT leads to increased groove ratings. Given that there is a true positive effect, possible alternative explanations for this result could be that: (1) participants did not perform the task correctly, (2) the MT was not perceived, (3) the MT made the MEs sound unnatural, and (4) participants perceived the MT but were unable to appreciate its groove-related properties (e.g., because of poor musical skill). Related to the first two points, participants systematically rated MEs to be slower on the speed rating scale as a function of MT, except for Jazz music style. Ratings decreased linearly as a function of MT magnitude, and differed on the order of 1 rating scale unit between 0 and 40 ms for all MT styles except Jazz MT ($p < .05$ according to confidence intervals). This is consistent with previous research showing that music with higher event density is perceived as faster even though the beat tempo is the same (e.g., Madison & Paulin, 2010). The crucial property of MT is that the pattern is particular, rather than the general patterning afforded by the metrical levels, and that it hence may induce stronger or more frequent grouping of events. Conceivably, more grouping of discrete events into larger chunks will lead to lower subjective event density, and hence lower ratings of speed. Of course, the present data offer no evidence that this is actually what is going on, but that is of no consequence for the argument that systematic effects of MT magnitude on both speed and groove ratings strongly indicate both that the participants perceived the MT and performed the task correctly.

That the effects of these quite subtle manipulations give such strong effects attest that listeners exerted acute perceptual ability, contrary to the idea that they did not perform the task accurately. To assess the second point we surveyed the naturalness ratings, as seen in Figure 4a-c. The naturalness ratings are not in general low indeed, a mean rating of about 7 on an 11-point scale must be considered very high. This indicates that the MEs were perceived as valid examples of real music. The ratings largely follow the same pattern as for groove, and decrease up to four scale steps for Samba and unsystematic MT applied to Samba and Funk. The decrease for Jazz style was less pronounced, but from a lower mean rating for the deadpan version. While this can be interpreted such that MT makes the MEs sound unnatural, it nevertheless shows that deadpan MEs were perceived as quite natural.

The third point that participants did not perceive the MT is refuted by the previous two arguments. The fourth point cannot be addressed by Experiment 1, because although participants obviously perceived the MT, they might nevertheless lack the musical skill required to appreciate the specifically groove-related properties of the MT patterns applied.

Experiment 2

METHOD

In order to address the principal discussion point arising from Experiment 1, that of whether nonexpert listeners were unable to appreciate the subtle microtiming deviations present in the MEs, we repeated the experiment with expert participants.

Participants. Eighteen participants (1 female, 16 male, 1 preferred not to say) took part in the experiment. Ages ranged from 22 to 71 years ($M = 38.1, SD = 11.5$). The expert participants recruited to take part in the experiment were known to the authors and targeted specifically based on their musical expertise, in particular for having a background in musical performance in at least one of the main music styles present in the experiment. Overall, the expert participants had $M = 22.5$ years of music training and performance.

In contrast to Experiment 1 where a call for participation was sent to various email lists, each expert participant was recruited by a personal email. As before, care was taken not to reveal the aim of the experiment beyond it being about rhythm perception. The expert participants were not paid for their participation.

Stimuli and design. The MEs, rating scales, and experimental design of Experiment 2 was identical to that of Experiment 1.

Procedure. Because sufficient numbers of expert listeners could not be recruited to take the listening in person in Porto (where Experiment 1 took place), we chose to conduct a web-based version of the experiment.³

To create similar listening conditions to those in Experiment 1, the participants were instructed to take the listening test seated in a quiet environment using high quality headphones and to complete the experiment in one session. Prior to starting the experiment, participants were given the same set of instructions to

³ While concerns have been raised over the use of web-based experiments (e.g., see Honing & Reips, 2008, for discussion), we took great care to exactly replicate the main functionality of the software used in the Experiment 1. We undertook substantial tests across different web-browsers to ensure the web interface operated correctly for all users.

