Technische Universität Berlin
Fakultät I - Geisteswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

‹ Masterarbeit ›

Prediction Of Continuous Emotional Responses To Music Using Hierarchical Linear Models

Wissenschaftliche Arbeit zur Erlangung des akademischen Grades

Master of Science (M.Sc.)

im Studiengang

Audiokommunikation und -technologie

Vorgelegt von:
Christoph Graefe
Matr.-Nr. 215891
✉ christoph(punkt)graefe(at)gmx(punkt)de
☎ +49 (179) 750 473 1

Gutachter:
Dr. Phil. Hauke Egermann
Prof. Dr. Stefan Weinzierl

Abgabedatum:
04.09.2014

Eidesstattliche Erklärung

Ich versichere,

- dass ich die Masterarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe.
- dass ich dieses Masterarbeitsthema bisher weder im In- noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe.
- dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Berlin, 04.09.2014

....

Ort, Datum Unterschrift

Acknowledgement

I owe sincere thanks to Hauke Egermann for the great support. I thank my family for support and encouragement. I thank Annika for patience and love. I dedicate this work to my father Albert Graefe.

Abstract

Previous research indicates that the prediction of perceived emotion while listening to music is possible based on linear regression models and few interval-scaled feature time series as predictors. This work presents the secondary analysis of a dataset collected during a live concert experiment with solo flute performance using a new type of autoregressive hierarchical linear modelling. It could be demonstrated that the time series from continuous ratings of subjective components (*valence, arousal*) and the continuous measurements of peripheral arousal (*skin conductance*) are predictable by a group of the following predictors: (a) four low level features (*RMS, brightness, roughness, spectral centroid*), (b) two mid level features (*tempo, melodic contour*) and (c) two high level features (subjective *unexpectedness* ratings (group means from the measurements), *information content* (calculated by computational model providing an estimation for the pitch of each note in the piece based on the previous pitches)). The results are expected to contribute to a deeper exploration of the mechanisms underlying musical emotions. They could also be interesting for purposes of music therapy, help raising *musical fit* in audio branding and increase the efficiency of music for advertising aims.

Zusammenfassung

Bisherige Forschung konnte zeigen, dass wahrgenommene Emotionen während der Musikrezeption mit Hilfe linearer Regressionsmodelle und weniger intervallskalierter Zeitreihen musikalischer features als Prädiktoren vorhersagbar sind. Diese Arbeit präsentiert die Sekundäranalyse eines Datensatzes, der während eines Konzertexperiments mit live dargebotener Soloflöte erhoben wurde. Die Analyse erfolgte mit Hilfe eines neuartigen Ansatzes autoregressiver hierarchischer linearer Modelle. Es konnte gezeigt werden, dass Zeitreihen kontinuierlicher Bewertungen subjektiver Komponenten (*valence, arousal*) sowie kontinuierlicher Messungen peripherer Erregung (*skin conductance*) mit einer Gruppe folgender Prädiktoren vorhersagbar sind: (a) vier low level features (*RMS, brightness, roughness, spectral centroid*), (b) zwei mid level features (*tempo, melodic contour*) und (c) zwei high level features (Bewertungen subjektiver *Unerwartetheit* (Gruppenmittelwerte aus den Messungen), *Informationsgehalt* (berechnet mit Computermodell, welches eine Schätzung für die Tonhöhe jeder Note eines Stücks basierend auf den Tonhöhen der vorhergehenden Noten ausgibt)). Die Ergebnisse können zur weiteren Erforschung der Mechanismen beitragen, die musikalischen Emotionen zugrunde liegen. Auch können sie für Zwecke der Musiktherapie von Interesse sein und dabei helfen, das *musical fit* beim audio branding zu steigern und die Effizienz von Musik in der Werbung zu erhöhen.

Contents

Acronyms

2DES two-dimensional emotion space 11, 12, 31, 80

ACF autocorrelation function 17, 55, 56, 82
AIC Akaike's information criterion 49
AICC Hurvich and Tsai's criterion 49
ANS autonomic nervous system 5, 12
AR1 first-order autoregressive covariance structure with homogeneous variances 48–51
ARH1 first-order autoregressive covariance structure with heterogeneous variances 49–51, 55
ARMA11 first-order autoregressive covariance structure with first-order moving average 49

BIC Schwarz's Bayesian criterion 49
BPM beats per minute 24
BRECVEM framework of seven psychological mechanisms 6, 8, 70, 74, 80, 81, 84

CAIC Bozdogan's criterion 49
CNS central nervous system 5

dV dependent variable 6, 15, 31–35, 42–44, 46–48, 52, 55, 57, 70, 71, 79, 80, 82, 84
dV_ARO arousal 44, 47, 49, 50, 52, 54–61, 64, 65, 67, 68, 70, 71, 74, 75, 78, 79
dV_SCR skin conductance response 45, 47, 50, 52, 54–56, 58–60, 62, 64–71, 73–75, 78, 79
dV_VAL valence 44, 47, 50, 52, 54–61, 64, 66–71, 74, 78, 79

EDA electrodermal activity 13, 14

FT Fourier transform 20

HL_ICH high information content 17, 24, 39–41, 45, 62, 82
HL_ICL low information content 17, 24, 39–41, 45, 82
HL_UEH high unexpectedness 17, 25, 40, 41, 44, 45, 82
HL_UEL low unexpectedness 17, 25, 40, 41, 44, 45, 82
HLM hierarchical linear model 15, 17, 41, 44–46, 48–50, 52, 57–61, 65–68, 79, 84
HR heart rate 7, 80

IC information content 2, 17, 25, 28, 29, 39, 40, 44, 64, 69, 78, 82
IFT inverse Fourier transform 20
IOI inter-onset interval 25
ITPRA Imagination–Tension–Prediction–-Reaction–Appraisal theory 10, 82
iV independent variable 15

LL_BRN brightness 16, 21, 37, 45, 72, 82
LL_RGN roughness 16, 22, 38, 45, 70, 72, 82
LL_RMS RMS 16, 21, 37, 45, 70, 72, 82
LL_SPC spectral centroid 16, 20, 21, 37, 45, 72, 82
LMM linear mixed model 46, 48, 50, 51
LTM long-term model 25

ML_MEL melodic contour 17, 23, 26, 38, 40, 45, 72, 82
ML_TMP tempo 17, 23, 26, 39, 45, 82

NaN not a number 32, 33, 42
NES neuro-endocrine system 5

PACF partial autocorrelation function 55, 56, 82

PANAS positive and negative affect schedule 12
PANAS-X expanded positive and negative affect schedule 12
PC principal component 46, 47, 49, 70, 71
PC_BRN brightness 47, 64, 67, 71–74, 76, 77
PC_ICH high information content 47, 52, 57, 64, 67, 71, 74, 77–79
PC_ICL low information content 47, 52, 59, 61, 62, 64, 69, 71, 73, 74, 77–79
PC_MEL melodic contour 47, 69, 71, 73, 74, 77
PC_RGN roughness 47
PC_RGN_BRN roughness, brightness 47
PC_RGN_BRN_RMS roughness, brightness, RMS 47, 52, 57, 61, 62, 71–74, 76, 77
PC_RGN_RMS roughness, RMS 47, 59, 60, 71–74, 77
PC_SPC spectral centroid 47, 52, 61, 62, 71, 73, 74, 77
PC_SPC_BRN spectral centroid, brightness 47, 60, 71, 73, 74
PC_SPC_MEL spectral centroid, melodic contour 47, 64, 67, 70, 71, 73, 74, 76, 77
PC_SPC_RMS spectral centroid, RMS 47, 69, 71, 73, 74, 77
PC_TMP tempo 47, 61, 64, 69, 71, 74–77
PC_TMP_MEL tempo, melodic contour 47, 52, 57, 59, 60, 71, 73–77
PC_UEH high unexpectedness 47, 52, 57, 59–61, 64, 71, 74, 77–79
PC_UEL low unexpectedness 47, 52, 57, 59–61, 64, 71, 74, 77–79
PCA principal component analysis 26, 45–47, 70

REML residual maximum likelihood estimation 50
RespR respiration rate 7
RMS root mean square 17, 28, 37

SC skin conductance 1, 4, 6, 13, 33, 73, 77
SCR skin conductance response 2, 33, 34, 42, 59, 62, 75, 82
SNS sympathetic nervous system 14
SoNS somatic nervous system 5
SSN stimulus sample number 47, 52
STFT short time Fourier transform 18, 20
STM short-term model 25

TP Toeplitz 49, 51
TPH heterogen Toeplitz 49, 51

UE unexpectedness 2, 17, 25, 28, 29, 40
UN unstructured 51

VC variance components 51

1 Introduction

The power of music to wake human's emotions is an old and widely researched field. Evolutionary development of ancient psychological mechanisms are essential to human survival on the one hand and cultural learning on the other, both contributing to the development of musical emotions that have been the subject of numerous studies. Interdisciplinary linked research has been conducted in musicology and psychology as well as in neuroscience, neurobiology and physiology. As Darwin already cites in 1871, "We can concentrate [. . .] greater intensity of feeling in a single musical note than in pages of writing." (Darwin, 1871, p. 335-336). As this statement implies impressively, the musical content of one single note may indwell such high density of musical information and complexity that research on the mechanisms of musical emotions and the link between emotional responses and musical structure remains widely fruitless (cf. Aramaki et al., 2012, p. V). Nevertheless, research is continued and the aim of this thesis is to contribute here expanding an approach of Schubert (2004) and enriching a recent analysis conducted by Egermann et al. (2013).

Schubert (2004) could show in an experiment with 67 listeners and four pieces of romantic music that the prediction of perceived musical emotion[1] (valence and arousal) is possible based on linear regression models and time series of few musical features as predictors. These predictors were loudness, tempo, melodic contour, texture and spectral centroid. The univariate linear regression models could explain between 33% and 73% of variation in the valence and arousal time series. The relevant findings were that (a) loudness and tempo changes had a positive impact on arousal changes, whereas the impact of loudness was stronger, and that (b) the overall emotional response time of test persons was between 1 and 3 seconds, except for sudden changes in loudness where this time could even decrease down to 0 seconds.

Egermann et al. (2013) conducted a live concert experiment with solo flute music and measured subjective components (valence and arousal) as well as different peripheral psychophysiological effects continuously (here, only the change of skin conductance (SC) is relevant). Moreover, they followed two approaches

[1] Here, the *cognitivist* position has been tested. See section 1.1.1 for explanation and distinction towards *emotivist* position.

to analyse musical expectations: on the one hand, a computational simulation was used to calculate the information content (IC) for each note in the melody (divided into high vs. low IC data) and on the other hand, human continuous ratings of unexpectedness (UE) (also divided into expected vs. unexpected data). The results relevant for this work are that

- the onset of high IC events is followed by higher UE ratings,
- arousal ratings increased at moments of high IC and at unexpected moments,
- arousal ratings decreased at expected moments,
- valence ratings decreased at moments of high IC,
- skin conductance response (SCR) increased at moments of high IC,
- SCR increased and decreased at both unexpected and expected events,
- valence showed no relevant changes with regard to expected/unexpected peak moments.

The research questions arising from these preliminary considerations and the summarised study results were: Are continuous ratings of *arousal* and *valence* and continuous measurements of *skin conductance* predictable by different musical features and furthermore by IC from a computer simulation and UE from continuous listener ratings? As well as, is it possible to predict these ratings/measurements *independently* with musical features on the one hand and IC/UE on the other hand? To answer these questions, a secondary analysis of the data gathered during the live concert experiment for the study of Egermann et al. (2013) was carried out.

To prevent confusion, a first definition of terms has to be undertaken at this point. In literature, numerous ambiguous designations for the field of features are found, ranging from *audio features*, *sound features*, *instantaneous features*, *psychoacoustic cues* etc. For this purpose, a clear definition of terms shall be given for the further progress of this work: the term *low level features* will always refer to the musical features *spectral centroid*, *RMS* (representing loudness), *brightness* and *roughness* whereas the term *mid level features* will denote *melodic contour* and *tempo*. IC and UE will be denoted as *high level features*. To refer to all three and the field of features in general, the terms *feature* or *musical feature* will be used as collective terms. This work is divided as follows: an introduction to the subject of music and emotion will be given in section 1.1. All musical features used as predictors will be introduced in section 1.2. The method section includes a description of the live concert experiment, the measurements and feature extractions in section 2.1 as well as a presentation of the statistical analysis process in section 2.2. The results will be presented in section 3 and the work will finally be completed by a discussion in section 4.

1.1 Music and Emotion

The establishment of *music and emotion* as independent field of research was not easy. The formation of psychology itself as a new emerging field at the end of the nineteenth century can be considered as a starting point for this process. Parallel to the development of general psychology, the first experimental and descriptive studies in the new area of music and emotion were published, usually based on western classical music. This new subdivision had to emancipate itself from the main area of music psychology with its early pioneering studies (e.g. Hevner, 1935) mainly concentrating on psychophysical and perceptual processes. Juslin and Sloboda (2012, p. 584-586) mention the books of Leonard Meyer (*Emotion and Meaning in Music*), Daniel Berlyne (*Conflict, Arousal and Curiosity*) and Manfred Clynes (*Sentics: The Touch of Emotions*) as key works of their respective time of origin, but conclude that none of these works could seriously support the field of music and emotion to gain in popularity. It was only in the early 1990s that the emancipation was successful, particularly due to the fact that studies of music and emotion were freed from the restrictions to explore cognitive processes in closed laboratory contexts. These limitations were lifted to study music and emotions in an extended everyday context. However, new approaches like neuroimaging studies and performance studies contributed to reaching its peak development with the publication of Patrick N. Juslin and John A. Sloboda (*Music and emotion: Theory and research*) that is completely dedicated to the field of music and emotion. As noted by Juslin and Sloboda (2012), views on music and emotions vary these days: many musicians are introverted and do not like to talk about their musical emotions, and a lot of listeners fear the musics' spirituality to be destroyed by talking about it. Furthermore, critics may decry the misuse of music for manipulative purposes in advertising or an omnipresence of music by medial distribution, often driven by commercial interests with a supersaturation coming along. That is why authors propose future research to target "the secrets of music and emotion, as a counterweight to commercial interests" (p. 583) and to aim at the uses for social well-being, mental wellness and physical health.

To go deeper into the topic of *emotion*, the next section will first present current definitions and models (1.1.1) and then associate the emotional theory to the emotional effects relevant here, valence and arousal (1.1.2) as well as SC (1.1.3).

