KTH Computer Science
and Communication

On the beat: Human movement and timing in the production and perception of music

SOFIA DAHL

Doctoral Thesis
Stockholm, Sweden 2005

TRITA-TMH 2005:5
ISSN 1104-5787
ISBN 91-7178-134-X

KTH School of Computer Science and Communication
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i tal- och musikkommunikation med inriktning på musikakustik torsdagen den 29 september 2005 klockan 10.00 i Salongen Biblioteket Entréplanet, Kungl Tekniska högskolan, Osquars backe 31, Stockholm.

Tryck: Universitetsservice US AB

Abstract

This thesis addresses three aspects of movement, performance and perception in music performance. First, the playing of an accent, a simple but much used and practiced element in drumming is studied, second, the perception of gradually changing tempo, and third, the perception and communication of specific emotional intentions through movements during music performance.

Papers I and II investigated the execution and interpretation of an accent in drumming, performed under different playing conditions. Players' movements, striking velocities and timing patterns were studied for different tempi, dynamic levels and striking surfaces. It was found that the players used differing movement strategies and that interpreted the accent differently, reflected in their movement trajectories. Strokes at higher dynamic levels were played from a greater average height and with higher striking velocities. All players initiated the accented strokes from a greater height, and delivered the accent with increased striking velocity compared to the unaccented strokes. The interval beginning with the accented stroke was also prolonged, generally by delaying the following stroke. Recurrent cyclic patterns were found in the players' timing performances. In a listening test, listeners perceived grouping of the strokes according to the cyclic patterns.

Paper III concerned the perception of gradual tempo changes in auditory sequences. Using an adaptive test procedure subjects judged stimuli consisting of click sequences with either increasing or decreasing tempo, respectively. Each experiment included three test sessions at different nominal tempi (80, 120, and 180 beats per minute). The results showed that ten of the eleven subjects showed an inherent bias in their perception of tempo drift. The direction and magnitude of the bias was consistent between test sessions but varied between individuals. The just noticeable differences for tempo drift agreed well with the estimated tempo drifts in production data, but were much smaller than earlier reported thresholds for tempo drift.

Paper IV studied how emotional intent in music performances is conveyed to observers through the movements of the musicians. Three players of marimba, bassoon, and saxophone respectively, were filmed when playing with the expressive intentions Happiness, Sadness, Anger and Fear. Observers rated the emotional content and movement cues in the videos clips shown without sound. The results showed that the observers were able to identify the intentions Sadness, Anger, and Happiness, but not Fear. The rated movement cues showed that an Angry performance was characterized by jerky movements, Happy performances by large, and somewhat fast and jerky movements, and Sad performances by slow, and smooth movements.

Papers included in the thesis

The papers will be referred to by their roman numbers.

Paper I:
Dahl, S. 2000 The playing of an accent - preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research* **29**(3) 225-233.
Printed with permission of Swets & Zeitlinger Publishers.

Paper II:
Dahl, S. 2004 Playing the accent - comparing striking velocity and timing in an ostinato rhythm performed by four drummers. *Acta acustica united with Acustica* **90**(4) 762-776.
Printed with permission of S. Hirzel Verlag.

Paper III:
Dahl, S. and Granqvist, S. 2005 Ability to determine continuous drift in auditory sequences: Evidence for bias in listeners' perception of tempo.
Submitted for publication in *Journal of Acoustical Society of America.*

Paper IV:
Dahl, S. and Friberg, A. 2005 Visual perception of expressiveness in musicians' body movements.
Revised version submitted for publication in *Music Perception.*

Author's contribution to the papers

Paper I:
This work was carried out entirely by author S. Dahl.

Paper II:
This work was carried out entirely by author S. Dahl.

Paper III:
The major part of the work (experiment, analysis and writing) was carried out by author S. Dahl. Co-author S. Granqvist performed the software programming and participated in editing the manuscript.

Paper IV:
The major part of the work was carried out by author S. Dahl. Co-author A. Friberg participated in the statistical analysis and in editing the manuscript.

Other related publications by the author

Peer-reviewed articles:

Dahl, S. and Granqvist, S. (2003). Estimating Internal Drift and Just Noticeable Difference in perception of continuous tempo drift. *The Neurosciences and Music, Annals of the New York Academy of Science*, 999:161–165.

Book chapters:

Dahl, S. (In press) Movements and analysis of drumming. In *Music, Motor Control and the Brain.* To be published by Oxford University Press.

Bresin, R. and Dahl, S. (2003). Experiments on gestures: walking, running, and hitting. In Rocchesso, D. and Fontana, F., editors, *The sounding object.*, pages 111 –136. Mondo Estremo, Florence, Italy.

Bresin, R., Falkenberg Hansen, K., Dahl, S., Marshall, M., and Moynihan, B. (2003). Devices for manipulation and control of sounding objects: The Vodhran and the Invisiball. In Rocchesso, D. and Fontana, F., editors, *The sounding object.*, pages 271 –296. Mondo Estremo, Florence, Italy.

Conference Proceedings:

Dahl, S. and Friberg, A. (2004). Expressiveness of musician's body movements in performances on marimba. In Camurri, A. and Volpe, G., editors, *Gesture-based Communication in Human-Computer Interaction*, volume 2915 of *Lecture Notes in Artificial Intelligence*, pages 479–486. Springer Verlag.

Rinman, M.-L., Friberg, A., Bendiksen, B., Cirotteau, D., Dahl, S., Kjellmo, I., Mazzarino, B., and Camurri, A. (2004). Ghost in the Cave - An interactive collaborative game using non-verbal communication. In Camurri, A. and Volpe, G., editors, *Gesture-based Communication in Human-Computer Interaction*, volume 2915 of *Lecture Notes in Artificial Intelligence*, pages 549–556. Springer Verlag.

Dahl, S. and Granqvist, S. (2003). Looking at perception of continuous tempo drift - a new method for estimating Internal Drift and Just Noticeable Difference. In Bresin, R., editor, *Proceedings of Stockholm Music Acoustic Conference 2003*, volume II, pages 595–598, Stockholm, Sweden.

Dahl, S. and Friberg, A. (2003). What can body movements reveal about a musician's emotional intention? In Bresin, R., editor, *Proceedings of Stockholm Music Acoustic Conference 2003*, volume II, pages 599–602, Stockholm, Sweden.

Rinman, M.-L., Friberg, A., Kjellmo, I., Camurri, A., Cirotteau, D., Dahl, S., Mazzarino, B., Bendiksen, B., and McCarthy, H. (2003). ESP- an interactive collaborative game using non-verbal communication. In Bresin, R., editor, *Proceedings of Stockholm Music Acoustic Conference 2003*, volume II, pages 561–564, Stockholm, Sweden.

Dahl, S. (2001). Arm motion and striking force in drumming. In Gerber, H. and

Müller, R., editors, *Proceedings of XVIIIth congress of the International Society of Biomechanics, 8-13 July 2001*, CD–ROM, Zürich, Switzerland.

Dahl, S. (2001). Arm motion and striking force in drumming. In Bonsi, D., Gonzalez, D., and Stanzial, D., editors, *Proceedings of the International Symposium on Musical Acoustics, Musical Sounds from Past Millennia, 10-14 September 2001*, volume 1, pages 293–296, Perugia, Italy.

Dahl, S., and Bresin, R. (2001). Is the player more influenced by the auditory than the tactile feedback from the instrument? In Fernström, M., Brazil, E., and Marshall, M., editors, *Proceedings of the Cost-G6 Conference Digital Audio Effects, (DAFx01), 6-8 December 2001*, pages 194–19, Limerick, Irland.

Bresin, R., Friberg, A., and Dahl, S. (2001). Toward a new model for sound control. In Fernström, M., Brazil, E., and Marshall, M., editors, *Proceedings of the Cost-G6 Conference Digital Audio Effects, (DAFx01), 6-8 December 2001*, pages 45 –49, Limerick, Irland.

Speech, Music and Hearing Quarterly Progress and Status Report:
Dahl, S., and Friberg, A. (2004). *Expressiveness of a marimbaplayer's body movements.* Speech, Music and Hearing Quarterly Progress and Status Report, TMH-QPSR 46(4):75–86, KTH, Royal Institute of Technology, Stockholm.

Dahl, S., Granqvist, S., and Thomasson, M. (2000). *Detection of drift in tempo.* Speech, Music and Hearing Quarterly Progress and Status Report, TMH-QPSR 41(4):19–28, KTH, Royal Institute of Technology, Stockholm.

Dahl, S. (2000). *Timing in drumming – some preliminary results.* Speech, Music and Hearing Quarterly Progress and Status Report, TMH-QPSR 39(4):95–102 KTH, Royal Institute of Technology, Stockholm.

Dahl, S. (1997). *Measurements of the motion of the hand and drumstick in a drumming sequence with interleaved accented strokes - a pilot study.* Speech, Music and Hearing Quarterly Progress and Status Report, TMH-QPSR 38(4):1–6, KTH, Royal Institute of Technology, Stockholm.

Dahl, S. (1997). *Spectral changes in the tom-tom related to striking force.* Speech, Music and Hearing Quarterly Progress and Status Report, TMH-QPSR 38(1):59–65, KTH, Royal Institute of Technology, Stockholm.

Thesis works and essays:
Dahl, S. (2003). *Striking movements: Movement strategies and expression in percussive playing.* Published licentiate thesis, KTH, Royal Institute of Technology, Stockholm, Sweden. ISBN 9-7283-480-3, TRITA-TMH 2003:3.
Dahl, S. (1996). *Kvinnor och slagverk.* Essay in musicology, Stockholm University, Stockholm, Sweden.

Abbreviations and definitions of recurrent terms

ANOVA analysis of variance, basic statistical technique for analyzing experimental data

articulation the manner in which notes are joined to one another by the performer. Playing *legato* means to tie the notes together, while *staccato* means to separate the notes clearly. The articulation can be calculated as $\frac{duration}{IOI}$. Values for staccato and legato vary between instruments and performances (see e.g. Bresin and Battel, 2000)

BPM beats per minute, commonly used for indicating musical tempo

continuation continued production of events (taps, strokes etc) after an initial synchronization to a metronome that is turned off

DAF delayed auditory feedback. A delay, Δt, is introduced between the onset of a played note and the auditory feedback to the player.

Δt a time change, positive or negative, applied either to change subsequent IOIs (tempo change or drift), alternatively applied to make feedback asynchronous with action (in DAF)

EMG electromyography, recordings of the electrical waves associated with the activity of skeletal muscles

drift span range of different continuous tempo drifts

ff fortissimo, extremely loud, musical definition for very high dynamic level

ID internal drift, the *perceived* isochrony relative real isochrony (used in Paper III)

IOI inter-onset interval, the time between the onset of one sound and the following

IOI$_{avs}$ average IOI across a whole recorded sequence

IOI drift gradual changes of IOI over time. Adding Δt to each interval results in a decreasing (decelerating) tempo, while subtracting Δt results in an increasing (accelerating) tempo.

IOI$_{rel}$ relative IOI, normalized to the average IOI. $\mathrm{IOI}_{rel} = \mathrm{IOI}/\mathrm{IOI}_{avs}$

isochrony an isochronous sequence has equally separated events and all IOIs are equal

JND the Just Noticeable Difference, commonly used to estimate perceptual threshold

LAST method for estimating ID or JND using the final drifts in an adaptive test track (used in Paper III)

LRM method for estimating ID or JND using Logistic Regression Model fitting (used in Paper III)

mf mezzo forte, half loud, musical definition for moderate dynamic level

non-discriminable drift span =2×JND, range of continuous tempo drifts that are too small to be perceived, not necessarily centered around isochrony (used in Paper III)

pp pianissimo, very soft, musical definition for very low dynamic level

perceived drift the perception of the physical drift, not necessarily of the same magnitude or sign as the physical drift

physical drift physical lengthenings or shortenings of subsequent intervals.

PRMD playing-related musculoskeletal disorder

proprioceptor special nerve endings in the muscles and tendons and other organs that respond to stimuli regarding the position and movement of the body

synchronization tapping, drumming etc in synchrony to a pacing signal (e.g. a metronome)

tempo drift gradual changes of tempo over time $1 + tempo\ drift \approx 1 - IOI\ drift$ for small drift magnitudes

Contents

Part I

Introduction

Chapter 1

Playing musical instruments: Interaction – Perception – Communication

Across the world, throughout the ages, people have invested time and effort to become musicians. Indeed, time and effort are needed. Few, if any, other human activities have as high demands on time-ordered motoric skill as playing music. The road from beginner to expert requires many, many hours of practice before the playing technique necessary for mastery of an instrument is obtained. Furthermore, it is not enough for a professional musician to perform specialized, complex motor tasks under a strict time constraint. For a performance to be truly successful it should also convey something to the listener. To excel in the playing of an instrument is to be able to understand and express the structure and meaning of the music. These practice hours are a necessary investment in order to have the means to convey what the music is about.

The fact that music is a form of communication puts musicians in a somewhat different position compared to many other experts in, for instance, sport and industry. For a musician, the evaluation of a movement has less to do with measurable units of time or distance and more to do with communication through musical sounds. The assessment of a music performance differs from that of many other types of skills. For example, a sprinter's success can be measured in terms of distance and time. In contrast, music performance tends to require a more subjective evaluation, and therefore present more of a challenge.

This thesis highlights a few areas in the complex world of music performance where motor tasks are executed with skilled timing and are integrated with the overall musical expression.

1.1 Music performance

As noted by Gabrielsson (1999, 2003), research in music performance has increased considerably during past years. The following section concentrates on reviewing studies that are relevant to the work in this thesis. In particular, the review focuses on studies of tempo, timing, and work involving studies on drumming. For a more encompassing survey of the field the reader is referred to Palmer (1997), Gabrielsson (1999, 2003), and Krumhansl (2000).

Meter and tempo

Music, melody and rhythm can be viewed as a series of events occurring at specific moments in time. Many of the movements involved in music making would be trivial if it were not for the timing. If striking a drum is easy, doing so at the right moment is not. To understand how music performances are planned and executed it is necessary to consider the overall time constraints governing most music performances, i.e. *tempo* and *meter*.

Tempo can be defined as the rate of regularly repeated events. In reality, however, the times separating the events are far from equal and tempo is known to change considerably within phrases and musical excerpts. Basically, there are three types of tempo in music performance (Gabrielsson, 1999): *mean tempo* – averaged across the whole piece of music, *main tempo* – the prevailing tempo (minus final ritards and so on), and *local tempo* – tempo that is maintained only for a short time. In some musical genres, tempo is constant enough for these three categories to merge.

Meter is a more abstract concept than tempo. Meter is defined by a systematic occurrence of strong and weak *beats*, i.e. time points without durations. This forms a hierarchical structure, with different periodicities at each level. According to Lerdahl and Jackendoff (1983), "a beat at a larger level must also be a beat at all lower levels". When asked to tap along to a piece of music, most people agree on one metrical level. Lerdahl and Jackendoff call it *tactus*.

Listeners' choice of the metrical level to tap along to is influenced by the main tempo of the music. In general, people tend to settle for an intermediate rate, usually not as fast as the melody, but not too slow either (Fraisse, 1982; Drake et al., 2000). This has led researchers to believe that there is a characteristic period that people find most convenient. This *preferred tempo* has been suggested to occur somewhere in the range 100–133 *beats per minute* (*BPM*) (e.g. Fraisse, 1982; Van Noorden and Moelants, 1999; Moelants, 2002).

The idea of a preferred tempo also frequently occurs in discussions on what mechanisms are responsible for human timing. One way to investigate the timing behavior of subjects is to conduct finger tapping experiments (e.g. Michon, 1967; Wing and Kristofferson, 1973b). In a commonly used approach, subjects are asked to first tap in synchrony with a metronome (synchronization) and then continue to tap after the metronome has been turned off (continuation). Many studies have reported that even though people strive to keep a constant tempo, gradual changes of tempo over time, *tempo drift*, are normally present (e.g. Michon, 1967; Ogden and Collier, 1999; Madison, 2001b; Chen et al., 2002). A drift in tempo means that the *inter-onset intervals* (*IOIs*) are successively increased or decreased. A linear tempo drift would mean that a fixed time unit is added or subtracted to each interval. In practice, however, tempo drift is seldom linear in longer sequences.

From several tapping experiments, Madison (2001b) estimated that drift was responsible for between 7 and 20% of the variation in IOI. The maximum variation occurred for low tapping rates, in the IOI range 1.1–1.4 s (54–43 BPM). The average linear drift in both directions (increasing and decreasing tempo) ranged between 0.05 and 0.3% per interval.

Since tempo changes occur so frequently in musical performances, several studies have investigated which direction of change is most easily detected by listeners. Some studies have reported that subjects detect decreasing tempi faster and more accurately than increasing tempi (e.g. Kuhn, 1974; Madsen, 1979; Sheldon and Gregory, 1997; Pouliot and Grondin, 2005). Other studies have found no, or negligible, differences between the two drift directions (e.g Madison, 2004; Ellis, 1991).

Concerning the perceptual thresholds of tempo drift, there are large differences between methods used in different studies. The studies that have investigated continuously changing tempo are very few. Instead of continuously changing each IOI, some studies (e.g. Wang, 1983; Wang and Salzberg, 1984; Sheldon and Gregory, 1997) have modulated the tempo by one or several BPM per time unit (measures or seconds).

A study that investigated the perceptual threshold of linear tempo drift was conducted by Madison (2004). Madison reported *Just Noticeable Differences (JNDs)* of about 2–3% for sequences of 5 intervals and nominal IOIs ranging between 300–1100 ms. Although not primarily concerned with JND, the results of Vos et al. (1997) suggest the threshold for 2–3 intervals is about 3.5% for nominal IOIs of 500 ms. These results for auditory thresholds of tempo drift are of about the same magnitude as the threshold reported for stepwise changes in tempo and other perturbations of short isochronous sequences (e.g. Friberg and Sundberg, 1995; Drake and Botte, 1993).

Timing and grouping

With regard to timing in musical performance, there have been an overwhelming number of research accounts (see reviews in Gabrielsson, 1999; Clarke, 1999). The majority of the studies have concerned performances on piano, but the systematic variations in the timing of rhythmical patterns have also been investigated for drumming (e.g. Gabrielsson, 1974; Friberg and Sundström, 2002).

A commonly reported feature is that notes are almost never played according to their nominal (notated) IOI value. Although small integer ratios (1:1, 1:2, and 1:3) are fundamental in musical notation, the durations of notes with these ratios are systematically lengthened or shortened by players.

For example, in early studies of performances of Swedish folk melodies, Bengtsson et al. (1969) found long-short relations between the first and second half of the bars, and also between half bars. Between two eighth notes, however, the first was played shorter than the second. The short-long feature between several notes with nominally equal duration occurs in a number of reported studies. In particular, the last note in a group of, say, four eighth or sixteenth notes, is lengthened (c.f. Clarke, 1982, 1985).

The examples above illustrate how musicians use systematic timing deviations from the notated score to emphasize grouping of notes and overall structure. The systematic timing deviations tend to be very similar between performances by the same player, although there may be marked differences between players (e.g. Repp, 1992a). Musicians seem to agree, however, on several general features in the system of deviations. During the last decades several models of this *expressive timing* behavior have been proposed (e.g. Sundberg et al., 1991; Friberg, 1995, see De Poli, 2004 for survey).

However, not all variations in timing are systematic. Embedded in the timing of music performance lie random variations. These random timing variations are not always easily distinguishable from intended lengthening or shortenings of individual note durations. In order to estimate the influence of random components on music performance, studies of isochronous finger tapping can offer valuable insights (e.g. Wing and Kristofferson, 1973b,a). In tapping, the short-term random variations in timing generally appear as variations with negative first-order dependency (alternating long and short durations between onsets, see overview in Madison, 2001a). Reported standard deviations for isochronous tapping (synchronization or continuation) usually are between 3 and 6% (see comparisons in Juslin et al., 2002). In recordings of isochronous tapping sequences performed by three professional drummers, Madison (2000) observed an average standard deviation in mean IOI of 2.8%.

The perception of variations in timing depends on the type of perturbation and when it occurs. For instance, at phrase boundaries, listeners expect performers to make large deviations from the nominal note durations. However, the same lengthening of a note that passes undetected by listeners at a phrase boundary can be detected when appearing in the middle of a phrase (Repp, 1992b). A survey of the JND in different experiments on time discrimination in short isochronous sequences has been made by Friberg and Sundberg (1995). For tempi with IOIs between 200 and 1200 ms, JND varies between 2 and 9% of IOI, depending on the type of timing manipulation done.

