Analysis of the acoustics and playing strategies of turntable scratching

Kjetil Falkenberg Hansen* Marco Fabiani
Roberto Bresin

Computer Science and Communication, KTH
Lindstedtsvägen 24, 10044 Stockholm, Sweden
Electronic mail: kjetil@kth.se

Abstract

Scratching performed by a DJ (disk jockey) is a skillful style of playing the turntable with complex musical output. This study focuses on the description of some of the acoustical parameters and playing strategies of typical scratch improvisations, and how these parameters typically are used for expressive performance. Three professional DJs were instructed to express different emotions through improvisations, and both audio and gestural data were recorded. Feature extraction and analysis of the recordings are based on a combination of audio and gestural data, instrument characteristics, and playing techniques. The acoustical and performance parameters extracted from the recordings give a first approximation on the functional ranges within which DJs normally play. Results from the analysis show that parameters which are important for other solo instrument performances, such as pitch, have less influence in scratching. Both differences and commonalities between the DJs' playing styles were found. Impact that the findings of this work may have on constructing models for scratch performances are discussed.

1 Introduction

Scratching is a striking feature to witness in DJ performances—it is the way the disk jockeys manipulate playback speed by pushing and dragging vinyl records, thus using the turntable as an instrument. DJ music is generally divided into distinct playing styles, where the main styles are (i) mixing records, (ii) playing a lead instrumental solo role with scratching, and (iii) constructing rhythms (with 'beat-juggling'). Scratch performances are both visually and musically compelling, and they demonstrate a high level of skills achieved through intensive exercise on the instrument ♪:1. [1]

The acoustical properties of scratching have to the writers knowledge not been thoroughly studied until now. In this paper, the primary aim is to explore DJ scratching by looking more systematically at the acoustical characteristics of the instrument. A second aim is to identify possible boundaries of prevailing playing strategies, and investigate which musical parameters are used in expressive performances. Also, we are interested to see if DJs use the same code as other musicians do, for instance with regards to changing performance cues such as tempo, sound level, timbre, and articulation in combination to communicate emotions (Juslin, 2001). Emotionally expressive DJ performances were therefore recorded in order to analyze playing strategies as they appear in a musical context. Encouraging expressive performances also ensures that the musicians use their whole artistic range.

Scratching started in the seventies, and has become the most expressive DJ playing style of a well-established instrumental practice. The turntable is now a popular instrument-of-choice for many young people starting to play music. However, the intended use of turntables was listening to music, not creating music. Consequently, there has been no formal instrumental training for DJs, although this is now changing with an increase in educational material. Both the background of the instrument and the current commercial realization of DJ products make it an interesting case study of a new, successful musical instrument.

The dominating way to use the turntable for scratching is through so-called *scratch techniques*. These are various formalized gesture combinations of synchronized record and crossfader hand movements, and are considered to be common vocabulary for scratch DJs. Performances can be decomposed by means of such techniques, like music notation is used for describing tones of an instrument, although they could also be compared to for instance guitar techniques like string bend, tapping, sweeping, tremolo, or even short licks,

[1] Sound and multimedia examples are marked ♪:*nr* and link to http://www.speech.kth.se/~kjetil/aaa10.php (date last viewed July 6, 2010; during review, username=aaa and password=review).

as one single performed technique seldom produce only one tone. In previous works some of the most common techniques were recorded and analyzed in isolation and not in their musical context (Hansen, 2002). For the current study, scratch techniques are not examined as isolated events, but in a natural musical context in which they are used in combination with each other.

In the next section, some characteristics of the DJ instrument are discussed. Then details of the data collection for the experiment are explained. The "Analysis" section describes the steps in the data analysis and feature extraction, including mechanical, acoustical and performance-related features. In the "Results" section, the data from the feature extraction are presented and discussed, both with a general perspective and with a focus on performance and expressive aspects. The paper finishes with some conclusive remarks on the feature extraction and data analysis, and also proposes some future directions.

2 Background

During the years, DJ equipment has reached a standard in its basic configuration. Turntables have a strong direct-drive motor, a stable platter, needles that endure direction changes, and a slip-mat between the record and the platter to keep a constant low friction. Audio mixers have a low-friction crossfader that is placed horizontally and close to the musician, see Fig. 1. Standard crossfaders have a 45 mm physical run, but the fading curve is often adjusted so as the span from silence to full signal covers only around 1 mm at the end positions, see Fig. 1. This makes the crossfader work basically as an on/off switch for sound.

Scratch music has strong elements of both rhythm and melody, but cannot easily be narrowed down to having just a percussive or melodic function. Even defining "a tone" can be surprisingly tricky. For solo instruments in general, it is common to consider a tone as consisting of an onset, a stable pitched part, and an offset. The tone is often associated with a performance gesture, for instance pressing down and releasing a piano key. However, there are a few exceptions worth noting. For example, to produce a sustained tone in percussion, a continuous roll can be played that has a number of onsets, but is perceived as a single long tone. Contrarily, the triple-tongue technique in wind instruments produces, from one breath and one tongue 'gesture', three onsets that are perceived as three separate tones. In written music

Figure 1: Left: A DJ in the recording session. A rotational sensor is centered above the vinyl. The left hand grips the crossfader on the mixer. Right: Audio mixer with the crossfader at the bottom. The *C.F. CURVE* knob zoomed in from the top right area adjusts the fading curve from linear to very steep.

such coupled or divided onsets have well defined functions and allow the player to overcome physical restrictions of the instrument. Scratching has however no tradition of notation, nor has a theoretical framework yet been established, and thus any similar coupling or separation of onsets will be uncertain.

A trained DJ knows the location on the vinyl of the sound sample that is used at any time, aided visually by the printed record label or by placing stickers on the vinyl—but the control is not perfect. Therefore, a gesture intended to produce a tone onset might coincide with a silent area on the record, thus resulting in a late onset or no onset at all. Furthermore, the sample itself might contain several onsets. As a consequence, there might not be a one-to-one correspondence between tone onsets and gestures. When such correspondence is not guaranteed, new methods for segmenting recorded DJ performances should be considered. The authors propose methods based on the combination of gestural data and the audio signal. These methods are described further below (in Section 4), and the features are listed in Table 1.

Little work has been done until now in the area of DJ scratching, and the music genre may still be unfamiliar to many. It can be interesting to notice that there are similarities between scratching, speech, the singing voice, and other instrumental sounds. For instance, Palmer & Hutchins (2006) review how speech prosody and musical prosody from traditional instrumental music can be compared. Similarities with the singing voice can be found for example in the humorous improvisation of scat singing, and naturally, in the vocal elements of hip-hop culture, *rapping* and *beatboxing* (Toop, 2000).

Similarities can also be found in several musical instruments. Percussive instruments, such as ethnic drums like the *kanjira*, resemble scratching in the use of dense, pitch-varying rhythmic patterns. There are also musicians who explicitly mimic DJs, for instance, some guitarists create scratch-like sounds with a volume switch and a pitch-shifter pedal ♪:2.

3 Data collection: Recordings with DJs

Recordings made by three Swedish professional DJs were collected. Their task was to improvise freely over a looped beat, using only a single sound sample. Because music notation has not been developed in this genre, improvisation is customary. Several performances were collected with instructions to express anger, sadness, self-confidence, a cool attitude, joy, and also to produce a neutral performance. Emotions were described both in English and Swedish, and each participant judged whether a performance was to be included. In total, more than 70 performances were collected ♪:3.