1.1.1 Definitions and Models

Numerous attempts have been made to classify emotions and to determine their structure, but there is neither a generally accepted definition of emotion nor an exhaustive classification of different emotions (Bellebaum et al., 2012, p. 127). As a subcategory of the collective term *affect*[2], emotions are inter alia grouped together with *moods*, *preferences* and *personality dispositions*. Emotions are characterised to "[. . .] involve intentional objects: they are 'about' something" (Juslin et al., 2010, p. 619) that either helps or disturbs an organism to achieve a specific evolutionary important goal and to be more intense than moods. However, moods are affective states with lower intensity, have much longer half-life periods than emotions and are not clearly object-oriented whereas a preference is a long-lasting state of constant evaluation of an object (Juslin and Västfjäll, 2008, p. 561). The *component process model of affective states* by Scherer (1987) is one of the best known emotion models and also provided the basis for the analyses of Egermann et al. (2013). Following this model, an organism experiences an *emotion episode* as a "response to the evaluation of an external or internal stimulus event" (Scherer, 1987, p. 7) that (the event) has a special importance for the organism. The emotion episode consists of synchronised and mutually linked status changes in the five subsystems listed in table 1 with its respective processing areas and emotional functions (see also Scherer, 2004). Closely following the component process model and on the basis of Izard (2009), Juslin and Sloboda (2012) also give a current working definition of emotions adding the social aspect of event triggers:

[2] Frijda and Scherer (2009) describe the term *affect* as "a mental state that is characterized by emotional feeling as compared with rational thinking" and as "being touched or moved by an event to a greater extent than what a normal perception or thought would entail" (p. 10).

Emotion function	**Organismic subsystem and major substrata**	**Emotion component**
Evaluation of objects and events	Information processing (CNS)	Cognitive component (appraisal)
System regulation	Support (CNS, NES, ANS)	Neurophysiological component (bodily symptoms)
Preparation and direction of action	Executive (CNS)	Motivational component (action tendencies)
Communication of reaction and behavioral intention	Action (SoNS)	Motor expression component (facial and vocal expression)
Monitoring of internal state and organism-environment interaction	Monitor (CNS)	Subjective feeling component (emotional experience)

Table 1

Relationships between organismic subsystems and the functions and components of emotion (from Scherer, 2005, p. 698) (central nervous system (CNS), neuro-endocrine system (NES), autonomic nervous system (ANS), somatic nervous system (SoNS)).

> Emotions are relatively brief, intense and rapidly changing responses to potentially important events (subjective challenges or opportunities) in the external or internal environment, usually of a social nature, which involve a number of subcomponents (cognitive changes, subjective feelings, expressive behaviour, and action tendencies) that are more or less ’synchronised’ during an emotion episode. (p. 587)

Another definition from Juslin and Västfjäll (2008) describes the subcomponents slightly differently and with examples, namely as:

> (a) cognitive appraisal (e.g., you appraise the situation as "dangerous"), (b) subjective feeling (e.g., you feel afraid), (c) physiological arousal (e.g., your heart starts to beat faster), (d) expression (e.g., you scream), (e) action tendency (e.g., you run away), and (f) regulation (e.g., you try to calm yourself) [. . .]. (p. 562)

It becomes possible to measure emotions by measuring the subcomponents from these approaches and, concerning musical emotions, it is suggested that if responses to music appear in all these subcomponents, music is supposed to be able to evoke real emotions. Juslin and Västfjäll (2008, p. 562, tab. 2) prove this by listing corresponding studies. The neurophysiological component SC and the two subjective feeling components valence and arousal have especially been captured and used as dependent variables (dVs) in the analyses presented here.

To explain these emotional reactions, some of the psychological mechanisms described in Juslin et al. (2010)[3] will be taken into account. The authors propose a framework of seven psychological mechanisms (BRECVEM) that are interdisciplinary and not limited to music to analyse the emotion induction process. These mechanisms may be "activated by taking music as their 'object'" (p. 619), and more precisely, it is assumed that these mechanisms are triggered by special musical features. Especially the lower order mechanisms *brain stem reflex*, *rhythmic entrainment*, *emotional contagion* and the higher order mechanism *musical expectancy* are of interest and will be further explained in the following (cf. tab. 2 for an overview of the mechanisms). The mechanisms *visual imagery*, *episodic memory* and *evaluative conditioning* are not supposed to help explain the results from the upcoming analyses because the musical features used and explained later are not assumed to be related to them.

Brain stem reflex The brain stem is described as an ancient brain structure that influences many motor and sensory functions as well as the auditory system. It may work upon instructions from the amygdala, the hypothalamus and the orbitofrontal cortex, but it is also able to react reflex-like while bypassing the prior analyses of the brain regions mentioned. This is especially the case when an audio input is interpreted to warn from dangers, namely if it is loud, appears suddenly, is fast in tempo or dissonant (Juslin and Västfjäll, 2008, p. 564). Levitin (2006) remarks that

> [⋯] the brain stem and the dorsal cochlear nucleus - structures that are so primitive that all vertebrates have them - can distinguish between consonance and dissonance; this distinction happens before the higher level, human brain region - the cortex - gets involved. (p. 72).

The very low level characteristics of the brain stem reflex shall be underlined by a short excursus to neuroanatomical signal flow: every single acoustic frequency

[3] Revised version of Juslin and Västfjäll (2008).

(1. neurons) is represented in a particular fiber of the nucleus cochlearis that transmits the signals from the hair cells of the organ of Corti to the brain stem. The nucleus cochlearis anterior that the signal is switched to (2. neurons) is the route where the very first basic signal analysis takes place. For instance, this analysis targets on beginnings and endings of signals or changes in frequency. Before being switched further (3. and 4. neurons) for higher level processing, directional hearing and localisation of sources is still performed at a low level, namely at the upper nucleus olivaris (cf. Garzorz, 2009, p. 105). Juslin et al. (2010) link this mechanism to general arousal and unpleasantness vs. pleasantness (see tab. 2).

Rhythmic entrainment Basically, rhythmic entrainment denotes the process of two independently working oscillators (or rhythmical units) that interact with or even synchronise to each other. For the live concert experiment in particular, this means that internal physiological rhythms of the listeners like heart rate (HR) and/or respiration rate (RespR) synchronise to the musical rhythm of the performance. If one of those internal physiological rhythms is synchronised, it may be transferred to other emotion-related regions of the body via proprioceptive feedback (Scherer, 2004, p. 7) and thus lead to increased arousal. Humans and animals seem to have a disposition for entrainment. In this context, it is presumed that brain neurons have a key function in the perception of metre and timing (cf. Juslin et al., 2010, p. 621). For further remarkable details to this oscillator theory, see Large (2002) who developed a computational model to simulate a system of listeners' internal oscillators that synchronise to the musical rhythm. This model could track tempo successfully and supported the thesis that listeners perceive regular musical tempo based on temporal expectancies that adapt during an interaction with an input that varies in tempo. Juslin et al. (2010) link this mechanism to general arousal and pleasant feelings of communion.

Emotional contagion First mentioned by Scheler in 1913, the term denotes a process during which one person imitates the emotional expression of somebody else internally. First studied in conjunction with facial expression, the most important aspect of emotional contagion concerning music is emotional speech. It is suggested that if we listen to music, a brain area that actually is responsible for the processing of emotional speech is triggered because music uses similar emotion-

Mechanism	Induced affect	Induction speed	Degree of volitional influence
Brain stem reflex	General arousal, unpleasantness vs. pleasantness	High	Low
Rhythmic entrainment	General arousal, pleasant feelings of communion	Low	Low
Evaluative conditioning	Basic emotions	High	Low
Emotional **C**ontagion	Basic emotions	High	Low
Visual imagery	All possible emotions	Low	High
Episodic memory	All possible emotions, although especially nostalgia	Low	Medium
Musical expectancy	Surprise, awe, pleasure, 'thrills', disappointment, hope, anxiety	Low	Low

Table 2
Mechanisms of the BRECVEM framework (adapted from Juslin et al. (2010, p. 626), shortened/modified). See also Juslin and Västfjäll (2008), p.570 for an extended table with references for all items.

ally relevant acoustic patterns as the human voice. Juslin and Laukka (2003) have worked out these similarities and propose that "listeners can become 'moved' by music performances through a process of emotional contagion" (p. 803). This may be illustrated by the picture of music inducing emotion being 'emotional speech in another dress'. The reason for this mechanism to work is suggested to be mirror neurons: Di Pellegrino et al. (1992) could show neuronal activity in the inferior premotor cortex of monkeys performing certain hand movements. The same neurons were also activated when the monkeys were passively looking at the experimenters performing the hand movements. Furthermore, it could be indicated that these neurons may not only be activated by the movements based on the stimulus, but also activated by the *meaning* of the movement. In this context, Bellebaum et al. (2012) refer to an interesting study researching empathy with the help of functional imaging: Singer et al. (2004) exposed two subjects being relatives to pain stimuli and showed that the so called *affective pain matrix* (consisting of and interplaying between different brain regions) was activated whilst the pain was felt, but also whilst seeing a signal that informs about only the other subject being exposed to the pain stimulus. However, the *sensory pain matrix* was only active while being exposed to the pain stimulus. Bellebaum uses the german term *geteilte Repräsentation* that could enrich reflections on emotional contagion. Another recent study demonstrates the potential of emotional contagion concerning not only single persons, but as a mass phenomenon: it was conducted by Kramer, Adam D. I. et al. (2014) with data collected from a large social network while manipulating the News Feeds of selected users. The authors could show that if positive contents were reduced in the News Feeds, users created fewer positive posts and that if negative contents were reduced, users created fewer negative posts. Juslin et al. (2010) link this mechanism to the basic emotions.

Musical expectancy For the author, philosopher and composer Leonard B. Meyer, musical expectancy has been the main field of research. His dissertation from 1954 resulted in the key work on this subject already mentioned above. In Meyer's theory, the emotional power of music is the result of violations and confirmations of listeners' expectations. These are rhythmic, melodic or harmonic expectations[4] that refer to syntactical relations between different sections of musical structure (musical grammar). Meyer's analyses are based on Gestalt psychology, aesthetics, music theory and music history among others and were further developed by his student Eugene Narmour in *The analysis and cognition of basic melodic structures: The implication-realization model* (cf. Gjerdingen, 2012, pp. 695-696). A comprehensive extension of Meyer's work has been undertaken by Huron (2006). He highlights that musical expectations originate from past experiences stored in different types of memory. As recalling past events from memory can limit evolutionary advantage for an organism, namely optimising future actions, he guesses that memory mainly serves to prepare for future events by triggering arousal, attention or emotion and compares memory to "readiness circuits" (p. 219). More concretely, Huron defines four different types of listener expectations that are all linked to three different kinds of memory: if a listener hears a known stimulus, *veridical expectations* towards the development of this stimulus arise from the episodic memory.[5] The term goes back to the psychologist Bharucha who also

[4] In this work, just rhythmic and melodic expectations are considered.
[5] Huron notes that a clear distinction between *expectation* and *memory* should be maintained.

distinguished it from *schematic expectations*: if expectations arise from the knowledge of different musical styles and an interplay with the semantic memory, they are called *schematic expectations*. However, *dynamic expectations* are formed and readjusted permanently in the short-term memory during listening to a stimulus. This adaptive character of expectations may develop as a response to relatively short triggers from the music. As a counterpart to the three previous kinds of expectations who are mostly unconscious or preverbal (Huron, 2006, p. 235), *conscious expectations* arise from conscious reflections about how the musical structure will develop. Regarding memory during the listening process, Schubert (2004) illustrates that "[. . .] listener's emotional response is not just a function of musical features, predisposition, preferences, personality traits, and a range of other variables [. . .], but also of memory during the listening process." and further that "[. . .] every moment in the listening experience is, to some extent, related to every other moment of listening to the same piece." (p. 581).

Moreover, Huron developed Meyers theses further into his own Imagination–-Tension–Prediction–Reaction–Appraisal theory (ITPRA) describing the core mechanisms of expectation. It postulates five different emotion response systems associated with listener expectations (see fig. 1.1). The first two are pre-outcome

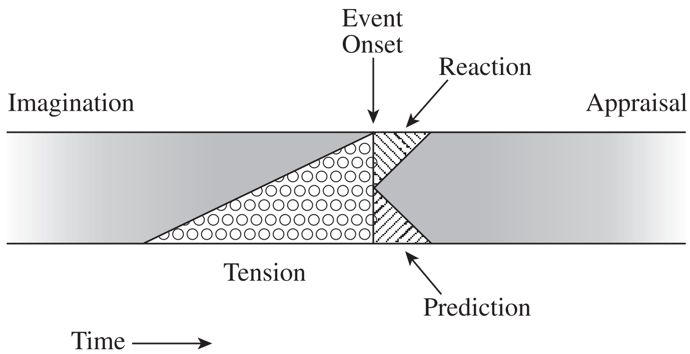


Figure 1.1. ITPRA theory of expectation (adopted from Huron, 2006, p. 17).

events: the function of the *imagination response* is to get an individual to react in a manner that results in a profitable outcome. The purpose of the *tension response* is to make suitable amounts of arousal and attention available to achieve optimum preparation of a forthcoming musical event. Immediately after the event, the last three systems take effect: evaluating the assumptions formed during the first two responses, the *prediction response* triggers some early (punishing or rewarding) feelings to improve future predictions. It happens in sync with the very fast "quick and dirty" (Huron, 2006, p. 13) *reaction response* during which a worst-case scenario for the outcome is considered. The final outcome is then evaluated during the slower and unconscious *appraisal response* that performs a more precise analysis of the outcome.

Juslin and Västfjäll (2008) link musical expectancy to "Surprise, awe, pleasure, 'thrills', disappointment, hope, anxiety" (p. 571).

Emotivism vs. Cognitivism At this point, another intensively discussed topic concerning music and emotion must be introduced: the difference between the *emotivist* and the *cognitivist* position towards musical emotions. In the difference between these two terms, the following question manifests itself: do I want to measure emotions that the listeners *feel themselves*, or do I want to measure emotions that the listeners think the *music expresses*? Tracing back to Peter Kivy, the emotivist position proclaims that music may *represent* or *express* specific emotions that listeners perceive and, additionally, is able to induce real emotions into listeners. However, the cognitivist position rejects the possibility of emotion induction, but keeps the representation or expression ability. To illuminate the confrontation of these two terms regarding the psychological mechanisms, Juslin et al. (2010, p. 632-633) note that the ability of a certain piece of music that expresses/represents a certain emotion to also induce exactly this emotion (or a different emotion or no emotion) to a listener highly depends on the psychological mechanisms being active during listening. For example, emotional contagion will very probably induce the same emotion the piece may express (because it transfers the emotion by definition), but our episodic memory[6] may remind us of a sad situation via a happy piece of music.