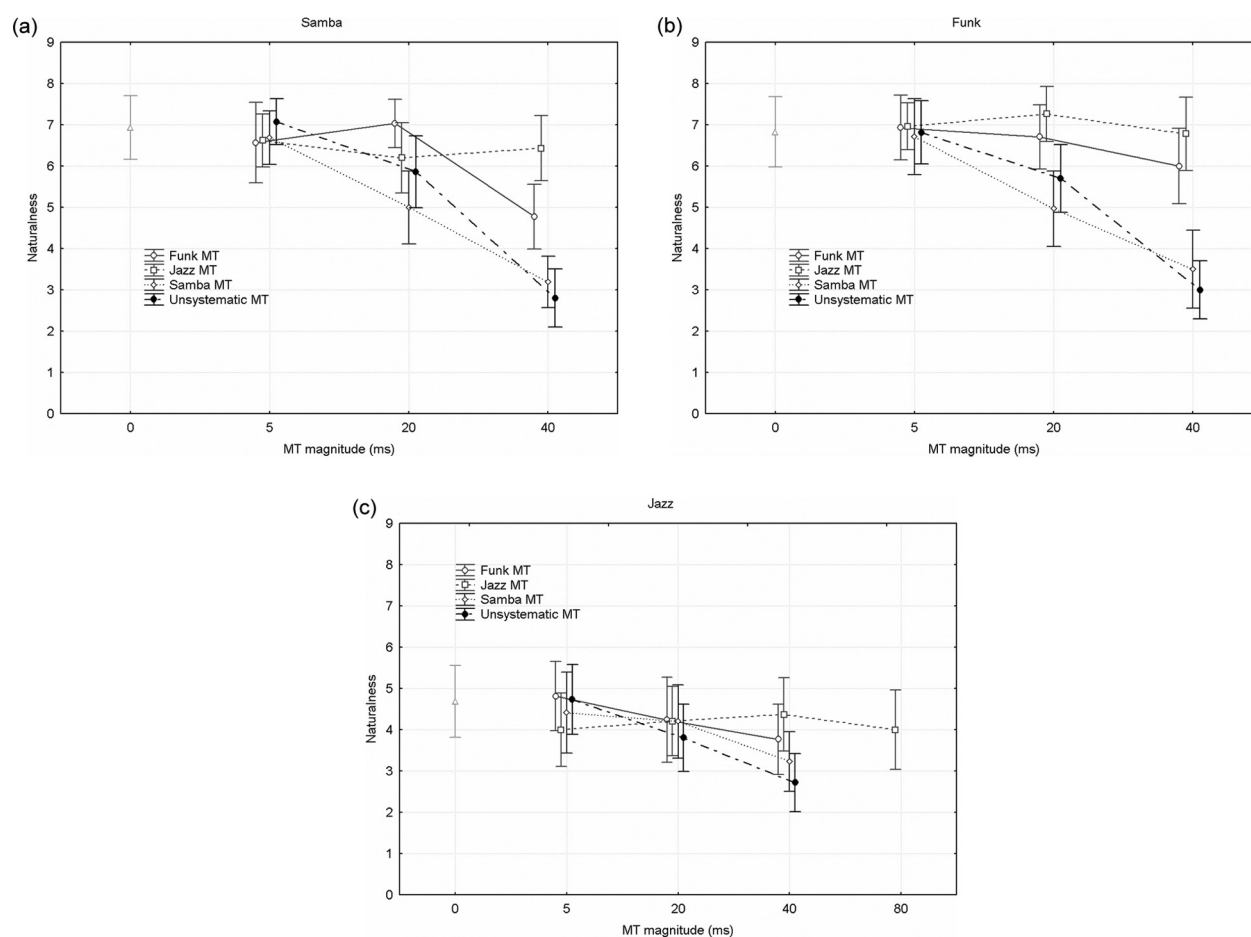


FIGURE 4. Mean naturalness ratings for nonexperts as a function of MT style and microtiming magnitude for (a) Samba, (b) Funk and (c) Jazz. The quantized condition has an MT magnitude of 0 ms.

read from Experiment 1, and also were informed that if they did not wish to complete the experiment they were free to stop at any time. In addition, the participants were asked to rate their level of musical expertise (“layman,” “knowledgeable,” or “expert”) in the following music styles: Jazz, Rock, Funk, Pop, Samba, Reggae, and Folk. The additional music styles not used in the experiment were included to disguise the main three M-styles used to create the MEs.