Multi-track recordings (96 kHz sample rate) of performances included the vinyl movement, the crossfader output level, the cross-faded record signal, the 'raw' scratched record signal (bypassing the mixer), and the background beat. The vinyl movement was tracked using a rotational sensor[2] placed above and affixed to the vinyl. This was not considered by the DJs as distracting, obstructing their gestures, or otherwise influencing their playing. Fig. 1 shows one of the subjects performing on the equipment. The sensor gave three output audio streams: the rotation speed, the directional changes, and a reference point for each complete round.

Both the record player (Vestax PDX-2000 MkII) and the audio mixer (Vestax PMC-05 ProIII) are standard models in professional use. The mixer had the possibility to adjust the crossfader characteristics. The sound sample, "ahhh", was the only sound used in the experiment ♪:4. It comes from a sound collection record (DJ 1210 Jazz, 2001) and is one of the most commonly used sounds for scratch solos. The sample reveals no distinct pitch in itself, but resembles a bandpass filtered noise with a bandwidth of about 1000 Hz centered around 1 kHz, see Fig. 2. The sound is not quite stable: there is a tendency of falling pitch towards the end, measured to span 330 cents over

[2]Fritz Kübler 8.5802.2171.1000 K1942-13, 1000 pulses/rotation.

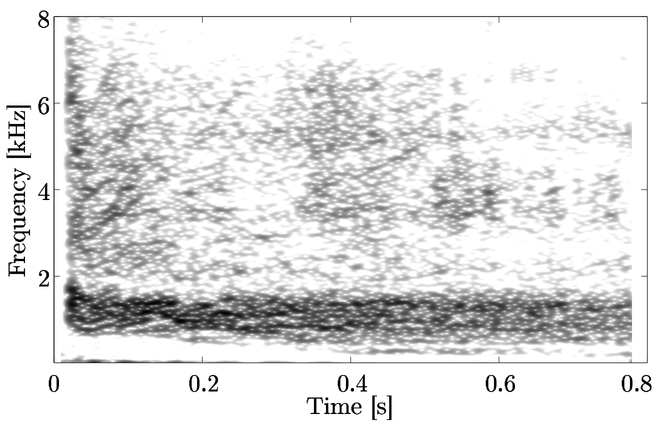

Figure 2: Spectrogram of the original sample used in the recordings. Most of the energy is concentrated in a 1000 Hz-wide band centered around 1 kHz without clear harmonic components.

the total duration of 0.7 s (see Sec. 4.2.5). The timbre is also changing, as can be discerned in the spectrogram.

The crossfader output level was measured by feeding a 25 kHz sine wave through the mixer, inaudible to the performer. In this way, although the DJs were free to adjust the mixer knobs, the crossfader curves, and the crossfader direction to their liking even during performance, the crossfader signal was always reliable. The DJs were encouraged to use standard techniques as they would do in a normal performance, but none were mentioned as being undesirable. Incidentally, unconventional techniques did not occur, thus facilitating the analysis.

The reference drum beat, chosen by one of the subjects, had a distinct 4/4 pattern at 90 beats per minute (bpm), which is predominant in hip-hop. The loop was considered to be a typical rhythmical background, simple and neutral ♪:5.

4 Analysis: Feature extraction

It is fair to say that DJ scratching differs from most other lead instruments in popular music in several respects. Firstly, its restricted pitch capabilities hinder tone segmentation based on melodic content. Secondly, temporal properties characterized by a constant high number of tone onsets make metrical hierarchy-based categorization impractical. Standard feature extraction methods used for melodic music (Friberg *et al.* , 2007; Lartillot & Toiviainen, 2007) proved to be very difficult and mostly unreliable for the extraction of features such as duration, timbre, attack properties and rhythm. A solution to this problem was found in the use of the additional gesture data: ad-hoc analysis of the signals acquired using various sensors, and their combination, allowed more reliable estimates of the feature values, compared to those extracted from the actual audible performance. A list of the extracted features is presented in Table 1.

4.1 Mechanical features

4.1.1 Crossfader

To extract crossfader onsets and offsets, the crossfader signal was first squared and then low-pass filtered (6^{th} order Butterworth, $f_c = 24\,\text{kHz}$) to obtain an amplitude envelope of the 25 kHz tone. A threshold to detect onsets and offsets was set at one third of the maximum envelope amplitude.

4.1.2 Record direction and speed

The signal recorded by the rotational sensor has a spike for each change in direction: between a negative and a positive spike, the record turns clockwise (cw) or forward, whereas between a positive and a negative spike, it turns counterclockwise (cc). Approximate spike positions were detected by searching for peaks in the time derivative of the signal. Positions were then refined by running a local maxima (cc) or minima (cw) search around the approximate values. The turning position values have a resolution of one sample.

The rotation sensor emitted an impulse every 0.36 degrees of rotation of the record ($N = 1000$ impulses for a full round). To determine the angular speed, the distance in seconds Δt between two successive impulses was computed. The angular speed (absolute value) is thus

$$|\omega_{\text{rec}}| = \frac{0.36}{\Delta t} \tag{1}$$

The module of the angular speed was finally combined with the direction of rotation data to obtain the angular speed vector.

4.1.3 Sample position

The rotational sensor was attached to the record in a fixed position. Parallel to the rotation signal, another signal was recorded that emitted an impulse each time a certain point on the record, placed approximately in the center of the sample, was trespassed. By taking this impulse as a reference, and by knowing the direction of rotation, the position $p(t)$ on the record was estimated for each spike in the rotational sensor signal ($1 \leq p(t) \leq 1000$, where 1000 is the total number of markers). The sample spanned over approximately 400 markers, corresponding to an 144° sector.

4.2 Acoustical features

4.2.1 Tone onsets and offsets

To detect tone onsets and offsets, a standard method based on a sound level envelope and a time-varying threshold, similar to that in Friberg *et al.* (2007), was first tested. This method proved unreliable because of the nature of the scratching performance: very short tones, and very different and sometimes not well-defined attacks. Instead, a new approach was designed, based on sensors data only.

Three types of onset/offset are possible in a scratching performance: *sample* onsets/offsets (actual onsets and offsets produced when passing from silence to sound on the record, based on the sample position $p(t)$ and the direction of rotation); *crossfader* onsets/offsets (produced by turning the crossfader "on" and "off"); *direction change* onsets/offsets (produced by the short silence interval caused by stopping the record to change direction)[♪:6], see Fig. 3. (Two less common onset types are needle drop and channel volume fade-in; these did not occur in the data set.) Knowing already the positions of all the three types separately, simple logic was used to combine them into a single sequence (On_{tone} and Off_{tone}): a sample onset/offset occurred in the audible performance if the crossfader was "on" at that instant; a crossfader onset/offset occurred if the record position $p(t)$ at that instant was inside the sample range; a direction change onset occurred if the crossfader was "on" and the position was inside the sample. For the direction change onsets, a corresponding offset was inserted 5 ms before the onset, a rough estimate of the silence due to a direction change.