1.1.2 Subjective feeling: Valence and Arousal

For the measurements of subjective feeling described later in section 2.1.2, the model of a two-dimensional emotion space (2DES) was used. It was developed by (Russell, 1980) and could account for a large part of variance in affective states of self-report data. The two main dimensions *valence* and *arousal* are arranged on

[6] Another mechanism not mentioned above, but introduced here for reasons of clarity.

two orthogonal axes (cf. fig. 1.2), whereas Russell (1980) originally denotes the extremes of the horizontal bipolar valence dimension as "pleasure-displeasure" and the extremes of the vertical bipolar arousal dimension as "arousal-sleep" (p. 1163). Besides valence and arousal, the figure shows the four additional dimensions *excitement*, *contentment*, *depression* and *distress* that are not intended to represent independent dimensions, but are helpful to interpret the free quadrant spaces. Based on the theory that emotions may be reduced to a few important dimensions, the affective dimensions shown are suggested to interact very closely. Salimpoor et al. (2009) note that the physiological effect of emotional arousal is an

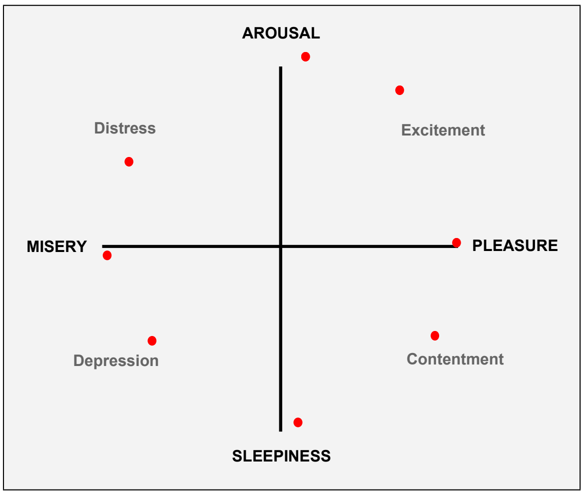


Figure 1.2. Eight affect concepts in a circular order (Russell, 1980, p. 1164, modified).

increased activity of the sympathetic branch of the ANS that the organism cannot control. Moreover, they describe physiological arousal as a proved indicator of emotional arousal. It has also been suggested to distinguish between the two different types *energetic* arousal (awake-tired) and *tense* arousal (tense-calm) (Thayer, 1989; Schimmack and Rainer, 2002). First is described as "subjective sensations of energy, vigor, or peppiness" and second as "feelings of tension, anxiety, or fearfulness" (Thayer, 1989, p. 6). This distinction was used for example in Ilie and Thompson (2006). However, Ellsworth and Scherer (2003) assume that the evaluations of the valence dimension[7] are based on psychological mechanisms that are more likely innate than learned. Furthermore, it seems to be closely connected

[7] Also denoted as "intrinsic pleasantness" with the two extremes "liking or attraction" and "dislike or aversion" by the authors (p. 577).

to the properties of the triggering stimulus itself and may provide an explanation for the formation of *aesthetic emotions*. Different variants of the 2DES model have been made, for example rotation of the orthogonal axes of the two dimensions by 45° leading to the positive and negative affect schedule (PANAS) by Watson et al. (1988) and the expanded positive and negative affect schedule (PANAS-X) by Watson and Clark (1999). Apart from the dimensional emotion models mentioned, there are models of discrete emotions based on basic emotions such as *happiness, sadness, fear, anger* and *tenderness* that are not relevant here[8].

1.1.3 Physiological Arousal: Skin Conductance

Electrodermal activity (EDA) was first discovered over 100 years ago in the laboratory of the french neurologist Jean Charcot. Working in this laboratory, Féré (1888) could show that the skin was a better conductor for applied electricity via two electrodes when subjects were exposed to external stimuli. Féré and the Russian physiologist Tarchanoff who discovered a change of SC without external electricity are considered the originators of the two most important recording techniques for EDA. Regarding human reactions to visual, auditory, gustatory or olfactory stimuli, Tarchanoff notes:

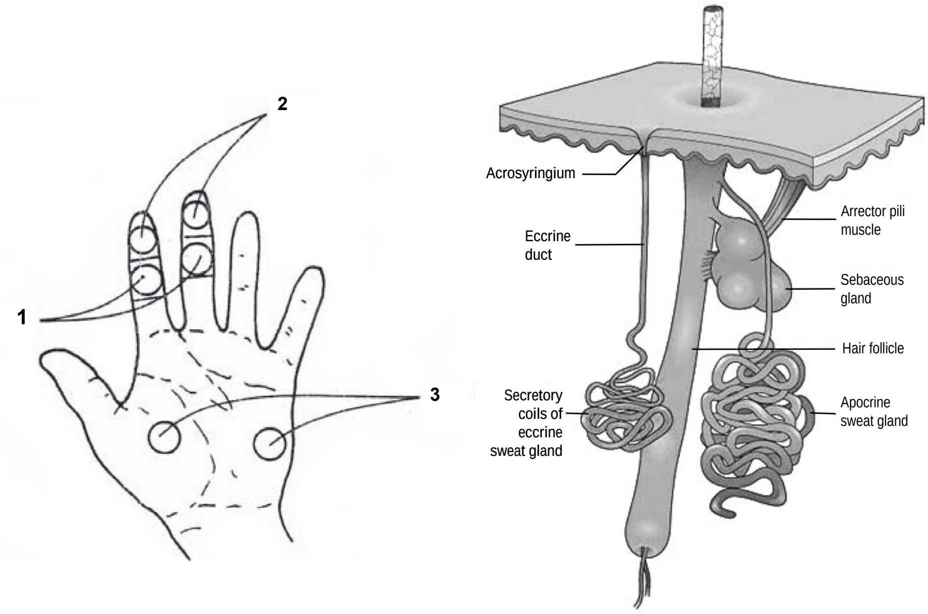
> Ferner Reizungen anderer Sinnesorgane: des Ohres - durch den Schall der elektrischen Glocke, durch lautes Rufen, Händeklatschen; Reizung der Nase durch Essigsäuredämpfe, Ammoniak u. dgl.; der Zunge - durch verdünnte Essigsäure, Zucker; des Auges - durch Licht, bewirken qualitativ in mehr oder weniger hohem Grade denselben galvanischen Effect, wie das Kitzeln der Haut. (Tarchanoff, 1890, p. 48)

His approach, capturing the skin potential response without external current, is called the *endosomatic* method, whereas the second most popular approach of measuring the skin resistance response with a very small voltage applied is known as the *exosomatic* method. This approach was used for the measurements during the live concert. The most important time of EDA research can be dated to the early 1970s and is associated with Lykken and Venables proposing standardised recording techniques for EDA (cf. Dawson et al., 2007, p. 159).

[8] The interested reader may see Vuoskoski and Eerola (2011) and Zentner and Eerola (2010) for descriptions and comparisons of different emotion models.

As EDA is based on the activity of sweat glands, a short description of anatomical and physiological basics is given in the following.

There are two types of sweat glands existing, the *apocrine* and the *eccrine* (cf. fig. 1.3(b)). Apocrine sweat glands emit odours and are therefore important for



(a) (b)

Figure 1.3. Three possible electrode placements for recording electrodermal activity, pos. 1: volar surfaces on medial phalanges, pos. 2: volar surfaces of distal phalanges, pos. 3: thenar and hypothenar eminences of palms (a, retrieved from Dawson et al., 2007, p. 163, modified), apocrine and eccrine sweat glands (b, retrieved from `http://cmapspublic2.ihmc.us/rid=1HQR4CFDC-1JBD8Z-7833/sweatgland.jpg`, modified).

social and sexual behaviour, but not of interest in this context. However, eccrine sweat glands contribute to a good heat balance: heat is taken from the body by evaporation of sweat whereby the body is cooling down while losing about 580 kilocalories per litre of sweat (Lippert, 2006, p. 282). As pointed out by Dawson et al. (2007), the glands located at the palmar[9] and plantar[10] surfaces are more important for grasping than for thermoregulation. However, they are thought to reflect mental processes by "psychological sweating" (p. 160) more than glands at

[9] Anterior of the hand.
[10] Foot sole.

other locations, also due to their high density in these areas. The authors give a vivid description of the eccrine ducts as part of every single gland as a system of resistors that are wired in parallel design. The ducts of a certain number of glands contain a sweat column that rises or falls depending on the activation of the sympathetic nervous system (SNS) innervating the glands. The level of sweat in the ducts and finally escaping them via the acrosyringium influences skin resistance and EDA.

In the next section, the musical features used as predictors for the hierarchical linear models (HLMs) of musical emotions will be introduced.

1.2 Musical Features

The process of musical emotion induction discussed in section 1.1 has been formalised by Scherer and Zentner (2001). Provided that both emotivist and cognitivist views should be taken into account, they claim experienced musical emotions to be a combination of (a) *structural* features, (b) *performance* features, (c) *listener* features and (d) *contextual* features. Within the formalisation, these features are defined as independent variables (iVs) whereas different emotional or affective states like musical preferences, emotions or moods[11] are defined as dVs. Some examples will be given for the four different feature categories following Scherer and Zentner (2001), pp. 362-365:

- structural features: acoustic properties of single tones or sounds, corresponding to phones in speech (segmental features); changes of whole musical sections over time, corresponding to intonation and amplitude contours in speech (suprasegmental features),
- performance features: all properties of the musical performance regarding the performer(s) and his identity (e.g. physical appearance), ability (e.g. interpretative skills) and performance state (e.g. interpretation and motivation),
- listener features: e.g. musical expertise or motivational state or individual sociocultural identity,
- context features: features concerning e.g. the location and all its properties.

[11] According to a design feature delimitation of various affective phenomena (see Scherer, 2000, pp. 140-141).

In this work, all musical features represent structural features and performance features of above-mentioned model. Listener features and contextual features are neglected. As it is hard to make a clear allocation of different musical features to either performance or structural features, this topic will be resumed after the introduction of the concrete features used.

Musical features are used in numerous studies and diverse contexts. Watson and Mandryk (2012) have modelled valence and arousal with feature data collected from listeners' own listening environments. Coutinho and Cangelosi (2009) extracted musical features to model affective responses with means of neural networks. Two years later, Coutinho and Cangelosi (2011) could demonstrate that musical emotions are predictable from a set of six psychoacoustic musical features like loudness, pitch level, pitch contour, tempo, texture, and sharpness. They assess the relevance of musical features as "[...] good predictors of the subjectively felt experience of emotion in music (at least for the affective dimensions considered)."' (p. 934) and conclude that "'[...] low-level psychoacoustic cues encode primary information about the emotions perceived and felt"' (p. 935). In another study with musical excerpts ranging from Mahler to Manowar, Gomez and Danuser (2007) showed that relationships between musical features and experienced emotions match with relationships found between musical structure and perceived emotions. Further studies considered were Schubert (1999), Schubert (2004), Korhonen et al. (2006), Synak and Wieczorkowska (2005), Guhn et al. (2007), Nagel et al. (2008), Dean and Bailes (2010) and Dean et al. (2011).

1.2.1 Categorisation

The categorisation of musical features is difficult because the borders blur. Some well established classifications will be presented in the following.

Lartillot (2013) proposes to divide the different musical features according to the five main dimensions of the musical domain, as they are dynamics, rhythm, timbre, pitch and tonality (p. 78), which is also the classification that the modular MIRtoolbox 1.5 is based on. However, a categorisation in the computational domain is proposed by Lerch (2012, p. 32) distinguishing between the *time domain* and the *frequency domain*. Lerch also mentions the categorisation following the MPEG-7 standard into *basic, basic spectral, signal parameters, temporal timbral, spectral timbral* and *spectral basis features* and introduces the terms *instantaneous* *feature*, *short term feature* and *descriptor* as designation for measurements that calculate one value per block of audio samples (see sec. 1.2.3). As these kinds of measurements may not reflect musical content or musical expression at all, they are referred to as *low level features* in this work. The following low level features are used and will be described in detail in section 1.2.3 (each with short identifier[12]): spectral centroid (LL_SPC), RMS (LL_RMS), brightness (LL_BRN) and roughness (LL_RGN).

As Hedblad (2011) points out, a group of *mid level features* may be ordered besides low level features. They often result as a combination of several low level features. For example, higher level pitch related features like key, harmony or melodic contour are calculated on the basis of sets of algorithms. Melodic contour, for example, could be calculated with a simple autocorrelation function (ACF)[13] to detect the fundamental frequency f_0, but also needs another component for the note onset detection to include the time dimension. Here, (the change of) root mean square (RMS) is important. Every single component has a very limited meaning regarding musical properties, but in combination, their meaning can have higher importance. The musical features from this group are often derived from music theory, for example tempo, modality and articulation. Melodic contour (ML_MEL) and tempo (ML_TMP) are the mid level features to be used in the following and are described in 1.2.4.

To complete this categorisation, *high level features* can be designed from combinations of low and mid level features. They may enable deeper analytical methods of musical semantics and musical perception. For the function `miremotion` of the MIRtoolbox 1.5 (described in section 2.1.4), Eerola et al. (2009) used a combination of 29 different musical features to analyse the three dimensions *activity*, *valence* and *tension* to enter the higher level of musical emotions[14]. UE and IC are also included into the group of high level features because both are evaluated by higher level cognitive processes. The following four high level features are used in this work and will be characterised in detail in section 1.2.5: high information content (HL_ICH), low information content (HL_ICL), high unexpectedness (HL_UEH) and low unexpectedness (HL_UEL).

[12] LL = low level, ML = mid level, HL = high level.
[13] Pitch detection algorithm, computes the value of τ for the first significant maximum following the value of $\tau = 0$ (global maximum of the ACF).
[14] Based on ground truth data of film soundtracks.

In the next section, the selection of musical features used as predictors for the HLMs of musical emotions will be described.

1.2.2 Selection

The first task was to figure out which low level features were "perceptually relevant and objectively quantifiable" (Schubert, 2004, S. 566) and could be represented as interval-scaled variables in the data set. Moreover, the musical features should fit the music material of the stimuli. Therefore, musical features like harmonic progression or texture[15] were excluded because they can be extracted from polyphonic material only.