To summarize, tempo and timing in music functions on several levels. Both systematic and random variations occur frequently in overall tempo as well as short-term timing. While the short-term variabilities are in correspondence with reported perceptual thresholds, reports concerned with tempo changes appear to be more inconsistent.

Accents

According to Cooper and Meyer (1963) an accent is "a stimulus (in a series of stimuli) which is marked for consciousness in some way". Examples of how such a marking may be done are changes in timing, sound level, or both.

A number of different types of accents are defined in the literature. Parncutt (2003) divided accents into *performed* accents that are added by the musician, and *immanent* accents that are perceived as accented even in a nominally performed score. The performed accents frequently coincide with the immanent accents.

Lerdahl and Jackendoff (1983) defined *phenomenal* (i.e. performed), *structural* and *metrical* accents. The phenomenal accent is a local intensification, i.e. intensity changes or changes in register, timbre, duration, or simultaneous note density. Structural accents are connected to the structure, e.g. a cadence arrival or departure in the music that causes the note to be perceived as accented. The metrical accent is perceived as accented because of its metrical location (position within the measure).

In a number of studies on music performance, performed accented tones have been found to be played lengthened, legato (tied to the following note) and with increased loudness (see e.g. Gabrielsson, 1999, 1974; Drake and Palmer, 1993; Clarke, 1988). A study by Gabrielsson (1974) included rhythms, with and without notated accents, performed on snare (side) drum, bongo drum, and piano. On drums, all rhythms were played with the highest peak amplitude for the first sound event in the measure (metrical accent). This also applied to rhythms played on a single note on the piano. The only exceptions were found for rhythms with a notated accent, where the accent received higher or equally high peak amplitude.

Accents have also been studied in finger tapping experiments. Similar to music performance, Billon et al. (1996) found that an accented finger tap was performed with higher force and longer contact duration. The inter-tap intervals were prolonged after, but shortened before, an accented tap. The movement times for an accented tap were shorter than for other taps, and were initiated from a higher position.

Musical expression

Playing music is not just a matter of mastering a playing technique. We also expect the music to move and engage us, to be *expressive*. Being closely related to mood and feeling, the essence of expressivity is equally difficult to define. Nevertheless, it is possible to study how expressive communication is carried out (Juslin et al., 2002). To date there have been a great deal of studies on expressivity and how expression is conveyed to listeners (for surveys, see e.g. Deutsch, 1999; Juslin and Sloboda, 2001).

Specifically for *emotional* expression, Gabrielsson and Juslin (Gabrielsson and Juslin, 1996; Juslin, 2000, 2001) have listed acoustical cues, i.e. the pieces of information extracted from the sound that help listeners detect emotional intentions.

The most important cues used are tempo, sound level, tone attack, timbre, and *articulation* (the manner in which notes are joined to one another). For instance, a sad performance is characterized by slow tempo, legato articulation, and low sound level, while a happy performance is characterized by fast mean tempo, staccato articulation, and high sound level.

Some attempts have been made to isolate cues in order to find when and how they have the highest influence on the perceived expression. Juslin and Madison (1999) manipulated piano performances with differing emotional intentions (Happiness, Anger, Sadness, and Fear) and asked listeners to rate the expressiveness of these performances. The results showed that the decoding accuracy for the intentions Anger and Sadness suffered greatly when variations in tempo and dynamic level were removed and substituted for the mean tempo and mean key velocities across all performances. The Happy and Fearful performances, however, seemed to rely more on variability in articulation. When the articulation was kept constant throughout performances (relation between note durations and IOIs = 0.7), Happy and Fearful performances were recognized to a lesser extent.

In an experiment designed to identify which acoustic cues contribute to the perceived "expressiveness" of a performance, Juslin (1997) explored 108 cue combinations in synthesized performances. The most expressive combination was (cues in order of strength) legato articulation, 'soft spectrum'[1], slow tempo, high sound level, and slow tone attacks. Juslin noted that there seemed to be a strong relationship between the rated expressiveness and the means to express Sadness/Tenderness.

So, how are the acoustic cues mentioned above related to the instrument played? For instance, legato playing with slow tone attacks can be a problem in percussion playing. Are there any limitations to expressivity for non-tonal rhythms? Will listeners recognize a sad drum performance? Or, in other words, can the same emotions be expressed through percussion instruments as through other instruments?

When Behrens and Green (1993) asked listeners to rate Sad, Angry and Scared solo improvisations performed on timpani, Anger seemed to be readily recognized, while the Sad and Scared improvisations were rated much lower. Other instruments included in the study (violin, trumpet, and voice) were much more successful in conveying the Sad intention. Fear was best recognized when performed on the violin. Unfortunately, no acoustic measures of the performances were included in their study.

Laukka and Gabrielsson (2000) combined investigations of listeners' discriminations of different emotional intent in performed rhythm patterns with acoustic measurements. They found that the emotions Happy, Sad, Angry and Fearful were more easily communicated than Tender and Solemn. The sound levels, timing variations and tempi used by two drummers were compared for different intentions. The

---

[1] Defined as a soft timbre generated through decreased energy in the range above 3 kHz (measured by long-time average spectrum).

softest sound levels were found for Sad and Tender performances, and the loudest for Angry, which was performed about 10 dB louder. A happy swing performance was played at a mean tempo more than three times that of the Sad performance (192 compared to 61 beats per minute). Fearful performances varied so much in tempo that the authors felt it was meaningless to talk about a mean tempo for that intention.

In summary, previous studies show that tempo and dynamic level are important cues in decoding emotion in musical performance, and that the acoustical cues reported for emotional expression in percussion performances seem similar to the cues found for other instruments.

1.2 Studying Movement

Musicians' movements tend to be evaluated subjectively by listening, making movement analysis difficult. Compared to performance in music, there has been a more extensive use of measurement techniques to study performance in sports and athletics. The reason is without doubt that in these areas an improvement is often easily measured in physical quantities like seconds or meters. Regardless of the field of expertise, however, skilled performers (e.g. musicians, sportsmen, fly-fishermen, bricklayers, etc) are able to be (Abernethy et al., 1997):

- fast, yet accurate,
- consistent, yet adaptable,
- maximally effective, yet with a minimum of attention and effort.

When watching experts perform, difficult tasks frequently appear to be easy. Specialized movements performed by a skilled person are known to appear smooth, characterizing efficient energy exchanges (Winter, 1990).

Motor skill and motor learning

The question of how a skilled performer is able to bring about the desired end result with such certainty, and yet at such low energy cost, has been pondered upon during the past century. Much research has been dedicated to the understanding of motor control and several theories, from various perspectives, of how skilled motor behavior is learned and controlled have been proposed (see e.g. Schmidt and Lee, 2005; Turvey, 1990; Rosenbaum, 1991; Shumway-Cook and Woollacott, 1995; Newell, 1991; Huys et al., 2004, for reviews).

Two fundamental concepts in motor control are the *closed-loop* or *open-loop* systems. The closed-loop model of motor control is basically a regulatory system where sensory feedback is sent to the central nervous system and compared to an internal reference. If there is a difference between expected and received feedback, an adjustment needs to be made. Otherwise the movement is considered correct. A closed-loop model suffices to explain ongoing tracking movements, such as keeping a vehicle at a constant speed or maintaining overall tempo in music performance (Schmidt and Lee, 2005). However, for fast movements, e.g. a drum stroke or a stroke in baseball, the feedback loop would be too slow.

In the open-loop model, a movement is specified in a motor command which is carried out without alteration or comparison to feedback. Feedback is employed only after each command is completed and the movement is executed (Schmidt and Lee, 2005). The prestructured motor commands allow for the execution of very rapid movements. Some stereotypical rhythmic movement patterns, like walking or trotting, can also occur without sensory feedback altogether. The idea of abstract *motor programs* proposes central pattern generators, specific neural circuits devoted to a specific type of behavior. However, this idea does not allow for the fact that the nervous system must take into account environmental factors, or the positions of limbs, in order to produce the right movement.

Open- and closed-loop models can describe isolated movements or behavior but fail to account for the fact that we control actions, not isolated muscle contractions. Skilled movements are determined by environmental factors as well as the goal of the movement. Furthermore, one does not tend to pay attention to the order in which muscles or limbs are to be moved. In fact, detailed control over all the individual muscles in a human body would be an incredibly complex affair. This is known as the *degrees of freedom problem*, first posed by Bernstein (1967). Given the approximately 10^2 joints, 10^3 muscles, and 10^{14} cells, it becomes apparent that the central nervous system could not possibly store motor programs that could account for the movement of each individual item (Turvey, 1990).

A solution to the degrees of freedom problem would be to have motor programs control classes of movements, rather than individual movements. Instead of learning and storing every single movement, actions could then be generated from *generalized motor programs* (Schmidt and Lee, 2005). In this organizational perspective, the central nervous system can, by adjusting the necessary parameters, use the same program to generate several different actions. Movements belonging to the same class and generated by the same program will have some aspects in common. Examples would be writing a signature on a cheque and on a blackboard. Although the movements used to produce the signatures are quite different, the signatures mainly differ only in size.

The idea of generalized motor patterns is to reduce the number of dimensions the motor system has to control. However, the number could be reduced even more if groups of joints and muscles worked in synergies (Bernstein, 1967). Instead of controlling each part with an individual command, like a marionette, parts could cooperate in groups forming a unit. If certain muscles are constrained to work together in coordinative structures, the marionette would need fewer strings, i.e. motor commands (Turvey, 1990). This “Bernstein” approach does not only take the nervous system into account, but also considers the body more from its mechanical properties.

The concept of cooperative structures, or synergies, has been further developed alongside the idea of the body as a *dynamic system* (Turvey, 1990; Huys et al., 2004). According to this theory, coordination and control arise spontaneously in high-dimensional open systems that are in permanent contact with the environment, that is, they are *self-organized*. The self-organization perspective stands in clear contrast to the theories where the brain is the controller (e.g. open and closed-loop theories). The system produces patterns that are more efficient under the conditions at hand. As the conditions change the system changes behavior over time. Stable movement patterns are formed, become unstable, and new patterns are formed.

Presently, none of the theories above is able to account for all characteristics of motor behavior, and how skilled movements are learned and developed is still not fully understood. What is known is that practice makes better, if not perfect. Motor skills are also specific to the trained task and are not necessarily easily transferred to other tasks.

Motor control in music performance

In order to excel as a performer on a musical instrument, the mechanical system defined by the combination of body and instrument needs to be fine-tuned and refined during many years of practice. Ericsson et al. (1993) estimated that by the age of 21 the best musicians have spent over 10,000 hours practicing their instrument.

What makes the acquirement of musical skill unique is the learning and refinement of complex motor patterns partly under the surveillance of the auditory system. The coupling between sensory-motor and auditory processing leads to a strong link between the auditory and sensory-motor cortical regions of the brain for musicians (Altenmüller and Gruhn, 2002). Musicians tend to have less asymmetry between left and right sided motor areas and also enlarged sensory areas. These structural changes in brain anatomy are more prominent if training began at an early age. Musicians who have started playing before the age of seven show enlarged motor areas and enhanced interactions between the two hemispheres compared to non-musicians (e.g. Altenmüller and Gruhn, 2002; Altenmüller, 2003).

An example of how a change in perception alters the produced movement patterns is seen in a study by Halsband et al. (1994). In their study, pianists performed sight-read music according to different grouping instructions. From key velocities and movement trajectories of the hands it was seen that the pianists reprogrammed their performance in correspondence to the prescribed grouping. Halsband et al. concluded that the formation of motor patterns was affected by the instructed rhythmic grouping. The change was most evident for the dominant hand, indicating that the process was mainly under the left (dominant) hemisphere control.

Because of the time-constraint imposed on the performance, movements have to be prepared in ample time so that note onsets can occur when planned. Such preparatory movements used by musicians have been reported in several studies.

Similar to coarticulation[2] in speech, musicians use anticipatory behavior when the tempo or piece demands it. For instance, Engel et al. (1997) reported pianists to prepare a "thumb under" movement about two notes before the thumb-played note was to occur. Similarly, Baader et al. (2005) found that violinists used preparatory movements in transitions between left hand positions. Fingers were placed in position for the coming note before the current note had finished. These preparatory movements were only seen during transitions between descending notes where the current note would not be disrupted by the early positioning.

Typically, movement strategies are consistent within players but vary between players. The studies by Engel et al. (1997) and Baader et al. (2005) both reported larger inter-individual (between performers) differences in kinematics and timing compared to intra-individual (within a performer) differences. Similarly, measured muscle activity patterns of string players have been found to be alike between repetitions for the same player, but to vary between players (Fjellman-Wiklund et al., 2004b).

Movement disorders in musicians

A sustainable career as a musician requires, amongst other things, an awareness of ergonomic issues. Musicians perform many specified movements over long periods of time and can suffer from strain-related injuries. Fishbein et al. collected questionnaire responses from 2,212 symphony and opera musicians and reported that no less than 76% suffered from at least one serious medical problem affecting their playing (Fishbein et al., 1988, ICSOM-study). A large part of these medical problems are *playing-related musculoskeletal disorders* (*PRMDs*). In a review of 24 studies of medical problems among musicians, Zaza (1998) reported a prevalence of PRMDs that ranged from 39 to 87% of injuries. Musculoskeletal disorders can be of acute or chronic character; pain in the arm, neck and back being among the most frequent complaints. Most of these problems occur after some sort of overplaying of the instrument (Brandfonbrener and Kjelland, 2002).

Risk factors for developing PRMDs are repetitive and/or forceful movements (Dawson, 2002; Chong et al., 1989) and playing for extended periods without taking breaks (e.g. Zaza and Farewell, 1997; Chong et al., 1989). PRMDs are also related to gender and the instrument played (e.g. Zaza and Farewell, 1997; Zetterberg et al., 1998; Dawson, 2002). According to Dawson (2001), the typical patient is "likely to be a female pianist or string player". In an examination of risk factors in classically trained musicians, Zaza and Farewell (1997) reported four times the risk for string players to develop PRMDs compared to that for non-string players. Protective factors included using a musical warm-up and taking breaks, elements that Zaza and Farewell suggested should be incorporated into musical training to a greater extent than at present.

---

[2] The adjustment of a speech sound to the context. In fluent speech, speakers change the pronunciation of phonemes depending on what is to follow.

Applications in teaching and performance evaluation

It is not necessarily so, that having learned control from extensive training, we are consciously aware of the movements we make. Furthermore, the movements can be too small or too fast to be observed with the human eye. In many cases teachers and students have to rely on metaphors or intuition to induce the right 'gut feeling'. But to be able to actually *see* what is happening and what movements are used can bring valuable insights. Music professor H. Winold explains why she used high-speed cameras to study cello performances (Winold, 1984):

> "High speed photography intrigued me first when I saw a small segment of a Heifetz[3] film in slow motion. Suddenly I could see preparatory movements, reaching for groups of notes, and minute adjustments required by particular passages in Heifetz's hand, a hand that had seemed almost motionless at regular speed."

Many of Winold's recorded subjects were surprised to see that they were playing a vibrato in a different way than they had thought.

In order to assess and evaluate movements and techniques used in music performance, methods like *electromyography* (EMG) have been used (see e.g. Kjelland, 1992, for a survey). Many studies have utilized EMG to investigate control in the playing of string instruments, quite likely due to the over-representation of string players among musicians suffering from PRMDs.

Guettler (1992) found that using EMG could be useful for visualizing techniques for vibrato playing on the double bass. Thiem et al. (1994) used EMG to evaluate whether a playing technique using "rhythmic cuing" could help to relieve stress on the fingers during cello playing. The effects were not significant, possibly due to the short training period (two weeks) between pre- and post-tests.

Similarly, Fjellman-Wiklund et al. (2004a) investigated the effect of basic body awareness therapy on the EMG activity for a group of string players. The group of subjects receiving body awareness therapy experienced being more relaxed and reported improvements in postural control and concentration. However, the study did not show any significant differences in EMG between groups. The authors concluded that an eight week training period was not enough to change the behavior of the experienced players.

Because established motor patterns tend to be robust, several authors state the importance of learning the right types of movements from the beginning (Fjellman-Wiklund et al., 2004a; Altenmüller and Gruhn, 2002; Brandfonbrener and Kjelland, 2002), especially since the brain develops somewhat differently for musicians starting to play before 7 years of age (Altenmüller, 2003). With this follows, of course, a great responsibility and a close collaboration between teachers, physicians, and scientists. What are the right types of movements that should be taught? Further

---

[3] Jascha Heifetz (1901-1987), Russian-American violinist, considered one of the greatest violinists ever.

studies of musicians' movements are important for self-monitoring and didactic reasons, but also for investigating relationships between learned movement patterns and PRMDs.

Visual perception of movement

Musicians move their bodies in ways not always directly related to, or needed for, sound generation. Changes in posture, large body sway, or other types of movements (conscious or unconscious) are frequently seen in performing musicians. Wanderley (2002) refers to these movements as *ancillary*, *accompanist*, or *non-obvious* movements. Some of these movements are, however, clearly intended to be expressive, or used for communication. For this reason these movements can be viewed as a kind of body language. Four aspects that can influence this body language have been suggested by Davidson and Correia (2002): (1) Communication with co-performers, (2) individual interpretations of the narrative or expressive/emotional elements of the music, (3) the performer's own experiences and behaviors, and (4) the aim to interact and entertain an audience.

Humans are apt at extracting information even from very sparse visual displays. In the early seventies, Johansson demonstrated that even point-light displays, where lights or reflective markers fastened onto a person are filmed in a darkened room, producing a clear impression of human movement (Johansson, 1973). Even when human movement is reduced to the motions of a small set of points, subtracting all other characteristics of a person, it still is possible to recognize various properties of the movement (see e.g. Pollick, 2004, for review). For instance, observers are able to recognize the sex of a person walking (Cutting and Kozlowski, 1977; Kozlowski and Cutting, 1977), and the weight of a lifted box (Runeson and Frykholm, 1981). Furthermore, observers are able to use the information in body movements to discriminate between different expressive intentions, emotions or affects. This has been shown both for music and dance performances (e.g. Walk and Homan, 1984; Dittrich et al., 1996; Davidson, 1994), but also for every-day arm movements like drinking and lifting (e.g. Pollick et al., 2001).

The elements of importance for conveying the information to observers are still not fully recognized. Some work has been devoted to the investigation of the types of movements that provide information needed to distinguish the moods of a performer, i.e. the *movement cues* used.

Paterson et al. (2001) found a relationship between the velocity of the wrist movement and how observers rated knocking and lifting movements. High velocity resulted in high ratings of affects with high activation; Angry, Excited, Happy and Strong. Manipulating the original movement stimuli by altering the average velocity affected the observers' ratings. For the three original affects, Sad, Neutral and Angry, there was a clear effect on the classification and intensity ratings as movement duration increased. However, even with much lower mean velocity, Angry movements were seldom categorized as Sad. Paterson et al. concluded that movement velocity plays a role in the perception of affect from movement, but that there are other properties of importance that are not controlled by velocity.

Boone and Cunningham (2001) found that children as young as 4 years old systematically varied their expressive movements when moving a teddy bear to Angry, Sad, Happy and Fearful music. The children used more upward movement, rotation, force, shifts in movement patterns and a faster tempo for Happy compared to Sad music.

De Meijer (1989, 1991) and Boone and Cunningham (1999) proposed several movement cues considered important for detecting emotional expression (see overview in Boone and Cunningham, 1998). These cues include frequency of upward arm movement, the amount of time the arms were kept close to the body, the amount of muscle tension, the amount of time an individual leaned forward, the number of directional changes in face and torso, and the number of tempo changes an individual made in a given action sequence. The proposed cues closely matched the findings by De Meijer concerning viewers' attribution of emotion to specific body movements (1989; 1991). For instance, he found that observers associated actors' performances with Joy if the actors' movements were fast, upward directed, and with arms raised. Similarly the optimal movements for Grief were slow, light, downward directed, and with arms close to the body.