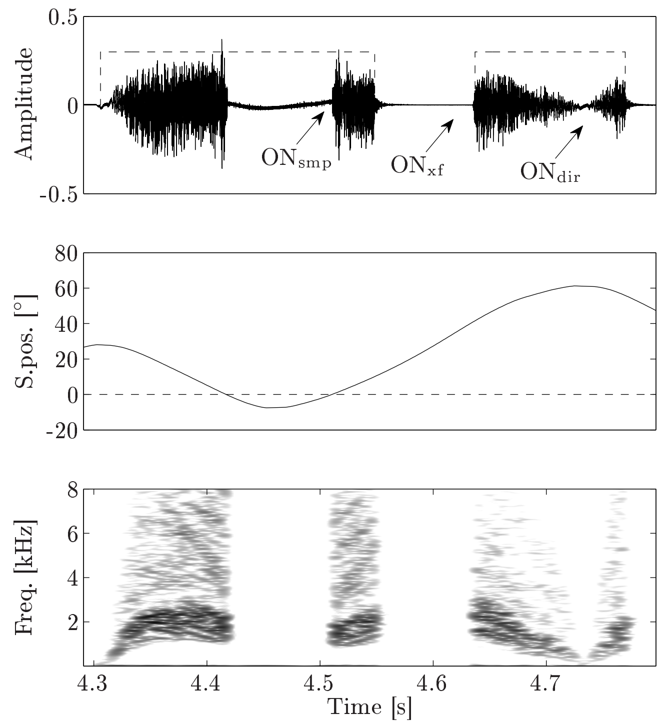

Figure 3: Performance excerpt containing all the three onset types of scratch tones: sample onset, crossfader onset, and direction change onset. The plots show the amplitude of the signal and crossfader (dashed line), the sample position (silence below 0°), and the spectrogram.

The attack slope AS for an onset was defined as

$$AS = \frac{\Delta SL}{t_{\text{att}}} \tag{2}$$

where ΔSL is the sound level difference between the beginning and the end of a tone's attack, and t_{att} is the attack time. The attack boundaries (and thus the attack time t_{att}) of a tone were determined by finding the zero-crossing points in the derivative of the amplitude envelope SL_{env}, in the vicinity of an onset.

4.2.2 Duration, IOI and articulation

For each tone (and in the same way for crossfader and record turning gestures), duration, *IOI* (inter-onset interval), and articulation were defined as

$$\begin{aligned} Dur_{\text{tone}}(n) &= Off_{\text{tone}}(n) - On_{\text{tone}}(n) && (3) \\ IOI_{\text{tone}}(n) &= On_{\text{tone}}(n+1) - On_{\text{tone}}(n) && (4) \\ Art_{\text{tone}}(n) &= Dur_{\text{tone}}(n)/IOI_{\text{tone}}(n) && (5) \end{aligned}$$

where On_{tone} and Off_{tone} are the positions of audible tone onsets and offsets, respectively. The IOI_{rec} (record direction *IOI*) was defined as the distance between two successive changes in direction. After an inspection of the audio signal, a 10 ms value was deemed suitable to exclude possible artifacts that did not qualify as real tones, thus all IOI_{rec} shorter than 10 ms were removed.

4.2.3 Sound level

An amplitude envelope (SL_{env}) was extracted by summing the amplitudes of the STFT (∼5.5 ms long frames, 50% overlap) for each time frame. The amplitude curve was then converted to decibel scale and smoothed using a low-pass filter. The sound level *SL* of a tone was defined (as in Friberg *et al.* , 2007) as the upper quartile of the sound levels of each time frame comprising that tone.

4.2.4 Spectral centroid

The spectral centroid *SC* of a single time frame was defined as

$$SC = \frac{\sum_{n=0}^{N-1} f(n)H(n)}{\sum_{n=0}^{N-1} H(n)} \qquad (6)$$

where $f(n)$ and $H(n)$ are the center frequency and magnitude of the n^{th} frequency bin of the Fourier's transform of the signal. To reduce the influence of outliers, the median, instead of the mean, was used as a single value of spectral centroid for each tone.

4.2.5 Pitch

As mentioned in Section 3, the sample does not have a well defined pitch itself. Most of the energy is concentrated in a relatively narrow band (about

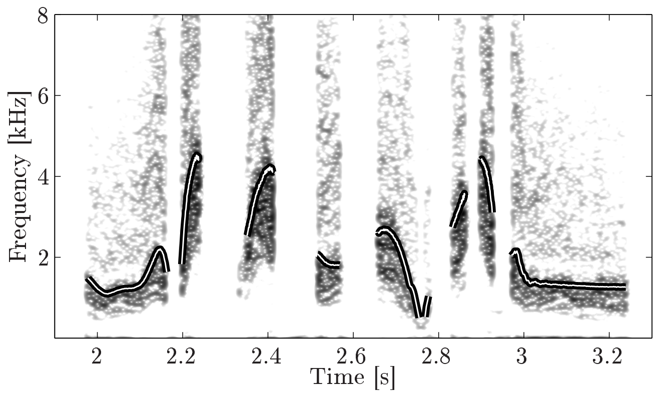

Figure 4: Pitch contour plotted as a black+white line on the spectrogram of a performance excerpt.

1 kHz wide) with a center frequency oscillating from about 1200 to 1500 Hz. This can be seen in Fig. 2. Bandpass filtered noise has been shown to induce a relatively weak sense of pitch (Fastl & Zwicker, 2007), and pitch perception is also disturbed by timbre changes (Vurma & Ross, 2007). There is a linear relation between the center frequency of the narrow band noise in the sample and the record playback speed. As a consequence of the playback speed variations, an effect of continuously changing pitch can be perceived in the scratch performances[♪:7].

Because of the noisy nature of the signal, any interpolation to obtain a more precise estimate of the peak's frequency was not considered necessary. A second order polynomial curve (the black+white line in Fig. 4) was fitted to the data to obtain a smoother contour (f_{smp}). Then, frequency was estimated at every time instant as

$$f(t) = \frac{|\omega_{rec}(t)| f_{smp}(p(t))}{|\omega_0|} \tag{7}$$

where ω_0 is the free-turning speed of the record ($\sim$180°/s). A single value for each tone was taken as the median of the curve between the tone's onset and offset.

4.3 Overall performance features

4.3.1 Gesture coordination

An interesting aspect of DJ scratching is the coordination between the different movements (record and crossfader) to produce tones, and the feeling of the position of the sample on the record. Four significant parameters were identified to study this phenomenon: ΔXF_{smp} and $\Delta REC_{\mathrm{smp}}$, the distance between a sample tone onset and the preceding crossfader and direction change, respectively; ΔREC_{xf}, the distance between a crossfader tone onset and the preceding direction change; ΔXF_{rec}, the distance between a direction change tone onset and the preceding crossfader onset. These values were computed only if the gesture was performed between the tone onset and the preceding tone offset.

4.3.2 Quantity of motion

Several quantity of motion indicators were defined from the extracted onsets. Three density measures were defined as number of onsets per second: tone density (δ_{tone}), crossfader gesture density (δ_{xf}), and record direction change density (δ_{rec}).

XF:*rec* represents the number of crossfader onsets per record turning direction, or how many times the crossfader was turned on–off during one record gesture. *REC*:*xf* instead represents the number of record direction changes observed while the crossfader was "on". *XF*:*rec* and *REC*:*xf* are discrete values computed for each record and crossfader movement respectively, and therefore not reciprocal.

5 Results and discussion

In this section, the extracted features are analyzed to (i) describe acoustical and gestural characteristics of scratch sounds, (ii) compare how the DJs used different playing strategies, and (iii) describe how features were varied in expressive performances. Only an approximated overview can be given with three subjects; nevertheless, the musicians showed distinct individual performance characteristics.