As there are innumerable musical features available, those features having been of importance in previous studies to make a founded selection of the most appropriate features to be used in this work were searched for. For this purpose, a ranking for the five main groups of features was prepared: two points were assigned to a single feature that contributed significantly to study results and had a main impact on it, one point was given to a feature that contributed significantly to study results in interaction with other features, and features that had no significant impact to study results were given zero points. Table 3 shows an overview of the ranking results[16]. Finally, the attempt has been made to choose musi-

Group	Related musical features	Points
1 - Dynamics	**Loudness**, **RMS**	20
2 - Rhythm	**Tempo**, Accentuation, Rhythmic articulation	18
3 - Tonality	**Melodic contour**, Mode, Harmonic progression/complexity	12
4 - Timbre	Sharpness, Frequency range, **Spectral centroid**, Spectral flatness, **Roughness**	11
5 - Pitch	Mean pitch (spectrum centroid), Pitch level, Pitch variation (STFT flux)	6

Table 3

Ranking for the selection of musical features.

cal features from every group to cover an effective range of features. Whereas the

[15] Estimate of the number of tones simultaneously perceived in a sound.

[16] For further details see the three-part table at the beginning of appendix F.

dynamic musical features were the most important in previous studies, two representatives of the group have been included. As rhythmic articulation of the single flute signal seemed hard to be analysed[17], the choice was limited to the tempo feature from this group. Necessarily rejecting features requiring polyphonic musical material, melodic contour seemed to be well suited to cover both the tonality group and the pitch group, the latter somehow also represented by spectral centroid which officially is domiciled in the timbre domain. From here, brightness was chosen to complement spectral centroid and roughness was chosen to include the important aspect of consonance/dissonance that is closely related to valence.

1.2.3 Low Level Features

Before the musical features selected as predictors in this work will be described in brief (sec. 1.2.3-1.2.5), a short introduction to low level feature extraction shall follow.

Extraction Basics The extraction of low level features is usually based on block by block processing of the input signal vector. The principle is shown in figure 1.4. Here, i denotes the index of the input sample, n indexes the block to be processed and $\mathcal{H}$ denotes the hop size, i.e. the processing overlap of two successive blocks each with a length of $\mathcal{K}$ samples. Thus, the block overlap ratio o_r is defined as

$$o_r = \frac{\mathcal{K} - \mathcal{H}}{\mathcal{K}}. \tag{1.1}$$

The block length may vary between the extreme cases of $\mathcal{K} = 1$ and the number of samples corresponding to the complete length of the audio file, but is usually set to powers of two between $2^5 \leq \mathcal{K} \leq 2^{14}$ (Lerch, 2012, p. 19). During the block-based processing, it is important to know the location of the beginning and ending of the respective block. These values can be accessed via their sample indices calculated as

$$i_s(n) = i_s(n-1) + \mathcal{H} \qquad \text{(start sample)} \tag{1.2}$$

and

$$i_e(n) = i_s(n) + \mathcal{K} - 1 \qquad \text{(stop sample).} \tag{1.3}$$

[17] For instance in S3 with extremely high variability in lengths of breaks.

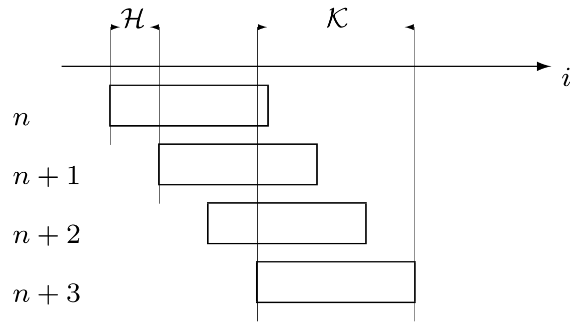


Figure 1.4. Principle of block-based processing (Lerch, 2012, p. 19).

Because the calculation of timbral musical features usually includes applying STFTs to signal blocks, the STFT is defined as

$$x(k,n) = \sum_{i=i_s(n)}^{i_e(n)} x(i)e^{-jk(i-i_s(n))\frac{2\pi}{\mathcal{K}}}. \tag{1.4}$$

As a basis for the STFT and for the sake of completeness, the Fourier transform (FT) is defined as

$$X(j\omega) = \mathfrak{F}[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt \tag{1.5}$$

as well as the inverse Fourier transform (IFT) as

$$x(t) = \mathfrak{F}^{-1}[X(j\omega)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\omega)e^{j\omega t}d\omega. \tag{1.6}$$

As the MIRtoolbox 1.5 (see sec. 2.1.4) does not perform in real time, the restrictions for real time processing[18] do not need to be applied to all the parameters mentioned above. Parameters set during the feature extraction concerning block-based processing are specified in detail in section 2.1.4.

In the following, low level features used will be presented with short identifiers and definitions.

LL_SPC The spectral centroid may be described as the spectral centre of gravity. Lerch (2012) defines the most common version as the frequency-weighted sum of

[18] In this case, the processing time for one block may not last longer than the time duration of the block.

the power spectrum normalised by its unweighted sum

$$LL\_SPC(n) = \frac{\sum_{k=0}^{\mathcal{K}/2-1} k \cdot |X(k,n)|^2}{\sum_{k=0}^{\mathcal{K}/2-1} |X(k,n)|^2}. \tag{1.7}$$

A high correlation was found between LL_SPC and the timbral properties *sharpness* or *brightness* (Krimphoff et al., 1994). The property considered as priority is the correlation between LL_SPC and brightness (Schubert et al., 2004, p. 656). As already mentioned in the introduction, Guhn et al. (2007) have shown that chill passages[19] of classical music are accompanied "by an expansion in its frequency range in the high or low register" (p. 480). They describe this expansion on the basis of the instrumentation, e.g. violins playing one octave higher than during the previous phrase or the double bass section adding a lower octave. Besides mapping the typical tonal character of the flute, LL_SPC may especially help to capture transient respiratory sounds in the pauses and strong blowing sounds emitted from the mouthpiece.

LL_RMS This musical feature was chosen to represent loudness. It is calculated by taking the root average of the square of the amplitude

$$LL\_RMS(n) = \sqrt{\frac{1}{\mathcal{K}} \sum_{i=i_s(n)}^{i_e(n)} x(i)^2}. \tag{1.8}$$

As demonstrated by Schubert (2004), loudness is positively related to arousal and "[...] sudden increases in loudness may even partially bypass the cognitive processing system because of the activation of a more primal, startle effect [...]" (p. 579). This startle effect is characterised as an automated alarm reaction to threatening unexpected loud sounds. The brain stem reflex described in sec. 1.1.1 controls it, but this controlling may be influenced by the amygdala if the signal is judged as aversive by the organism (Peretz, 2010)[20]. Dean et al. (2011) presented further interesting findings showing the strong influence of loudness on subjective

[19] Indicated via button presses.
[20] See also Yeomans and Frankland (1996) for a description of the acoustic startle reflex in animals.

arousal: they changed the intensity envelopes of different pieces of classical music and inverted it completely for one piece. While all other properties of the piece were unchanged, the perceived arousal ratings were also inverted. As already mentioned in 1.1.1, Guhn et al. (2007) showed the importance of loudness in connection with chill passages.

LL_BRN Brightness means the relative amount of energy above a certain cut-off frequency as the percentage of the total signal energy (see fig. 1.5) and is described as "one of the main correlates of perceived timbre" (Juslin, 2000, p. 1802). The default value for the cut-off frequency is 1500Hz. Brightness was

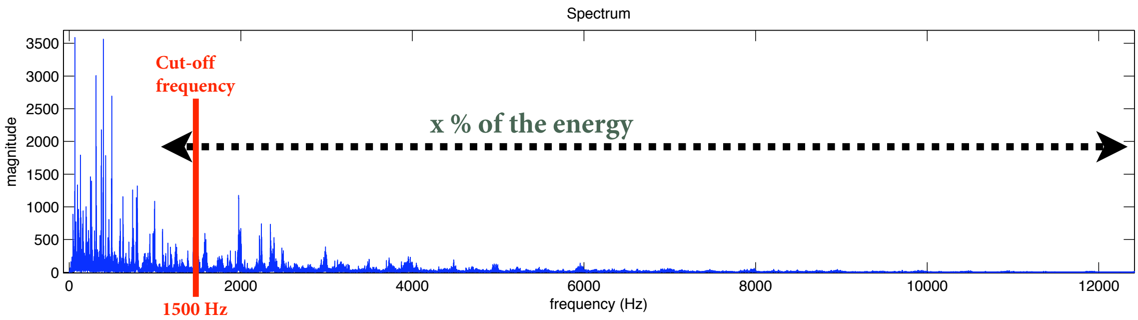


Figure 1.5. Calculation of brightness (Lartillot, 2013, p. 121, modified).

also included in the aforementioned study of Watson and Mandryk (2012) where it could contribute to a classification accuracy of 67% for arousal and 75% for valence together with other features and listening context features.

LL_RGN Consonance and dissonance of two tones depend on the difference of their frequencies and on the ability of the auditory system to resolve the tones. Played simultaneously, the tones are evaluated as dissonant when their frequencies are located in the same critical band, meaning that their individual vibrations at the basilar membrane overlap. If the tones are more than a critical band apart, their vibrations on the basilar membrane do not influence each other and consonance is perceived (Plomp, 1965; Gelfand, 2010).[21] The first processing step of the algorithm used from MIRtoolbox 1.5 considers this principle of critical bands

[21] See Weinzierl (2008, pp. 56-58), Fastl and Zwicker (2007, pp. 149-172) and Gelfand (2010, chapter 10) for more information on critical bandwidth/frequency analysis abilities of the auditory system.

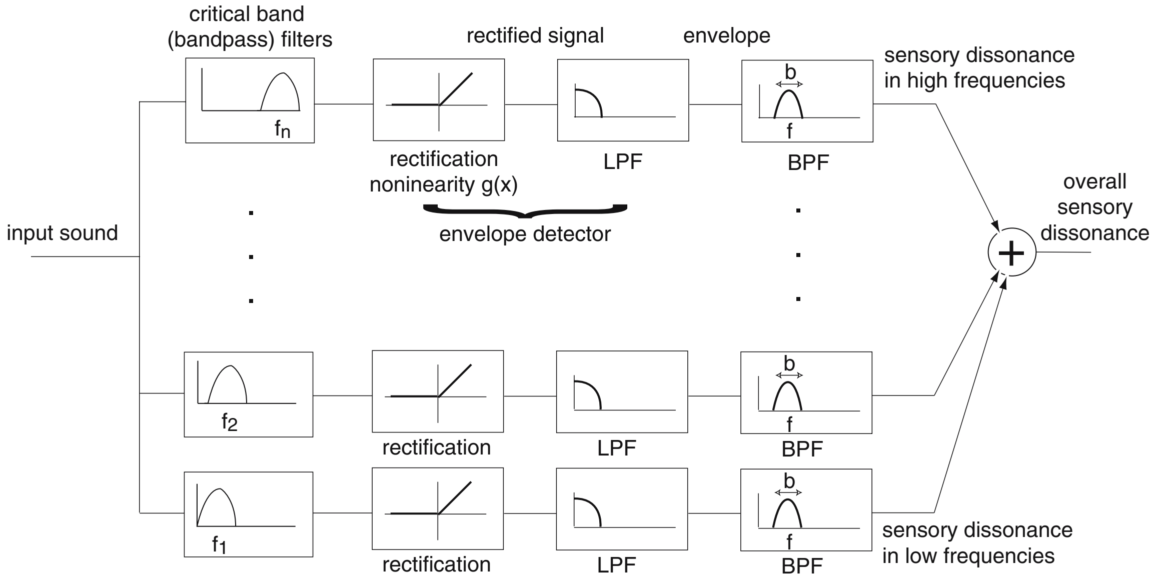


Figure 1.6. Model for the extraction of sensory dissonance (Sethares, 2005, p. 356).

(see fig. 1.6). After the critical band filters section, the algorithm rectifies the signal to simulate nonlinearity introduced by the behaviour of the hair cells on the basilar membrane and performs a bandbass filtering to remove the lowest and highest frequencies (corresponding to slow, pleasant beats and the sensation of loudness/fundamentals, overtones, and summation tones). Finally the 'individual sensory dissonances' between all pairs of sines extracted by picking peaks from the magnitude frequency spectrum are added up to get the total *sensory dissonance* value. For further details of the model see Sethares (2005), pp. 355-359.

1.2.4 Mid Level Features

In the following, mid level features used will be presented in short.

ML_MEL Following Schubert (2004), ML_MEL is defined here as the derivation of the melodic pitch stream, i.e. the change of pitches from note to note is considered in the analyses. Schubert notes that emotional responses to melodic contour remain still unclear while referring to a study by Gerardi and Gerken (1995). This study with 5-year-olds, 8-year-olds and college students showed that ascending melodic contour was rated significantly more positive than descending melodic contour, thus this feature is positively related to valence.

ML_TMP Following Lerch (2012), the perceived tempo for a section of music is based on its *tactus* which also provides the base for the foot tapping rate. To derive this term, Desain and Honing (1993) shall be cited here who define tempo as the "rate at which beats occur" and as "closely linked to the metrical structure"; *metre* is again defined as a ratio between at least two different time layers: the first layer is the beat period as referential layer and the second layer is the measure, defined as a higher order layer based on a determined number of beat periods[22]. From these time layers, the one at which "the units (beats) pass at a moderate rate [...] around the 'preferred' or 'spontaneous' tempo of about 100 beats per minute" (p. 125) is called the *tactus*. The *absolute* tempo in beats per minute (BPM) is thus calculated as

$$\mathfrak{T} = \frac{\mathcal{B} \cdot 60s}{\Delta t_s} \text{ [BPM].} \tag{1.9}$$

where Δt_s defines the length of the musical section in seconds and $\mathcal{B}$ denotes the number of bars in this section. However, the *local* tempo for stimuli with a dynamic tempo can be calculated based on the onset time of the beats t_b and their neighbours indexed with j and $j+1$ as

$$\mathfrak{T}_{local}(j) = \frac{60s}{t_b(j+1) - t_b(j)} \text{ [BPM].} \tag{1.10}$$

As with loudness, tempo seems to be positively related to arousal (Schubert, 2004). Moreover, interesting effects of tempo were observed by Husain et al. (2002) who investigated how tempo and mode of a Mozart sonata influenced listeners' performance of a spatial ability task, their arousal and mood.[23] Among others, it could be shown that listeners who had heard the piece in a faster tempo performed better in the following spatial ability test. The positive relationship between tempo and arousal was also confirmed.

1.2.5 High Level Features

In the following, high level features used will be presented in short.

[22] At this point, the oscillator theory of Large (2002) mentioned in section 1.1.1 (rhythmic entrainment) may be referred to again.

