Exploring human perception of melodic boundaries in musical improvisation

Marco Fiorini
Aalborg University Copenhagen, Sound and Music Computing, 7th Semester
Music Perception and Cognition

***Abstract*—This research explores the possible existence of common cognitive mechanisms for recognizing structure in improvised music pieces. An existing model for musical events has been used and extended, defining hierarchically melodic phrasing boundaries as Temporal Gestalt-units. Eight musically trained participants took part to the experiment, listening to a piano improvisation while making continuous judgement of melodic phrasing (a measure correlated with perceived musical structure) using a sliding fader as the performance was presented. The data analysis investigated if it was possible to find similarities in the phrasing structure judgement of the piano improvisation within the participants. The results showed a strong correlation between inter-participant rating resulting trajectories, supporting the proposed hypothesis. The investigation is aligned with existing studies on human music cognition and communication, involving different modalities of information transfer and musical structure definition. It is hoped that this research can work as a foundation for future studies in the field of music cognition and perception of improvised music.**

I. INTRODUCTION

Improvised music appears to be a multidimensional concept, including technical, expressive, and social elements. It is a structured activity, requiring several specific skills.

Results produced employing the quantitative methodology adopted in [1] by Sawyer indicated that the most important ability employed during improvisation is the anticipation of musical events, a procedure that needs cognitive effort involving long-term foresight of the development of the whole solo. The experience of a flow state was also considered important, a state that motivates musicians to improvise and allows them to extensively express their abilities [2]. Toiviainen in [3] considered that one of the basic skills involved in music improvisation in many musical genres, such as jazz, blues, and rock, is the ability to identify the underlying harmonic structure.

As reported by Vines et al. in [4], "theoretical accounts of music and language posit that they both have structure, that this structure is hierarchical, and that smaller elements are grouped to form larger phrases with clear and specifiable temporal boundaries. Investigators have used an analogy between music and speech to collect judgments of musical phrase structure." In their study [5] Deliège & El Ahmadi reported segmentations marked during active listening of a contemporary music piece by depressing a key on a computer console. The segmentations of both musicians and non-musicians were in general agreement with a reference analysis provided by two composers. Clarke & Krumhansl [6] investigated listeners' perceptions of the segmentation of a piece of atonal piano music, the location of segments extracted from the piece, and the duration and structural qualities of each segment. Additionally, researchers have found that musicians' interpretations of a musical score, with respect to phrasing boundaries in particular, are revealed by adjustments in performance timing (e.g. rubato), and that such adjustments are recognized by listeners as cues for musical boundaries. Judgments of structure and phrasing in music have proven to be consistent across levels of musical skill [4].

In this research my intent was thus to extend the Temporal Gestalt model proposed by Tenney & Polansky in [7] and adopt the experimental protocol utilized by Vines et al. in [4] to investigate my hypothesis. The following model and experimental protocol are relevant to this research because augmented the knowledge of human communicative processes and explored the ways in which auditory information convey structure (as indexed by phrasing).

A. *Existing model for phrasing judgment*

A piece of music can thus be considered a hierarchically ordered network of sounds, motives, phrases, passages, sections, movements, etc.; time-spans whose perceptual boundaries are largely determined by the nature of the sounds and sound-configuration occurring within them [7]. Such time-spans (and the events or processes which define them) have been previously called *Temporal Gestalt-units* (or *TGs*) by Tenney & Polansky [7].

Nevertheless, the model proposed in [7] and investigated through experimentation in [4] have some limitations in terms of the kind of music it can deal with (it can only work with *monophonic* music), as well as the musical factors which it consider. The model does not even consider *harmony* (or harmonic relations between pitches or pitch-classes) and *shape* (pattern, motivic/thematic relations) of the melodic phrases. In their study [8] Bruderer et al. suggested that people tend to segment a piece in much the same way whether they are given just the melody or a full polyphonic texture. A plausible hypothesis about polyphonic grouping, then, is that listeners first extract the melody, segment it, and then impose that grouping on the full texture [9]. Most models have also been somewhat limited in that they have addressed only a single low level of grouping (roughly speaking, the level of the phrase). As noted earlier, grouping is usually assumed to be hierarchical, with smaller units combining into larger ones; large sections of an extended piece might be several minutes in length. Listening to polyphonic music also involves grouping notes into lines or voices. This, too, presents a challenging problem, one known by various names such as voice separation, stream segregation, and contrapuntal analysis.