Camurri et al. (2003) also found a connection between the intended expression of dance and the extent to which the limbs are kept close to the body. In their study, automatic movement detection was used to extract cues in rated dance performances with the expressive intentions Joy, Anger, Fear and Grief. The cues studied were amount of movement (Quantity of motion), and how contracted the body was, that is how close the arms and legs are to the center of gravity (Contraction index). They found that performances of Joy were fluent, with few movement pauses and with the limbs outstretched. Fear, in contrast, had a high contraction index, i.e. the limbs were often close to the center of gravity.

The fact that the position of the arms is an important cue is somewhat contrasting to the conclusion that the visual information is enough for observers to distinguish between music performances with different expressive intentions (e.g. Davidson, 1993, 1994; Sörgjerd, 2000). Musicians primarily use their arms for sound generating movements and have small opportunities for raising them in the air during playing. Therefore, the cues used by observers to identify expressive intentions in music performances either have to appear in other parts of the body, or *coincide* with the playing movements.

In studies of movements in clarinet performances, Wanderley et al. (2005) reported that the players tended to decrease their movements in active or technically demanding passages in the piece, while easier passages often were accompanied by more movements. Players would also sometimes shift the onsets of their movements with respect to the score, either by slightly anticipating the music or following it. Such phase shifts between movement and sound also affected how subjects perceived the phrasing. For instance, subjects who both saw and heard performances rated the phrase as longer if the players' gestures extended into the silence.

It is a well-known fact that visual information influences the perception of auditory signals, the classic example being the "McGurk effect" (McGurk and MacDonald, 1976). However, similar cross-modal interference has also been reported for non-speech sounds. Saldaña and Rosenbaum (1993) found a significant effect of video information on how subjects rated cello tones as either plucked or bowed. Recent work has shown a similar "visual pull" on the judgments of marimba tones (Schutz, 2004). The cross-modal interference is not restricted to the visual information influencing the perception of auditory events. The reverse effect, a visual illusion induced by sound, has also been demonstrated (e.g. Shimojo and Shams, 2001; Shams et al., 2002).

To summarize, visual information in human movement helps observers to extract various types of information about a performance. Results from several studies suggest that visual information constitutes an important channel of additional information in music performance.

1.3 Sensory feedback in music performance

Several different types of sensory feedback are available to the musician during performance; visual feedback from seeing one's own or others' movements, tactile feedback from the physical contact with the instrument, kinesthetic feedback from the proprioceptors in joints and muscles, and, of course, auditory feedback. Of all our senses, hearing is the most accurate in terms of temporal precision (Levitin et al., 1999).

One might assume that for a musician, auditory feedback is critical for performance, but this is not always the case. Although pianists learn melodies better with auditory feedback, the number of errors produced in performance of recalled pieces does not increase if auditory feedback is absent (Finney and Palmer, 2003). Finney (1997) showed that no auditory feedback is better than the wrong type of feedback. Repp (1999) reported minor differences between piano performances with and without auditory feedback. In a listening test, expert listeners were only able to tell apart the modes of performance apart in 64% of cases. These findings suggest that once the musical piece is learned, performance is not dependent on auditory feedback.

In the context of synchronization, the role of sensory feedback becomes more intriguing. The ability to synchronize demands an anticipation of a stimulus onset instead of a reaction to it, making the act of synchronization essentially different from many other types of behavior. In order to synchronize to other events (or musicians), the nervous system has to estimate when an event is about to occur and to initiate the action that is to match it. Afterwards, feedback on the outcome (accurate or too early/late) can be used for correcting subsequent matching attempts. The question is how is this done. It has long been known that when synchronizing to a metronome, subjects typically tap 30 ms *ahead* of the beat (see e.g. Fraisse, 1982; Mates and Aschersleben, 2000; Wohlschläger and Koch, 2000;

Aschersleben et al., 2001). This *negative asynchrony*, or *anticipatory error*, is very robust. Only after being informed about the direction and magnitude of the error, are subjects able to produce taps in exact synchrony (Aschersleben, 2002).

The negative asynchrony is maintained because subjects perceive it as being exactly on the beat. If the pacing signal of the metronome is exchanged with the auditory feedback of the subject's own taps, tempo tends to increase. This can be interpreted as a tendency to reestablish the negative asynchrony, even though the taps could not be more synchronized than they already are (see Aschersleben, 2002). By including auditory feedback from the taps, the negative asynchrony decreases. Musical skill also seems important. On average, musically trained subjects display about 10 ms less error compared to untrained subjects (Aschersleben, 2002).

By altering the pacing signal to a tactile or visual stimulus, the negative asynchrony decreases. Increased amplitude of the tapping movement, resulting in increased force, also reduces error (Aschersleben, 2002). Increased negative asynchrony is observed when subjects tap with their foot, knee or toe instead of their finger (Wohlschläger and Koch, 2000), or when tactile feedback is excluded using local anesthesia (Aschersleben et al., 2001). These findings have been taken as evidence that it is not the movement onset, but the sensory feedback from the action, that is matched with the stimulus (Aschersleben, 2002). However, since the negative asynchrony predominantly occurs for synchronization to "empty" intervals and is reduced by subdivision or added tones, Wohlschläger and Koch (2000) proposed that it is simply an error in time estimation.

Delayed Auditory Feedback.

Most acoustic instruments give an almost instantaneous response[4]. The player uses this response as a major source of feedback in overseeing the performance. Electronic instruments, dependent on electronic amplification, can introduce delays in the player–instrument loop. The delays can be due to signal processing in the synthesizing process, or even due to too large distances between player and loudspeaker.

Several studies have investigated the role of latency in response by delaying the auditory feedback to the player: *Delayed Auditory Feedback* (*DAF*). It has been shown that DAF causes players to decrease tempo, produce more errors, and increase timing variability (e.g. Gates and Bradshaw, 1974; Finney, 1997; Pfordresher and Palmer, 2002). Pianists playing with DAF also have been reported to produce increased key velocities, implying that they play harder so as to increase the tactile feedback on their performance (Finney, 1997).

In rhythmic tapping, Finney and Warren (2002) reported that subjects made the most errors when the delayed feedback coincided with the subsequent tap. Also for pianists, performances are considerably impaired when delay values are close to IOI, so that the feedback of the preceding event appears at the onset of

---

[4] Although organists may have to cope with delays of up to several hundred milliseconds.

the following one (Pfordresher and Palmer, 2002; Pfordresher, 2003). By contrast, timing variation decreases for delay values that coincide with subdivisions of the performed tempo (Pfordresher and Palmer, 2002).

Most studies on DAF have manipulated feedback with delays of 100 ms or more, which is well above the JND for tempo perturbation (typically 2–5% of IOI, see Friberg and Sundberg, 1995, for a survey of JNDs). Although the maximum disruption in performance seems to occur for values corresponding to IOI (Finney and Warren, 2002), the latencies that are acceptable for performance still need to be studied further. In a study of ensemble clapping, performed under DAF and no visual feedback from co-performers, Gurevich et al. (2004) observed that for delays of 20 ms or more tempo decreased. However, for no or short delay values, 2 and 5 ms, subjects instead *increased* tempo. In addition, recent research indicates that compensatory behavior occurs also for timing perturbations below the perceptual thresholds (e.g. Repp, 2000). It appears as though the motor system may have access to timing information which does not reach perceptual levels.

To summarize, delays of the same order as IOI, causing the auditory feedback to coincide with the following event, leads to maximum disruption in performance. Large delays frequently cause a decrease in tempo, while it appears that small or no delays in ensemble synchronization can lead to increases in tempo.

1.4 Playing percussion instruments

All musicians strive to master rhythm and timing in their performances, but for percussionists these words carry a special importance. In many ensembles, the function of the percussionist/drummer is to be *the* timekeeper. Keeping a steady rhythm and tempo are fundamental elements in any percussion training. As part of the work presented in this thesis focuses on the movements and timing in percussion playing, this section will give a background to the playing of percussion.

All percussion instruments share some properties that distinguish them from most other instruments (e.g. winds or strings).

- **Diversity – not one instrument but many.** Percussion includes a huge variety of instruments, both membranophones (e.g. drums) and idiophones (e.g. cymbals, wood blocks, etc). No other instrument family can so easily fill up a whole encyclopedia (e.g. Olsson, 1985). Normal percussion playing requires that the player performs the same rhythm on different instruments with differing physical properties[5]. A change of instrument changes the kinesthetic feedback to the player, who must adapt to the conditions at hand. In general, this will lead to many compromises, as a single drumstick or mallet may not be equally suitable for all the instruments that are to be played.

---

[5] For instance, world-renowned solo percussionist Evelyn Glennie has a collection of 1400 instruments and travels with, on average, 1.7 tons of equipment.

- **Short interaction times.** Percussion instruments generally produce sounds with impulse-like characteristics. Normally the note onsets are well defined, the duration of the excitation is short, and in general the player has little control over the tone, once initiated. The note can be shortened by dampening (e.g. by forcing the mallet to stay in contact with the drumhead after the hit), but it cannot be lengthened. While, for instance, players of wind instruments have close control of the vibrating air column during the full duration of a note, the percussion player's direct contact with the instrument is limited to a few milliseconds. This implies that whatever the resulting striking force and dampening effect the percussionist wants to induce would need to be integrated in the entire striking gesture. The mallet will strike the drumhead (or some other object) with the velocity and effective mass supplied by the player's movement, and the same striking gesture will also determine the contact duration.

- **Changing pitch and timbre is awkward.** In order to modify the timbre or the fundamental frequency a percussion player can strike at different positions, or choose a mallet with a different mass, hardness and/or shape. The stiffness and tension of a drumhead can also be adjusted. However, in normal playing there is seldom time for such types of adjustments. A change of timbre or pitch is typically achieved by striking another instrument with its own specific characteristics and limitations. This can lead to conflicts in several cases. For instance, optimal position might not be possible to reach in optimal time. From this follows a need to plan ahead in order to reach the instruments that are positioned at varying distances from the player.

Playing techniques

Players of percussion instruments strive to acquire playing techniques that can be adapted to the feedback from the instrument. Most techniques share a common ground regarding the quality of "efficient" playing. This general consensus can be formulated as three guidelines of how to facilitate playing:

1. Let the drumstick/mallet do most of the work.
2. Take advantage of the rebound that is given by the instrument
3. Plan stick positions ahead

All the three points above rely on the wrist to work as a hinge, relaxed and flexible. The acquirement of relaxed and flexible movements is one of the main issues in learning drumming techniques. According to Cook (1988), the striking movement should be "like waving good-bye or bouncing a ball." With a cramped grip the drumstick does not have enough freedom to move and accelerate. As a result, the command 'Relax!' frequently occurs in instruction manuals.

Cracking a whip. The resulting excitation of the instrument ('shock spectrum') is related to the history of the contact force. A high velocity at impact results in a high and short force pulse. This corresponds to a high sound level and a rich spectrum with strong high-frequency partials. An alternative to striking with the bare hand is to use a hard tool, which makes it easier to excite even more partials. A stick or mallet can help a player to excite an instrument more vigorously by increasing the striking velocity at low physiological cost. At low dynamic levels this is easy enough; just by letting the stick fall there will be a sound. As dynamic level is increased more force is needed, something that can be achieved by providing a sufficient "runway" during which the stick can be accelerated.

The general method of beating a single stroke can be described as follows by Shivas in "The Art of Tympanist and Drummer":

> "The actual stroke may be quite aptly likened to the action in cracking a whip. The tip of the stick is held about an inch above the drumhead and the stick is flicked *upwards* and then 'cracked' downwards till it strikes the head, which will, by its elasticity, throw the stick back again in an upward direction. " (Shivas, 1957)

An example of how a single stroke is played is shown by the 'stick figure' in Figure 1.1. The figure displays the three-dimensional positions of five markers on a player's stick, hand, wrist, elbow and shoulder during a stroke. The three loops mark the trajectories of the tip of the drumstick (leftmost loop), the hand (middle loop) and the wrist joint (right loop). The figure shows how the hand and wrist lead the drumstick. The preparation for the stroke is initiated by raising the wrist with the stick lagging behind, tip pointing down. The stick is first flicked up to a vertical orientation and then flung down to the drum head, gaining velocity over the full height of the stroke.

In a three-dimensional tracking study of drumming movements, Trappe et al. (1998) found differences between non-drummers, beginners, students, and professional drummers. The motion patterns of the professionals were found to be flexible and whiplash-like. The students showed similar patterns, but the calculated angles between segments (drumstick, hand, lower and upper arm) showed less control of the stick compared to the professionals. Compared to the drummers, the wrist movements of the non-drummers were stiffer and less flexible.

The strategy for prolonging the arc during which acceleration occurs can be found also in other contexts. Players of tennis, baseball and golf all display similar striking movements when they have time to do so (Wickstrom, 1983; Abernethy et al., 1997).

Letting it bounce. A relaxed grip of the drumstick also opens up the possibility of taking advantage of the rebound from the surface. In drumming, particularly in snare drum playing, the player tries to take advantage of the fact that a normal drumhead is elastic and therefore will "send the stick back to where it came from"

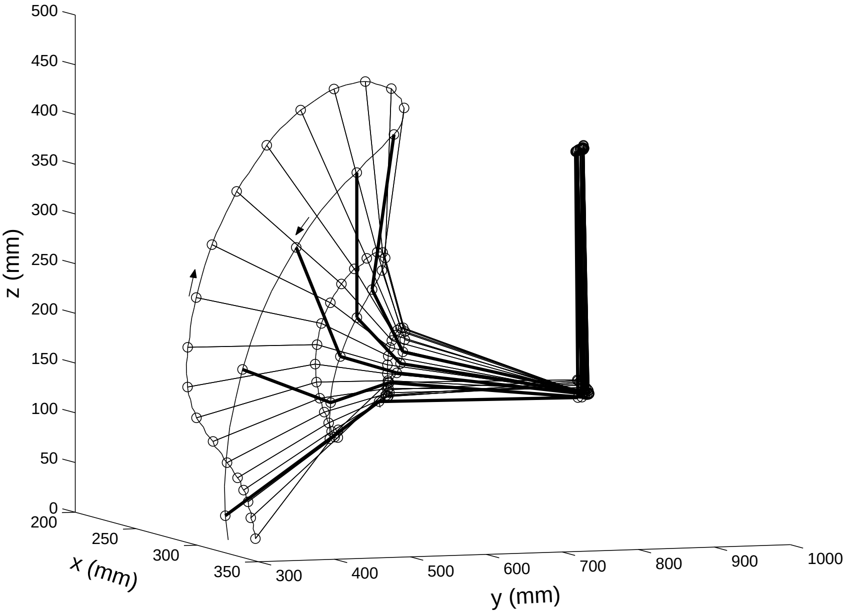


Figure 1.1: Schematic representation of the movement of the drumstick and the player's arm movement during a stroke. The circles mark the positions of five markers on the drum stick, hand, wrist, elbow and shoulder. The leftmost, largest loop describes the trajectory of the tip of the drumstick, while the two inner trajectories represent the path the hand and wrist take, respectively. The preparation for the stroke starts with an upward movement of the wrist, dragging the drum stick upwards. The stick is flicked upward, passing the hand and taking a vertical orientation. Also during the actual down stroke (here indicated by thicker lines) the wrist leads the downward movement. The drumstick follows in a whiplash manner, gaining high velocity from the elongated arc produced by the preparatory motion. (The figure was generated using 3D data from Waadeland, 2003. Time separation between frames about 1 ms).

(Shivas, 1957). The rebound throws the stick away from the surface quicker than the player could him/herself. By controlling the rebound it is possible to play more than one stroke per movement gesture, an energy-efficient technique that is used for fast rolls.

The use of the rebound leads to a convenient "working area" with combinations of tempi and dynamics that are easier to master than others. As when bouncing a rubber ball, certain heights and tempi fit well with the natural periodicity of the ball. However, as the ball is forced closer to the ground it becomes increasingly difficult to maintain a slow tempo. Similarly, at soft dynamic levels, the rebound from the drum head can be a bit too soft to give the desired "flow" of movement at some tempi. On the other hand, loud dynamic levels give a rebound which can be too strong to master. Both these cases impose difficulties that have to be dealt with within the given time constraint. As a result, pushing the boundaries to increase the working area of playing demands more control of the player.

Planning ahead. By initiating a transition between one position and another as soon as possible, the player strives to maximize the time between events. One example of how this is done can be found in the playing of accented strokes. In drumming, one of the most commonly used features is to accent a stroke by increasing loudness and much time and effort is spent on mastering the playing of accented rhythm patterns.

To facilitate the playing of strokes at different dynamic levels, some of the playing techniques taught specify at what height the stick should be *at the end* of the stroke. The terms tap, upstroke, downstroke, and full stroke are commonly used to help the performer plan and carry out the right movements (see e.g. Moeller, 1956; Famularo, 1999). Using these basic strokes the player can execute sequential strokes of different heights. This is a prerequisite for playing various combinations of strokes at different dynamic levels, including accents. An example of their use for the playing of an accented stroke is shown in Figure 1, Paper I.

Auditory vs. tactile feedback in drumming

The time the drumstick stays in contact with the drum head is very short (usually between 5 – 10 ms, see Dahl, 1997), but it determines the timbre and quality of the stroke. The time is so short that the player has no opportunity to adjust something once the drumstick hits the drum head. It seems reasonable to assume that to get the desired result the movement must be right throughout the whole stroke, from beginning to end, including the final position after the rebound. This implies that the auditory and tactile feedback from the short contact time is important for the player to plan and oversee the performance.

When a player is performing on electronic drum pads and synthesized drums, the relationship between the tactile and the auditory feedback is somewhat different than for acoustic drums. The recorded or synthesized sounds make it possible to change the acoustical properties of the drum very easily, without affecting the physical properties of the instrument. A change of sound is no longer necessarily connected to a corresponding change in the tactile feedback. Consequently the sounding stroke may not correspond exactly to the characteristics of the played stroke. For the player, the mismatch between sound and tactile feedback introduces a conflict. This feature would not be possible when playing acoustical drums, and hence to play electronic drum pads means conquering a different instrument.

A computer-based physical model of an instrument responds naturally to gestures during playing. Such a model, however, can still introduce a delay in the response. It is therefore important to know to what extent such perturbations can be tolerated.

Delayed auditory feedback in drumming

In a series of experiments, Dahl and Bresin investigated drumming with DAF. In a first test, subjects used an electronic conducting device[6] to synchronize with a metronome. By gradually increasing the auditory delay, it was observed that subjects strived to match sound with sound by consistently increasing the negative asynchrony, causing strokes to occur earlier (Dahl and Bresin, 2001).

In the following experiment, Bresin and Dahl (2003) studied effects of DAF on drummers playing with and without a metronome. Ten drummers played single strokes at two tempi, 120 and 92 BPM (nominal beat separation 500 and 652 ms, respectively) on an electronic drum pad. Five of the players synchronized to a metronome throughout the test session. For the other five drummers, the metronome was turned off after a short synchronization phase. A delay was introduced in steps from 0 to Δt and then back to 0. For *upward* sessions Δt was incremented by 5 ms each time the delay was introduced, going from small Δts to 180 ms. In *downward* sessions, Δt first started at 180 ms and was decreased. An example of how the changes in auditory feedback were introduced to the player during an upward session is shown in Figure 1.2.

General impressions during the experiment were that players easily coped with short delays. In the beginning of the first upward session, playing appeared easy and there were but minor disturbances in timing. As auditory delay increased, subjects tended to increase the striking force to achieve more tactile feedback from the drum pad, in line with the behavior of pianists' DAF performances (see Finney, 1997). Drummers performing without metronome generally decreased tempo throughout the session. In both groups there were also players who tended to either increase or decrease tempo during the periods where the delay was applied.

Comparing the performances for all subjects, some observations could be made. For instance, large differences between adjacent IOIs did not necessarily coincide with the times where the delay was applied or removed. Rather, it would normally take about 4–5 strokes after the shift before the player would 'lose track' in the rhythm (see Figure 1.2). Strokes at approximately double the tempo were frequent,

---

[6] Max Mathew's radio-baton (Boulanger and Mathews, 1997).