[23] The study was also an investigation of the *Mozart effect*.

HL_ICH, HL_ICL At this point, a review of the paragraph on musical expectancy in section 1.1.1 is necessary. As stated by Huron (2006), Narmour characterised melodic expectations to be innate and based on Gestalt theory. Contrary to this assumption, Huron states experimental research to prove that "expectations are learned, and that the predictive heuristics used by listeners are the product of sustained exposure to commonplace musical patterns." (p. 95). One of these experiments is a statistical model constructed and evaluated by Pearce (2005) which was applied to the stimuli used in this work. This model calculates probability distributions for the pitch or the onset time or both for each note in a melody. The latter setting was the case in this work. The model remembers what it has learned about the melody from the previous notes and includes this knowledge to the calculated expectation towards the next note. Furthermore, it is of variable order, i.e. the length of the melodic context may vary (Pearce, 2005), and it may be applied as long-term model (LTM) or short-term model (STM) or a combination of both. As LTM, the model is applied successively to all stimuli which equals a listeners' long-time exposure to music. However, the STM is applied to one stimulus in real-time and resets before the next stimulus which equals a short-time exposure to music. For the present research, a combination of both approaches was applied. The final result of the model is a value for IC in Bit for each note in a melody. For the sake of completeness, it has to be mentioned finally that the model is able to use different viewpoints not used here, but described in detail in Pearce (2005), chapter 5.

HL_UEH, HL_UEL In principle, UE conveys a similar meaning as IC. The main differences are that (a) IC is calculated by a computer and UE is the group mean of continuous ratings of 24 listeners and (b) IC considers expectations exclusively related to the *pitch* dimension and a time dimension, the inter-onset interval (IOI), whereas UE relates to many more dimensions: during the live concert experiment, listeners were instructed to rate the UE of musical events they hear. This instruction will finally lead to multidimensional ratings of UE related to harmonic progression, dynamics, timbre or even location-specific dimensions like reverb times or the like. Similarly for IC, looking back to the paragraph on musical expectancy in section 1.1.1 is also necessary to characterise UE. Juslin and Västfjäll (2008) assign the processing of sudden *unexpected* events to the brain stem reflex while they relate events of broader contexts (like the unexpectedness of liking the actual stimulus more than the previous) to the *cognitive appraisal* mechanism. As outlined by Ellsworth and Scherer (2003), the basic process of all appraisal theories is that an individual evaluates a current, remembered or imagined situation whereby this evaluation is important for the development of specific emotions and to distinguish between them. Unexpectedness ratings were collected via iPod touch® devices during the live concert (see also fig. 2.1(a)).

To now establish a connection between the musical features and the mechanisms described above, the *lens model of musical communication* (fig. 1.7) shall help. Originally referred to visual perception by Brunswik in 1956, Juslin and Sloboda (2012) use this model for a possible explanation as to why there is a general accordance of listeners on the emotional expression of a certain piece of music although the same musical features may vary strongly in other pieces with the same expression. The key message of this model is that mediations of emotions result from *additive* processes of different features. The listener performs inner combinations of certain features that lead to emotional judgements. Moreover, the model postulates that the recognition of specific emotions is possible despite strongly varying conditions like for example individuality of the performance or variability of pieces and instruments. This principle will also be considered in the interpretation of results, especially to analyse the interplay between the different features as the result of the principal component analysis (PCA) to be described in section 2.2.1. As announced above, the assignment of musical features used here to either structural or performance features will be resumed at this point: Scherer and Zentner (2001) classify both mid level features ML_TMP and ML_MEL to the group of structural features. However, if the performer varies ML_TMP slightly (maybe unconsciously due to a lack of ability to keep the tempo), it will, at least partly, also be a performance feature. This also applies to ML_MEL: if the performer plays one wrong note, ML_MEL is no longer only a structural feature but also turns into a performance feature. Thus, it seems to be quite clear that a strict distinction according to Scherer's formalisation approach is not possible. Figure 1.8 shows a summary of the theoretical part up to now, and to complete the introduction, the following hypotheses were formulated based on the previous considerations:

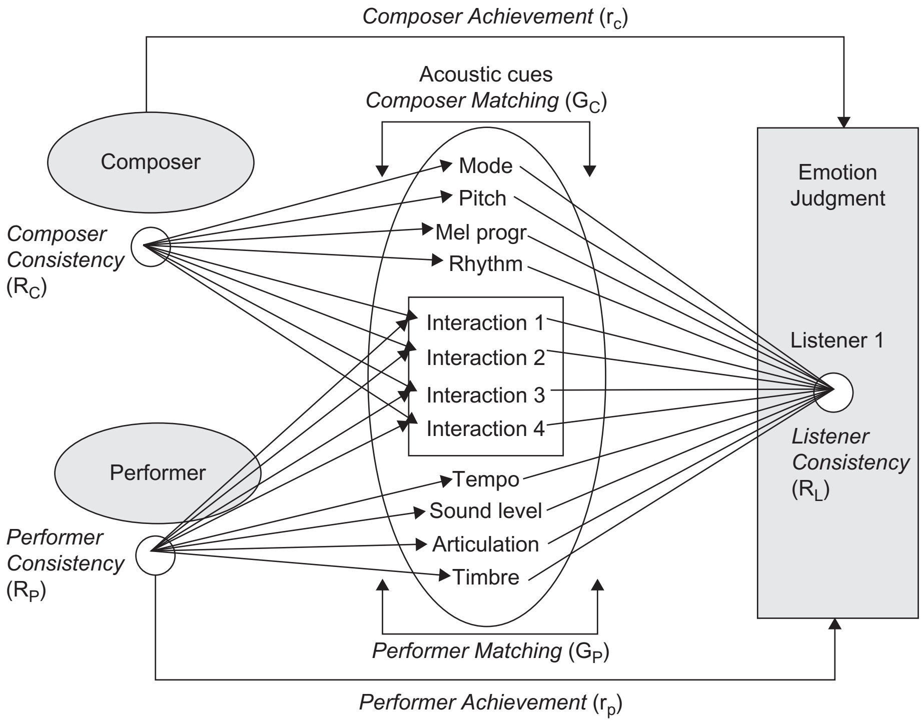


Figure 1.7. Expanded lens model of musical communication (Juslin and Sloboda, 2012, p. 599).

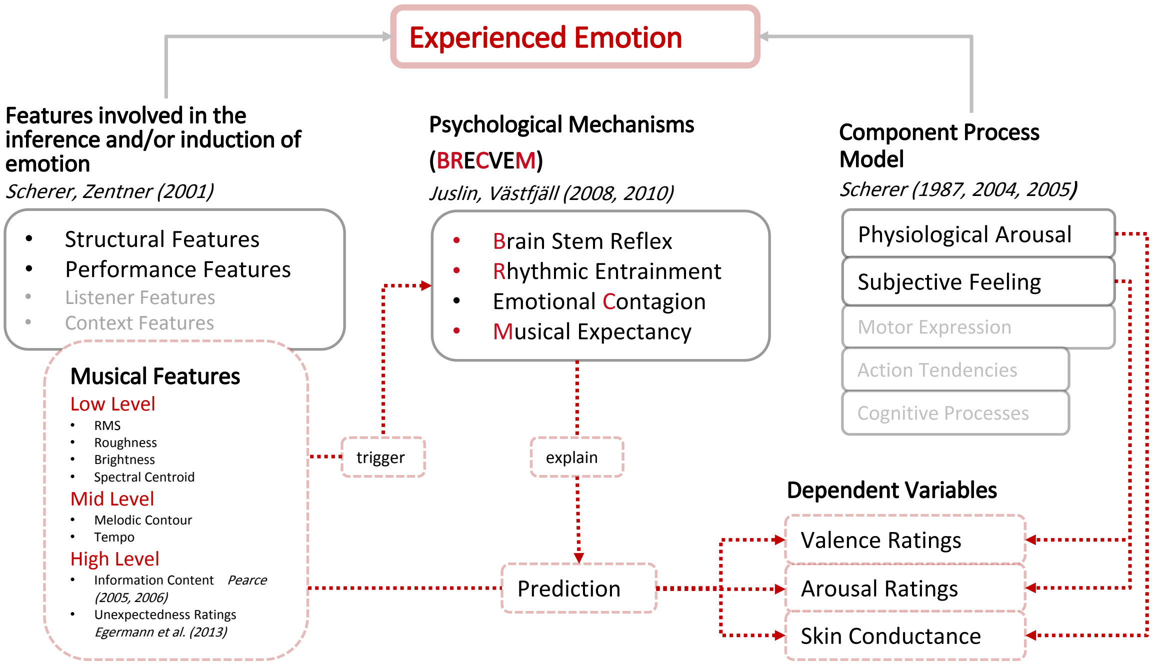


Figure 1.8. Graphical summary of the theoretical part.

- ***H*$_1$**: Time series from continuous ratings of the subjective component *arousal* are predictable by four low level features (*spectral centroid*, *RMS*, *brightness*, *roughness*), two mid level features (*melodic contour*, *tempo*) as well as four high level features (the group mean of continuous *UE* ratings (high UE/low UE), *IC* of each note calculated from an information-theoretic cognitive model (high IC/low IC)).

- ***H*$_2$**: Time series from continuous ratings of the subjective component *valence* are predictable by four low level features (*spectral centroid*, *RMS*, *brightness*, *roughness*), two mid level features (*melodic contour*, *tempo*) as well as four high level features (the group mean of continuous *UE* ratings (high UE/low UE), *IC* of each note calculated from an information-theoretic cognitive model (high IC/low IC)).

- ***H*$_3$**: Time series from continuous measurements of peripheral arousal (*skin conductance*) are predictable by four low level features (*spectral centroid*, *RMS*, *brightness*, *roughness*), two mid level features (*melodic contour*, *tempo*) as well as four high level features (the group mean of continuous *UE* ratings (high

UE/low UE), *IC* of each note calculated from an information-theoretic cognitive model (high IC/low IC)).

- ***H_4***: Low and mid level features on the one hand and high level features on the other hand are independently capable to influence predictions of subjective continuous ratings (*valence* and *arousal*) and measurements of peripheral arousal (*skin conductance*).

2 Method

The method part is divided into two sections. In section 2.1, the live concert experiment settings and the measurements[24] executed during this event will be described, followed by a description of the statistical model building process in section 2.2.

2.1 Experiment and Measurements

2.1.1 Participants and Stimuli

The live concert experiment was conducted with an audience of $N = 48$ music students (21 female). The average age of participants was 23 years. All subjects were selected with the help of an online questionnaire to make sure they were familiar with classical music and had a preference for it. Further criteria to be met were (a) normal hearing, (b) willingness to be filmed, (c) willingness to abstain from makeup/forego makeup, (d) willingness to shave, (e) speak English well, (f) no classical music dislike and (g) a familiarity with the stimuli (ideally 50% familiar, 50% unfamiliar). During the concert, six classical monophonic flute pieces were presented. They were chosen to cover a wide range of musical styles and have been picked from the flutists' repertoire (Egermann et al., 2013). The first two pieces were played from recordings as an introducing practice so that the listeners would become familiar with the rating devices and get used to the concert environment, followed by a live performance of the remaining four pieces on stage[25]. The stimuli are presented in table 4.

Now crossing over to the measurements, it has to be mentioned that conventional studies of music and emotion mostly used ratings that were carried out *after* performances. As outlined by Schubert (2010), the assumption during the last 80 years was that musical emotions could be explored collecting emotional responses in *post-performance* ratings like questionnaires. But he argues that "[...] given the intricate connection between music and emotion, it stands to reason that emotions occurring in response to unfolding music should also be time dependent" (p. 223). The distinction between *post-performance ratings* and *during-performance*

[24] Only the measurements relevant for this work.
[25] See appendix A for the scores and appendix G for audio data.

	Title	Composer	Presentation mode	Duration (min:s)
S1	Acht Stücke für Flöte Allein: VI. Lied, leicht bewegt	Paul Hindemith	recorded	00:38
S2	Un Joueur de Flûte Berce les Ruines	Francis Poulenc	recorded	01:18
S3	Density 21.5	Edgar Varèse	live	03:30
S4	Syrinx	Claude Debussy	live	02:35
S5	Solo Partitas for Flute, A-minor: 2nd Movement "Corrente"	Johann S. Bach	live	01:53
S6	Solo Partitas for Flute, A-minor: 3rd Movement "Sarabande"	Johann S. Bach	live	02:11

Table 4
Music stimuli presented.

measure was stated more precisely by Sloboda and Lehmann (2001) studying listeners' evaluation of expressivity and other performance aspects of previously recorded different versions of Chopin's Prelude op. 28 no. 4. Here, the authors used both post-performance ratings and continuous ratings. This notion is the background for the development of continuous self report methods that were also applied for recording the dVs considered in this work and described in the following.

2.1.2 Valence and Arousal

All participants were randomly assigned to two groups. The members of group 1 (N = 24) had to rate their subjective feelings continuously on a 2DES of *valence* and *arousal*, whereas the members of group 2 (N = 24) had to rate the *unexpectedness* of the music continuously on a single slider. Participants of group 1 were instructed to indicate continuously during listening to the pieces how the music affects their feelings:

> By moving your finger from left to right you can indicate how pleasant the music is to you (left = negative and unpleasant; right =

(a) (b)

Figure 2.1. (a) Rating devices, (b) Procomp Infiniti Encoder (from MediTECH Electronic GmbH, 2004).

> positive and pleasant). By moving your finger from top to bottom you can indicate your degree of emotional arousal during listening to the music (top = excited; bottom = calm). (cited from questionnaire to be found in appendix G)

However, participants of group 2 were instructed as follows:

> By moving your finger from top to bottom you can indicate how unexpected the music events you are hearing are (top = very unexpected; bottom = very expected). The position of your finger should reflect at each moment the unexpectedness of the events as you are listening. You need to constantly monitor your expectations for every musical event in order to keep your finger at the corresponding position. (cited from questionnaire to be found in appendix G)

48 iPod touch® served as rating devices (fig. 2.1(a)) and were fixed to the participants' thighs. The measurement time series for both variables were extracted from the existing file `preprocdataNEEDED_live.mat`, differentiated using the `diff` function and finally z-transformed. The z-transform had to be calculated with the Matlab® functions `nanmean` and `nanstd` following

$$z = (x - \texttt{nanmean}(x))/\texttt{nanstd}(x). \tag{2.1}$$

This considers not a number (NaN) values in the data due to (a) missing listener ratings, (b) NaN values appearing in the time series of the dVs introduced by the time lag structure to be described in section 2.1.6 and (c) possible NaN values resulting from feature extraction.[26] Differentiation and z-transform were applied to *absolutely all* the following time series before entering the statistical analyses, thus they all showed the *change* of what they represented. The purpose of the z-transform was turning b-coefficients into standardised β-coefficients, and the differentiation step was intended to help reduce autocorrelation (described in detail in sec. 2.2.2 and sec. 2.2.4).