Around 6500 tones, 6300 record movements and 3100 crossfader on–off gestures were identified and segmented in the collected data. A record movement is defined as a direction change, so while it could be argued that a typical gesture is a forward–backward movement cycle of the record, 'record gesture' will henceforth be used to describe the action sequence of starting, moving and stopping the record.

It is important to consider that the following analysis is limited by using only the "ahhh" sound sample. Manipulating another sample would reasonably yield different results for most or all of the regarded features, as DJs have to adapt their playing to accommodate the different sounds, like a keyboardist would do.

The extracted data, listed in Table 1, support many of the findings in Hansen & Bresin (2004). However, in that previous study the sensors used to record gesture data were less accurate. Also, it was limited to a case study of one performance.

5.1 General observations for extracted features

5.1.1 Sample position

The sound sample was used similarly in all the performances. Fig. 5 shows a summary of which sample part was played, or the needle's position in the sound at any point. As seen in the curve for Total, which is averaged across all performances, there is a clear preference for using the beginning of the sample. About 88% of the played sounds come from the first half of the sound, 22% from the very beginning up to 10° (corresponding to moving the record 1–1.5 cm under the needle from the spot where the sample was located on the vinyl).

Most scratching is done using samples with duration similar to the sound used here. However, even shorter sounds are common, such as single spoken syllables or percussive sounds, and this stimulates a practice of staying close to the sample's beginning regardless of the duration. Another reason for the bias on the beginning is to produce tones with decreasing energy, which are common for many traditional instruments, the voice, and non-musical sounds, and thus also common in the records used by DJs.

5.1.2 Sound level and timbre

Dynamic variation, together with tone duration and tempo, is one of the most important parameters in music performance, yet for some instruments

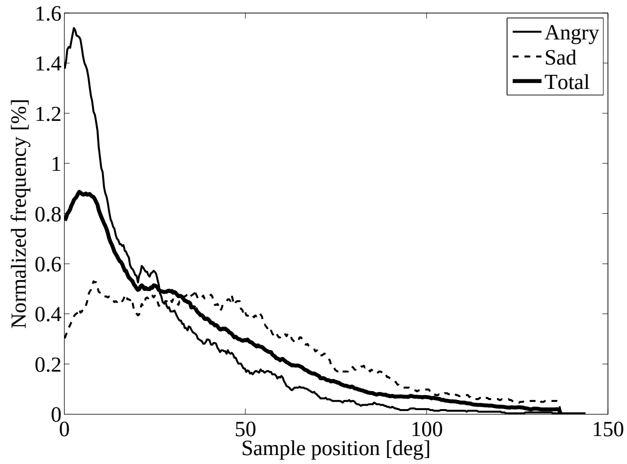

Figure 5: Normalized frequency in percent of played sample position, $p(t)$, in degrees for the averaged total for all performances (bold line), the "angry" condition (thin line), and the "sad" condition (dashed line).

such as church organs and bagpipes, physical constraints make dynamics less variable, and for others like rock guitar, stylistic conventions have the same effect. Although SL can easily be adjusted with the channel volume fader, this is quite rarely done in a systematic manner in scratching—apart from making for instance echo-effects—and none of the subjects did.

In the collected data, SL variations (mean$=30$ dB, SD$=7$) can to a large extent be accounted for by the speed variations (the correlation between ω_{rec} and SL was .8). Another explanation for the observed variations in SL is the sample position, as the original sample sound's pitch, energy and timbre characteristics are not constant. It can also be seen in Fig. 6 that SL is affected by tone duration: very short durations, below 50 ms (or about $1/64$ notes at 90 bpm), have the widest variation in SL, while longer durations, above 150 ms (or $1/16$ notes), are fairly constant and reflect the SL decrease in the original sample. Reasonably, it is easier to control the SL in the 50–150 ms duration range.

For the analysis of timbre, the spectral centroid (SC) serves as a measure of brightness. SC and pitch vary almost equivalently, but also the sample position can explain small variations in SC as the timbre changed in the original sound. Undoubtedly, timbre difference is less used as a parameter than for instruments where control of tone quality is more direct. The anal-

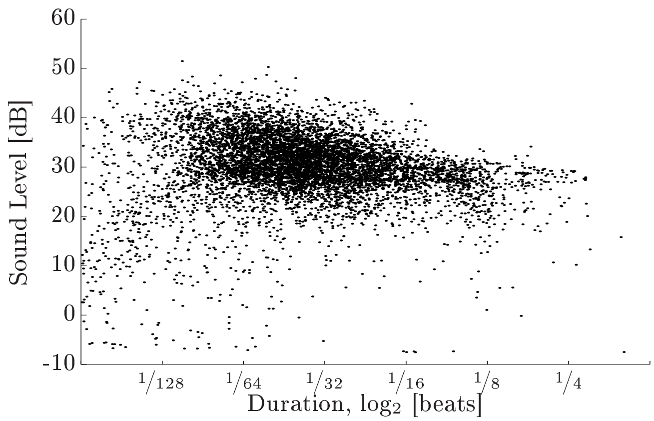

Figure 6: Tone duration vs. sound level for all tones, plotted corresponding to note durations from 1/256 notes (10 ms) to quarter notes (667 ms).

ysis does not discern whether different parts of the sample were employed to attain timbre qualities or for other reasons, such as tone attack qualities.

5.1.3 Attack properties

The three main types of tone onset produce attacks with distinguishing properties. The attack times measured with a *SL* definition were almost twice as long for direction change onsets than for crossfader onsets, see Fig. 3 and Sec. 4.2.1.

Direction change onsets, in total 28% of the total number of tones, allow the most nuanced attack control. However, these attacks render a highly stereotypical sound with closely connected and gliding tones; for an example, see the spectrogram in Fig. 3. Direction changes are most often muted by the crossfader, and assumably, this serves a particular musical purpose. Sample onsets normally produce very sharp attacks, according to the sample's own onset characteristics. Although the DJs mostly use the first part of the sample, only 19% of the tones have a sample onset. Crossfader onsets, in total 54%, give little variation in attack time because of the steep fading curve. The crossfader is used to produce several tones from one record gesture, and it also allows better control for timing onsets.

Given the fundamentally different nature of the attack types, a direct comparison of attack onset times seems inapplicable. In particular, it is very difficult to establish when a direction change tone attack is perceived as completed, as it might depend on both *SL* and pitch.

5.1.4 Pitch

The mean pitch as defined in Section 4.2.5 for the material was 1532 Hz, close to the sample's played at nominal speed, with less than one octave standard deviation. The estimate is only based on an energy measure of the signal, and it is debatable if we perceive the pitch to be as high as the measured frequency. In the following, pitch span is defined in terms of cents, calculated from the measured range for each tone. By inspecting the pitch curve in the audio data (an example is plotted in Fig. 4), it appears that very few tones have a stable pitch. Even the original sample played at nominal speed would not yield a stable pitch. The average pitch span and variation in cents for each tone were overall very high (mean= 2109 cents, SD= 1575), suggesting that pitch stability is not pursued, nor, possibly, even attainable. The maximum span observed for one tone was more than seven octaves.

Tones with direction change onset or offset will be characterized by such large pitch spans. For example, a pitch glide of several octaves is generated when the vinyl speed is varied between nominal speed and rest. As a consequence, tones composed of both a direction change onset and offset—7.6% in the data set—have two glides.

Pitch perception was not approached in the analysis. However, several studies in perception have shown that listeners will, depending on the conditions, either separate a gliding tone into several pitches or blend into one (Nábělek *et al.* , 1973). Also, studies have shown when in a glissando the pitch is perceived, and how different glides are discriminated (for example Madden & Fire, 1997).