B. Problem Analysis

Based on these previous researches, I decided then to focus on the gesture of the right hand of a piano improvisation, in temporal and structural relation to the accompanying left hand. Expanding the presented model of TGs, the terms presented here are *motif* and *period*. A motif relates to a short chunk of right hand melody, delimited by short temporal and interval boundaries. Two or more motifs, interleaved by short temporal units of silence in the right hand, form a period, considered as a larger logical segment.

Using these terms and the proposed expanded model, one can assume it is possible to follow the melodic arch of a phrase and its development over time, in terms of minor (motifs) and major phrase boundaries (periods). The selected melodic phrase belongs to the right hand of a piano improvisation, whereas the left hand acts as an harmonic and timing context. It was hoped that this would lead to a common perceived evaluation in the structure of the improvisation, in terms of melodic phrasing of the right hand in relation to the left hand.

II. Experimental Protocol

A. Stimulus

The stimulus consisted in the first 56 seconds of *Part I* of Keith Jarrett's *The Köln Concert*, as transcribed in Figure 4 (in Appendix) and available at [1]. Besides the shared musical relevance described in Appendix A, the excerpt was chosen because it is considered one of the most influential recordings of piano solo and piano improvisation ever, with sales of more than 3.5 million copies [10]. The phrase boundaries displayed as vertical lines in Figure 2 and Figure 3, determining the duration of detected motifs and periods in the piece over time, were annotated by a professional musician (A. del Sordo, personal communication, December 2021).

B. Hypothesis

Based on the previous assumptions and on the author's experience and background in performing improvised non-idiomatic music, the goal of this study was to investigate if people could recognize and follow the melodic structure development in an improvised piece of music. Participants rated melodic motifs and periods in a selected piece of improvised music based on continuous phrasing judgement, generating rating curves that represented the perceptual organization of the development of a melodic structure over time.

Hypothesis: *There will be similarities in the judgement evaluation curves, showing the possible existence of common cognitive mechanisms for recognizing structure in improvised music pieces.*

Two pilot tests were executed before the experiments were conducted, to prove the correct formulation of the hypothesis and avoid any misunderstanding of the task.

C. Participants

8 people (6 m, 2 f, aged 26-36, 28.5 $\pm$ 3.3 y/o) participated in the experiment on different days. All participants were asked if they could understand and recognize the use of two hands in a piano performance. None of the participants were compensated for their participation.

D. Experimental Setup

For all participants, the rating was captured using a continuously adjustable linear slider on a Korg Nanokontrol (Korg Inc. Tokyo, Japan) while the performance was presented. The location of the slider was sampled once every 100 ms by a computer program in the Max/MSP programming environment [11] as shown in Figure 1.

E. Procedure

The participants were asked to wear headphones for the test. Participants read the following on-screen instructions before performing the task:

You are going to hear a piano improvisation. In the recording, the left hand acts as a context, accompanying and defining the timing for the phrasing of the right hand. Use the full range of the slider to express the phrasing arch you experience in the pianist's right hand, in terms of MOTIFS (short temporal and melodic chunks) and PERIODS (larger chunks, made of more motifs). Begin with the slider all the way down. Move the slider upward as a period is entered and all the way downward as a period is exited. During the melodic arch of a motif, the slider should be dynamically moving between the top (peak of the Period) and the bottom (end of the Period). Use the full range of the slider.

Participants became familiar with the task and apparatus during a separate practice round.

[1] https://open.spotify.com/track/0T4KV1pj8as2xvdHZAP5ae?si=7c3bb53798744e00

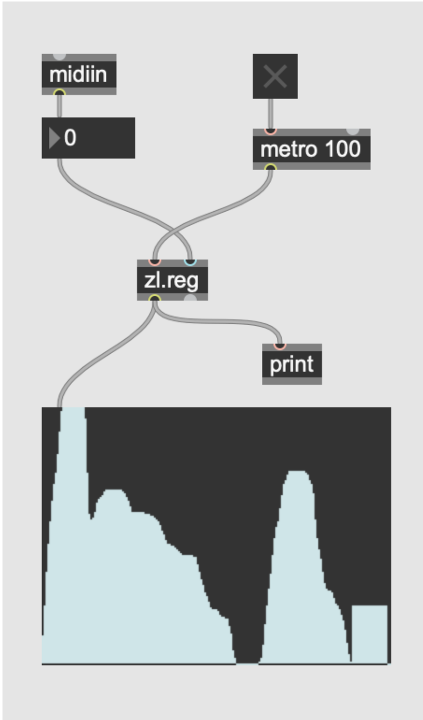

Fig. 1. Max/MSP patcher for fader data sampling. The midiin object listens to a specific MIDI port and outputs the raw MIDI data received, in a [0, 127] range. The zl.reg object functions as a register that holds a list. A list received in the right inlet is stored, whereas a bang received in the left inlet sends the stored list out the left outlet. Thanks to the object metro, a bang is sent to zl.reg every 100 ms (10 Hz). The values of the list are then printed in the Max console through the object print and drawn as a function.