2.1.3 Skin Conductance

The physiological reactions of all participants were measured continuously via a custom audience response system that was developed especially for the live concert experiment. Besides SC that is relevant for this work, more measurements[27] were made with five electrodes and a respiration belt per subject connected to several Procomp Infiniti® devices (fig. 2.1(b)). The reciprocal of the skin resistance response is the SCR being present in the data as dV. It was measured in micro-Siemens (μS) at the phalanx distalis of the index and middle finger at the non-dominant hand as shown in figure 1.3(a), pos. 2 with f_s = 256 samples. It was also extracted from the existing file `preprocdataNEEDED_live.mat`. In order to remove e.g. slow drifts over long recording time, a linear detrending of the SCR time series was performed with break points every 60 seconds to prevent accumulation effects. Trends may appear due to charge accumulation at the conjunction between skin and electrodes (see Salimpoor et al., 2009). Moreover, the first seconds of the uncompressed SC time series deserved careful consideration. As listeners were resting before the pieces[28], a typical shape of the SC graphs could be expected. Dawson et al. (2007) describe the typical behaviour of SC to gradually decrease while the organism is resting. As soon as it is exposed to a *new* stimulus, SC increases quickly and will gradually drop again when the *same* stimulus is repeated. Figure 2.2 shows this typical fast increase lasting up to

[26] The complete Matlab® code written and used in this work can be found in appendix G.
[27] Heart rate, blood volume pressure, respiration and electromyography of corrugator and zygomaticus muscles.
[28] During 45 seconds of baseline recording before (not considered in this work) and filling out questionnaires after each piece.

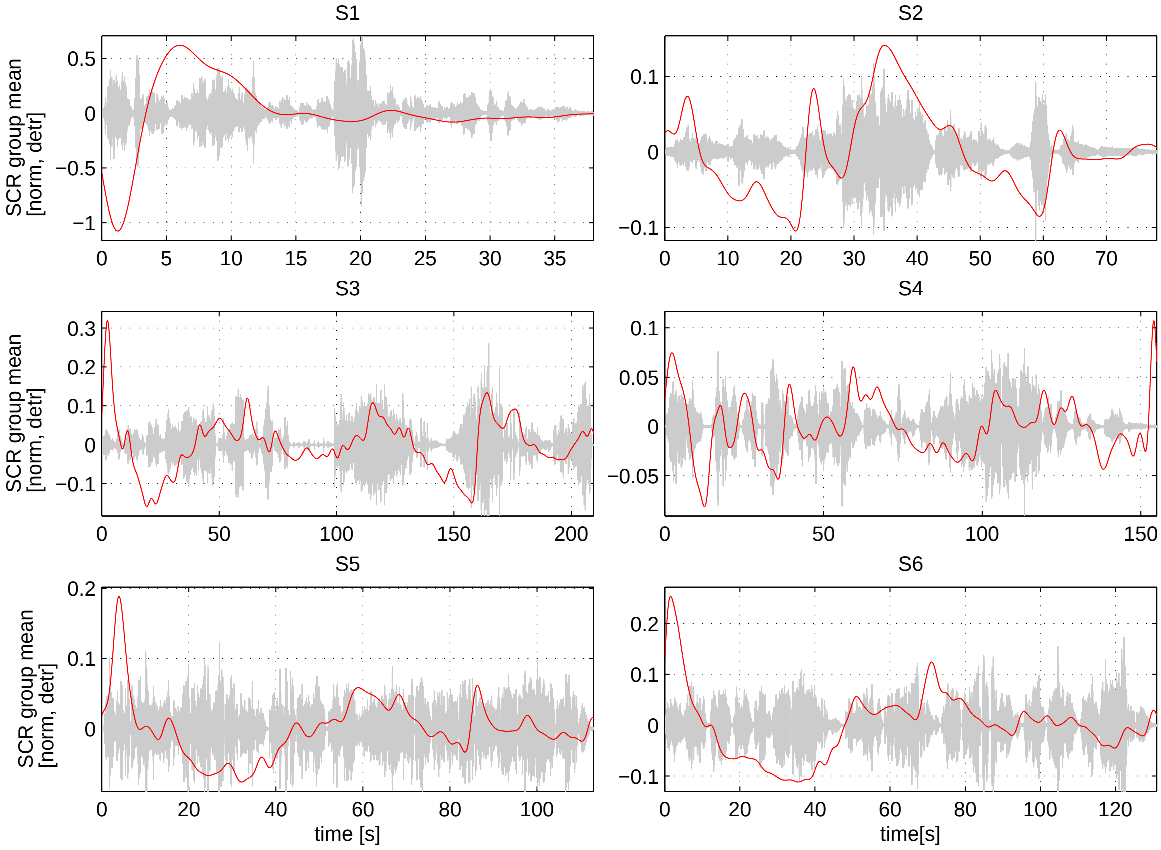


Figure 2.2. Uncompressed, but normalized and detrended SCR time series (S1-S6).

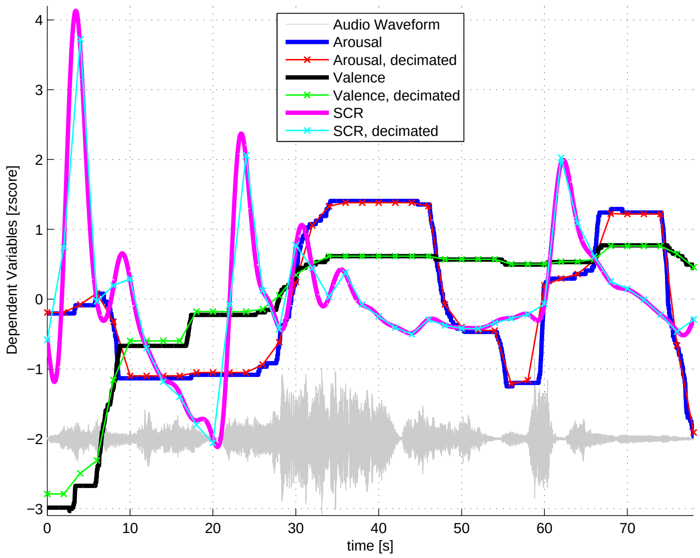


Figure 2.3. Original and decimated time series of the three dVs (exemplary for S2, subject 24).

≈ 10 seconds for all stimuli. Whereas these values are expected to cause distortions of the modelling results, they were removed between 0 – 10 seconds. Finally, all SCs time series of were also decimated from the original sampling frequency of $f_s = 256Hz$ to $f_s = .5Hz$ using the `decimate` function in Matlab® with `fir` setting. A comparison between the original and decimated time series for all dVs is shown in figure 2.3 for S2, subject 24 as an example.

2.1.4 Extraction of Low and Mid Level Features

Before the feature extraction was performed, suitable software tools had to be researched. Research results are presented in table 5. All low level features were extracted with MIRtoolbox 1.5, whereas Sonic Visualiser 2.3 and MidiToolbox were used for mid level feature extraction. This choice was based on different criteria. A high degree of reliability, ease of use and good integration possibilities into the Matlab® environment were desired. For example, some of the applications are available only as source code. Thus a huge effort was made to compile a binary command line version from the Marsyas C++ source code, a pitch extraction was performed, compared to results from MIRtoolbox and rejected due to worse results, awkward handling, lack of flexibility and finally a low cost-benefit ratio. However, the MIRtoolbox fulfils the criteria with a great Matlab® integration, includes many musical feature extractors, is flexible and has adjustable block-wise processing and finally simple and adaptive syntax (Lartillot, 2013). In the following, feature extraction with regard to individual extraction settings for the stimuli on hand will be described.

First, all mono and normalised audio data vectors were created importing the stimuli with the Matlab® `wavread` function. These vectors had to be converted to `mir`-objects using the MIRtoolbox `miraudio` function because those objects are accepted as inputs only. In the next step, all low level features were extracted with $\mathcal{K} = 2000$ samples[29], $\mathcal{H} = 1000$ samples (50%) and a typical value of $o_r = .5$ (as proposed by Lerch (2012)) and finally decimated by factor 48 to $f_s = .5Hz$ using the Matlab® function `smooth`. This process introduced some information loss (see fig. 2.3) that has to be accepted with the aim to reduce autocorrelation (see 2.2.4). Processing was made in loops to include all 48 subjects and all six stimuli S1-S6

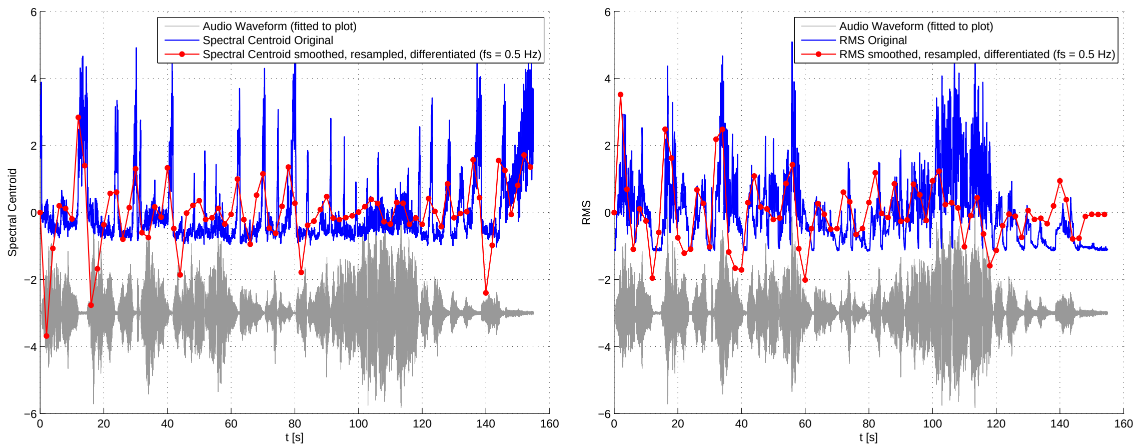
[29] Due to the properties of the analyses/dataset, there was no necessity to use a power of two value here.

Software	Author/s	Short description
aubio 0.4.1	Brossier	Extraction of annotations from audio signals
Auditory Toolbox 2	Slaney (1998)	Matlab® toolbox, auditory modelling
CAMEL	Sanden et al. (2010)	A framework for audio analysis
FEAPI	Lerch et al. (2005)	Low fevel feature extraction plugin API
jAudio	McEnnis et al. (2005, 2006)	Software package for extracting features from audio files
LibXtract 0.4.5	Bullock (2007)	Library of audio feature extraction functions
Maaate 0.3.0	Pfeiffer	C++ frequency domain audio analysis toolkit
Marsyas 0.5.0	Tzanetakis and Cook (2000)	Music analysis, retrieval and synthesis for audio signals
MIDI Toolbox	Eerola and Toiviainen (2004a,b)	Matlab® toolbox, analyze and visualize MIDI data
MIRtoolbox 1.5	Lartillot and Toiviainen (2007); Lartillot (2013)	Matlab® toolbox, modular framework of different algorithms for music information retrieval
openSMILE 2.0-rc1	Eyben et al. (2013)	The Munich versatile and fast open-source multimedia feature extractor
PRAAT 5.3.68	Boersma, Weenink	Speech analysis
PsySound3	Cabrera (1999); Ferguson et al. (2008)	Matlab® toolbox, psychoacoustic feature extraction
SCMIR 1.0	Collins	SuperCollider 3 library, audio content analysis
Sonic Visualiser 2.3	Cannam et al. (2010); Cook and Leech-Wilkinson (2008)	Stand alone application for viewing and analysing the contents of music audio files
Timbre Toolbox 1.2	Peeters et al. (2011)	Matlab® toolbox, extract audio descriptors from musical signals
VAMP Plugins	Queen Mary University London	Audio analysis plugin system
YAAFE 0.64	Mathieu et al. (2010)	Yet another audio feature extractor

Table 5
Software for audio feature extraction.

effectively.

LL_SPC The spectral centroid was extracted using the `mircentroid` function. As the spectral centroid is not defined for silence, it needs special attention for the stimuli of this work. A decision was needed on removing spectral centroid values in pauses or not because the spectral centroid of a low level noise floor may take very high values. Whereas the pauses of the flute performance are almost completely filled with loud, fast and impulsive respiratory sounds that are seen belonging to the performance[30], the pause values were kept. This was realised by setting the parameter `'MinRMS'` to 0. An example plot of LL_SPC extracted from S4 is presented in figure 2.4(a).



(a) (b)

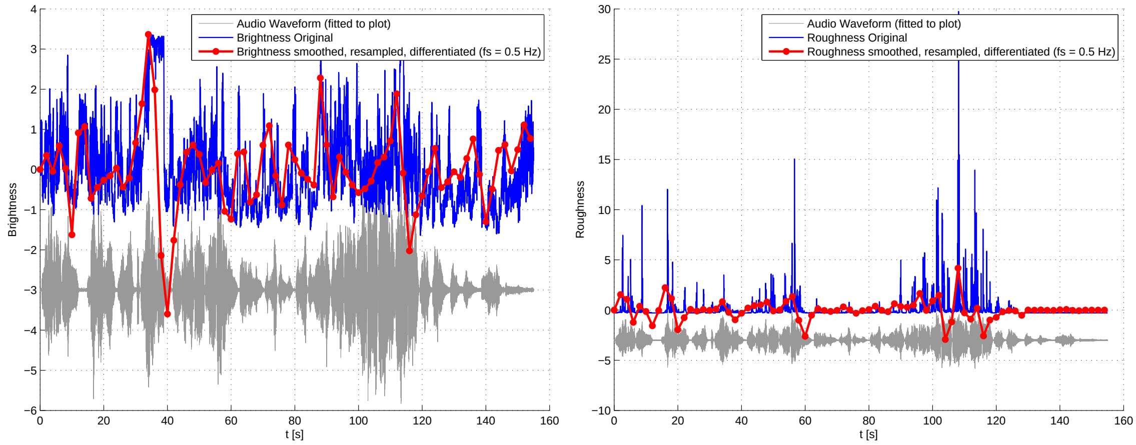
Figure 2.4. Low level features (a) spectral centroid and (b) RMS (exemplary for S4).