5.1.5 Hand gestures and record speed

Record movements in the data set had an average span of 36°, with a local maximum in the distribution around 10° due mainly to techniques which were played with trembling movements on the record. The *scribble scratch* (Hansen, 2002) [♪:8] is the most common of these techniques and it affected other findings in the present study (by for instance producing a high number of direction change onsets). Few movements were longer than 75°. The analysis does not discern the hand position on the vinyl; the distance at the vinyl perimeter is three times longer than at the label perimeter for the same

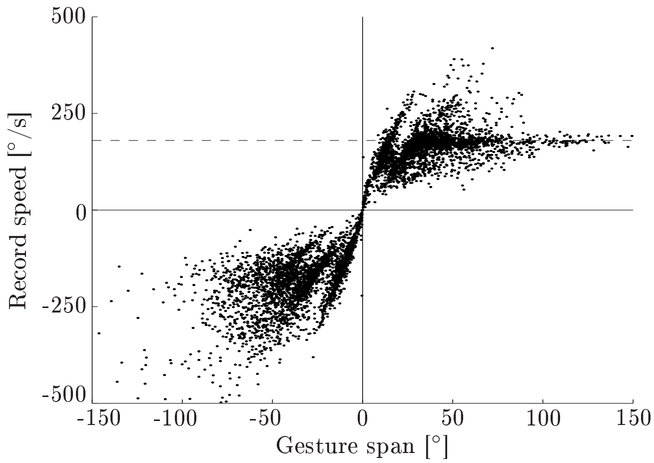

Figure 7: Record gesture span vs. record speed for all tones. The dashed line is the turntable's motor speed.

record movement.

In the part of the movement that corresponded to audible output, around 78% of the gestures ranged between 116 and 276°/s, which would suggest that there is a preferred ω_{rec} used throughout the performances (mean$=170°$/s, SD$=66$). The mean speed was also close to the nominal speed. The highest record speeds (ω_{rec}) found were around four times larger than the nominal speed at 180°/s (the motor speed was pitched down about 10% from 33 r/min by the subjects). Only around a tenth of the gestures were found to be slower than 116°/s, which indicates that there is a practical lower speed limit.

The distribution of gesture span and speed values is plotted in Fig. 7. In the positive quadrant, the speed of tones with a long span (forward movements with span above 75°) approximates 180°/s, i.e. the nominal rotation speed of the motor. This was caused by releasing the record, a typical gesture to end a phrase[♪:9], and naturally this tendency was not observed for the backward movements above 75°. Also discernable from the plot are two possible trends (crossing the x-axis at appr. $-40°$ and $-82°$) which show that the record speed and gesture span varies, but the duration is constant (73 ms and 155 ms respectively, corresponding to $^1/_{32}$ and $^1/_{16}$ in 90 bpm).

In general, faster movements were used for returning the record; the counterclockwise record speed was in average 11% higher than the clockwise. This might be done both to compensate the motor force, and also to reposition the sample in time. Significantly more tones were generated from forward movements than backward movements. Of the total sounding material measured in time, 60% came from clockwise movements.

5.1.6 Gesture timing

Each of the three onset types implies a different gesture coordination pattern for timing to rhythmical events. As the performances were improvised, the musical or compositional intentions were not revealed and it is assumed that all onsets were rhythmically timed events. It should be noticed here that undetectable errors are introduced when the DJs misjudge the sample position.

For the sample onsets, the crossfader must be switched on and the record must be moved before there can be an onset. In average, both the crossfader and the record were moved 51 ms before the tone onset. Also for crossfader onsets, the record movement started 51 ms before the onset, suggesting that the record should reach a certain speed before the crossfader produces the attack. The crossfader is probably most convenient for timing precise onsets: the gesture is larger for record movements, the sample onsets depend on the sample's position on the record, and direction change onsets produce longer attack times.

Apart from the mechanical advantages of using the crossfader for timing, it could be that certain techniques are more "acoustically suited" for timing with accented beats. In techniques like e.g. *stabs* the crossfader is switched on immediately after a record gesture is started, while in *chirps* the crossfader is switched off directly after the record gesture is started (Hansen, 2002) [♪:10], and timing gets much harder.

The record and crossfader gestures could thus be seen as preparatory movements for sound production. A previous study found that accented musical events are often evenly spaced in time, which implies that musicians match upcoming events to a regular beat (Jungers *et al.* , 2002). Two possible strategies for precise timing in scratching are to use crossfader onsets, or to use techniques with a regular movement pattern, such as *stabs*.

5.1.7 Durations and articulation

In all the material analyzed, the longest durations were around 800 ms (a quarter note duration in 90 bpm is 667 ms). 60% of the tones had *IOI* shorter than 167 ms. Record movements had a dominance of eighth notes, which indicates that the record hand can be used to keep a pulse (mean= 213 ms, SD= 130). Crossfader *IOI*s were distributed with peaks at $^1/_{32}$, $^1/_{16}$ and $^1/_8$ note durations.

It is beyond the scope of this study to discriminate if the perceived tones correspond strictly to the segmented tones, or if there are psychoacoustic effects like fused tones. The dominance of very short durations is naturally also a strong explanation why pitch is difficult to apprehend (Moore, 1973).

Articulation varied over the whole range from short staccato to almost connected tones (mean= 59%, SD= 26), and 68% of the tones were within the staccato–detached range (i.e., tone duration between 25–75% of *IOI* (Bresin & Widmer, 2000)) or shorter. Direction change onsets (28% of the onsets) have as discussed above minimal *SL* gaps, and those tones were classified as legato. Arguably the tone quality caused by pitch glides does not give a legato impression, and legato style could instead be performed by using long, slow record movements and short crossfader cuts. However, it seems that connected tones were not endorsed by the DJs. Only 54 % of the total performance time consisted of sound.

5.1.8 Event density

Density of tone events, here also including gestures, can serve as a simple measure of musical complexity. As the pitch and tonality is already equivocal compared to traditional music, it would make little sense to apply existing measures of melodical complexity to the data, such as tone transitions or weighting according to tone duration (Eerola *et al.* , 2001).

Almost all scratch techniques involve both hands, however many techniques emphasize the crossfader activity more than the record activity. Typically, a technique has one simple forward–backward movement with the record, synchronized with a number of carefully timed crossfader movements. The number of record gestures while the crossfader was on was higher than the number of crossfader onsets during a record gesture (*REC*:*xf* mean= 1.27 vs *XF*:*rec* mean= .75; recall that forward and backward record movements are treated as individual gestures). One explanation is the aforementioned *scribble scratch* which produces many directional changes with short *IOI*s and span. It can be maintained for longer durations, sometimes up to several seconds, and can generate around 20 onsets/s. In Hansen & Bresin (2004) it was found that the scribble accounted for one third of the direction change attacks.

The tone density was high (mean= 5.7 tones/s, SD= 1.2, comparable to a stream of sixteenth notes in 90 bpm; however, the durations were not regular). To achieve this, crossfader and gesture densities need to be high. The crossfader was switched on–off 3.7 times/s in average over the whole data (SD= 1.7), while the record changed direction 4.8 times/s (SD= 1.0).

5.2 Performance strategies

Consistency was observed in the three subjects' use of gesture preparation time, but they showed individual differences in timing. DJ A had preparation times around 50 ms in average, while DJ C had 35% shorter and DJ B 20% longer times. As neither gesture speed, attack slope, pitch, nor pitch span reflected these proportions, it is likely that individual preferences or skills can explain the preparation times more than aesthetic intentions.