F. Data Analysis

The data recorded in Max/MSP were then stored in a CSV file and analysed using MATLAB. Here, the data was truncated to a length corresponding to 56 seconds. Each time series was then smoothed using a second order Butterworth IIR Lowpass filter with a cutoff frequency of 0.75 Hz. The data values, originally ranged in MIDI format [0, 127] were normalized in a [0,1] range. The curves of all individual participants were averaged to one mean curve. Furthermore, Pearson correlation values were obtained, measuring the degree of association between every participant, as well as mean and standard deviation.

III. Results

A. Phrasing

This structure analysis focused upon temporal coincidence (or lack thereof) in judgments, as well as similarities in contour across the performance. Upon visual inspection the mean curve shown in Figure 2 aligns with lines delineating minor and major phrase boundaries (as well as in Figure 3). The low and high peaks in the mean curve were aligned across the whole performance duration. The phrasing judgments revealed what points in time participants perceived phrase boundaries in the music as well as their sense of phrasing contour, as shown in Figure 3 in Appendix A.

B. Correlation Analysis

Table I shows the inter-participant Pearson correlation coefficients, with a mean $\pm$ STD of 0.48 ± 0.15.

TABLE I
Correlation matrix, showing only the upper value components. The values are all positive, ranging from 0.23 to 0.72

	P1	P2	P3	P4	P5	P6	P7	P8
P1	1	0.51	0.60	0.56	0.62	0.72	0.27	0.49
P2	–	1	0.36	0.24	0.54	0.48	0.24	0.23
P3	–	–	1	0.47	0.46	0.57	0.26	0.39
P4	–	–	–	1	0.47	0.71	0.36	0.62
P5	–	–	–	–	1	0.66	0.53	0.65
P6	–	–	–	–	–	1	0.29	0.61
P7	–	–	–	–	–	–	1	0.55
P8	–	–	–	–	–	–	–	1

IV. Discussion

The goal of the research described in this paper was to evaluate the rating of phrasing judgment for a particular piece of music improvisation, revealing the possible existence of common cognitive mechanisms for recognizing structure in improvised music pieces. In this research, phrasing rating judgments of a particular piece of improvised piano music were collected.

A. Discussion of results

Though the magnitudes of the individual curves differed, all of them followed a contour that lead to a mean curve aligning within all the phrasing boundaries over the course of the measurement, as shown in Figure 2. Each major phrase boundary occurred with, or was followed shortly by dips in the phrasing fitted mean. On average, all participants followed the high-level period structure annotated by the professional musician. Note also the agreement for registering motif boundaries, as evidenced by synchronous low peaks within the major phrases. The only exception is noticeable for the first motif, where the dip appears slightly later than the others, and the high peak is lower in magnitude. This can be a consequence of the right hand melody starting almost immediately in the recording, and then influencing the judgement of the participants in terms of expectation and reaction time, as well as overall confidence with the proposed task. Furthermore, the correlation values displayed in Table I and the mean presented there and in Section III-B indicate a strong correlation within all the participants' rating curves. By applying the proposed model, described in [7] and expanded in I-A, the participants rated the structure development of this