As with Experiment 1 there was a short training phase to allow participants to familiarize themselves with the operation of the experimental interface, and to set the playback volume to a comfortable level. On completion of the training phase, the participants then took the main listening experiment. Since it was not possible to interview the participants on completion of the experiment, a text window was included on the final page of the web-interface to allow participants to provide feedback on their experience.

To allow reproduction of the web-based version of the experiment, complete source code is available online.⁴

RESULTS AND DISCUSSION

Since the experiment with expert listeners was conducted via the Internet, it was not possible to interview participants in the same way as had been done with the nonexperts in Experiment 1. However, through web interface feedback, participants raised questions over the definition of naturalness given that all MEs were clearly synthetic. This observation was not made by the nonexpert listeners, but beyond this comment, none reported the experiment as tiring or difficult. The summary plots of groove ratings for the expert listeners for Funk, Samba and Jazz are shown in Figure 5a-c.

Inspection of the mean groove ratings for each M-style shows the same overall effect as for the

⁴ <https://github.com/SMC-INESC/weve>

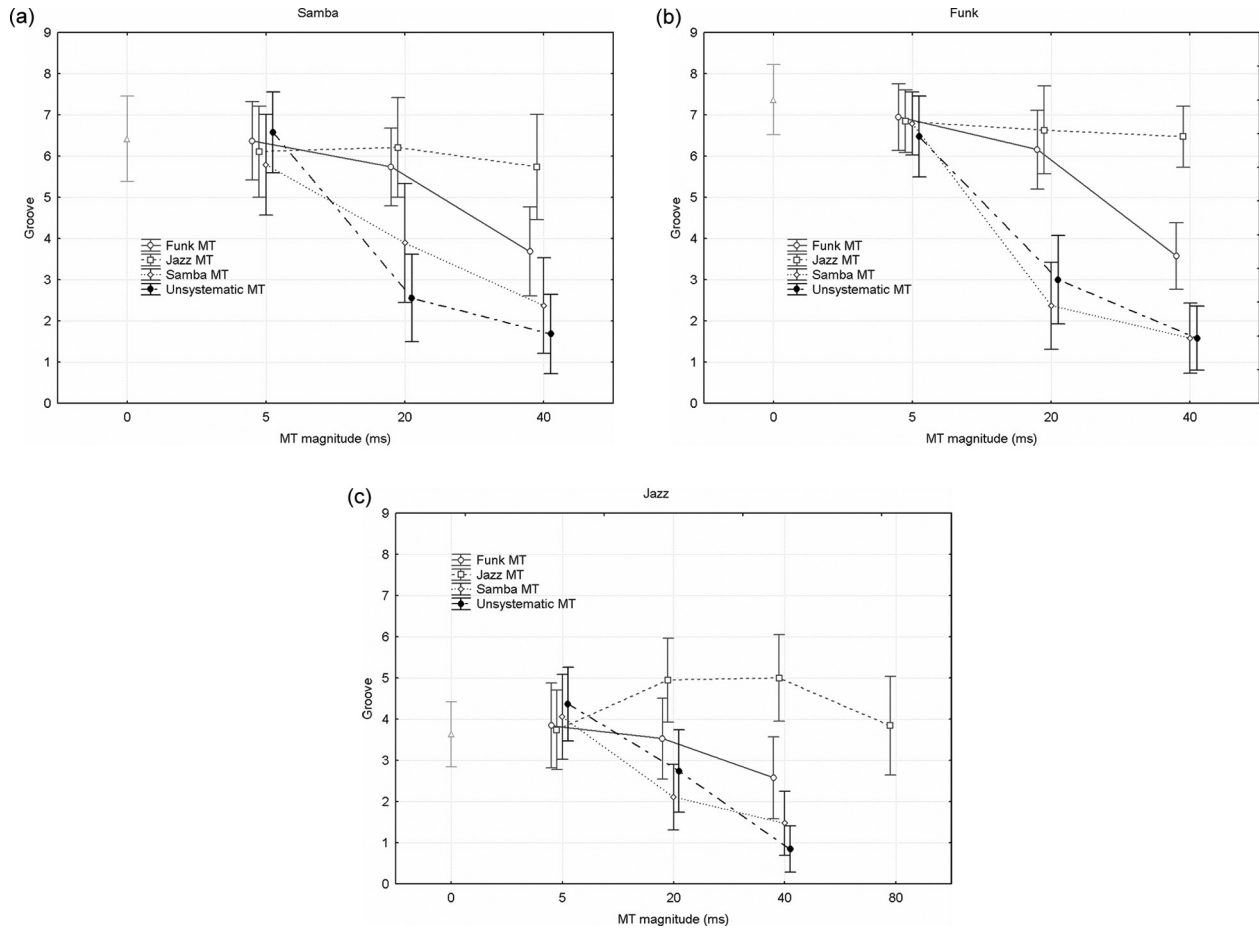


FIGURE 5. Mean groove ratings for experts as a function of MT style and microtiming magnitude for (a) Samba, (b) Funk and (c) Jazz. The quantized condition has an MT magnitude of 0 ms.

nonexpert listeners, that, except for Jazz, groove ratings decreased as MT-magnitude increased. For the 40 ms magnitude across all MT-styles, the MEs were considered less groovy than the quantized conditions (for Funk M-style, $d' = 1.93$ for Funk MT-style, 0.52 for Jazz, 3.22 for Samba, and 3.55 for Unsystematic MT; for Jazz M-style and Funk MT $d' = 0.51$, 1.30 for Samba, and 1.68 for Unsystematic; for Samba M-style and Funk $d' = 1.24$, 1.76 for Samba, and 2.23 for Unsystematic). Moreover, significant decreases were also found for 20 ms for Funk and all MT-styles except Jazz (Funk = 0.67, Samba = 2.38, Unsystematic = 3.05), for Jazz M-style and Samba MT-style ($d' = 1.35$), and for Samba M-style and Samba MT-style ($d' = 1.01$) and Unsystematic ($d' = 1.61$). The Jazz MT pattern again was different from the other MT-styles, remaining largely unaffected by the magnitude, except when Jazz MT was applied to Jazz. A one-way ANOVA showed a positive groove effect between the 40 ms magnitude and quantized condition,