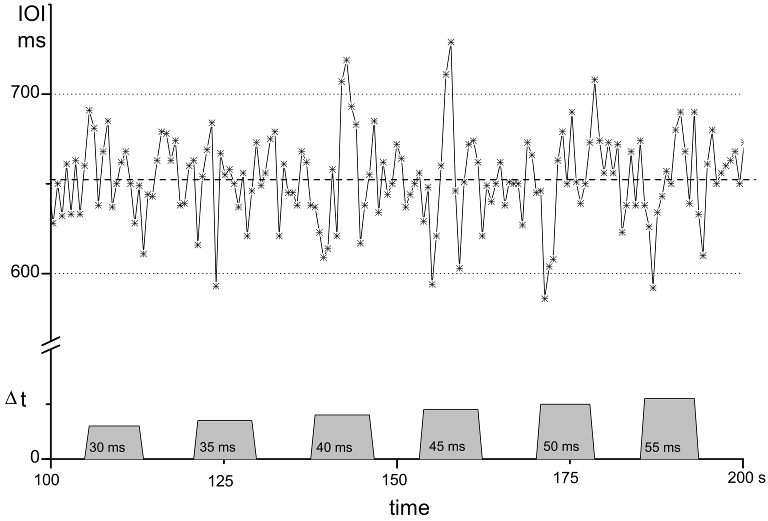


Figure 1.2: Example of raw data from one of the subjects synchronizing to a metronome. The top curve shows the successive IOIs encircling the nominal tempo, 92 BPM (nominal IOI 652 ms) indicated by the dashed line. At the bottom of the figure the amount of delay and the onset times can be seen. It can be noted that the larger differences between adjacent IOIs do not necessarily coincide with the shifts in Δt.

causing large variations in standard deviations and mean IOIs. Although errors, these mistakes indicate that the subjects tried to maintain tempo by subdivision. There were, however, plenty of strokes that were not close to any subdivision.

Standard deviations were generally higher for the synchronizing drummers (mean value 5.2%) compared to those doing the continuation task (mean 3.1%). The maximal standard deviations, reaching 14%, also occurred for the subjects playing with a metronome. This was not surprising, considering that synchronizing to a metronome is in itself no trivial task (Madison, 1992).

The average spread in IOI for lower delay values was about 3%, close to the reported JND for perturbations in isochronous sequences, (see Friberg and Sundberg, 1995). However, for certain values of Δt it was more difficult for the drummers to maintain tempo and avoid errors. Figure 1.3 displays an abrupt increase in mean standard deviation for delay values of more than 100–110 ms for the synchronizing group, indicating a shift in performance. For the drummers doing the continuation task no clear shift appeared. For this group, however, gradual drifts in tempo were a more dominant feature.

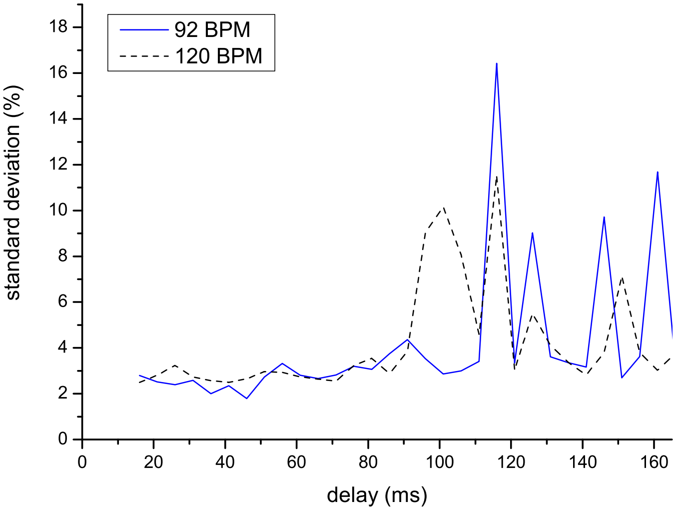


Figure 1.3: Relative spread in IOI for drumming with delayed auditory feedback. The curves show the standard deviations across all normalized IOIs and drummers synchronizing to a metronome for different delays. The two tempi are indicated by line shape; 92 BPM (solid) and 120 BPM (dashed). The standard deviations for delays up to about 100-120 ms remain close to 3%, in correspondence to the JND for perturbations in isochronous sequences (see Friberg and Sundberg, 1995).

Summarizing, the experiment with DAF in drumming suggests that it is possible to continue playing for short delay values (of the order of 50 ms). However, the players experienced increasing difficulties, as reflected in occasional large timing errors (synchronizing group) or tempo drift (continuation group). For delays above 120 ms players tended to reject the auditory feedback and rely on the tactile feedback of the stroke.

1.5 Objectives and specific questions for the thesis

This chapter has reviewed some of the research done on the interaction, perception, and communication involved in music performance. The field is large and complex and in general it is not possible to deal with more than small sections of it at one time. With respect to the complexity of skilled music performances, research has to bring together knowledge from many different disciplines to form a better understanding of the processes involved.

The work presented in this thesis investigates different aspects of music performance. Diverse as they may seem, each aspect concerns motoric and perceptual skills performed under the time-constraint of tempo. General aims for the work in this thesis were three-fold.

1. To gain increased knowledge on how the planning and control of separate, but musically relevant, elements in musical performance is altered in response to changes in tempo and other playing conditions (e.g. tactile feedback of the instrument). Specifically, an aim was to investigate whether drummers acquire playing strategies applicable for different combinations of tempi, dynamic levels, and rebounds.

2. To expand the knowledge of how listeners relate to changes in tempo. Questions were whether the perception of a "steady tempo" is shared among listeners, and whether their opinions change with test occasion and nominal tempo. Specifically, the aims were to estimate JND for perception of tempo drift; to investigate whether there are individual biases in the perception of steady tempo, and, if so, if the bias could affect overall threshold estimates.

3. To isolate and describe the intrinsic communicative components in musicians' movements during music performance. Specific questions were whether important information on emotional expression is conveyed through movements directly involved in sound generation, or appear in other parts of the players' bodies, and how much of this communication is affected by player and instrument. An additional aim was to investigate if the communication could be described in terms of movement cues.

In addition to these aims, the study of body movements in music performance can provide valuable insights into how musicians perform and communicate when playing music. In particular, four objectives for studying musicians' movement can be outlined:

- To describe playing techniques and their effect on the production of sound.
- To increase our understanding of sensory-motor integration in the optimization of skilled motor behavior with auditory feedback

- To increase the knowledge about ‘efficiency’ and ‘ergonomics’ in skilled movements used for artistic, as opposed to competitive, purposes.
- To explore visual communication through movements in music performance.

Chapter 2

Contributions of the present work

2.1 Original contribution and importance of the present work

The papers included in this thesis deal with different aspects of human movement and music performance, representing three layers of interaction and perception:

1. At the lowest and most “automated” level, there is the actual tone production in playing, using highly specialized movements that have been learnt specifically for controlling basic musical elements.
2. The medium level concerns the overall temporal context, including meter and tempo, in which the small musical elements must be placed correctly to be meaningful, among other things marking the structure of the piece.
3. At the highest level, the low-level note production as well as temporal envelopes become means that the player uses to shape the musical performance, and to express and communicate emotions and intentions (personal as well as composed).

Presently, none of these three layers are fully understood. As described in Section 1.2, there has been some earlier work on musicians’ playing movements (e.g. Baader et al., 2005). However, none of these have systematically investigated the effect of different playing conditions on motor behavior. There are hardly any previous works on movements in percussion playing (Trappe et al., 1998, 2005; Waadeland, 2003).

With respect to human perception and production of timing, there has been a lot of work devoted to the understanding of the mechanisms involved (see e.g. Desain and Windsor, 2000). For short-term timing perturbations (e.g. displacement of single events), the reported just noticeable differences are supported by the reported values of spread in production. This is not the case for previous work on long-term variations. When comparing data from production of time series with reported perceptual thresholds there are substantial differences. As noted in Chapter 1, the JNDs of gradual changes in tempo (Madison, 2004) have been reported to be about the same as the JND for stepwise changes in tempo, about 2% at 120 BPM (see survey in Friberg and Sundberg, 1995). This implies that in order to detect a continuous change in tempo, the successive change would have to be of about the same magnitude as the JND for a single, stepwise change in tempo. A change of 2% per interval for a sequence of nine intervals results in a total tempo change of 16%, or 80 ms for a nominal tempo of 120 BPM. What could possibly explain this paradox?

Many theories on the mechanisms involved in timing behavior share the common idea that the system strives towards a *stable* tempo, tacitly assuming that the "ideal" tempo is isochronous (e.g. Vos et al., 1997; Ogden and Collier, 1999; Collyer et al., 1994). However, it seems that an isochronous internal reference fails to explain many of the recurrent features in production and perception of tempo. Is it reasonable to assume that isochrony has a special status in our perception of temporal events?

As described in Section 1.1 earlier work on music performance has increased our knowledge about the communication chain between listener and musician. Earlier work (see Section 1.2) has also shown that musicians' movements can convey expression to observers. However, the specific cues that make this communication possible have not previously been investigated. Basically, the movements of musicians' arms and hands are primarily dedicated to sound production, and constrained to specific time events. The expressive movement cues used by the observers to detect emotional expression must thus either appear in other parts of the body, or *coincide* with the actual playing movements. What types of a musician's movements are important to convey expression to an observer?

The original contributions of the work in this thesis can be summarized as:

Paper I describes movement and timing patterns involved in the playing of accented strokes in drumming. The results show inter-individually different movement patterns between players, and that brief cyclic patterns in the timing of strokes can convey grouping information to listeners.

Paper II investigates how differing playing conditions affect players' movement strategies for playing interleaved accents in drumming. The results show that players maintained their individual movement strategies for most playing conditions. The dynamic level has a strong influence on the height to which the drummers lift the drumstick in preparation for a stroke, on the striking velocity, and on the timing of strokes.

Paper III suggests a novel approach to estimate the perceptual threshold of tempo drift. The results show that listeners are more apt at hearing gradual changes in tempo than what has been previously reported. The found values correspond well to values of tempo drift in production data. Most listeners display an individual bias in their perception of tempo drift.

Paper IV deals with characterization of movement cues important for the communication of specific expressive intentions in music performance. The results show that such a communication is possible, and that there is a clear relationship between the movement cues used and the audio cues reported for music performance.

2.2 Paper I: The playing of an accent

In relation to the lowest, most automated, low-level aspects of music performance, four percussion players' strategies for performing an accented stroke were studied. The main objective was to investigate what kind of movement strategies the players used when playing interleaved accented strokes. An additional objective was to study the timing performance of the players and how listeners perceived it.

Three professional percussionists and one amateur (the author) played on a force plate with markers on the drumstick, hand, and lower and upper arm. The movements trajectories were recorded using a motion detection system (Selspot), tracking the displacement of the markers at 400 Hz. The rhythmic pattern – an ostinato with interleaved accents every fourth stroke – was performed at three dynamic levels (*pp*, *mf*, and *ff*), and at three tempi (116, 160, and 200 BPM).

Movement trajectories.

The analysis showed that the four players used individual movement strategies to prepare for the accented stroke. The movement strategies were maintained consistently within each player, but differed considerably between players both with respect to the phase between limb segments and to what extent the arm was used in the movement (see Figure 2.1). The accented strokes were initiated from a greater height than the unaccented strokes. Subject S2 displayed the largest difference in movement trajectories between the two types of strokes.

The timing of strokes

With regard to the percussionist as time keeper, the timing of the strokes ought to be of special concern when studying playing strategies. If we hypothesize that one of the reasons for the preparatory movement prior to the accent is to deliver the accented stroke “on time” the natural question to ask is how the drummers performed in timing. None of the subjects were, however, informed that their timing performances would be analyzed.

The timing analysis concentrated on the strokes recorded by the force plate, and their separation in time, the *inter-onset intervals (IOI)*. Inter-onset interval number n is defined as the interval between stroke number n and $n + 1$

All subjects displayed both long-term changes in tempo (drift), and short-term variations between adjacent IOIs. In general, every fourth interval, i.e. the one beginning with the accented stroke, was prolonged. In addition, recurrent cyclic patterns with regularly replicated IOIs during three to four measures could be observed. An example of the timing in one of the recorded sequences can be seen in Figure 2.2 (top sequence). A particularly clear example of a prolongation of the fourth interval can be seen in intervals No. 36–50.

Some of the recorded sequences showed a drift in tempo that was either increasing or decreasing (see Figure 4, Paper I). The drift was estimated as linear and

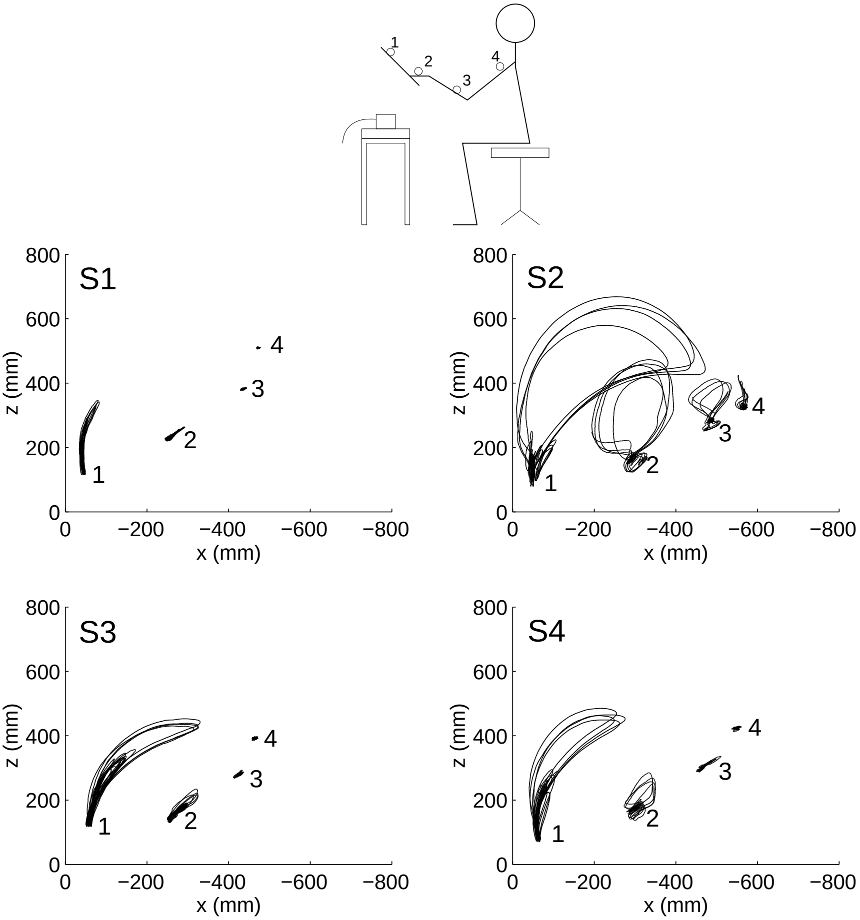


Figure 2.1: Movement trajectories captured from four markers: on the drumstick, and on the subjects' hand, lower, and upper arm. Side view from the players' left side; vertical direction (z-axis) vs. horizontal direction (x-axis). Subject S1 (top left), S2 (top right), S3 (bottom left), and S4 (bottom right). Each panel includes approximately four measures at *mf*, 200 BPM, played on the normal surface. The preparatory movements for the accented stroke can be seen as a larger loop compared to that of the unaccented strokes. The players' drumstick, hand lower and upper arm are involved to different extent in the movements.

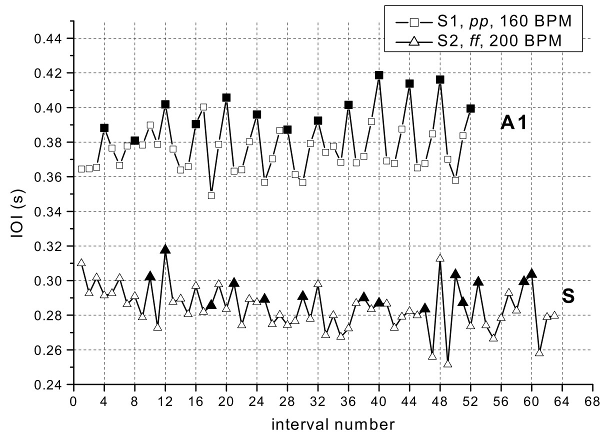


Figure 2.2: Example of timing performance for regularly (top) and irregularly (bottom) occurring accents, indicated by filled symbols. The timing of the two recorded sequences were used in the listening test in Paper I. In sequence A1, (S1 *pp*, 160 BPM on the normal surface) cyclic patterns with every fourth interval prolonged can be discerned, most clearly seen in the intervals No. 36–50. Sequence S (S2, *ff*, 200 BPM on the normal surface) is played with irregular accents.

typically amounted to 0.2%/measure, or 0.05%/interval (median), when normalized relative to the initial IOI of each 20-s sequence.

On average, the differences between adjacent IOIs across sequences ranged between 10 and 14 ms. Occasionally these differences were substantial, clearly above the reported JNDs (c.f. Friberg and Sundberg, 1995). In the topmost sequence in Figure 4, Paper I, the difference between successive intervals reached 65 ms, or 12% of the associated tempo.

Listening test.

The appearance of a lengthened fourth interval in cyclic patterns was seldom maintained throughout a whole 20-s recording. The question thus emerged whether these patterns still could convey some information on the grouping of strokes to listeners. To resolve this question a listening test was designed, including modified and synthesized sequences.

Method. Four of the recorded sequences with interleaved accents every fourth stroke were used in the test (sequence A1 is shown in Figure 2.2, top, A2, A3, and A4 in Figure 7, Paper I). A performance of irregular accents (sequence S in

Figure 2.2, bottom) was also included. All recorded sequences were notated in 4/4 measure.

In order to isolate the timing aspects of the recorded sequences, the original strokes were replaced with a standardized stroke. The measured IOIs for the recordings were used as timing templates and a recorded stroke was inserted at each position. In this way differences in striking force between strokes and overall dynamic level were removed, leaving only timing information.

In addition to the stimuli based on the recordings, two fully artificial files were synthesized; one completely isochronous (sequence I), and one with every third interval prolonged by 7% (sequence W). A detailed description of the seven stimuli used in the test can be found in Table 1, Paper I.

Two groups of listeners participated in the listening test: 10 professional percussionists or advanced students ("percussionists"), and 10 listeners without any musical training ("novices"). The listeners were asked to sort the stimuli sequences according to the perceived grouping of strokes; groups of 2, 3, and 4 strokes, or no grouping at all. When sorting the stimuli, listeners also indicated how confident they were in their sorting.

The expected result was that sequences originating from the recorded accented 4/4 rhythm (A1, A2, A3, and A4) would be perceived as groups of four strokes, that the synthesized sequence W would be perceived as groups of three strokes, and that sequences without pronounced cyclic patterns (S and I) would be perceived as no grouping.

Results. The results of the listening test for the two groups of listeners can be seen in Table 2.1. The table displays the relative occurrence of the listeners' placement of the stimuli into the four categories. An occurrence of 100% would mean that all versions of the same stimulus were placed in the same category. In general the listeners sorted the stimuli as expected. Sequence A3, with the most persistently maintained cyclic pattern, was placed in the 4 group with the highest occurrence. However, it does not seem to be necessary for the player to consistently maintain the patterns throughout the 20 seconds in order to convey grouping to listeners. Also sequences A1, A2, and A4 were repeatedly placed in the same group, although the two latter also appeared in other groups – quite in line with the timing patterns. An unexpected result was that the fully artificial sequences I and W were sorted into the "wrong" category to some extent. Even the percussionists repeatedly placed the artificial sequences in categories other than expected.

Figure 2.3 shows the percussionists' (top) and the novices' (bottom) confidence ratings in their placements. The figure displays the mean ratings for the seven stimuli files in each category. The listeners' ratings follow the relative occurrences of placements well, but with somewhat lower values. The percussionists seem more confident in their placements than the novices.