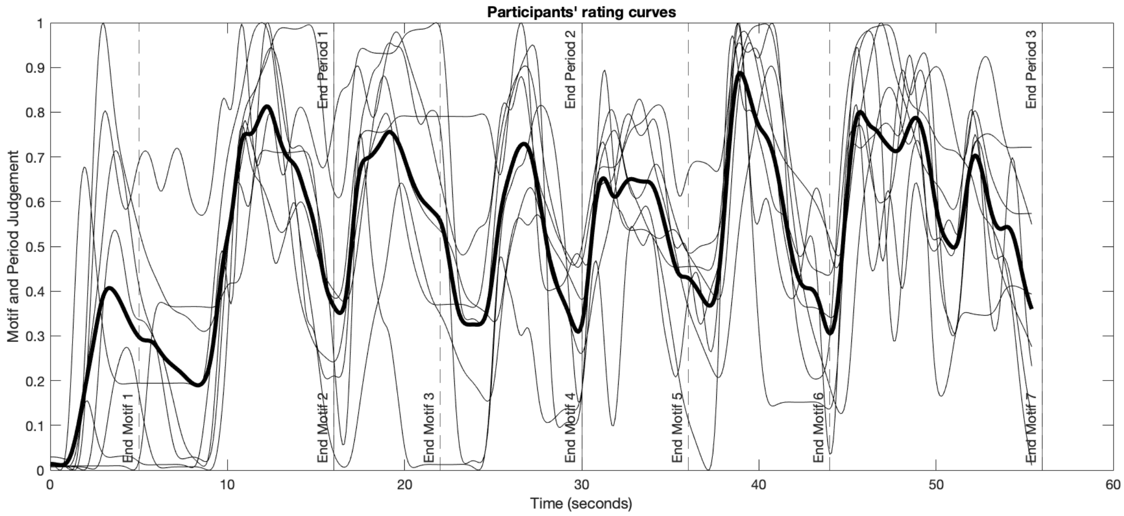

Fig. 2. Participants' rating curves. Each line represents the phrasing judgment of a single participant. Thicker line represents the mean for all participants' rating. Individual lines are also presented in Figure 3 in Appendix A. Vertical lines delineate Motifs and Periods in the piece as previously defined in Section II-E. X axis shows time in seconds, Y axis represents scaled values from 0 to 1.

improvised piece of music according to the model, as shown by the data reported in Figure 2, Table I and Section IV-A. The results observed in this research are aligned with existing theories proposed in previous studies [4] [5] [6] [7], fulfilling the proposed hypothesis and thus supporting the possible existence of common cognitive mechanisms for recognizing structure in improvised music.

B. Limitations

In evaluating other relevant aspects of this research, the experiments were not performed in a completely controlled environment. Consequently, this could have influenced the participants' confidence with the proposed task. A few participants pointed out that the slider used on the Nanokontrol was quite small and therefore they had problems in managing the range over the measurement. I used only one piece of music and this led to limited results. Thus, if more pieces were used, findings could be generalized to a wider range. The experiment was also short and there were a small sample of participants, all of them with a background in music or interested in sound and music computing. Therefore, a future work might focus on choosing a different kind of participants, focusing only on non-musicians. Another interesting development of the project could consider a richer polyphonic texture, involving more instruments and focusing on deepening the TGs model here presented into higher levels of structure abstraction and definition.

V. Conclusions

The present study on musical perception contributes to a larger context of studies of cognition, ranging from auditory streams to structure recognition.

It is hoped that the research done during this study can work as a foundation for future investigations in the field of music cognition and perception of improvised music.

References

[1] R. K. Sawyer, "Improvised conversations: Music, collaboration, and development:," *https://doi.org/10.1177/0305735699272009*, vol. 27, pp. 192–205, 6 2016.

[2] M. Biasutti and L. Frezza, "Dimensions of music improvisation," *Creativity Research Journal*, vol. 21, pp. 232–242, 2009.

[3] P. Toiviainen, "Real-time recognition of improvisations with adaptive oscillators and a recursive bayesian classifier," *International Journal of Phytoremediation*, vol. 21, pp. 137–147, 2001.

[4] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin, "Cross-modal interactions in the perception of musical performance," *Cognition*, vol. 101, pp. 80–113, 8 2006.

[5] I. Deliege and A. el Ahnmadi, "Mechanisms of cue extraction in musical groupings: A study of perception on sequenza vi for viola solo by luciano berio:," *https://doi.org/10.1177/0305735690181003*, vol. 18, pp. 18–44, 6 2016. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0305735690181003

[6] E. F. Clarke and C. L. Krumhansl, "Music perception perceiving musical time," *the Spring*, vol. 7, pp. 213–252, 1990.

[7] J. Tenney and L. Polansky, "Temporal gestalt perception in music," *undefined*, vol. 24, p. 205, 23 1980.