$F(1, 18) = 4.76, p < .05 (d' = 0.62)$, and also for the 20 ms magnitude and the quantized condition, $F(1, 16) = 4.58, p < .05 (d' = 0.59)$. These positive effects can be considered special cases since the quantized condition for Jazz is not so much an example of no MT but rather a lack of expected MT. These positive effects over the quantized condition were not present for all MT-magnitudes, with both the 5 ms and 80 ms conditions not having a significant difference in ratings from the deadpan version. Overall for Jazz MT applied to Jazz, the 40 ms (ideal) condition was rated the most groovy.

Regarding the other hypothesized ideal conditions, Samba MT at 20 ms for Samba and Funk MT at 20 ms for Funk, we see two different pictures. For the Samba case this is clearly less groovy than the quantized condition, but for Funk the mean rating is moderately lower than the quantized condition. In general, Funk MT was preferred over Samba MT across all M-styles—the latter of which was rated similarly to the unsystematic pattern.

One further result consistent across all M-styles was the lack of a significant difference for any MT-style with 5 ms magnitude compared to the respective quantized cases. This result is consistent with the existing literature concerning the lower limits for perceptual timing thresholds (e.g., Madison, 2004).

As with the nonexperts ratings, the correlation between groove and naturalness was high. One notable difference occurred for 40 ms Jazz MT applied to Samba. This was rated high in terms of groove (5.7) but low for naturalness (3.4), suggesting that while groovy for the expert participants, it was not a normal sounding Samba performance, as Samba is not typically characterised by a shuffle pattern. The same was not true of Funk, with a groove rating of 6.5 and a naturalness rating of 6.1 and where the “Funk-shuffle” is a well-known drumming style, of which Funk drummer Bernard Purdie is perhaps the most recognized musician.