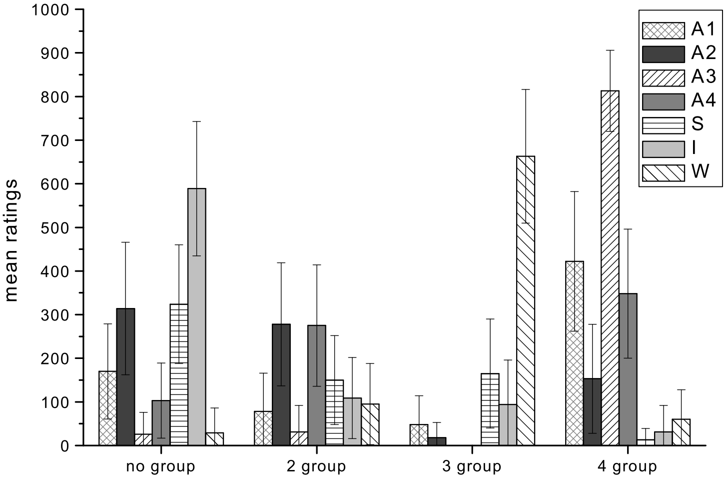
Table 2.1: Subjects' categorizations of the sequences in the listening test in Paper I. The table shows the relative occurrence, in percent, for percussionists (top) and novices (bottom). An occurrence of 100% would mean that all subjects placed all three versions of a sequence in the same category. The categories receiving the highest ranking for each sequence is shown in bold. The expected grouping according to the score (recorded files) and to the specified timing patterns (synthesized files) respectively is indicated in italics.

		PERCUSSIONISTS			
sequence	rhythm characteristics	no group	2 group	3 group	4 group
A1	$\frac{4}{4}$ rhythm	27	10	7	***57***
A2	$\frac{4}{4}$ rhythm	37	**43**	3	*17*
A3	$\frac{4}{4}$ rhythm	3	3	0	***93***
A4	$\frac{4}{4}$ rhythm	17	37	0	***47***
S	irregular	***50***	27	20	3
I	isochronous	***67***	17	13	3
W	$\frac{3}{4}$ rhythm	3	13	***73***	10

		NOVICES			
sequence	rhythm characteristics	no group	2 group	3 group	4 group
A1	$\frac{4}{4}$ rhythm	23	23	17	***37***
A2	$\frac{4}{4}$ rhythm	30	13	10	***47***
A3	$\frac{4}{4}$ rhythm	7	0	17	***77***
A4	$\frac{4}{4}$ rhythm	0	**67**	13	*20*
S	irregular	***53***	10	10	27
I	isochronous	***47***	30	7	17
W	$\frac{3}{4}$ rhythm	13	27	***47***	13

Discussion.

The four players all initiated the accented stroke from a greater height compared to the unaccented strokes. However, they clearly displayed different movement strategies in their preparatory movements. The spread in IOI encompassed both short term deviations in IOI and long-term tempo drift. The magnitudes of the timing variations were somewhat surprising. It was expected that the players would be able to maintain the indicated tempo, despite the short break between the turn-off of the metronome and the beginning of the recording. While the players were not explicitly instructed to maintain tempo, many of the larger tempo variations would be audible. However, integrated in the short-term variations there was also information on the rhythmic grouping identified by listeners.



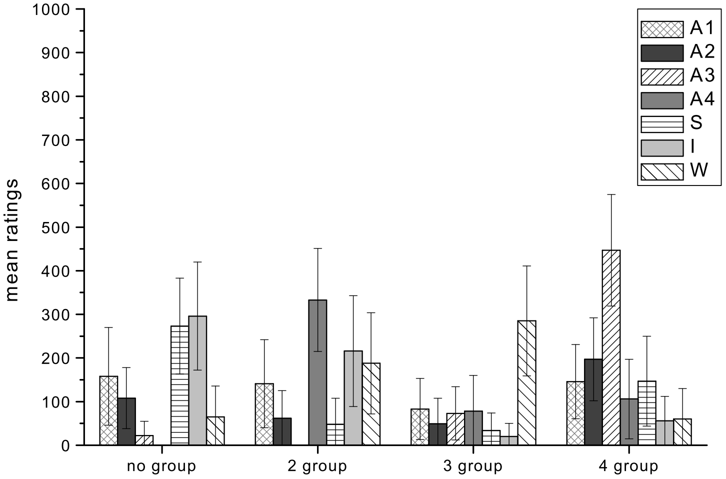


Figure 2.3: Mean ‘confidence’ ratings for the ten percussionists (top) and ten novices (bottom) in the listening test. For each of the seven stimuli (A1-A4, S, I, and W) the mean ratings for the four possible responses (no group, 2, 3, or 4 group) are shown by bars. A value of 1000 would correspond to a case in which all 10 subjects were fully confident on their placement of the sequence in one category. The vertical error bars indicate the 95% confidence intervals.

2.3 Paper II: Comparing striking velocity and timing

In Paper II the low-level aspect of the control of accented strokes was further explored through a more detailed movement analysis of the data presented in Paper I. The questions addressed were whether there would be any differences in movement patterns or striking velocities depending on playing conditions; dynamic level, tempo, or striking surface.

Subjects and recordings were the same as in Paper I. An analysis of performances on three different striking surfaces (soft, normal, and hard, placed on top of the force plate) was also included.

The movement analysis was concentrated on the vertical displacement of the drumstick at the initiation of a stroke, the *preparatory height*, and the vertical velocity before impact, the *striking velocity*. Both these measures were extracted from the vertical displacement of the marker on the drumstick.

The timing analysis was similar to that described in section 2.2, but here results were presented as IOI normalized to the average IOI. The average IOI across a whole sequence is denoted by IOI_{avs} , and normalized IOIs as $\mathrm{IOI}_{rel} = \mathrm{IOI}/\mathrm{IOI}_{avs}$

Special attention was paid to the *metric location* of a stroke, that is, the position of the stroke in the measure. The mean values were therefore calculated and presented with respect to the metric location.

Kinematics

As noted already in Paper I, all subjects raised the stick to a greater height before the accented stroke. In Figure 2.4 the average preparatory height for the four subjects is seen. The figure shows how the players increased the preparatory heights with increasing dynamic level and in preparation for the accented stroke (beat No. 4).

The characteristics of the players' individual movement patterns were reflected in the striking velocities. The observed preparatory heights corresponded well to the striking velocities. The most influential condition on the movement patterns was the dynamic level, resulting in higher striking velocities at higher dynamic levels. When comparing the striking surfaces, the players tended to increase striking velocity when playing on the soft surface, and to decrease striking velocity for the hard surface.

The main difference between the playing styles of the drummers was the emphasis on the accented stroke as compared to the unaccented stroke. This difference was noted both for preparatory heights and striking velocities.

Figure 2.5 shows a comparison between how players S1 and S2 emphasized the accented stroke compared to the unaccented, at different tempi and dynamic levels. The figure shows a plane fitted to the measured striking velocities for all unaccented strokes (stroke No. 2, bottom plane), and the accented strokes (stroke No. 4, top plane) for players S1 and S2 performing on the normal surface. As illustrated by the different inclination of the planes in the figure, tempo and dynamic level

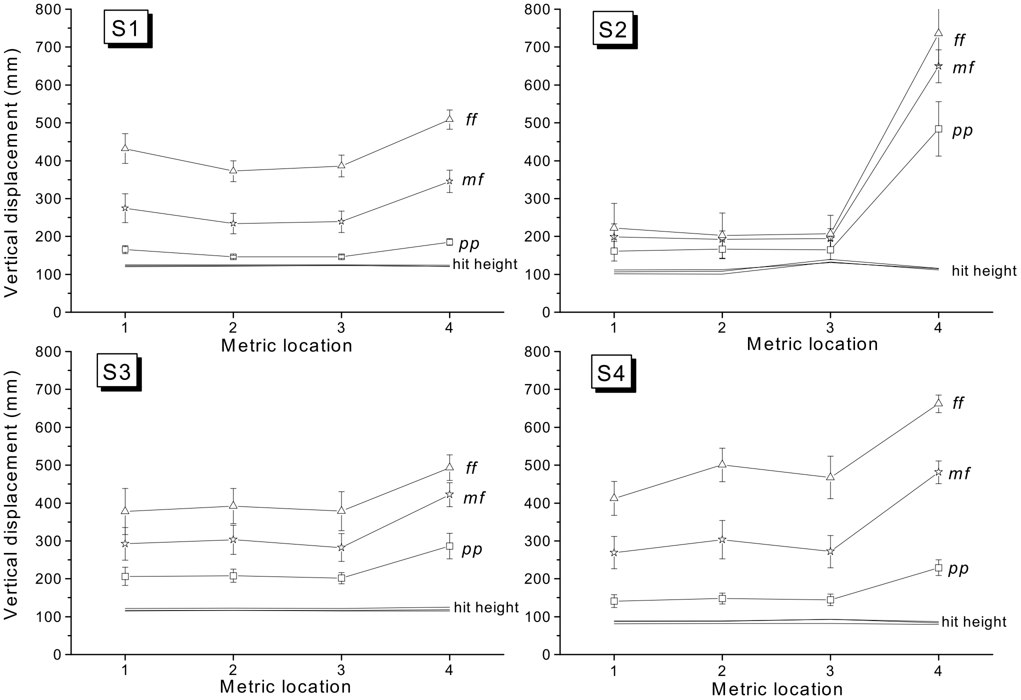


Figure 2.4: Average preparatory heights and hit heights for the four subjects playing on the normal surface. Each panel shows the preparatory height and hit height for the three dynamic levels (*pp* – *mf* – *ff*), averaged across each of the four metric locations. The error bars indicate standard deviations. The subjects increased the preparatory height with increasing dynamic level. The panels reflect the individual strategies of the subjects, shown in the movement trajectories (Figure 2.1).

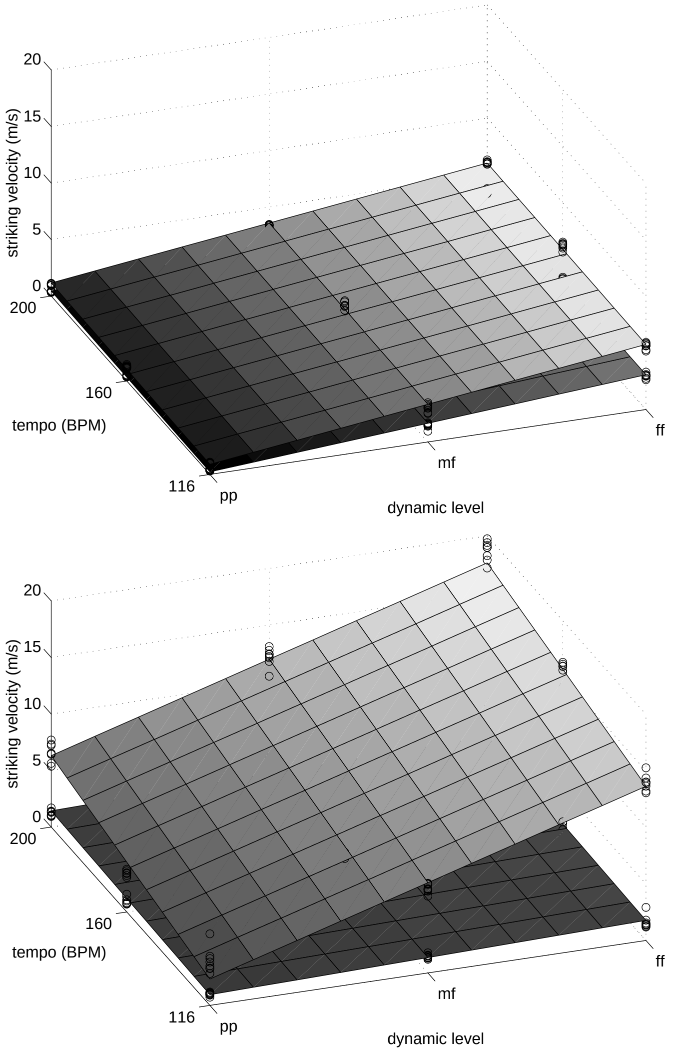


Figure 2.5: Striking velocities for player S1 (top panel) and S2 (bottom panel) playing at three tempi and three dynamic levels on the normal surface. The graphs show a plane fitted to the measured striking velocities for an unaccented stroke (stroke No. 2, bottom plane), and the accented strokes (stroke No. 4, top plane). The eight measured data points for each case are indicated by circles.

have different influences on the players' emphasis on the accented strokes. S1 (top panel) maintains approximately the same difference between the unaccented and accented strokes for all dynamic levels. By contrast, S2 (bottom panel) uses the same striking velocity for unaccented strokes, regardless of dynamic level, while the accented strokes are highly influenced by both tempo and dynamic level.

Two of the players (S2 and S4) also showed a slight decrease in striking velocity for stroke No. 3, the stroke preceding the accent (see Figure 10, Paper II). The explanation can be found in the preparatory movements. Both these players initiated the preparation for the accented stroke by moving the hand and wrist upwards, letting the stick follow. To reach the height from which the accented stroke is initiated in time, the hand starts the upward movement even before the preceding hit (see Figure 5, Paper II).

Timing and accent prolongation.

The standard deviations of IOI for the first 8 measures of all recorded sequences ranged from 2% to about 7% (normalized to the mean IOI across each sequence). The spread in IOI compares rather well to that reported by Madison (2000), considering that the influence of the lengthening due to the accent in the rhythm is included. In general the standard deviations in IOI_{rel} decreased with increasing dynamic level. For some subjects also tempo had an influence on the spread in IOI_{rel}, with higher standard deviations at a higher tempo.

The stroke that was displaced in order to contribute to the lengthening differed between players. On average, subject S1 and S2 shortened the first two intervals and prolonged the last two, while subjects S3 and S4 shortened all the first three intervals and prolonged only the accented interval. The average lengthening across all subjects and playing conditions reached 3%.

Figure 2.6 illustrates the influence of the accent on striking velocity and IOI_{rel}. The figure, in which each data point is averaged across all subjects (288 strokes in all), shows the average IOI_{rel} vs striking velocity for each metric location and dynamic level. As seen in the figure, the lengthening in IOI_{rel} for the interval beginning with the accented stroke decreases with increasing dynamic level while the increase in striking velocity for the accent is about a factor of 2 for all dynamic levels.

Discussion.

The movement patterns of the players were clearly reflected in the striking velocities. It is likely that the players' backgrounds could explain part of the differences in their emphasis on the accented stroke and movement strategies. Player S1 and S3 are mainly active in the symphonic and military orchestral traditions, while S2 and S4 mainly play the drum set in the Afro-American music tradition. In orchestral playing, an accent, although emphasized, should not be *over* emphasized. In contrast, many genres using drum set playing often encourages large dynamic

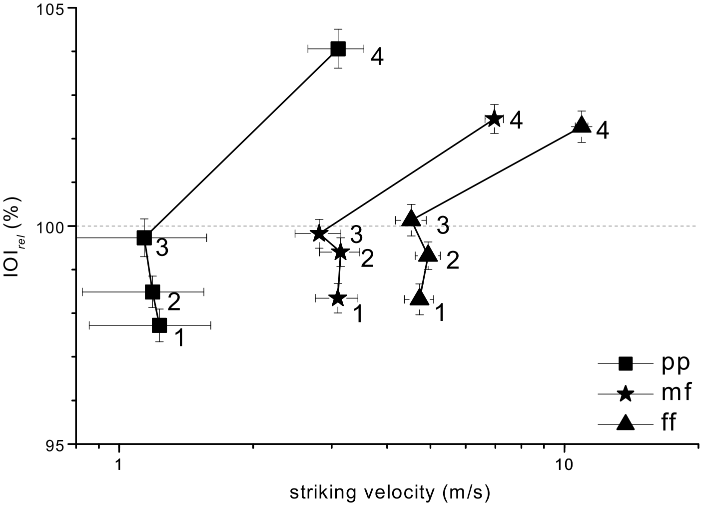


Figure 2.6: The relation between striking velocity and IOI_{rel} for the different metric locations and dynamic levels. The figure shows the grand average for the metric locations (numbered 1, 2, 3, and 4), at *pp* (squares), *mf* (stars) and *ff*(triangles), across all subjects and playing conditions. Each data point represents 288 analyzed strokes. The 95% confidence intervals are indicated by error bars. As seen in the figure, the increase in striking velocity for the accented stroke is by about a factor of 2 for all the dynamic levels. The lengthening in IOI_{rel}, however decreases with increasing dynamic level.

differences between accented and unaccented beats. In fact, unaccented notes are sometimes played as soft as possible; "ghost notes".

The tendency for lengthening every fourth interval, the accented note, was evident for all four players. It is worth noting that it is not the interval involving the preparatory movement before the accented stroke that is prolonged, but the following interval. Although the preparatory movements used for the accented strokes are larger (sometimes much larger, such as when player S2 involved the whole arm), the accented stroke arrives on time. It is the following stroke (the first stroke in the next measure) that is delayed. The lengthening not only emphasizes the accent, but also allows the oscillations of the rebounding stick to decay.

2.4 Paper III: Perception of drift in auditory sequences.

The overall temporal context of the middle layer was addressed in Paper III. As observed in Papers I and II, subjects tended to drift in tempo across sequences, typically 0.05% per interval. This tempo drift was in the same range as drifts reported by Madison (e.g. 2000) but small compared to the threshold estimates for tempo drifts reported (e.g. Madison, 2004; Vos et al., 1997), which indicate values of about 2–3%/int. at nominal IOIs of 500 ms.

However, the fact that several of the listeners participating in the test in Paper I commented on drift in tempo for some of the sequences prompted an investigation of the perceptual thresholds for tempo drift.

Model and method

A novel approach to investigate perception of tempo was proposed. The method was designed to investigate the presence of an internal representation of a "steady tempo", without ruling out the possibility that this representation itself can have an inherent tempo drift. Such a drift would mean that some listeners could perceive an increasing tempo when, in fact, the physical tempo is decreasing, and vice versa.

The model used for perceiving continuous (linear) tempo drift was based on the assumption that there exists range of non-discriminable tempo drifts, centered around an *Internal Drift* (ID). ID is the bias of the listener's perception of a 'steady tempo', relative to isochrony. Surrounding this ID is an interval corresponding to twice the just noticeable difference (JND). A graphic representation of the model used is seen in Figure 2.7.

An adaptive procedure was used for the listening tests. The method was similar to the Parameter Estimation by Sequential Testing (PEST, see Taylor and Creelman, 1967) and the Transformed up-down staircase method (Levitt, 1970), but was modified by our requirements to track *two* borders. In each test session, stimuli consisting of click sequences were presented to the listener. Depending on the listener's response (correct or incorrect) the magnitude of the tempo drift is modified for the next presentation.

In both Experiment I and II, stimuli of 10 clicks (9 intervals) were used. At each of the investigated tempi subjects completed three 40-minute test sessions. After each presentation of a stimulus, subjects indicated whether the tempo had increased or decreased.

Experiment I: Empirical test of method.

In Experiment I an intermediate tempo 120 BPM was studied. The results showed that 10 of the 11 subjects participating in the test tended to perceive the tempo changes "biased" towards either increasing or decreasing tempo. The subjects were rather consistent between test sessions. An example of the responses during three test sessions is shown in Figure 2.8.

To evaluate the method, a comparison between sensitivity to termination criteria of different estimates of ID and JND was made. Two estimates were used in the comparison: 1) ID and JND calculated using the last presented stimulus level (LAST) and 2) ID and JND obtained using a Logistic Regression Model (LRM).

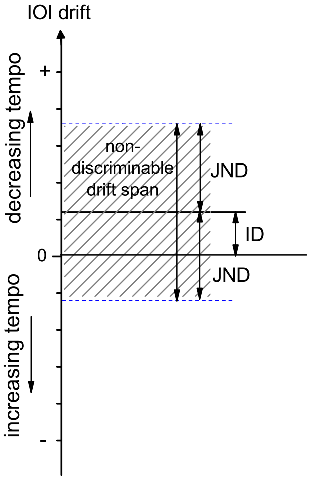


Figure 2.7: Graphic representation of the model for tempo discrimination used in Paper III. The non-discriminable drift span is twice the Just Noticeable Difference (JND) and centered around the Internal Drift (ID). The lower and upper borders of the non-discriminable drift span are tracked by the method. Note that in this model the magnitude of ID is relative to isochrony, while the magnitude of JND is relative to the ID.

The influence of different termination criteria on these estimates was tested by obtaining three data sets for each subject 1) All responses made under a complete 40-minute test session 2) Data truncated after 40 presentations 3) Data truncated after 4 presentations with minimum step size (0.2 ms)

An analysis of variance revealed that the termination criterion was important for estimating JND, an expected result. There were no significant differences between the two types of estimate for either ID or JND. However, the LAST-estimate for JND was somewhat more sensitive to the type of termination criterion than the LRM-estimate. It was therefore decided that ID and JND should be estimated using all the available data points from three test sessions in a logistic regression model for each subject.