Although techniques are not specifically approached in this study, the distributions and proportions of crossfader and gesture events suggest that the DJs do not vary their techniques in the same way. DJ A had half the number of gestures per crossfade ($REC{:}xf = .6$) than the others, but twice the number of crossfades per gesture ($XF{:}rec = 1.4$). DJ B had a dominance of 150 ms crossfader *IOI*s, while the others had a strong peak at 75 ms. While tone density (δ_{tone}) was similar for the subjects, between 5.5 and 6.0, record and crossfader event densities (δ_{rec} and δ_{xf}) varied. δ_{rec} decreased from DJ C and DJ B to DJ A. On the contrary, DJ A had almost two times higher δ_{xf} than the others, and even 25% more crossfader onsets.

Eerola *et al.* (2001) found that tone density is a good descriptor for predicting musical similarity between folk music melodies. Likewise, it is reasonable to believe that the gesture density measures, which were different for the three subjects, are good variables for predicting variations in playing strategy between performers, even when tone density is nearly the same.

Performances of the *chirp* technique were identified for each of the DJs; in Fig. 8, chirps from DJ A and DJ C were taken from the angry condition, DJ B from the 'neutral' condition [♪:11]. DJ B played two instances of the chirp, while DJ A and DJ C did several in their faster versions (a plausible explanation for the differences is the intended emotion). Chirps are considered to be hard to learn, as they require precisely coordinated gestures and control of sample position. To produce a chirp, only a fragment of the start and ending of a record gesture should be heard, and not the direction changes.

Because of speed variations, the examples in Fig. 8 sound slightly different, but they are made with the same gestures, and there are commonalities between the DJs in the production of the scratch. All used only the very beginning of the sample, up to 20° of the whole length of 144°. Record direction changes always occurred halfway between two tones (at the minima and maxima of the sample position curve). The tones have almost equal duration despite the timing difference, which was accommodated by adjusting the crossfader-off duration. The gesture speed, however, was varied (cf. the spectrogram plots in Fig. 8); DJ A had a lower speed than DJ C, while DJ B modulated the speed more. The use of techniques as seen with the chirp reminds of how musical phrases provide a structure for the performance which is indeed followed by players although with stylistic variations within the phrase boundaries, as in the case of piano performance (see Repp, 1992).

5.3 Cues for expressivity

It is out of the scope of this work to verify the success in the communication of intended emotions, but informal listening tests conducted by the authors revealed that it is reasonable to take the DJs' interpretation of the instruction given as ground truth for each emotion category. It follows that it is risky to collapse the data from all subjects for each emotion. For the same reason, only a few interpretations are discussed. Most of the observations are supported by results from other instruments and disciplines. Table 1 shows some of the parameter values found for performances with different emotional intentions.

Many of the parameters listed in Table 1 can be considered as indicators of either energy or activity, and several of them are physically related to each other. High-frequency energy, for example, is associated with louder and brighter tones, produced by faster record movements. Using faster movements implies shorter time intervals to cover the functional range of the record area—thus shorter *IOI*s—and plausibly a higher activity or event density. As seen in previous studies of musical expression, energy and activity play key roles in communicating emotions through music (for example Bresin & Friberg, 2000; Juslin, 2001; Juslin & Laukka, 2003), thus similar tendencies are expected.

To determine the existence of significant differences between the emotional intentions in the recorded performances, features that related to energy and activity were organized on a low–high energy and activity scale, respectively. After normalizing those parameters assigned to the energy cat-

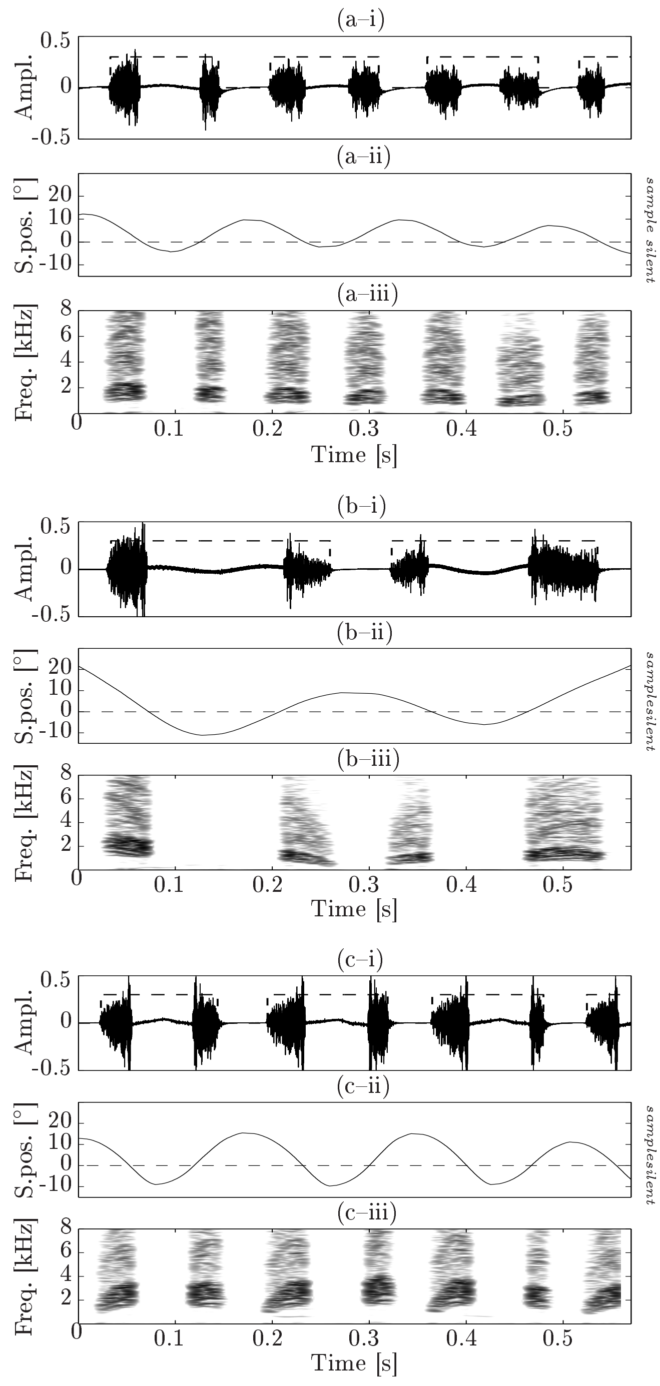

Figure 8: (a–c) *Chirp* scratches performed by three DJs. Plots show (i) Amplitude of the signal and crossfader (dashed line), (ii) the sample position (silence below 0°), and (iii) the spectrogram.

Table 1: Measured values for the extracted parameters grouped in acoustic, gestural, and performance features. Mean and standard deviations for all performances (total), and mean for emotional intentions.