Comparing Results Across Experiments

Comparing the groove ratings of experts with the non-experts, we can observe some differences despite the similarity in overall trends. The most immediate difference is that the experts listeners made wider use of the ratings scale, in particular when MEs were not considered groovy (low ratings), whereas the nonexperts tended to deviate less. In this sense we may infer that the expert listeners were better able to differentiate between the various MT conditions and demonstrated greater consistency in their ratings. These observations were confirmed by a mixed four-way 2 Group (experts vs. nonexperts) x 3 M-style x 2 MT-style x 3 MT-magnitude ANOVA with groove rating as the dependent variable. Note that the design was non-orthogonal, so the lowest level of MT-magnitude (0 ms) and the extra 80 ms condition for Jazz were omitted from the ANOVA. This was deemed appropriate because there was no significant difference between 0 and 5 ms MT magnitude, and because the outcome of the Jazz MT was atypical of the results. No three-way or higher interactions were significant ($p > .05$), so only two-way interactions and main effects were considered. First, there was no main effect of group; the mean rating across all conditions was not different between the groups. The Group x MT-style interaction was significant, $F(3, 147) = 17.06, p < .00001$, as was the Group x MT-magnitude interaction, $F(2, 98) = 16.01, p < .00001$, whereas Group x M-style was not, $F(2, 98) = 2.59, p = .08$. The Group x MT-style interaction accounts for the overall lower ratings the experts gave to the Samba and unsystematic MT-styles, and the Group x MT-

magnitude interaction reflects the experts' greater decrease in groove as MT-magnitude increased. A significant M-style x MT-style interaction, $F(6, 294) = 6.10, p < .00005$, confirmed that all MT-styles except Jazz affect Jazz M-style the least, Funk the most, and Samba to an intermediate level. The central findings of the study are buttressed by main effects of M-style, $F(2, 98) = 36.19, p < .000001$, MT-style, $F(3, 294) = 95.33, p < .000001$, and MT-magnitude, $F(2, 98) = 106.57, p < .000001$, and the trivial derivations M-style x MT-magnitude, $F(4, 196) = 10.24, p < .000001$, and MT-style x MT-magnitude, $F(6, 294) = 29.36, p < .000001$.

General Discussion

The purpose of the present study was to examine the role of style-appropriate microtiming for the experience of groove. Specifically, we hypothesized that groove ratings would be higher for style-appropriate patterns and magnitudes of microtiming than for deadpan versions. The consistent result across both nonexpert and expert listeners is that, within the confines of our experiment, microtiming does not increase groove for any combination of M-style, MT-style, and MT-magnitude, with one exception. Rather than causing an increase in groove ratings, the ratings exhibit a clear negative effect, which provides strong evidence against our main hypothesis. The exception is that the experts did rate the 20 and 40 ms shuffle patterns that constitute the Jazz MT slightly higher than the other three MT-magnitude levels.

The results for the Jazz MT are quite different from the other three types of MT. First, groove ratings remain largely unaffected by MT-magnitude. A possible reason for this difference is that the Jazz MT is both familiar and predictable for listeners; furthermore, it is the most simple, manifested by the same “long-short” pattern repeated for each beat. The Funk and Samba patterns are repeated only at the bar level and the unsystematic patterns did not repeat at all. Second, this “tolerance” for the Jazz MT transfers to the other music styles, such that groove ratings remain close to the level corresponding to the deadpan MEs even for Funk and Samba M-styles. In other words, the groove ratings of the present rhythms were not improved by the presence of shuffle.

Among the MT-styles that caused a negative groove effect, the Funk pattern decreased less than the Samba and unsystematic patterns. While both the Samba and Funk MT-styles were derived from existing research, they were constructed in different ways. The Funk pattern was derived from a specific performance, that of “Funky Drummer” by James Brown (McGuinness, 2005), whereas the Samba pattern was derived as the

average across many Samba pieces (Naveda et al., 2011). To this end, the Funk pattern may have greater internal consistency than a Samba pattern that is an aggregate of many performances but that is not specific to any individual piece. To test whether specific Samba MT patterns cause higher groove ratings than the aggregate pattern used here remains as a topic for future work.