Figure 2.9 shows the resulting ID with surrounding JND for all eleven subjects, obtained by a logistic regression analysis (LRM estimate) of data from the three test sessions. As seen in the figure most of the subjects display ID≠0. Five of the subjects show positive IDs (ranging from 0.05 to 0.35%/int) meaning that these subjects will consider a (slightly) decelerating tempo as isochronous. Five of the subjects have negative IDs (-0.05 to -0.35 %/int.) and would thus consider a slightly accelerating tempo to be isochronous. The remaining subject, N5, shows no bias (ID=0), but a fairly large JND. The grand average ID across subjects was close to zero.

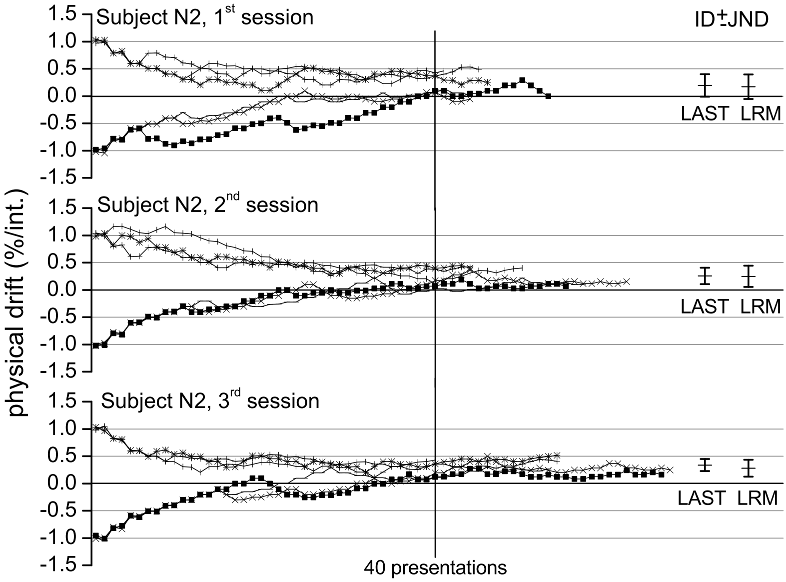


Figure 2.8: Examples of consistency between test sessions and estimates of ID and JND. The three panels show the response curves produced by one subject during three test sessions. Each test session consisted of three tracks with an initial physical drift of -1%/int. ('increasing'), and three tracks with an initial drift of 1%/int. ('decreasing'). To the far right in each panel the resulting IDs surrounded by ±JND are shown, as estimated by the last presented stimulus drift (LAST), and fitting by a logistic regression model (LRM). The vertical line crossing all response curves indicates 40 presentations.

JNDs ranged from 0.18 to 0.60%/interval with a grand average of 0.32%/interval. Lengthening each interval by 0.32%/int. (1.6 ms/int. at 120 BPM) would mean a decrease to 114.5 BPM after four measures in 4/4 beat.

Experiment II: Comparisons between tempi.

The objective for Experiment II was to investigate how nominal tempo affected subjects' ID and JND. Five of the subjects from Experiment I performed additional test sessions at two other tempi; 80 and 180 BPM (nominal IOIs 750 and 333 ms). As in Experiment I ID and JND were obtained by logistic regression and the estimates were compared to the values found for 120 BPM (Experiment I).

An analysis of variance revealed that tempo had an effect on JND, but not ID. The average JNDs in Experiment II were 0.37, 0.25, and 0.30%/int. at 80, 120

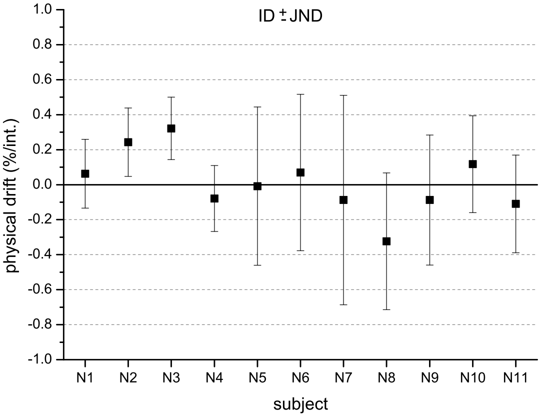


Figure 2.9: Estimated internal drifts, ID (squares) for the eleven subjects in Experiment I surrounded by ±JND. The grand average of JNDs (vertical bars) across all subjects was 0.32%/int., corresponding to 1.6 ms/int.

and 180 BPM respectively. The increasing JND at 80 BPM is partly explained by the fact that the time of each test session was fixed to 40 minutes. The increased presentation times for each stimulus at 80 BPM (7 s as compared to 5 and 3 s) thus resulted in a reduced number of presentations at the slowest tempi.

Discussion

The results showed ID$\neq 0$ for most subjects. In contrast to the study by Vos et al. (1997) there was no effect of nominal tempo on ID. However, the tempi studied (80, 120, and 180 BPM) were not quite as extreme as those used by Vos et al. (60, 120, and 240 BPM).

The found JNDs agree well with values of drift reported for production, but were considerably smaller compared to those reported in previous studies, by almost a factor of ten. Possible explanations for the differences between studies are mainly related to the methods used. To ask a listener to indicate whether or not there is a change, or in which direction the change occurs, are two separate things. In contrast to other studies, the model used here puts JND in relation to the internal drift, ID, and not relative isochrony. However, if the grand average JND in Experiment I (0.32%/int. or 1.6 ms/int.) is added to the largest observed ID for all subjects, it still does not amount to more than 0.64%/int., corresponding to 3.2 ms/int. at 120 BPM.

Whether these results are valid for other stimulus lengths and other tempi would have to be further investigated. Experiments with non-linear types of tempo drift could also provide additional insights into the mechanisms involved in tempo perception.

2.5 Paper IV: The perception of players' movements.

The highest layer, concerned with musical expression, was investigated through two studies on the perception of expressive intentions in performances on the marimba, bassoon and saxophone. The objective was to explore (1) to what extent the movements conveyed the intended emotions, (2) if particular features or certain parts of the player's body were more important than others in the communication, and (3) what movement cues were used by observers to discriminate between expressions.

Experiment I: Ratings of marimba performances

In Experiment I, a professional percussionist was video recorded when performing a short piece of music with the intentions Sadness, Anger, Happiness and Fear. From the video recordings, stimuli clips were generated showing different parts of the player in four *viewing conditions*: *full* (showing the full image), *no-hands* (the player's hands not visible), *torso* (the player's hands and head not visible), and *head* (only the player's head visible). The clips were shown without sound, and edited so that facial expressions would not be visible.

Twenty subjects watched the videos individually and rated the emotional content on a scale from 0 (nothing) to 6 (very much) for the emotions Fear, Anger, Happiness, and Sadness. The subjects were also asked to mark how they perceived the movements. The ratings were done on bipolar scales (from 0 to 6) for the cues:

Amount: none - large
Speed: slow - fast
Fluency: jerky - smooth
Regularity: irregular - regular

The ratings for the emotions showed that the intentions Sadness, Happiness, and Anger were successfully conveyed, while Fear was not (see Figure 2, Paper IV). The intention receiving the highest number of correct identifications was Sadness. There was some confusion between the intentions Anger and Happiness, Anger rated as Happiness and vice versa, suggesting some features in common among these two emotional expressions.

The influence of viewing condition proved to be surprisingly small. However, the head of the player seemed to play a key role in some cases. For the Sad intention, all the conditions where the head was visible were rated high (means from 4.3 to 4.6), while the torso received much lower ratings (mean 3.1). The low torso ratings appeared in both performances of the Sad intention.

The ratings for movement cues showed that for the three recognized emotions Sadness, Anger, and Happiness, observers used movement cues similar to audio cues found for music performances. Anger was characterized by large, fast, uneven, and jerky movements; Happiness with large and somewhat fast movements, and Sadness with small, slow, even, and smooth movements.

Experiment II: Ratings of woodwind performances

To further explore the use of movement cues in the communication of specific emotions, and to investigate issues related to different types of instruments, a second experiment was carried out. In Experiment II observers rated video clips of two woodwind players, performing four musical excerpts with the emotional intentions Happiness, Sadness, Anger, and Fear. No viewing conditions were generated for these clips. Twenty subjects rated the emotional content of the four emotions as well as the movement qualities rated in Experiment I, Amount, Speed, Fluency, and Regularity.

The emotion ratings of the woodwind performances confirmed the results of Experiment I. Happiness, Sadness, and Anger were all well communicated to the observers, while Fear was not. There was less confusion between the intentions Happiness and Anger.

Compared to Experiment I, the movement ratings revealed fewer cues important for each emotion. Sadness was characterized by slow Speed and even Fluency, and Anger was characterized by jerky Fluency. Happiness was mainly characterized by fast Speed for the bassoon player, but large Amount for the saxophonist. The fact that the movement cues for Happiness were not overlapping with those for Anger, explains the lack of confusion between these two intentions compared to Experiment I.

Discussion

The combined results for the emotion ratings for Experiment I and II are shown in Figure 2.10. The bars in the figure show the mean ratings for all the rated emotions and intentions, averaged across all three performers and all subjects and cases. As shown by the figure the three intentions Happiness, Sadness, and Anger all receive a large portion of ratings corresponding to the intention. The ratings for the intention Fear, however, are evenly spread between the four rated emotions.

Also in the figure the percent correct for each intended emotion and player can be seen: bassoon player (circles), saxophonist (triangles), and the marimba player as shown in the full condition (squares). The values give the proportion of ratings for the intended emotion receiving the highest rating. While the values for Happiness, Sadness and Anger range between 56 and 83% correct, the values for Fear are much lower, for the wind players even below chance level (25%), indicating that the communication of Fear failed.

Rated movement cues could be used to characterize the different emotional intentions: Anger was primarily associated with jerky movements; Happy with large and somewhat fast and jerky movements; Sadness with slow and smooth

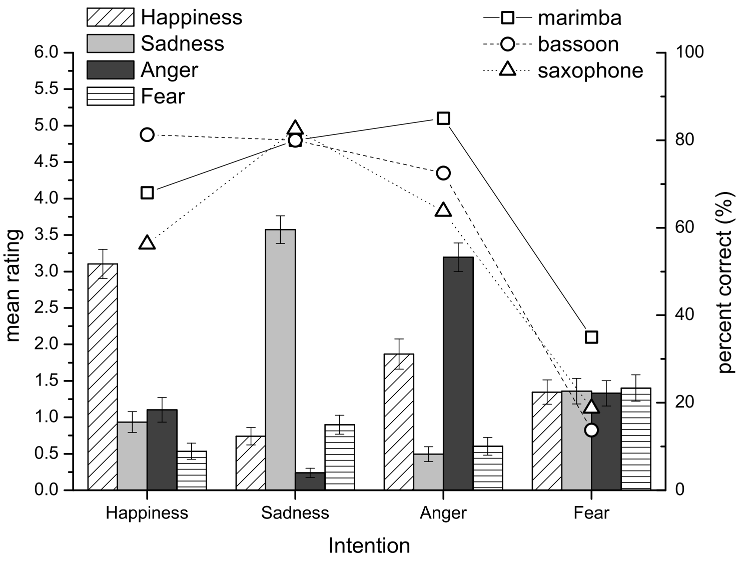


Figure 2.10: Mean ratings for the four rated emotions averaged across all players, subjects and cases in the two experiments in Paper IV. The bars show the mean ratings for each intended emotions, averaged across all subjects and rated video clips. Happiness, Sadness and Anger are all identified, while the ratings for Fear are evenly spread across the four emotions, suggesting that the communication of Fear failed. The lines show the percent correct identified intentions for each of the individual performers, bassoon player (circles), saxophonist (triangles), and the marimba player shown in the full viewing condition (squares). The transformation into percent correct was strict. A 'correct' response was obtained only when the original intention received higher ratings than the other rated emotions. The results of the transformed percent correct and the mean ratings are very similar. Happiness, Sadness and Anger obtain between 56 and 85% correct identifications, while the results for Fear are much lower, between 14 and 35%, meaning that two of the players even fall below chance level (25%).

Table 2.2: Comparison between the most prominent movement cues found in Paper IV, and the corresponding audio cues (in *italics); sound level, tempo, articulation, and tempo variability (selected from Juslin, 2001)*

	amount *sound level*	speed *tempo*	fluency *articulation*	regularity *tempo var.*
Happiness	large	fast	jerky	
	high	*fast*	*staccato*	*small*
Sadness		slow	smooth	
	low	*slow*	*legato*	*final ritard*
Anger			jerky	
	high	*fast*	*staccato*	*small*
Fear	small			irregular
	low	*fast*	*staccato*	*large*

movements; and Fear with small and somewhat irregular movements. However, since the communication of Fear failed, its characterization is questionable.

Table 2.2 displays a comparison between the movement cues from Experiment I and II, and the corresponding audio cues Juslin (selected from 2001).

The influence of viewing condition on the observers' ability to recognize the intentions was surprisingly small. In the cases where there was an effect of viewing condition, the main differences in ratings were related to either the head or the torso conditions. For the Sad intention, the viewing condition 'head' was actually identified as Sad to a higher degree than was the 'full' condition. One explanation for this surprising observation could be that the actual playing movements of the hands and arms had a distracting effect on the movements of the head. Some support for this can be found in an earlier study. Davidson (1994) measured the head and hand movements of a pianist performing the same piece as “deadpan”, “projected”, and “exaggerated”. She found that the extent of head movements differed significantly between performance conditions, while the extent of the hand movements was about the same.

Interestingly none of the players *have* to move in different ways in performances with different intentions – then why do it? A tempting answer is that these movements facilitate the interpretation of different moods in some way. A possible support for this hypothesis is that the movement cues, as rated by the observers in the viewing test, indeed show strong similarities to the acoustic cues for the same emotions in music performances. Another, more straightforward, explanation is that the players tried to enter the mood they intended to convey and moved accordingly. Despite the differences in movement patterns between the players, the similarities dominate.

2.6 Summary

The papers included in this thesis deal with three aspects of music performance: (1) the execution and interpretation of a single musical element (the accent), (2) the perceptual representation of the overall reference by which musicians have to relate to (the tempo), and (3) the interpretation and communication of specific emotional intentions (the expression). While the first aspect deals with structural elements on the micro-level, and the second focus on structure on a higher level, the third incorporates both lower levels for shaping high-level performance characteristics. A common aspect, however, is the need for the player to prepare and execute the correct movement at the right time.

In summary, the investigations showed that percussion players follow well-defined strategies in their performance of interleaved accents (Papers I and II). Listeners are sensitive to gradual changes in tempo and show differing preferences as to what they consider to be isochronous tempo (Paper III). Furthermore, body language alone may serve as an efficient way of communicating different emotional intentions (Paper IV).

Although a simple element in music, the accent receives a lot of attention in percussion training. From the results presented in Papers I and II, it is clear that the subjects have acquired a playing technique that allows them to produce an accent at the right time. Although there are the differences in playing styles, many common features can be observed. First, the accented stroke was always initiated from a greater height than the unaccented, which was directly reflected in the striking velocity for the stroke. Second, the interval beginning with the accented stroke was prolonged, primarily by delaying the following stroke (see Figure 2.6). However, it is also clear that players interpret the accent very differently. A possible explanation for the different performances can be sought in the musical context the players are used to performing in (symphonic orchestras, military bands, and jazz and rock bands).

The results in Paper III show that the opinions about a "steady tempo" differ between listeners. Most of the subjects displayed internal drifts, either in the positive or negative direction. Further, the results suggest that the internal drift does not vary with nominal tempo. The found JNDs were much smaller than in earlier perceptual studies, but agreed well with values for drift reported for production. These findings have two implications of special interest: 1) The inter-individual differences in internal drift is a plausible explanation for the non-consistent results on whether increasing or decreasing tempo is more easily perceived by listeners, as well as the large JNDs reported in other studies. 2) The intra-individual consistencies suggest the interesting implication that two performers with internal drifts of different magnitude and/or sign would perceive difficulties in playing together. This would have to be investigated further.

The aspects on a higher performance level addressed in Paper IV concern how different parts of the body contribute in conveying a specific emotional intention to an observer. The intentions Anger, Happiness, and Sadness were all identified by observers, while the communication of Fear failed. The movement cues used by observers to distinguish performances with different emotional intentions from each other have clear similarities to the audio cues found in music performances (Amount – sound level, Speed – Tempo, Fluency – articulation, Regularity – tempo variation). Interestingly enough there were only small differences between the viewing conditions, both for the ratings of emotions and for the ratings of movement cues. In the cases where one viewing condition received significantly different ratings than the others that viewing condition was either the torso or the head. The results from this study suggest that the head is especially important for conveying the intended emotion.

Chapter 3

Discussion

3.1 General discussion

This thesis has investigated different aspects of musicianship, representing three layers in the perception and interaction involved in music performance. Admittedly, it has not been possible to address more than a small portion of all the strategies and processes involved. Nevertheless, it is hoped the studies have contributed to an increased understanding of the vast and complex area that constitutes musical skill and communication.

The major part of the thesis focuses on percussion playing, an area of music performance that has not been much explored in previous work. Looking at research accounts made, the majority of studies in music performance have concerned piano playing (see e.g. Gabrielsson, 1999), no doubt due to the availability of midi-controlled instruments (e.g. Yamaha's Disklavier). At the same time, a lot of research has used tapping as a means to try to describe the mechanisms involved in human timing production and perception. It could be argued that studying drumming would be the natural step from isolated timing performance in laboratory settings towards time keeping used in more musical contexts. Drummers and percussionists are not isolated from other instrumentalists in terms of musical expertise. However, given that their changes in expression during playing are mainly restricted to timing and dynamic level, they constitute an interesting group. In addition, compared to instruments where individual finger movements are used to larger extent, the playing movements tend to be larger for percussion playing. The larger, more visible, movements are suitable for movement studies.

The results in Papers I and II show that drummers use well-defined movement strategies in preparation for an accented stroke. This was expected, considering that the accented stroke is a much practiced element in drumming. Each player maintained his own strategy for different playing conditions, but between players there were large differences in movement strategies and interpretation of how to emphasize the accent. This is consistent with results reported for other instruments such as piano and violin (e.g. Engel et al., 1997; Baader et al., 2005).

There are many possible reasons for the inter-individual differences in movement strategies. The perhaps most obvious explanation is the different backgrounds of the players, each having their main experience from differing music genres. However, even players that have been taught the same way by their teachers and performed in similar musical contexts would be likely to display differences in movement patterns. Considering that people come in all sizes, differently built with joints of varying mobility, each motor system has to find the optimal solution for the individual system. Furthermore, the instrument itself – in the case of percussionists the striking tool and striking surface – must be integrated if the system is to be optimized.

The results in Paper III showed that listeners have lower thresholds for perception of tempo drift than previously reported. The found results were about one tenth of the estimates reported by Madison (2004) and Vos et al. (1997), but relate well to estimates of drift from production data (e.g. Madison, 2001b). The large discrepancies between the threshold estimates in Paper III and those previously reported could have several possible reasons. To begin with, there are few studies truly investigating continuous tempo drift. Among the handful that we are aware of, there are considerable differences in the methods that have been used (Vos et al., 1997; Madison, 2004; Pouliot and Grondin, 2005). Pouliot and Grondin measured the reaction time for subjects to detect when a tempo change occurred. In their study they compared results for gradual and abrupt tempo changes in both musical excerpts and piano tones. Vos et al. (1997) applied continuous drift to the last two or three intervals in sequences of ten clicks and asked subjects whether the direction of tempo drift was increasing or decreasing. The study by Madison (2004) is the only one dedicated to finding the perceptual thresholds for tempo drift for differing tempi and number of intervals. Madison used an adaptive procedure, asking subjects to indicate whether they detected a change or not. Vos et al. (1997) argued that such a “yes-no” paradigm would be more influenced by subjects’ decision rather than perceptual processes. By forcing subjects to indicate the direction of the drift (“forced directional response” paradigm) the individual criterion for deciding whether the drift is heard or not is avoided. Mixing test tracks for several investigated factors (e.g. number of intervals and nominal tempo as done by Madison, 2004) could also affect subjects’ performances. A switch between a stimulus consisting of nine intervals and a subsequent stimulus only presenting two intervals could result in difficulties in switching “listening mode” for the subject.