LL_RMS RMS was calculated using the `mirrms` function. An example plot of LL_RMS extracted from S4 is presented in figure 2.4(b).

LL_BRN Brightness was extracted using the `mirbrightness` function. The default value of 1500Hz for the cut-off frequency was reduced to 1000Hz following Laukka et al. (2005), and the `MinRMS` value was set to 0 to capture all noises like

[30] According to Scherer and Zentner (2001), these sounds can be well classified as part of the performance features.

respiratory sounds and key noises of the flute. An example plot of LL_BRN extracted from S4 is presented in figure 2.5(a).

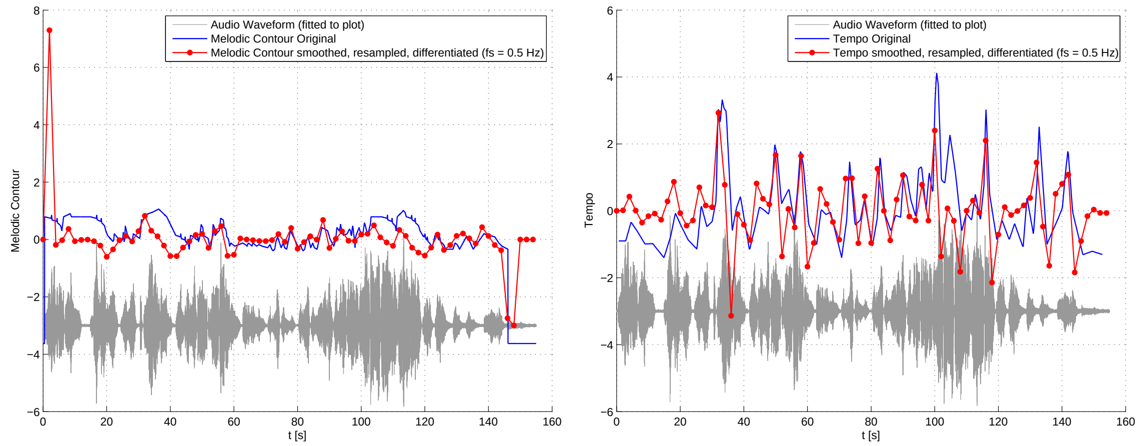


(a) (b)

Figure 2.5. Low level features (a) brightness and (b) roughness (exemplary for S4).

LL_RGN Roughness was extracted with the `mirroughness` function and the `Sethares` method (see Sethares, 2005, p. 356). An example plot of LL_RGN extracted from S4 is presented in figure 2.5(b).

ML_MEL The first step of creating the time series for ML_MEL included a manual annotation of MIDI notes along the audio recordings of the live concert with the help of *Sonic Visualiser 2.2*. These Midi notes could then be exported to MIDI files that again were extracted in Matlab® using the function `midi2nmat` from the MidiToolbox. Subsequently, MIDI pitches and onset times from the resulting `nmat` tables were interpolated to a time vector with $f_s > 2 \cdot f_{max}$ to meet the sampling theorem. At this point, it was also important to consider relating the last value of ML_MEL to the last note onset and not to the entire length of the piece. The resulting time series were smoothed using a moving average filter to keep as much information as possible during the following downsampling to the final sampling frequency of $f_s = .5Hz$. The last two steps were differentiating the time series using the `diff`-function and creating `.csv`-files for further use in SPSS®. The peak of the differentiated values at the beginning of S4 in figure 2.6(a) results from the big difference between 'no melody' and 'melody', vice versa at the end of the piece when melody turns off.



(a) (b)

Figure 2.6. Low level features (a) melodic contour and (b) tempo (exemplary for S4).

ML_TMP For this feature, *Sonic Visualiser 2.2* was again used to set tempo instants manually for all stimuli according to the respective scores (see app. A). During playback, markers could be set via key presses with the possibility of correction during repeated playbacks. To create tempo time series as precisely as possible, each main musical beat of a bar was annotated (i.e. each quarter note of a 4/4-bar). The local change of tempo was then calculated by the software based on equation 1.10. Resulting tempo values with associated time instants could then be exported to `.txt`-files that could again be imported in Matlab® for similar further processing as with melodic contour. An example plot of ML_TMP extracted from S4 is presented in figure 2.6(b).

2.1.5 Integration of High Level Features

HL_ICH, HL_ICL The data for the basic IC model was extracted from the file `IC_Signals.mat`[31]. This file contained the IC data for S1-S6 with note onset times (cell arrays `ICN`, `IC_withonset`, `IC_withonsetInt` and `model`) and the three vectors

[31] The data was provided by the first supervisor of this work and can be found in appendix G. The model has previously been applied to MIDI data of all six stimuli for the analyses of Egermann et al. (2013).

`input`, `model` and `viewpoints` allowing to specify the model. A basic model was extracted with the settings presented in figure 2.7. Further steps were an interpo-

```
mmodel = [3 2 1 3 2 3 2 1 3 2 1 3 2 1 3 2 1 3]
%1: stm
%2: ltm
%3: both <--

input = [1 1 1 1 1 3 3 3 3 3 3 2 2 2 2 2 2 2]
%1: ioi
%2: pitch <--
%3: pitch+ioi

viewpoints = [1 1 1 2 2 2 2 2 1 1 1 2 2 2 1 1 1 3]
%1: basic <--
%2: optimized
%3: a viewpoint linking scale degree and pitch interval

IC_withonsetBASIC{s}(:,1) = model{s}(:,2); % Onsets for every single ...
    note (in sec.)
IC_withonsetBASIC{s}(:,2) = ICN{s}(:,mmodel==3 & input==2 & ...
    viewpoints==1); % --> IC-Model #15
```

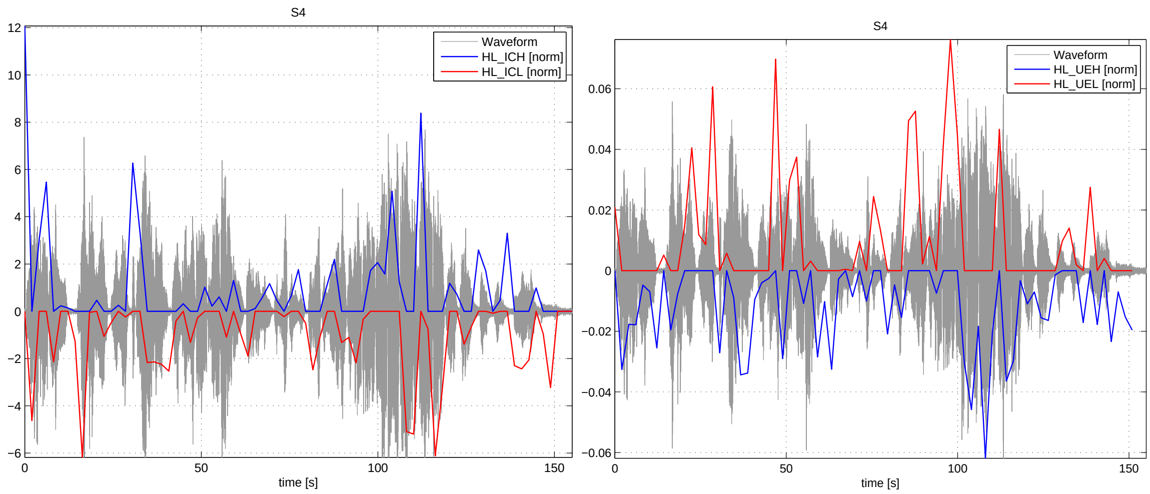
Figure 2.7. Matlab® code for IC model extraction.

lation using `interp1`[32], applying a moving average filter using `smooth`, resampling using `downsample` and differentiation using `diff`. The high and low versions were then calculated by setting negative respective positive values to 0, z-transforming following equation 2.1 to keep data consistent and finally saving `.csv`-files using `csvwrite` for further use in SPSS®. An example plot of HL_ICH and HL_ICL extracted from S4 is presented in figure 2.8(a).

HL_UEH, HL_UEL Both were calculated as group means from subjective unexpectedness ratings with the function `repeatedMeasures_spss_exp_mean.m` and `repeatedMeasures_spss_exp_mean_highlow.m`. Further steps were nearly similar to the steps for HL_ICH and HL_ICL. An example plot of HL_UEH and HL_UEL extracted from S4 is presented in figure 2.8(b).[33] As values are only considered

[32] Here, as with ML_MEL, it was also important to consider relating the last IC value to the last note onset and not to the entire length of the piece.

[33] Plots of IC and UE for all other stimuli can be found in appendix B.



(a) (b)

Figure 2.8. HL_ICH and HL_ICL (a), HL_UEH and HL_UEL (b) (exemplary for S4).

until the last note onset of IC, there are values missing at the end of the final time series.

The topic of musical features will be left at this point, and in the next section, the issue of time lags and 'emotional reaction times' of listeners will be outlined.

2.1.6 Time Lag Structure

As introduced in section 2.1.2, continuous self report methods gained an importance and were used to record arousal and valence ratings. Along with this approach, the new issue of time lags that were not relevant in conjunction with conventional (post-performance-)ratings became important. A time lag structure was included into the analyses of this work to consider the 'emotional reaction time' of listeners in the HLMs (section 2.2). Schubert (2004) defines this reaction time as

$$t(\text{auditory perception}) + t(\text{cognitive processing [including task decision]}) + t(\text{physical response}) \qquad (2.2)$$

where the first two terms represent the internal processing time of perception and cognition whereas the last term denotes the process of the physical rating action like moving the finger continuously on the iPod touch®. A time lag structure allows an examination of different temporal offsets of peripheral and subjective reactions of participants and enables information to be captured in a way that conventional methods neglecting the time dimension cannot. This is, for example, the possibility to record emotional responses to musical structure as a 'nonlinear' process that stretches, compresses and varies in intensity over time. One of these nonlinearities is the fact that sudden changes of musical features, for example sudden increases of loudness or high frequency power, reduce the response lag (as shown by Schubert, 2004). Another aspect that can be registered are orientation phases at the beginnings and endings of stimuli. In the present data set, these phases can be identified by NaN values and flat rating curves caused by hesitating, unsecure rating behaviour or even missing ratings. During the live concert experiment, S1 and S2 were intended as practice pieces to become familiar with the rating devices and to get used to the concert environment. Figure 2.9 shows that the first orientation phase at the beginning of S1 lasts $\approx 4-6$ seconds but becomes shorter and shorter for the following pieces finally lasting only $\approx 1-2$ seconds at the beginning of S6. For each of the three dVs arousal, valence and SCR, four further versions with time lags of 1, 2, 3 and 4 seconds were created. For this purpose, the data column of the uncompressed source data of each dV was duplicated four times. These duplicates were brought forward by the sample number according to the delay time values before any further processing. That is 256 samples[34] for the 1 second duplicate, 512 samples for the 2 second duplicate, 768 samples for the 3 second duplicate and 1024 samples for the 4 second duplicate. This step was carried out with the function `doLag.m` and introduced an unavoidable but necessary loss of the respective samples at the beginning of the time series. Figure 2.10 shows an example of the time lag structure for arousal, subject ID = 12 and S1 after further processing and downsampling to the final sampling frequency of $f_s = .5Hz$. The period of 4 seconds was chosen on the basis of previous research: Krumhansl (1996) used time lags between 0 and 3.25 seconds in models with tension judgements about piano music, whereas Sloboda and Lehmann (2001) mention reaction times of half to one measure resp. 2-4 seconds in a study with a Chopin piano prelude. Schubert (2004) confirms these results with lags of 2-3 seconds, but refines the issue. He suggests that *sudden*

[34] With $f_s = 256Hz$ as the original sampling frequency of the source data from the Procomp Infiniti® device.

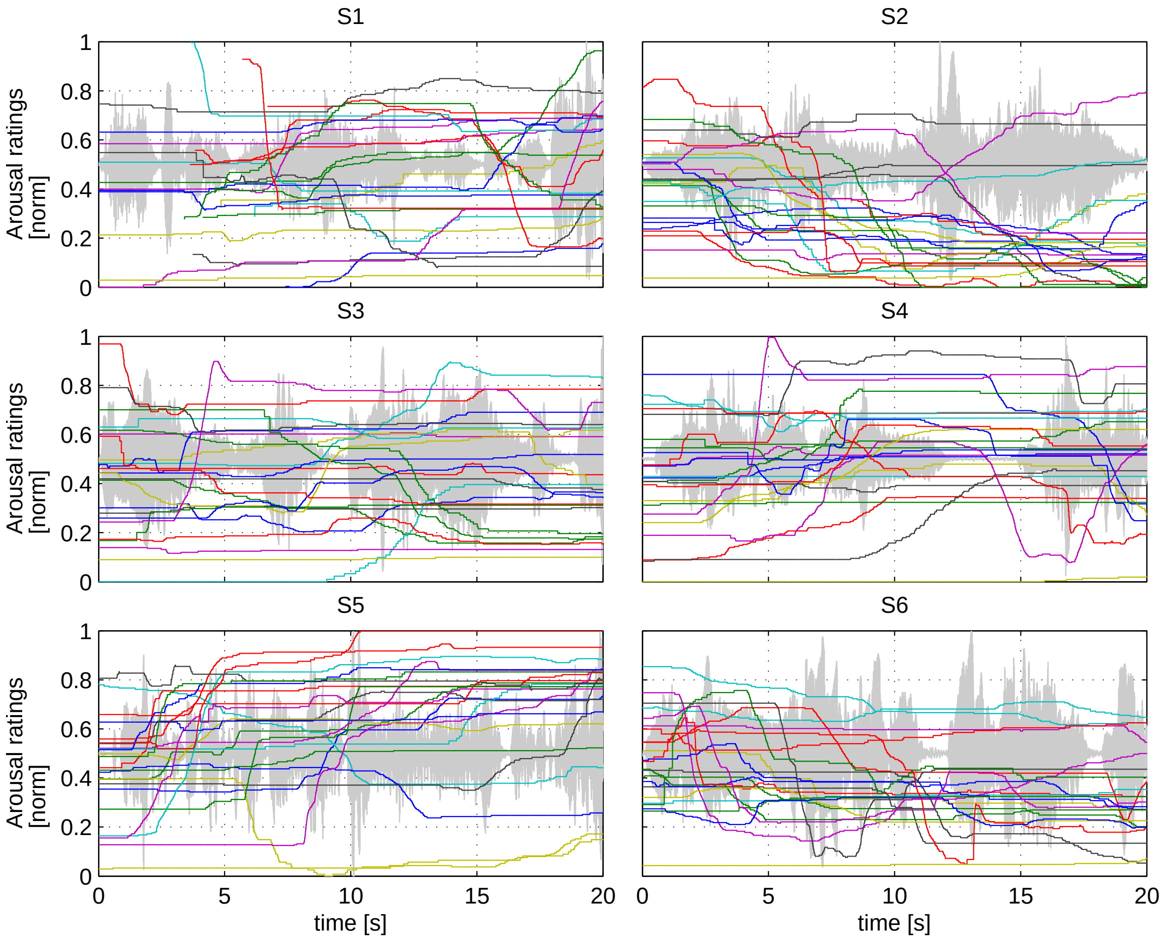


Figure 2.9. Orientation phases (the first 20 seconds of S1-S6).

changes in musical features lead to shorter response times, that is a 0-1 second time delay for the effects of sudden changes of loudness on arousal ratings.