Beyond specific differences in the construction of the MT patterns, it is useful to consider what further issues, common to all MEs might have either caused the negative groove effect or prevented a positive effect from emerging. One reason may have been the use of percussion music examples rather than real music examples. In the groove rating experiment conducted by Janata et al. (2012) participants rated a set of real music examples and a set of drum loops. Ratings for the real music examples spanned a much greater proportion of the 127 point scale than those of the percussion examples: 29.3 - 108.7 for the real music examples compared to 40.3 - 58.1 for the drum loops. This tendency for participants to rate the purely percussion examples at the lower end of the scale in the Janata et al. experiment (2012) was not found in Madison (2006). However, a percussion-only music example was rated highest in groove, in competition with several typical Jazz and groove-based MEs. Therefore there is no reason to presume that percussion patterns should be less groovy. By way of an informal comparison, the highest rated ME in our experiment—the deadpan Funk example—was rated 7.4 by expert listeners; if scaled proportionally this would be 93.6 on the 127-point scale used by Janata et al. (2012) and therefore very high in groove.

It is important to consider whether the participants were simply not up to the task; that a lack of familiarity with the music styles and microtiming patterns played a role in masking a positive groove effect for microtiming. Based on statistical analysis of the ratings, it would appear that music training plays a role in appreciating and evaluating microtiming. Stronger effects were present among the expert group compared to the nonexpert group, even though the expert group was smaller. The expert listeners were more decisive in their ratings than the nonexperts, and more willing to use very low ratings for MEs they considered were not at all groovy (e.g., Samba and unsystematic MT-styles with large MT-magnitudes). While it would have been theoretically possible, although somewhat impractical, to recruit, for example, a set of Samba experts, a significant positive effect for a specific subset of music experts would not be of great use for understanding musical properties that are considered groovy by the general population. That is to say, if the appreciation of microtiming for groove

requires high familiarity with music styles, or indeed specific microtiming patterns, then microtiming may not be the most significant contributor of groove for most listeners, even if it is considered important by musicians. Indeed the positive association of microtiming and groove by musicians may not be so much the presence of systematic microtiming but in fact the *lack* of unsystematic microtiming; in effect, the recognition of the skill required to play with precise timing (Freeman & Lacey, 2002). In the context of our experiment, this relationship holds for the unsystematic patterns but perhaps also with the systematic Samba and Funk patterns if they were not recognized as systematic by the participants.

Having considered potential reasons for the lack of support for microtiming to be associated with groove given that such an association is true, we now turn to potential reasons for it being false; that is, why microtiming might be bad for groove. One explanation is that microtiming makes music harder to predict and therefore to synchronize to. This resonates with the only positive responses being for the simplest and most recognizable Jazz microtiming pattern, where all other microtiming patterns were detrimental. Madison et al. (2011) suggested that the purpose of groove might be to promote movement in synchrony to music, and stated some physical properties that would in that case be expected to be associated with groove: the repetition of rhythmic patterns, the presence of fast metrical levels, and the density of events between beats. Accordingly, manipulations that make synchronization harder should not logically be good for groove, e.g., systematic microtiming patterns that are not clearly and obviously repetitive or unsystematic patterns that do not repeat at all. However, this hypothesis would need to be tested via a separate synchronization and microtiming experiment.

If we are to accept that the purpose of microtiming is not to increase groove in music, then we should consider why it is there in the first place. Repp (1998) suggests that microtiming is not the result of intentional behavior by musicians; rather, it is obligatory—that timing deviations are a function of musical structure. Indeed, the possibility of creating quantized music at all is a recent phenomenon reliant on the use of music sequencers. Certainly for Romantic music the role of expressive timing is not to create the sensation of groove. In this context the Vienna waltz is an interesting example since it is a dance whose rhythm is characterized by deviations in timing. However these deviations are both large and predictable, meaning that, like Jazz microtiming, the deviations do not hinder the ability of dancers to synchronize with the music. As part of an analysis of microtiming in Jazz drumming Butterfield

(2010) suggested that systematic deviations are used for expression, but only sparingly and in the context of near-isochronous playing.