The results in Paper III also showed the presence of a bias in listeners’ perception of a steady tempo (‘internal drift’). To our knowledge, evidence of such a bias has not been reported previously. Our study, which used a method specifically designed to allow for the presence of a bias, focused on individual thresholds. It may well be that individual biases in the perception of steady tempo are hidden in the averaged values of previous studies.

Among musicians, a common notion seems to be that decreasing tempo is easier to detect, suggesting that reducing main tempo is less accepted. Kuhn (1974);

Madsen (1979); Sheldon and Gregory (1997); Pouliot and Grondin (2005) reported subjects to be more apt at detecting decreasing tempi than increasing tempi. Other studies have found no differences between drift direction (e.g Madison, 2004; Ellis, 1991). Considering the inconsistencies in these results, an interesting question is whether individual differences in internal drift could have affected the results in these investigations.

It is also possible that the negative asynchrony observed in tapping could play a role in the perception of tempo drift. An effect of the tendency to place taps in advance of a stimulus is that subjects usually increase in tempo if the auditory feedback of their own taps replaces the original pacing signal (e.g. Aschersleben, 2002). By tapping at a faster rate, the subjects try to maintain their lead (which is perceived as synchrony) before the auditory signal. Furthermore, in another study Aschersleben showed that only with feedback on the magnitude and direction of their systematic error in tapping are subjects able to learn to tap in exact synchrony. However, the subjects then report that they achieve this by deliberately tapping “too late” (Aschersleben, 2002). It is possible that these effects of negative asynchrony could have an effect if listeners use tapping as a means to detect the direction of tempo drift.

The fact that even musically experienced subjects consistently displayed bias in their perception of tempo drift is somewhat intriguing. One cannot help but wonder how such differences in the opinion of a “steady tempo” would affect performances. The results suggest that two performers with largely differing bias would find it difficult to play together. It seems reasonably to assume that a situation where one player perceives the other to increase tempo and vice versa would be far from ideal.

Like Paper I and II, Paper IV also displayed inter-individual differences in overall movement patterns. The results in Paper IV suggested movement cues important for the communication of Happiness, Sadness, Anger, and (with some reservations due to the low recognition rate) Fear. Most of the cues were found for all three players, suggesting the use of a common body language. The cues also showed some similarities to movement cues found in studies of other types of activities (e.g. Boone and Cunningham, 2001; De Meijer, 1989).

In some cases, however, it was evident that the players moved differently. In particular, the cues Fluency and Regularity received different ratings between players in several cases. For instance, Sad performances were characterized by regular movements for the marimba player, irregular movements for the saxophonist, and neither regular or irregular for the bassoon player (c.f. Tables 6 and 9, Paper IV). The characterization of Happiness offers another example. The marimba player and the saxophonist both used fast, large and even movements for their Happy performances, while the bassoon player used intermediate amount, fast and somewhat jerky movements for the same intention The bassoon player was the most successful in the overall communication of Happiness, which explains why the jerky cue comes through in the regression of all the available ratings (Table 11, Paper IV).

Another origin of differences in the ratings of the players’ performances seems to be the choice of overall tempo. The duration of the video clips showing the saxophonist were shorter compared to those for the bassoon player, indicating that the average tempo was slower for this performer. Considering that mean tempo is one of the most important cues used for identifying emotional expression (see e.g. Juslin, 1997; Juslin and Madison, 1999; Laukka and Gabrielsson, 2000) this could have contributed to the lower ratings for the intentions Anger and Happiness for the saxophonist compared to the bassoon player. Other studies of visual recognition of affects through movements have also recognized the importance of movement Speed and changes in tempo (see e.g. Boone and Cunningham, 1998, 2001; Paterson et al., 2001).

The results in Paper IV show that musicians' movements are able to convey specific emotional intentions, Sadness being the most successfully conveyed intent. The fact that Sadness was so easily identified corresponds well with the early development of the ability to recognize and portray this emotion. Boone and Cunningham showed that children as young as 4-years old are able to identify Sadness (Boone and Cunningham, 1998) and also produce the relevant movement cues to have adult performers identify Sad performances (Boone and Cunningham, 2001). The ability to identify and portray Anger, Happiness, and Fear appears from the age of five.

The communication of the Fearful intention failed for all three performers, with a possible exception for the marimba performances in the full viewing condition. The unsuccessful communication for Fear was unexpected, especially since Fear has been identified in sounding music performances (e.g. Gabrielsson and Juslin, 1996; Juslin, 2000) and in visual displays of other activities (Walk and Homan, 1984; Pollick et al., 2001; Camurri et al., 2003). Specifically for visual communication of emotions in music performance, Sörgjerd (2000) found no differences in how accurately the intentions Anger, Sadness, Happiness and Fear were conveyed. However, it is not clear whether facial expressions were visible to the observers in her study.

One straightforward explanation for the failed communication of Fear is that all three players were generally unable to convey the Fearful intention. In that case a listening test with the corresponding audio recordings of the performances would yield similar results to the visual ratings. However, many of the comments by subjects participating in the test suggest that they were simply unable to associate a Fearful expression with a music performance. A pilot test, using many more rated emotions, indicated that the emotional intentions Disgust and Jealousy received low ratings due to a low expectancy level of the observers. Similarly, recognition of disgust has been poor in several studies using visual displays (e.g. Dittrich et al., 1996). Clarke et al. (2005) suggested that the face provides a better channel for recognition of some emotions while the body is better for others.

3.2 Methodological considerations

Papers I and II

Obviously, the experimental settings for the recordings in Papers I and II did not resemble a normal playing situation for the drummers. There was nothing musical about sitting in a motion analysis laboratory with markers attached, playing single strokes on a force plate. It could be that some of the timing variation was due to a context which perhaps, by the performer, was not considered as “proper playing”. However, much of deliberate instrumental practice is quite distant from musical situations. The practice of playing technique for a drummer usually involves extensive training on practice pads. Such pad are usually small rubber plates, resembling the one used for the ‘normal’ striking surface. A practice pad gives a good response in playing and is easy to transport. The main reason for using them, however, is the considerably lower sound level produced compared to a real drum (meaning that practice is possible without disturbing others). For this reason it can be assumed that the absence of a real instrument would be less of a problem for drummers compared to many other instrumentalists.

Another issue concerning spread in data is the order in which the playing conditions were altered. It is possible that the larger timing variation seen at *pp*-level is due to the fact that this dynamic level was recorded first at each tempo, meaning that the recordings at this dynamic level also coincided with a change in nominal tempo and sometimes also striking surface. One way to control for effects of initiating new playing conditions would be to randomize the order in which new conditions are introduced to the players. Such *counter balancing* between subjects would, however, make separation of individual differences and variations due to the experimental procedure more difficult. Clearly, the optimal solution for this problem would be to include more subjects and repetitions of each playing condition, something that hopefully will be done in future studies.

There were only two cameras available to track the movements of the drummers. When setting up the experiment, this was not considered a problem since the movement would be mostly restricted to one plane, perpendicular to the striking surface. However, for some of the different movement patterns used by the drummers in the experiment this was not always the case. Sometimes the arm was rotated during the larger, preparatory, movements causing one or several LEDs on the markers to be occluded. Such data gaps were most frequent during the initial part of the upward transition in preparation for the accented strokes. As long as the points of directional change in the vertical displacement component were unaffected the gaps did not affect data analysis, but in some cases interpolation was necessary.

Paper III

The key feature in the adaptive procedure used in Paper III is that the physical drift of test tracks is allowed to change sign regardless of their designated direction. This is our solution to track the two borders on each side of the non-discriminable drift span, without forcing a reference to isochrony. As several test tracks run in parallel during a test session, this feature could lead to situations where a particular answer to a stimulus (e.g. ‘decreasing’) could be both correct or incorrect, depending on which test track the stimulus belonged to. At first glance this may seem peculiar. The tracking of each threshold, however, is done individually for each test track and the multiple test tracks in a test session can be seen as separate tests. Running them in parallel is not much different from mixing other types of test tracks (such as direction of drift, number of intervals etc used by Madison, 2004). Ultimately, by estimating the perceptual thresholds using logistic regression, both track designations and information on correct/incorrect answers were discarded. Only the stimulus levels with the associated responses (increasing/decreasing) were used for the results. Analysis showed that the logistic regression also reduced side-effects of interruption criteria.

It appears that the method used in Paper III is capable of estimating internal drifts (IDs). The ID for each subject was very consistent between different test sessions, even with intermissions of two years (see Table III, Paper III). However, in its present form the method is not fully able to exclude other possible explanations for the found bias. Firstly, a linear change in tempo is not common in music performance and may violate general expectations for tempo changes. Listeners may well expect a tempo change to conform to a non-linear function, such as those found to be preferred for modeling phrasing (Juslin et al., 2002) or final ritardandi (Friberg and Sundberg, 1999). Should this be the case, the responses of the listener are likely to be affected by the discrepancies between the expected, non-linear function, and the linear stimulus. The fact that some of the subjects reported a perceived shift in the direction of drift in the middle of stimulus presentations supports this explanation.

Secondly, our model assumes a random response behavior in the non-discriminable drift span (see Figure 2.7), meaning that within this range subjects are as likely to choose the ‘increasing’ as the ‘decreasing’ button. In reality, however, subjects may well show a predisposition towards one of these answers, even though he/she truly does not perceive any tempo change. Such behavior is very difficult to separate from a “genuine” perceptual bias. Furthermore, a subject may also become aware of his/her own bias and try to change it. Most likely, this is what happened in Experiment II, when subject N8 changed response behavior between test sessions. A possible solution to the dilemma with non-random response behavior for non-discriminable drifts could be to complement the response for each stimulus with a confidence rating, indicating how sure of the choice of drift direction the subject is.

Paper IV

How to best measure expressive communication quantitatively is a recurrent problem in research. One of the strengths of musical communication is the possibility to directly convey emotion without the involvement of spoken language, which by default makes the quantification of the processes somewhat difficult.

Nevertheless, certain measures have been used more than others in studies of musical expression. A frequent approach is to ask subjects to, among a number of choices available, select the one they feel most appropriate (“forced choice,” e.g. Walk and Homan, 1984; Pollick et al., 2001). Another alternative is to have a scale for each emotion which the subject uses for indicating his/her response ("multiple choice schemata," e.g. De Meijer, 1991; Juslin et al., 2002; Resnicow et al., 2004). The motivation for using scales is that the responses more closely reflect how the subject perceives the stimulus. However, the results can also be harder to quantify since there is no "correct" answer and several scales can be rated equally high. A rating of zero on all scales is also accepted as an answer.

When using ratings on scales in factorial designs, statistical analysis is commonly performed for each rated adjective individually (e.g. Juslin et al., 2002). The achievement measure proposed in Paper IV offers an alternative to this approach, combining the values of several rated scales into one value for each individual presentation. By comparing the ratings to the original intention, a measure of how well the communication succeeded is obtained. Similar approaches for describing how successfully an expressive intention is conveyed have been used by Juslin (2000) and Resnicow et al. (2004). Compared to these studies, our achievement measure displays a main advantage of increased sensitivity to different magnitudes of the ratings, making it more suitable for estimating the achievement for individual cases.

3.3 Directions for future work

By studying behavior at different levels of the perception and production in music performance, this thesis has answered some questions and spurred more. Future work could take several different directions.

Acoustic characteristics of the contact between stick and drumhead. In order to investigate the interaction player – instrument in more detail, more studies of the acoustical and mechanical aspects would have to be included. For drumming the mechanisms controlling the impact and rebound would be of particular interest. Such a study could involve recordings of the history of the contact force and contact time between drumstick and drumhead, detailed relative movements of the stick and drumhead, and possibly also contact forces and resulting effective masses involved in the grip.

Perception and production of tempo drift. The result that most subjects participating in the listening test in Paper III consistently showed an internal drift suggest that the natural next step would be to investigate whether a perceptual bias for an individual is reflected in his/her production. If the ID is a "true" perceptual bias, an individual with a positive ID would tend to decrease tempo when intending to play or tap isochronous intervals.

As listeners most probably expect non-linear changes, further research on perception of other than linear tempo drifts would be needed. Although the approximation of drifts in tempo as linear seems adequate for describing short sequences of production data, it should be validated in perceptual tests.

The role of expressive movements in music performance. Considering the results from Experiments I and II in Paper IV, an interesting question is if the performance on some instruments somehow benefits from overall body movements. If so, there could be a link between expression in performers' movements and the control of the instrument.

Players of percussion and piano have limited influence of the tone once it is initiated. The time of direct control over the sound is much shorter than for, for instance, wind instruments or strings. Particularly, the onset times for notes differ substantially between these instruments. The well-defined onsets for piano and percussion result in high demands on the precision of the playing gestures. The player has essentially no opportunities to adjust anything once the note is initiated, except for the decay. Perhaps the playing of these short-contact, "ballistic" instruments is easier to learn and control by using large and "gestural" playing movements?

The number of musicians in Paper IV was too small for any conclusions on the influence of different instruments to be drawn. However, it is worth noting that two saxophonists were excluded from Experiment II in Paper IV because they simply did not move.

If the playing of certain instrument types is facilitated by larger playing gestures, visual recognition of different expressive intentions from movements would be higher for these instruments compared to e.g. wind instruments. In general, there would also be a high correlation between emotion ratings in visual tests and listening tests.

3.4 Conclusions

- Players of percussion use well-defined movement strategies to prepare for accented strokes. The movement strategies are maintained for changes in playing conditions, but strategies differ between performers.
- Drummers' individual interpretations of how to emphasize an accent in terms of dynamic level is reflected in the magnitude of preparatory movements, striking velocity, and timing.
- Listeners are able to retrieve information on the grouping of strokes even from brief cyclic patterns in timing.
- Listeners are able to detect considerably smaller tempo drifts than previously reported.
- Listeners tend to display a bias in their perception of steady tempo which is consistent between test sessions.
- The perceptual bias in steady tempo does not appear to be affected by nominal tempo.
- Musicians seem to share a common body language which enables them to convey specific emotional intentions.
- The visual communication of emotions to the observer can be described in terms of movement cues. These cues resemble the auditory cues reported for music performance.

3.5 Acknowledgements

This work has been a part of the project Music and Motion, supported by the Bank of Sweden Tercentenary Foundation. The work has also been supported by the European Union through the following projects: MEGA - Multisensory Expressive Gesture Applications[1]; MOSART IHP Network -Music Orchestration Systems in Algorithmic Research and Technology[2]; and SOb - The Sounding Object[3].

For all the help and support during the various stages in my work I thank the two Anders who have taught me to read and write, i.e. my supervisors. *Anders Askenfelt* is a committed supervisor and also one of the best teachers I have known. He has taught me to value the handicraft of writing. If something less than good is found in this thesis it is more likely due to my stubbornness than it was missed by his scrutinizing eye. In his characteristically humorous way, Anders has always taken the part of main supervisor seriously. Apart from that, all who know Anders know to expect the unexpected. *Anders Friberg's* thorough knowledge of the field of music performance has guided me in new directions in my own research. He has taught me the value of a thorough literature search (exclaiming Google to be his favorite website). Anders has the enviable quality of conducting exciting, new research with a laid-back attitude that I have found to take the edge off even the most stressed periods.

I am grateful that my first experiments were conducted with the help and patience of *Markku Haapakorpi* and *Virgil P. Stokes*, who also helped with the initial analysis of the movement trajectories. Other persons I wish to thank are: *Peta Sjölander* and *Becky Hincks* for their patient corrections of my English; *Raoul Huys* for guidance in the field of movement science, helpful comments, and for his hospitality; *Bruno Lucio Giordano*, for sharing comments and help on psychoacoustic test methods, and for being such an agreeable guy with a wicked laughter; *Giampiero Salvi*, for help with all sorts of script programming, critical thinking, sound advice, stimulating discussions and of course all the memorable times together.

For the inspiring collaboration and exchange of ideas I wish to thank Antonio Camurri and the INFOMUS group; Davide Rocchesso and his group at UNIVR, as well as *Nicola Benardini*, *Guy Madison*, and *Frank Pollick*.

I am indebted to all the drummers and percussionists who have contributed to this thesis: *Anders Loguin*, *Anders Holdar* and the others in *Kroumata*; *Alison Eddington*; *Petur Östlund*; *Joakim Anterot*; and *Fredrik Björlin*. Special thanks are due to all the subjects participating in the experiments without whom this work would not exist. I also owe much to my former drumming teachers *Per Hultén*; *Tommy Törner*, and *Hans Bruniusson*.

The analysis and presentation of data have greatly benefited from the existence

---

[1] IST-1999-20410 www.megaproject.org

[2] HPRN-CT-2000-00115

[3] IST-2000-25287 www.soundobject.org

of several free softwares, such as R[4], Virtual Dub[5], LaTeX 2ε[6], BibTeX and others[7]. I extend my gratitude towards the people and communities that have developed these softwares and formats, and made them available. Specifically I want to thank the enthusiasts who have assisted me in finding what I need (incidentally all Italians): *Giampiero Salvi*, *Carlo Drioli* and *Roberto Bresin*.

Thanks to Carl Haakon Waadeland for providing the data used for the figure in Section 1.4.

Magnus Wedberg and *Linus Hultstrand* did an amazingly quick arrangement and produced the pretty picture on the cover. Tack!

I wish to thank everyone at the Department of Speech, Music and Hearing for the genuinely friendly and open atmosphere which continues to be both stimulating and relaxed. Especially I want to thank all members, past and present, of the Music Acoustics group for creating such a wonderful (and just a tiny bit crazy) working environment. I am really grateful to have had the opportunity to get to know you all: *Johan* for creating the group and inviting me to join it, for sharing his spirit of wit and humor – and for always having the right words when words have failed me; *Peta* for being a good friend and mentor, sharing wine and whine, booking me massage when my muscles scream too loud, and for the painstakingly many hours invested in proofreading my manuscripts; *Svante* for inspiring collaboration, relaxed talks about life and research, and for always seeming to have a software ready and waiting to suit ones needs; *Monica* for sharing the laughs, doctorate hardships, and for her lovely voice sounding in the department stairwell; *Eva* for the positive spirit, the friendly words and hugs when needed, and for her daring encounter with the Hultsfred mud; *Kjetil* for cheerful humor, enthusiasm, the shared interest in weird music and for helping me with practicalities when in dire need; *Micke* for proving that with some determination and self-irony it's possible to take on the system, and debug some Matlab in passing; *Erwin* for all the help with debugging Matlab code, and for being so thorough; *Roberto* for his cheerful laugh and positive attitude, and his never-failing commitment to find a good place to eat; *Sten* for some rewarding brain-storming and for continuing the work and spirit of the group; *Anick* for the sunny smile, and for kindly making it clear that I'm not the only one with holes of my socks; *Erik* for the morning 'Godmorgon!' ritual and his ability to bring an element of play into serious research; *Eric* for his dry humor and for sticking to his guns in a gentlemanly way ...and *Leonardo* and *Heloisa* for tender memories of three years with a touch of Brazilian atmosphere in the group.

Since the start of this work my life has taken many turns, not all of them easy. I owe much to the following persons for sharing the fun and helping me through the hard parts. Although not directly involved in the writing process, this thesis

---

[4] www.R-project.org

[5] www.virtualdub.org

[6] www.latex-project.org

[7] sourceforge.net

simply would not have been written without their, direct or indirect, support. I count myself very lucky to have people like them holding me together, even upright when need be. But above all I wish to thank them for being who they are.

Tack för att ni finns!

Mirre for more than twenty-five years of close friendship and, although no longer living in the neighborhood, never further than a phone call away (even in the middle of the night); *Anki* for support and pep-talks, all the video dinners in her sofa, and for taking me out for brain-charging walks; *Zarah* for her almost uncanny sharp-sightedness and intelligent analysis of life; *Rasmus* for taking me out to see bands, skiing, and sailing in Ebba, and for telling me I shouldn't work too much when he tends to work too much himself; *Hedvig* for sharing the frog as well as laughs, drinks, and discussions on research; *Lena* for her frank, refreshing attitude, and for not succeeding in her original I-will-make-you-a-drunkard intention; *Elin* for her wit and refreshing remarks on life, the universe, and everything; *Alan*, because even the worst days seem to brighten up when I find a letter from him on my doormat; *Ingrid*, la mia principessa Napoletana, for all our food-excesses; The *Naglitsch* families for sharing all the passion and frustration involved in trying to make playing in bands work with life in practice; *Maggie* for all the emergency turn outs and good advice;

The guys in Aoria: *Erik*, *Erik*, and *Andreas*, for bringing back rock 'n roll in my life when I needed it dearly;

Moster Marie and the *Eriksson-Stenvall* family for their relaxed and friendly attitude and because Örvikens camping is the place to be; *Pigge* for being the bravest, kindest, best and most admirable 'storasyster kusin' one could possible wish for; the rest of the *Jagrén-Ohlsson* family for all support and for taking me out to Vettersö to pick blueberries (and ticks) from time to time.