[8] M. J. Bruderer, M. F. McKinney, and A. Kohlrausch, "The perception of structural boundaries in melody lines of western popular music," 2009. [Online]. Available: https://jyx.jyu.fi/handle/123456789/22611

[9] D. Temperley, "Computational models of music cognition," *The Psychology of Music*, pp. 327–368, 2013. [Online]. Available: /record/2012-14631-008

[10] C. da Fonseca-Wollheim, "A jazz night to remember - wsj," 2008. [Online]. Available: https://www.wsj.com/articles/SB122367103134923957

[11] "Max/MSP, Cycling '74." [Online]. Available: https://cycling74.com/

[12] K. Jarrett, "The Köln Concert," 1975. [Online]. Available: https://www.ecmrecords.com/shop/143038750696/the-koln-concert-keith-jarrett

[13] "The Köln Concert Original Transcription Piano." [Online]. Available: https://en.schott-music.com/shop/the-koeln-concert-no38222.html

Appendix

A. Background Information on Stimulus

Recorded in 1975 at the Köln Opera House and released the same year, nothing on this program was considered before he sat down to play. All of the gestures, intricate droning harmonies, skittering and shimmering melodic lines, and whoops and sighs from the man are spontaneous. There is a consensus among critics, listeners and musicians alike that this, Keith Jarrett's second solo piano concert recording (after *Solo-Concerts Bremen/Lausanne*), is a milestone not only of Keith Jarrett's work, but of the entire history of jazz. [12]

The concert was organized by 17-year-old Vera Brandes, then Germany's youngest concert promoter. At Jarrett's request, Brandes had selected a *Bösendorfer 290 Imperial* concert grand piano for the performance. However, there was some confusion by the opera house staff and instead they found another *Bösendorfer* piano backstage – a much smaller baby grand piano – and, assuming it was the one requested, placed it on the stage. The error was discovered too late for the correct *Bösendorfer* to be delivered to the venue in time for the evening's concert. The piano they had was intended for rehearsals only and was in poor condition and required several hours of tuning and adjustment to make it playable. The instrument was tinny and thin in the upper registers and weak in the bass register, and the pedals did not work properly. Consequently, Jarrett often used ostinatos and rolling left-hand rhythmic figures during his Köln performance to give the effect of stronger bass notes, and concentrated his playing in the middle portion of the keyboard. *ECM Records* producer Manfred Eicher later said: "Probably [Jarrett] played it the way he did because it was not a good piano. Because he could not fall in love with the sound of it, he found another way to get the most out of it." [10]

A notable aspect of the concert was Jarrett's ability to produce very extensive improvised material over a vamp of one or two chords for prolonged periods of time. For instance, in Part I, he spends almost 12 minutes vamping over the chords Am7 to G major, sometimes in a slow, rubato feel, and other times in a bluesy, gospel rock feel.

B. Individual participants' rating curves

A visual inspection of the individual rating curves presented in Figure 3 shows that certain participants, like number 1, 2, 5 and 6 had similar results, with wide and clear phrase curves. Other participants, like number 4 and 8 moved more in steps rather than curves, keeping certain phrases almost at the same Y value for a longer time-span. Furthermore, participants like number 3 and 7 performed a series of narrow and fast changes, resulting in edgy rating curves with more noise. A further exploration of the proposed model and a more specific study

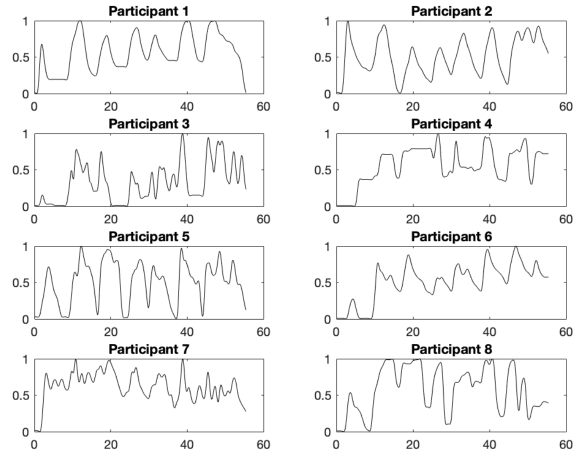

Fig. 3. Individual participants' rating curves. Vertical lines delineate Motifs and Periods in the piece as previously defined in Section II-E. X axis shows time in seconds, Y axis represents scaled values from 0 to 1.

of TGs could lead to a better way to define such TGs and present them to other participants (hopefully non-musicians) in order to collect a wider range of results.

THE KÖLN CONCERT

Part I

Keith Jarrett

Fig. 4. Piano transcription of the excerpt used in this study, from beginning to fourth bar of second page [13]


Ped.