Finally, if not microtiming, which musical features do correlate positively with groove? Janata et al. (2012) showed significant listener preferences both in terms of genre (soul/r&b was the most groovy and folk the least) and tempo (faster tempi were groovier than slower tempi). However, genre is not a factor that can be systematically controlled for in the same way we address microtiming in this paper and the tempo effects are perhaps confounded by genre. Referring again to the study of Madison et al. (2011), the two musical features that contributed most prominently to groove were event density and the presence of fast metrical levels. Unlike genre, these properties can be effectively controlled. Therefore, as part of future work we plan to explore both of these factors along with the role played by dynamics and syncopation in order to

better understand the musical factors that explain the sensation of groove.

Author Note

This work was financed by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project: PTDC/EAT-MMU/112255/2009-(FCOMP-01-0124-FEDER-014732). Part of this research was supported by a grant from the Bank of Sweden Tercentenary Foundation (P2008:0887).

Correspondence concerning this article should be addressed to Matthew E. P. Davies, INESC TEC - Instituto de Engenharia de Sistemas e Computadores do Porto, Campus da FEUP, Rua Dr. Roberto Frias, 378, 4200 - 465 Porto, Portugal. E-mail: mdavies@inescporto.pt

References

- ALÉN, O. (1995). Rhythm as duration of sounds in Tumba Francesa. *Ethnomusicology*, 39, 55-71.
- BILMES, J. (1993). *Timing is of the essence: Perceptual and computational techniques for representing, learning and reproducing expressive timing in percussive rhythm*. Unpublished master's thesis, Massachusetts Institute of Technology.
- BUTTERFIELD, M. (2010). Participatory discrepancies and the perception of beats in jazz. *Music Perception*, 27, 157-176.
- DANIELSEN, A. (2006) *Presence and pleasure: The funk grooves of James Brown and Parliament*. Middletown, CT: Wesleyan University Press.
- FREEMAN, P., & LACEY, L. (2002). Swing and groove: Contextual rhythmic nuance in live performance. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition*, Sydney (pp. 548-550). Adelaide: Causal Productions.
- FRIBERG, A., BRESIN R., & SUNDBERG, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, 2, 145-161.
- FRIBERG, A., & SUNDBERÖM, A. (2002). Swing ratios and ensemble timing in Jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19, 333-349.
- GABRIELSSON, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 501-602). New York: Academic Press.
- GREENWALD, J. (2002). The rhyme may define, but the groove makes you move. *Black Music Research Journal*, 22, 259-271.
- HELLMER, K. (2006). *The development of a drum machine using the Steinberg VST- specification: Implementation of extracted timing behaviors from human drum performances*. Unpublished master's thesis, Luleå University of Technology, Luleå Sweden.
- HENNIG, H., FLEISCHMANN, R., FREDEBOHM, A., HAGMAYER, Y., NAGLER, J., WITT, A., ET AL. (2011). The nature and perception of fluctuations in human musical rhythms. *PLoS ONE*, 6(10), e26457.
- Honing, H., & Reips, U-D. (2008). Web-based versus lab-based studies: A response to Kendall (2008). *Empirical Musicology Review*, 3, 73-77.
- HONING, H., & DE HAAS, W. B. (2008). Swing once more: Relating timing and tempo in expert Jazz drumming. *Music Perception*, 25, 471-478.
- IYER, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception*, 19, 387-414.
- JANATA, P., TOMIC, S. T., & HABERMAN, J. M. (2012). Sensorimotor coupling in music and the psychology of groove. *Journal of Experimental Psychology: General*, 141, 54-75.
- KEIL, C. (1995). The theory of participatory discrepancies: A progress report. *Ethnomusicology*, 39, 1-19.
- MADISON, G. (2004). Detection of linear temporal drift in sound sequences: Principles and empirical evaluation. *Acta Psychologica*, 117, 95-118.
- MADISON, G. (2006). Experiencing groove induced by music: Consistency and phenomenology. *Music Perception*, 24, 201-208.