Acoustic features		Scale	Total	(SD)	Angry	Joyful	Self-conf	Cool	Sad
SL	Sound Level	dB	**30**	(7)	33	30	30	29	27
SC	Spectral Centroid	Hz	**3633**	(993)	3949	3549	3795	3609	3345
On_{xf}	Crossfader onsets	%	**54**		51	55	53	54	64
On_{smp}	Sample onsets	%	**19**		31	14	22	16	10
On_{dir}	Direction change onsets	%	**28**		18	31	24	30	26
f	Pitch	Hz	**1532**	(676)	1849	1525	1634	1419	1344
	Pitch span	cents	**2109**	(1575)	1731	2190	1991	2170	2070
Dur_{tone}	Tone duration	ms	**93**	(89)	76	84	91	102	126
IOI_{tone}	Inter-onset Interval	ms	**170**	(163)	159	151	172	186	229
Art	Articulation, Dur/IOI	%	**59**	(26)	52	61	57	60	60
Gestural features		Scale	Total	(SD)	Angry	Joyful	Self-conf	Cool	Sad
ω_{rec}	Record speed	°/s	**170**	(66)	189	173	180	166	164
	Record gesture span	deg	**36**	(23)	34	36	36	39	41
ΔXF_{dir}	Crossfader offset time	ms	**51**	(37)	42	51	52	61	49
$\Delta REC_{\mathrm{xf,dir}}$	Record offset time	ms	**51**	(34)	40	51	48	62	55
Dur_{rec}	Record IOI	ms	**213**	(130)	181	199	200	235	258
IOI_{xf}	Crossfader IOI	ms	**288**		273	252	286	312	340
Performance features		Scale	Total	(SD)	Angry	Joyful	Self-conf	Cool	Sad
$XF{:}rec$	Crossfades/record gesture	n	**0.75**	(1.76)	0.66	0.81	0.68	0.73	0.74
$REC{:}xf$	Record gestures/crossfade	n	**1.27**	(1.36)	1.47	1.18	1.38	1.29	1.27
δ_{tone}	Tone density	n/s	**5.7**	(1.2)	6.3	6.5	5.7	5.3	4.2
δ_{xf}	Crossfader density	n/s	**3.7**	(1.7)	3.8	4.2	3.5	3.3	2.8
δ_{rec}	Record gesture density	n/s	**4.8**	(1.0)	5.7	5	5	4.3	3.9

egory

$$e(n) = \{SC, SL, AS_{\text{dir}}, AS_{\text{smp}}, AS_{\text{xf}}, |\omega_{\text{rec}}|, f_{\text{tone}}\}$$

and those assigned to the activity category

$$a(m) = \{Dur_{\text{tone}}, Dur_{\text{xf}}, Dur_{\text{rec}}, \delta_{\text{tone}}, \delta_{\text{xf}}, \delta_{\text{rec}}\}$$

the value for each emotion was calculated from

$$Energy = \frac{1}{N}\sum_{n=1}^{N} e(n) \quad (8)$$

$$Activity = \frac{1}{M}\sum_{m=1}^{M} a(m) \quad (9)$$

The values for the emotional intentions organized in the two dimensions are plotted in Fig. 9. There were significant relationships between both emotion and energy and emotion and activity for the calculated values of energy and activity (with $\eta^2 = .91$ and $\eta^2 = .89$, respectively). A further test, Tukey's HSD (Honestly Significant Difference), showed that for the energy dimension, all differences between emotional categories are significant except between self-confident and joyful ($\alpha = .05$). For the activity dimension, all differences were significant except between joyful and angry. It thus seems that the DJs used acoustical and gestural cues systematically in the communication of emotional intentions, and organizing features in energy and activity parameters would as expected reveal this.

The organization of the emotions in Fig. 9 is similar to the placement of emotions in activity–valence dimensions as reported by Russell (1980) and Juslin (2001). The emotional categories cool and self-confident were chosen as they are regularly used by DJs, and were not explicitly included in the two above studies. Cool is positioned approximately like calm in Russel's unidimensional scaling, while self-confident appears more like a neutral or base expression, positioned centrally in the plot.

The measures of activity or event density as described above are among the most important parameters in expressive DJ performances. It is impossible to have a high energy without a concurrent high activity, but contrarily the DJ can generate high event density even with lower energy. Fig. 9 shows that the emotions joyful and angry have the same levels of activity, while

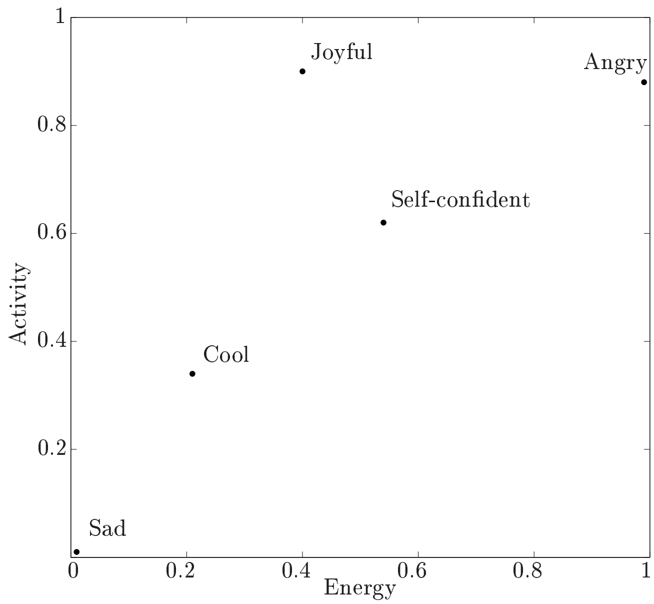

Figure 9: Organization of emotion categories for scratch performances in activity and energy dimensions based on normalized values of performance features.

joyful has lower energy. Balkwill & Thompson (1999) showed that emotional communication can be perceived based on musical complexity (among other psychoacoustic features) even in unfamiliar music cultures, and for scratching it indicates that event density can be a robust cue for untrained listeners.

In addition to the energy and activity connection, some performance parameters are linked due to physical properties of turntables. For instance, sadness could feasibly be portrayed with long and slow movements, and anger with short and fast ones, and all the gesture parameters would reflect this. Although angry performances had higher movement speed than sad performances (mean= 189 vs. 164°/s), and shorter durations (mean= 181 vs. 258 ms), the movements had almost the same span (mean= 34° vs. 41°; a difference corresponding to less than 1 cm on the record at the place of the needle). Thus, despite limiting the pitch variability, gesture span was not extensively used as a control parameter in the data set.

Even features that are unique to the turntable, for instance the sample position plotted in Fig. 5, seem to be varied systematically for expressive purposes. It is likely though that those curves resulted from tendencies of varying tone duration and onset type more than from carefully using sample position.

6 Conclusions and future studies

By applying methods for feature extraction based on audio and gesture data, recordings of expressive scratch performances were segmented and analyzed. Among the extracted features, event density, duration, sample position, gesture speed, gesture timing, and pitch span provided results for revealing characteristics of the gestures and acoustics of scratching in the present data set. In expressive performances, there are strong indications that DJs use the same parameters and in a similar way for conveying the same emotions. The success of such communication for improvised scratch performances was not approached in this study.

The acoustical analysis shows that the musical sounds of scratching are dissimilar from those of other solo instruments used in popular music. Perceptual studies are needed to disclose more about pitch, tone onsets, temporal apprehension and structuring of the short tones, and also how varying the original sample affect these aspects. For instance, the auditory process behind grouping onsets into rhythmical units is affected by experience

(Iversen *et al.* , 2008), which could mean that trained and untrained listeners experience tone density in scratching differently. It can also be relevant to compare scratching to areas with resembling characteristics, such as the mentioned musical prosody, which is culturally dependent (Thompson & Balkwill, 2006), and other musical instruments and vocal styles.