And last, or really, first: *Mamma* och *Pappa*, för att det inte är så många som har sån tur att de har en egen fanclub i sina föräldrar.

☥

Bibliography

Abernethy, B., Kippers, V., MacKinnon, L. T., Neal, R. J., and Hanrahan, S. (1997). *The biophysical foundations of human movement.* Champaign, Ill.:Human Kinetics.

Altenmüller, E. (2003). Focal dystonia: advances in brain imaging and understanding of fine motor control in musicians. *Hand Clinics*, 19:523–538.

Altenmüller, E. and Gruhn, W. (2002). Brain mechanisms. In Parncutt, R. and McPherson, G. E., editors, *The science and psychology of music performance. Creative strategies for teaching and learning.*, pages 63–81. Oxford University Press.

Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, 48:66–79.

Aschersleben, G., Gehrke, J., and Prinz, W. (2001). Tapping with peripheral nerve block. a role for tactile feedback in the timing of movements. *Experimental Brain Research*, 136:331–339.

Baader, A. P., Kazennikov, O., and Wiesendanger, M. (2005). Coordination of bowing and fingering in violin playing. *Cognitive Brain Research*, 23:436–443.

Behrens, G. A. and Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music*, 21:20–33.

Bengtsson, I., Gabrielsson, A., and Thorsén, S. M. (1969). Empirisk rytmforskning (Empirical rhythm research). *Swedish Journal of Musicology*, 51:49–118.

Bernstein, N. (1967). *The coordination and regulation of movements.* Pergamon, London.

Billon, M., Semjen, A., and Stelmach, G. E. (1996). The timing effects of accent production in finger-tapping sequences. *Journal of Motor Behaviour*, 28:198–210.

Boone, R. T. and Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology*, 34:1007–1016.

Boone, R. T. and Cunningham, J. G. (1999). The attribution of emotion to expressive body movements: A structural cue analysis. Manuscript submitted for publication.

Boone, R. T. and Cunningham, J. G. (2001). Children's expression of emotional meaning in music through expressive body movement. *Journal of Nonverbal Behavior*, 25(1):21–41.

Boulanger, R. and Mathews, M. (1997). The 1997 Mathews radio-baton and improvisation modes. In *Proceedings of the International Computer Music Conference 1997*, pages 395–398.

Brandfonbrener, A. G. and Kjelland, J. M. (2002). Music medicine. In Parncutt, R. and McPherson, G. E., editors, *The science and psychology of music performance. Creative strategies for teaching and learning.*, pages 83–96. Oxford University Press.

Bresin, R. and Battel, U. (2000). Articulation strategies in expressive piano performance. *Journal of New Music Research*, 29(3):211–224.

Bresin, R. and Dahl, S. (2003). Experiments on gestures: walking, running, and hitting. In Rocchesso, D. and Fontana, F., editors, *The sounding object.*, pages 111 –136. Mondo Estremo, Florence, Italy.

Camurri, A., Lagerlöf, I., and Volpe, G. (2003). Recognizing emotion from dance movements: Comparison of spectator recognition and automated techniques. *International Journal of Human-Computer Studies*, 59(1-2):213–225.

Chen, Y., Repp, B. H., and Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, 21:515–532.

Chong, J., Lynden, M., Harvey, D., and Peebles, M. (1989). Occupational health problems of musicians. *Canadian Family Physician*, 35:2341–2348.

Clarke, E. (1999). Rhythm and timing in music. In Deutsch, D., editor, *The Psychology of music*, pages 473–500. Academic Press.

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

Clarke, E. F. (1985). Some aspects of rhythm and expression in performances of Erik Satie's "Gnossienne No. 5". *Music Perception*, 2:299–328.

Clarke, E. F. (1988). Generative principles in music performance. In Sloboda, J. A., editor, *Generative processes in music*, pages 1–26. Oxford: Clardendon Press.

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., and Rose, D. (2005). The perception of emotion from body movement in point light displays of interpersonal dialogue. *Perception*, in press.

Collyer, C. E., Broadbent, H. A., and Church, R. M. (1994). Preferred rates of repetitive tapping and categorical time production. *Perception & Psychophysics*, 55(4):443–453.

Cook, G. D. (1988). *Teaching Percussion.* Schirmer Books, New York.

Cooper, G. and Meyer, L. B. (1963). *The Rhythmic Structure of Music.* University of Chicago Press.

Cutting, J. E. and Kozlowski, L. T. (1977). Recognize friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9:353–356.

Dahl, S. (1997). Spectral changes in the tom-tom related to striking force. Speech Music and Hearing Quarterly Progress and Status Report 1, KTH, Dept. of Speech, Music and Hearing, Royal Institute of Technology, Stockholm.

Dahl, S. and Bresin, R. (2001). Is the player more influenced by the auditory than the tactile feedback from the instrument? In Fernström, M., Brazil, E., and Marshall, M., editors, *Proceedings of the Cost-G6 Conference Digital Audio Effects, (DAFx01), 6-8 December 2001*, pages 194–19, Limerick, Irland.

Davidson, J. W. (1993). Visual perception and performance manner in the movements of solo musicians. *Psychology of Music*, 21:103–113.

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musician performers? *Journal of Human Movement Studies*, 6:279–301.

Davidson, J. W. and Correia, J. S. (2002). Body movement. In Parncutt, R. and McPherson, G. E., editors, *The science and psychology of music performance. Creative strategies for teaching and learning.*, pages 237–250. Oxford University Press.

Dawson, W. J. (2001). Upper-extremity overuse in instrumentalists. *Medical Problems of Performing Artists*, 16:66–71.

Dawson, W. J. (2002). Upper-extremity problems caused by playing specific instruments. *Medical Problems of Performing Artists*, 17:135–140.

De Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior*, 13:247–268.

De Meijer, M. (1991). The attritution of aggression and grief to body movements: The effects of sex-stereotypes. *European Journal of Social Psychology*, 21:249–259.

De Poli, G. (2004). Methodologies for expressiveness modelling of and for music performance. *Journal of New Music Research*, 33(3):189–202.

Desain, P. and Windsor, L., editors (2000). *Rhythm perception and production.* Lisse: Swets and Zeitlinger.

Deutsch, D., editor (1999). *The Psychology of Music.* Academic Press, 2nd edition.

Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, 6:727–738.

Drake, C. and Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, 54(3):277–286.

Drake, C. and Palmer, C. (1993). Accent structures in music performance. *Music Perception*, 10:343–378.

Drake, C., Penel, A., and Bigand, E. (2000). Why musicians tap slower than nonmusicians. In Desain, P. and Windsor, L., editors, *Rhythm perception and production.*, pages 245–248. Lisse: Swets and Zeitlinger.

Ellis, M. C. (1991). Thresholds for detecting tempo change. *Psychology of Music*, 19:164–169.

Engel, K. C., Flanders, M., and F., S. J. (1997). Anticipatory and sequential motor comtrol in piano playing. *Experimental Brain Research*, 113:189–199.

Ericsson, K. A., Krampe, R. T., and Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100:363–406.

Famularo, D. (1999). *It's Your Move. Motions and Emotions.* Warner Bros. Publications.

Finney, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Perception*, 15(2):153–174.

Finney, S. A. and Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & Cognition*, 31(1):51–64.

Finney, S. A. and Warren, W. H. (2002). Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift. *Perception & Psychophysics*, 64(6):896–908.

Fishbein, M., Middlestadt, S., Ottati, V., Straus, S., and Ellis, A. (1988). Medical problems among ICSOM musicians: Overview of a national survey. *Medical Problems of Performing Artists*, 3(1):1–8.

Fjellman-Wiklund, A., Grip, H., Andersson, H., Karlsson, J. S., and Sundelin, G. (2004a). EMG trapezius muscle activity pattern in string players: Part II – Influences of basic body awareness therapy on the violin playing technique. *Industrial Ergonomics*, 33:357–367.

Fjellman-Wiklund, A., Grip, H., Karlsson, J. S., and Sundelin, G. (2004b). EMG trapezius muscle activity pattern in string players: Part I – Is there variability in the playing technique? *Industrial Ergonomics*, 33:347–356.

Fraisse, P. (1982). Rhythm and tempo. In Deutsch, D., editor, *The Psychology of Music*, pages 149–180. New York: Academic Press, 1st edition.

Friberg, A. (1995). *A Quantitative Rule System for Musical Performance*. Published doctoral dissertation, Royal Institute of Technology (KTH), Stockholm, Sweden.

Friberg, A. and Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustic Society of America*, 98(5):2524–2531.

Friberg, A. and Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustic Society of America*, 105(3):1469–1484.

Friberg, A. and 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. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology*, 15:63–72.

Gabrielsson, A. (1999). The performance of music. In Deutsch, D., editor, *The Psychology of music*, pages 501–602. Academic Press.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music*, 31(3):221–272.

Gabrielsson, A. and Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, 24:68–91.

Gates, A. and Bradshaw, J. L. (1974). Effect of auditory feedback on a music performance task. *Perception & Psychophysics*, 16:105–109.

Guettler, K. (1992). Electromyography and Muscle Activities in Double Bass Playing. *Music Perception*, 9(3):303–310.

Gurevich, M., Chafe, C., and Leslie, Grace an Tyan, S. (2004). Simulation of networked ensamble performance with varying time delays: Characterization of ensamble accuracy. In *Proc. of the International Computer Music Conference (ICMC 2004)*, pages 29–32, Miami, United States of America.

Halsband, U., Binkofski, F., and Camp, M. (1994). The role of the perception of rhythmic grouping in musical performance: Evidence from motor-skill development in piano playing. *Music Perception*, 11:265–288.

Huys, R., Daffertshofer, A., and Beek, P. J. (2004). The evolution of coordination during skill acquisition: The dynamical systems approach. In William, A. M. and Hodges, N. J., editors, *Skill acquisition in sport. Research, theory and practice.*, pages 351–373. Routledge.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14:201–211.

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae*, 1:225–256.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6):1797–1813.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and theoretical framework. In Juslin, P. and Sloboda, J. A., editors, *Music and Emotion*, pages 309–337. Oxford University Press.

Juslin, P. N., Friberg, A., and Bresin, R. (2002). Toward a computational model of expression in performance: The GERM model. *Musicae Scientiae*, (Special issue 2001-2002):63–122.

Juslin, P. N. and Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception*, 17:197–221.

Juslin, P. N. and Sloboda, J. A., editors (2001). *Music and Emotion.* Oxford University Press.

Kjelland, J. M. (1992). Application of electromyography and electromyographic biofeedback in music performance research: a review of the litterature since 1985. *Medical Problems of Performing Artists*, 15:115–118.

Kozlowski, L. T. and Cutting, J. E. (1977). Recognizing the sex of a walker from a dyniamic point-light display. *Perception & Psychophysics*, 21:575–580.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126(1):159–179.

Kuhn, T. L. (1974). Discrimination of modulated beat tempo by professional musicians. *Journal of Research in Music Education*, 22:270–277.

Laukka, P. and Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music*, 28:181–189.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, 1st edition.

Levitin, D. J., Mathews, M. V., and MacLean, K. (1999). The perception of cross-modal simultaneity. In *Proceedings of the 7th International Journal of Computing Anticipatory Systems*, Belgium.

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *Journal of the Acoustic Society of America*, 33:467–476.

Madison, G. (1992). Drumming performance with and without clicktrack - the validity of the internal clock in expert synchronisation. In *Proc. Fourth Workshop on Rhythm Perception & Production*, pages 117–122, Bourges, France.

Madison, G. (2000). On the nature of variability in isochronous serial interval production. In Desain, P. and Windsor, L., editors, *Rhythm perception and production.*, pages 95–113. Lisse: Swets and Zeitlinger.

Madison, G. (2001a). *Functional Modelling of the Human Timing Mechanism.* Published doctoral dissertation, Uppsala University, Sweden.

Madison, G. (2001b). Variability in isochronous tapping: Higher order dependencies as a function of intertap interval. *Journal of Experimental Psychology: Human Perception and Performance*, 27(2):411–422.

Madison, G. (2004). Detection of linear temporal drift in sound sequences: empirical data and modelling priciples. *Acta Psychologica*, 117(1):95–118.

Madsen, D. K. (1979). Modulated beat discrimination among musicians and non musicians. *Journal of Research in Music Education*, 27:57–67.

Mates, J. and Aschersleben, G. (2000). Sensorimotor synchronization: the impact of temporally displaced auditory feedback. *Acta Psychologica*, 104:29–44.

McGurk, H. and MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature*, 264:746–748.

Michon, J. A. (1967). *Timing in temporal tracking.* Assen, the Netherlands: Van Gorcum.

Moelants, D. (2002). Preferred tempo reconsidered. In Stevens, C., Burnham, D., McPherson, G., Schubert, E., and Renwick, J., editors, *Proceedings of the 7th International Conference on Music Perception and Cognition*, pages 580–583. Adelaide: Causal Productions.

Moeller, S. A. (1956). *The Moeller book.* Ludwig Music Publishing Co.

Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology*, 42:213–237.

Ogden, R. T. and Collier, G. L. (1999). On detecting and modeling deterministic drift in long run sequenced of tapping data. *Communications in Statistics - Theory and Methods*, 28:977–987.

Olsson, S. (1985). *Kroumata.* Svensk Skolmusik AB.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–346.

Parncutt, R. (2003). Accents and expression in piano performance. In Niemöller, K. W., editor, *Perspektiven und Methoden einer Systemischen Musikwissenschaft (Festschrift Fricke)*, pages 163–185. Frankfurt:Peter Lang.

Paterson, H. M., Pollick, F. E., and Sanford, A. J. (2001). The role of velocity in affect discrimination. In Moore, J. D. and Stenning, K., editors, *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, Edingburgh*, pages 756–761. Laurence Erlbaum Associates.

Pfordresher, P. Q. (2003). Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5):949–964.

Pfordresher, P. Q. and Palmer, C. (2002). Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, 66(1):71–79.

Pollick, F. E. (2004). The features people use to recognize human movement style. In Camurri, A. and Volpe, G., editors, *Gesture-based Communication in Human-Computer Interaction*, volume 2915 of *Lecture Notes in Artificial Intelligence*, pages 10 – 19. Springer Verlag.

Pollick, F. E., Paterson, H. M., Bruderlin, A., and Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, 82(2):B51–B61.

Pouliot, M. and Grondin, S. (2005). A response-time approach for estimating sensitivity to auditory tempo changes. *Music Perception*, 22:389–399.

Repp, B. H. (1992a). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustic Society of America*, 92:2546–2568.

Repp, B. H. (1992b). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. *Cognition*, 44:241–281.

Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception*, 16(4):409–438.

Repp, B. H. (2000). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research*, 63:106–128.

Resnicow, J. E., Salovey, P., and Repp, B. H. (2004). Is recognition of emotion in music performance an aspect of emotional intelligence? *Music Perception*, 22(1):145–158.

Rosenbaum, D. A. (1991). *Human Motor Control.* San Diego: Academic Press Inc., 2nd edition.

Runeson, S. and Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology*, 7(4):733–740.

Saldaña, H. M. and Rosenbaum, Lawrence, D. (1993). Visual influences on auditory pluck and bow judgements. *Perception & Psychophysics*, 54(3):406–416.

Schmidt, R. A. and Lee, T. D. (2005). *Motor Control and Learning: A Behavioral Emphasis.* Champaign, Ill.:Human Kinetics, 4th edition.

Schutz, M. (2004). Influence of visual information on auditory judgements of marimba strokes types. Unpublished thesis, Northwestern University, Dept. of Music Technology, Evanston: Illinois.

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14:147–152.

Sheldon, D. A. and Gregory, D. (1997). Perception of tempo modulation by listeners of different levels of educational experience. *Journal of Research in Music Education*, 45(3):367–379.

Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, 11:505–509.

Shivas, A. A. (1957). *The art of the tympanist and drummer.* London: Dobson Books.

Shumway-Cook, A. and Woollacott, M. H. (1995). *Motor Control. Theory and Practical Applications.* Williams & Wilkins, Baltimore.

Sörgjerd, M. (2000). Auditory and visual recognition of emotional expression in performances of music. Unpublished thesis, Uppsala University, Dep. of Psychology, Uppsala: Sweden.

Sundberg, J., Friberg, A., and Frydén, L. (1991). Threshold and preference quantities of rules for music performance. *Music Perception*, 9(1):71–92.

Taylor, M. M. and Creelman, C. D. (1967). PEST: Efficiant estimates on probability functions. *Journal of the Acoustic Society of America*, 41(4):782–787.

Thiem, B., Greene, D., Prassas, S., and Thaut, M. (1994). Left arm muscle activation and movement patterns in cellists employing a playing technique using rhythmic cuing. *Medical Problems of Performing Artists*, 9:89–96.

Trappe, W., Katzenberger, U., and Altenmüller, E. (2005). Expertise-related difference in cyclic motion patterns in drummers: A kinematic analysis. *Manuscript in preparation.*

Trappe, W., Parlitz, D., Katzenberger, U., and Altenmüller, E. (1998). 3-d measurement of cyclic motion patterns in drummers with different skill. In *Proc. of the Fifth International Symposium on the 3-D Analysis of Human Movement.*, pages 97–99, Chattanooga: Tennessee.

Turvey, M. T. (1990). Coordination. *American Psychologist*, 45(8):938–953.

Van Noorden, L. and Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66.

Vos, P. G., van Assen, M., and Franěk, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research*, 59:240–247.

Waadeland, C. H. (2003). Analysis of jazz drummers' movements in performance of swing grooves - a preliminary report. In Bresin, R., editor, *Proceedings of Stockholm Music Acoustic Conference 2003*, volume II, pages 573–576, Stockholm.

Walk, R. D. and Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of Psychonomic Society*, 22:437–440.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In Wachsmuth, I. and Sowa, T., editors, *Gesture and Sign Language in Human-Computer Interaction*, pages 241–253. April, Springer Verlag.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34:97–113.

Wang, C. C. (1983). Discrimination of modulated music tempo by music majors. *Journal of Research in Music Education*, 31(1):49–55.

Wang, C. C. and Salzberg, R. S. (1984). Discrimination of modulated music tempo by string students. *Journal of Research in Music Education*, 32(2):123–131.

Wickstrom, R. L. (1983). *Fundamental Motor Patterns.* Lea & Febiger, Philadelphia, 3rd edition.

Wing, A. M. and Kristofferson, A. B. (1973a). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, 14(1):5–12.

Wing, A. M. and Kristofferson, A. B. (1973b). The timing of interresponse intervals. *Perception & Psychophysics*, 13(3):455–460.

Winold, H. U. (1984). High speed photography of cello playing. In Roehmann, F. L. and Wilson, F. R., editors, *The Biology of Music Making*, pages 180–182. Proceedings of the 1984 Denver Conference.

Winter, D. A. (1990). *Biomechanics and Motor Control of Human Movement.* New York: John Wiley & Sons, 2nd edition.

Wohlschläger, A. and Koch, R. (2000). Synchronization error: an error in time perception. In Desain, P. and Windsor, L., editors, *Rhythm perception and production.*, pages 115–127. Lisse: Swets and Zeitlinger.

Zaza, C. (1998). Playing-related musculoskeletal disorders in musicians: A systematic review of incidence and prevalence. *Canadian Medical Association Journal*, 158:1019–1025.

Zaza, C. and Farewell, V. T. (1997). Musicians' playing-related musculoskeletal disorders: An examination of risk factors. *American Journal of Industrial Medicine*, 158:292–300.

Zetterberg, C., Backlund, H., Karlsson, J., Werner, H., and Olsson, L. (1998). Musculoskeletal problems among male and female music students. *Medical Problems of Performing Artists*, 14:160–166.

Part II

Included papers