- MADISON, G., GOUYON, F., ULLÉN, K., & HÖRNSTRÖM, K. (2011). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1578-1594.
- MADISON, G., & PAULIN, J. (2010). Ratings of speed in real music as a function of both original and manipulated beat tempo. *Journal of the Acoustical Society of America*, 128, 3032-3040.
- MCGUINNESS, A. (2005). *Microtiming deviations in groove*. Unpublished master's thesis, Australian National University, Canberra, Australia.
- NAVEDA, L., GOUYON, F., LEMAN, M., & GUEDES, C. (2011). Microtiming patterns and interactions with musical properties in Samba music. *Journal of New Music Research*, 40, 225-238.
- PRESSING, J. (2002). Black Atlantic rhythmic: Its computational and transcultural foundations. *Music Perception*, 19, 285-310.
- REPP, B. H. (1995). Expressive timing in Schumann's "Traumerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98, 2413-2427.
- REPP, B. H. (1998). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research*, 61, 33-43.
- STEWART, A. (2000). 'Funky Drummer': New Orleans, James Brown and the rhythmic transformation of American popular music. *Popular Music*, 19, 293-318.
- WAADELAND, C. H. (2001). "It don't mean a thing if it ain't got that swing" Simulating expressive timing by modulated movements. *Journal of New Music Research*, 30, 23-37.
- WIDMER, G., DIXON, S., GOEBL, W., PAMPALK, E., & TOBUDIC, A. (2003). In search of the Horowitz factor. *AI Magazine* 24(3), 111-130.
- WINDSOR, W. L., & CLARKE, E. F. (1997) Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool. *Music Perception*, 15, 127-152.
- WRIGHT, M., & BERDAHL, E. (2006). Towards machine learning of expressive microtiming in Brazilian drumming. *Proceedings of the International Computer Music Conference* (pp. 572-575). Ann Arbor, MI: MPublishing, University of Michigan Library.
- ZEINER-HENRIKSEN, H. T. (2010). *The "PoumTchak" pattern. Correspondences between rhythm, sound and movement in electronic dance music*. Unpublished doctoral dissertation, University of Oslo, Oslo, Norway.

Appendix

TABLE 1. Summary Of Microtiming Deviations.

Metrical Position	Samba	Funk	Jazz
1/16	5 ms (0.27)	0 ms (0)	0 ms (0)
2/16	0 ms (0)	0 ms (0)	40 ms (1.00)
3/16	-18 ms (-1.00)	0 ms (0)	0 ms (0)
4/16	-14 ms (-0.78)	14 ms (0.67)	40 ms (1.00)
5/16	11 ms (0.61)	13 ms (0.62)	0 ms (0)
6/16	0 ms (0)	21 ms (1.00)	40 ms (1.00)
7/16	-18 ms (-1.00)	13 ms (0.62)	0 ms (0)
8/16	-14 ms (-0.78)	17 ms (0.80)	40 ms (1.00)
9/16	5 ms (0.27)	10 ms (0.47)	0 ms (0)
10/16	0 ms (0)	0 ms (0)	40 ms (1.00)
11/16	-18 ms (-1.00)	0 ms (0)	0 ms (0)
12/16	-14 ms (-0.78)	0 ms (0)	40 ms (1.00)
13/16	11 ms (0.61)	0 ms (0)	0 ms (0)
14/16	0 ms (0)	0 ms (0)	40 ms (1.00)
15/16	-18 ms (-1.00)	0 ms (0)	0 ms (0)
16/16	-14 ms (-0.78)	0 ms (0)	40 ms (1.00)

Note: Shown are the raw microtiming patterns for Samba, Funk and Jazz with deviations in ms for each metrical position. The normalized durations are shown in parentheses.