There were some different strategies between the musicians in the use of gestures. Specific techniques were not analyzed, but individual differences in the *XF*:*rec* and *REC*:*xf* measures (number of crossfader onsets during one record gesture and number of record gestures while the crossfader is on, respectively) indicate that their playing strategies involve also personal preferences of techniques. The present study only considers how the *record and crossfader* moved. The hand and finger positioning is a more personal choice if compared to traditional melodic instruments, affecting individual playing strategies and amount of preparatory movements needed.

With regards to measures of energy and activity, significant differences between all the emotion categories were found. Joyful and self-confident got similar values for many performance parameters, as did the sadness and cool categories, and the measured values were overall high (e.g. large pitch span, high tone density, a constant *SL*). This can possibly be attributed as genre-specific for hip-hop, where the emotion sadness is rarely expressed musically, while anger and self-confidence are common. Arguably, energy and activity are overlapping categories, and uniquely assigning a feature to one of these two involves some ambiguity. There were significant differences between joyful and self-confident only for the activity dimension, and between joyful and angry only for the energy dimension.

The study shows both differences and commonalities between musicians in their playing strategies. Some of the general findings can be used for refining digital controllers and computational models of scratching that have emerged recently. For systems that augment the record player with haptic feedback (Beamish *et al.* , 2003) or gestural control (Lippit, 2006), taking the typical durations for preparatory gestures into account can improve both the feedback and the control; for instance by offsetting the feedback or interpreting the gesture. For automated crossfader onsets (featured on some modern mixers), the findings in this study can contribute to make such controllers more accepted by DJs than they seem now; for instance, the starting and stopping motion of the record should be muted to get sharper onsets and offsets, and durations of automated crossfader movements should be set relative to the tempo of the music. In other systems, scratch techniques have been made more accessible for novices for instance by modeling the techniques (for example in Fukuchi, 2007; Hansen & Alonso, 2008; Hayafuchi & Suzuki, 2008)). Various parameter ranges, and in particular the relationship between movement span, movement speed, and gesture density, can provide valuable input. Such systems can also sound more realistic by attuning the output to the acoustical characteristics described in the present work.

Acknowledgments

This study was supported by the Braintuning project (FP6-2004-NEST-PATH-028570).

The authors would like to thank the Vestax Corporation for equipment, and the three participating DJs.

References

Balkwill, Laura-Lee, & Thompson, William Forde. 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, **17**(1), 43–64.

Beamish, Tim, MacLean, Karon, & Fels, Sidney. 2003. Designing the Haptic Turntable for Musical Control. *In: Proc. of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'03)*. IEEE.

Bresin, Roberto, & Friberg, Anders. 2000. Emotional Coloring of Computer-Controlled Music Performances. *Computer Music Journal*, **24**(4), 44–63.

Bresin, Roberto, & Widmer, Gerhard. 2000. Production of staccato articulation in Mozart sonatas played on a grand piano. Preliminary results. *Speech Music and Hearing Quarterly Progress and Status Report*, 1–6.

DJ 1210 Jazz. 2001. *Book of Five Scratches. Book 2*. Snickars Rec., SR1206. LP.

Eerola, Tuomas, Järvinen, Topi, Louhivuori, Jukka, & Toiviainen, Petri. 2001. Statistical Features and Perceived Similarity of Folk Melodies. *Music Perception*, **18**(3), 275–296.

Fastl, Hugo, & Zwicker, Eberhard. 2007. *Psychoacoustics*. 3rd edn. Berlin Heidelberg: Springer Verlag.

Friberg, Anders, Schoonderwaldt, Erwin, & Juslin, Patrik N. 2007. CUEX: An algorithm for extracting expressive tone variables from audio recordings. *Acta Acustica united with Acustica*, **93**, 411–420.

Fukuchi, Kentaro. 2007. Multi-track Scratch Player on a Multi-touch Sensing Device. *Pages 211–218 of: Entertainment Computing – ICEC 2007*. Lecture Notes in Computer Science, vol. 4740. Berlin Heidelberg: Springer Verlag.

Hansen, Kjetil Falkenberg. 2002. The Basics of Scratching. *Journal of New Music Research*, **31**(4), 357–365.

Hansen, Kjetil Falkenberg, & Alonso, Marcos. 2008. More DJ techniques on the reactable. *Pages 207–210 of: Proc. of the Conference on New Interfaces for Musical Expression*. Genova, Italy: Infomus, Casa Paganini.

Hansen, Kjetil Falkenberg, & Bresin, Roberto. 2004. Analysis of a Genuine Scratch Performance. *Pages 519–528 of:* Camurri, Antonio, & Volpe, Gualtiero (eds), *Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop*. LNCS 2915. Berlin Heidelberg: Springer Verlag. Selected Revised Papers.

Hayafuchi, Kouki, & Suzuki, Kenji. 2008. MusicGlove: A Wearable Musical Controller for Massive Media Library. *Pages 259–262 of: Proc. of the Conference on New Interfaces for Musical Expression*. Genova, Italy: Infomus, Casa Paganini.

Iversen, John R., Patel, Aniruddh D., & Ohgushi, Kengo. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America*, **124**(4), 2263–2271.

Jungers, Melissa K., Palmer, Caroline, & Speer, Shari R. 2002. Time after time: the coordinating influence of tempo in music and speech. *Cognitive Processing*, **1**, 21–35.

Juslin, Patrik N. 2001. Communicating emotion in music performance: A review and theoretical framework. *Pages 309–337 of:* Juslin, Patrik N., &

Sloboda, John A. (eds), *Music and Emotion*. New York: Oxford University Press.

Juslin, Patrik N., & Laukka, Petri. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, **129**(5), 770–814.

Lartillot, Olivier, & Toiviainen, Petri. 2007. A Matlab toolbox for musical feature extraction from audio. *In: Proc. of the 10th Int. Conference on Digital Audio Effects*.

Lippit, Takuro Mizuta. 2006. Turntable Music in the Digital Era: Designing Alternative Tools for New Turntable Expression. *Pages 71–74 of: Proc. of the Conference on New Interfaces for Musical Expression*. Paris: IRCAM, Centre Pompidou.

Madden, John P., & Fire, Kevin M. 1997. Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency. *Journal of the Acoustical Society of America*, **102**(5), 2920–2924.

Moore, Brian C. J. 1973. Frequency difference limens for short duration tones. *Journal of the Acoustical Society of America*, **54**, 610–620.

Nábělek, Igor V., Nábělek, Anna K., & Hirsh, Ira J. 1973. Pitch of sound bursts with continuous or discontinuous change of frequency. *Journal of the Acoustical Society of America*, **53**(5), 1305–1312.

Palmer, Caroline, & Hutchins, Sean. 2006. What is musical prosody? *Pages 245–278 of:* Ross, B. H. (ed), *Psychology of Learning and Motivation*, vol. 46. Amsterdam, The Netherlands: Elsevier Press.

Repp, Bruno H. 1992. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's “TrÃ¤umerei”. *Journal of the Acoustical Society of America*, **92**(5), 2546–2568.

Russell, James A. 1980. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, **39**(6), 1161–1178.

Thompson, William Forde, & Balkwill, Laura-Lee. 2006. Decoding speech prosody in five languages. *Semiotica*, **158**, 407–424.

Toop, David. 2000. *Rap Attack #3*. 3rd edn. London: Serpent's Tail.

Vurma, Allan, & Ross, Jaan. 2007. Timbre-Induced Pitch Deviations of Musical Sounds. *Journal of Interdisciplinary Music Studies*, **1**, 33–50.