For the predictors associated with the group mean of unexpectedness ratings (HL_UEH and HL_UEL up to here), a fix time lag of 3 seconds was introduced. This value is justified by results of (a) Pearce et al. (2010) who stated that response times to unexpected musical notes ranged from 1357 to 2710 milliseconds (tab. 2, p. 307) and (b) findings of Egermann et al. (2013) showing that an onset of high IC events caused increasing unexpectedness ratings reaching a peak at 5 seconds (fig. 3, p. 12). The higher value of 3 seconds (with regard to Pearce et al., 2010) was chosen to find a middle course here.

Finally, six `.sav`-files for S1-S6 had to be created. These files were assembled from all the single `.csv`-files created in Matlab® during the feature extraction and

Time lags in seconds

arousal_d_z_4	arousal_d_z_3	arousal_d_z_2	arousal_d_z_1	arousal_d_z	time_ArVa	Subj_ID_Ar...	piece_id
.		.	.52241	.52971	37	11	1,00
.	.	.	.	.52971	39	11	1,00
.	.	.	.	.	1	12	1,00
-.84537	-.92270	-.80391	-.87650	-.61130	3	12	1,00
-.84537	-.92270	-.80391	-.87650	-.61130	5	12	1,00
-.84537	-.92270	-.80391	-.87650	-.61130	7	12	1,00
-.84537	-.92270	-.80391	-.87650	-.61130	9	12	1,00
-.16907	-.16019	-.80391	-.87650	-.61130	11	12	1,00
2.05300	1.58270	-.12059	-.10956	-.61130	13	12	1,00
.12077	-.16019	2.12460	1.64340	.02888	15	12	1,00
-.55553	.16660	.17227	-.10956	2.13230	17	12	1,00
.79706	.39750	-.51106	.21913	.30324	19	12	1,00
2.05300	2.87680	.85559	.45137	-.33694	21	12	1,00
1.08690	.52143	2.12460	2.94510	.94342	23	12	1,00
-.84537	.13856	1.14840	.57602	2.13230	25	12	1,00
-.84537	-.92270	-.80391	.19092	1.21780	27	12	1,00
-.16907	-.16019	-.80391	-.87650	-.61130	29	12	1,00
-.26569	-.59591	-.12059	-.10956	-.61130	31	12	1,00
.12077	-.16019	-.21820	-.54781	.02888	33	12	1,00
.	.16660	.17227	-.10956	-.06257	35	12	1,00
.	.	.	.21913	.30324	37	12	1,00
.	.	.	.	-1.80020	39	12	1,00
.	.	.	.	.	1	13	1,00
-.14503	-.06780	-.01203	-.01686	.04817	3	13	1,00

Figure 2.10. Example for the time lag structure (dV arousal and S1, subject ID = 12).

integration processes and during the creation of the time series for the dVs. The merging process was performed with the help of SPSS® syntax[35]. This was the last step before the further analyses, described in detail in the next section, could be started.

2.2 Hierarchical Linear Modelling

There are some key advantages of mixed effects modelling presented in the following. First, it is possible to consider effects like learning or fatigue that develop during the course of an experiment which is obviously of great importance for a live concert experiment. Moreover, it is no longer necessary to average over subjects so that the information about individual responses of subjects survives and can be considered in the models (Baayen et al., 2008). The development of the HLMs is described in detail in the following sections.

[35] For an example portion of the assembling syntax, see appendix E. The six `.sav`-files and the single `.csv`-files including all time series for the following analyses can be found in appendix G.

2.2.1 Variables

Dependent variables The dVs in all of the following models are the measured variables introduced in section 2.1.2 and 2.1.3. They are denoted with new short identifiers for

- arousal (dV_ARO),
- valence (dV_VAL) and
- skin conductance response (dV_SCR)

from here on.

Predictors The predictors in all the following models are the time series of all low, mid and high level features introduced in sections 1.2.3, 1.2.4 and 1.2.5 with the short identifiers already assigned:

- spectral centroid (LL_SPC),
- RMS (LL_RMS),
- brightness (LL_BRN),
- roughness (LL_RGN),
- melodic contour (ML_MEL),
- tempo (ML_TMP),
- high information content (HL_ICH),
- low information content (HL_ICL),
- high unexpectedness (HL_UEH) and
- low unexpectedness (HL_UEL).

These predictors will be called *first order* predictors in the following to distinguish them from the predictors arising from the PCAs described in the next section. The first order predictors are *implicitly* used in the models.

Principal Component Analyses Of First Order Predictors Six PCAs were conducted on the time series of all first order predictors for each stimulus. This step was necessary for four reasons:
(a) The four high level features HL_ICH, HL_ICL, HL_UEH and HL_UEL could be separated as single components and their effects could be made distinguishable among themselves and compared to all other predictors. This was also needed to

test hypothesis H_4.
(b) Linear dependencies between predictors (multicollinearity)[36] could be completely removed. Multicollinearity can cause the following issues: first, the standard errors of the b coefficients in the model results rise with increasing collinearity, resulting in decreasing reliability of the βs and "instability of regression coefficients" (Bortz, 2010, p. 354) of the HLMs. Second, collinearity reduces R and R^2 values that again inform about the variance in the results that the predictors explain, and third it reduces the importance of predictors so that it can hardly be analysed which predictor had a significant influence in a model (cf. Field, 2009, p. 224). The collinearity of predictors was tested for each stimulus using correlation tables[37]. Although Field (2009) mentions values of above .8 – .9 as "very high correlation" (p. 224), the criterion for 'high correlation' was set to $r > .6$ based on discussions with and experiences of the first supervisor.
(c) PCA allowed to obtain a better understanding of the information structure and the effect hiding beneath the combinations of predictors.
(d) Data could be thinned and redundant data could be removed.

The number of components was gradually increased until the four high level features were separated each into one component without any other predictor. The resulting number of principal components (PCs) for each stimulus is indicated in table 6. The column entries below the stimuli abbreviations indicate the position of the respective PC in the original rotated component matrix[38]. Orthogonal varimax rotation was used and the total cumulative variance explained for the stimuli was 97,8% for S1, 92,3% for S2, 96,1% for S3, 98,7% for S4, 97,3% for S5 and 97,3% for S6. The Kaiser–Meyer–Olkin measure shows a mediocre sampling adequacy for the analyses (Field, 2009, p. 671) (KMO = .543 for S1, KMO = .633 for S2, KMO = .691 for S3, KMO = .525 for S4, KMO = .608 for S5 and KMO = .673 for S6). Table 6 shows an overview of the PCs resulting in the *second order* predictors for each stimulus together with the labels used in the following description of the model building process. The second order predictors are *explicitly* used in the models.[39]

---

[36] Strong correlation between two or more predictors in a regression model.
[37] See appendix D.
[38] Ibidem.
[39] Plots of the time series of second order predictors for all stimuli and dVs are placed in appendix C.

Label	S1	S2	S3	S4	S5	S6
PC_MEL	-	-	PC8	PC9	-	PC6
PC_BRN	-	-	-	PC4	PC7	-
PC_TMP	-	-	PC6	PC3	PC4	PC4
PC_TMP_MEL	PC2	PC6	-	-	-	-
PC_RGN	-	-	-	PC2	-	-
PC_RGN_RMS	-	PC1	-	-	PC2	-
PC_RGN_BRN	-	-	-	-	-	PC1
PC_RGN_BRN_RMS	PC1	-	PC1	-	-	-
PC_SPC	PC3	-	PC2	-	-	-
PC_SPC_MEL	-	-	-	-	PC1	-
PC_SPC_RMS	-	-	-	PC1	-	PC2
PC_SPC_BRN	-	PC2	-	-	-	-
PC_ICH	PC6	PC3	PC5	PC5	PC8	PC5
PC_ICL	PC7	PC4	PC3	PC6	PC3	PC8
PC_UEH	PC4	PC7	PC4	PC7	PC5	PC7
PC_UEL	PC5	PC5	PC7	PC8	PC6	PC3
$\sum$ PCs	**7**	**7**	**8**	**9**	**8**	**8**

Table 6

PCs for the six stimuli (S1: Hindemith, S2: Poulenc, S3: Varèse, S4: Debussy, S5: Bach 1, S6: Bach 2).

2.2.2 Model Building

The model building process will be explained on the basis of linear mixed models (LMMs) as a starting point. Then, the final use of HLMs (as the title of this work suggests) will be elaborated.

LMMs can be characterised as parametric linear models for clustered, longitudinal or repeated measurements data. In general, these kinds of models specify a relation between a continuous dV and different predictor variables. Here in particular, it establishes a link between each of the dVs and one particular group of second order predictors (c.f. tab. 6). The six `.sav`-files include repeated measurements made on $N = 24$ listeners for dV_VAL, $N = 24$ listeners for dV_ARO and $N = 48$ for dV_SCR. All listeners are *level 2* variables whereas the dVs (single samples along time for every subject) are *level 1* variables. The repeated measurements are specified by a time ID and a subject ID and were integrated one after another in column vectors that contain predictor time series, dVs time series, subject IDs and time IDs (see fig. 2.10). After the calculation of PCAs, the PC time series were also written into the `.sav`-files via the SPSS® syntax command `/SAVE` to make them available for the following `MIXED` calculations. The stimuli sample numbers (SSNs) were $SSN_{S1} = 19(\sim 38$ sec.), $SSN_{S2} = 39(\sim 78$ sec.), $SSN_{S3} = 104(\sim 208$ sec.), $SSN_{S4} = 77(\sim 154$ sec.), $SSN_{S5} = 56(\sim 112$ sec.) and $SSN_{S6} = 65(\sim 130$ sec.). For the first modelling attempts, time series of individual differences of listeners were also included.

LMMs may include *fixed* factors and *random* factors. Fixed factors are categorical/classification variables or qualitative covariates like gender, region, social stratum or being a musician/non-musician. Here, the second order predictors are inserted as fixed factors. However, random factors are taken into account to control for random variation and autocorrelation of a dV across levels of the random factors. Random factors are often the classification variables that identify level 2 and level 3 units in clustered and repeated-measures/longitudinal data.

Covariance Structure for the *R*-matrix As the covariance structure for the *R*-matrix[40] has been a central issue during modelling and has to be selected based on educated guesses from a wide range of possible structures, it will be focused on in more detail here. It was selected following (a) the *information criteria method*, (b) by *meaning* (Field, 2009, S. 738; Kincaid, 2005, S. 5) and (c) with the goal to expand the concept of Schubert (2004). Following West et al. (2007, p. 16), the general form of the R_i matrix is

$$R_i = Var(\varepsilon_i) = \begin{pmatrix} Var(\varepsilon_{1i}) & cov(\varepsilon_{1i}, \varepsilon_{2i}) & \dots & cov(\varepsilon_{1i}, \varepsilon_{n_i i}) \\ cov(\varepsilon_{1i}, \varepsilon_{2i}) & Var(\varepsilon_{2i}) & \dots & cov(\varepsilon_{2i}, \varepsilon_{n_i i}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\varepsilon_{1i}, \varepsilon_{n_i i}) & cov(\varepsilon_{2i}, \varepsilon_{n_i i}) & \dots & Var(\varepsilon_{n_i i}) \end{pmatrix} \tag{2.3}$$

where ε_i denotes a vector of n_i residuals. Each element of this vector is a residual observed at time point t for subject i. Here, the residuals related to multiple measurements taken on all subjects at the same time may be correlated, i.e. for example the residuals of arousal rating values of subjects $1 \dots 24$ at second 45

[40] For HLMs with random effects *and* repeated measurements, two different covariance structures may be defined. Because random effects were dropped in the final models, the respective *D*-matrix is not considered here.

in the piece are assumed to be correlated. Based on formula 2.3, the first-order autoregressive covariance structure with homogeneous variances (AR1) for the R_i matrix (used by Schubert, 2004) is

$$R_i = Var(\varepsilon_i) = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \dots & \sigma^2\rho^{n_i-1} \\ \sigma^2\rho & \sigma^2 & \dots & \sigma^2\rho^{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho^{n_i-1} & \sigma^2\rho^{n_i-2} & \dots & \sigma^2 \end{pmatrix}. \tag{2.4}$$

This matrix assumes homogeneous variances and correlations and helps controlling for autocorrelation because the correlation between residuals decreases quadratically over time with increasing distance. It only consists of the two parameters σ^2 and ρ with its $\{i, j\}$th element being $\sigma_{ij} = \sigma^2\rho^{|i-j|}$. Autocorrelation arises from continuous ratings and measurements: focusing for example on the finger movement on the iPod touch®, dependencies occur because the finger will never jump as fast from one value to the next and so the pre and the post-jump values will be independent from each other. Instead, the finger will move comparably slow and thus produce dependencies that are on the one hand largely dissolved by downsampling the data (as described in sec. 2.1.4) and, on the other hand, taken into account by using an appropriate covariance structure. Thus, the autoregressive property simulates an 'emotional memory' of the local musical context that influences the continuous listener ratings. From this starting point, the information criteria method involved running HLMs on the dependent variable dV_ARO without time lags on all stimuli for AR1, first-order autoregressive covariance structure with heterogeneous variances (ARH1), first-order autoregressive covariance structure with first-order moving average (ARMA11), Toeplitz (TP) and heterogen Toeplitz (TPH) only for fixed effects and PCs as predictors[41]. The lowest overall values for the information criteria *Akaike's information criterion (AIC)* and *Schwarz's Bayesian criterion (BIC)*[42] were reached with ARH1.

[41] Results of this test can be found in appendix F.1.

[42] Following Field (2009, p. 737), these are the two mostly used criteria besides further criteria like *Hurvich and Tsai's criterion (AICC)* and *Bozdogan's criterion (CAIC)*. All information criteria are adjusted versions of the *-2 Log Likelihood* criterion.

Following IBM Corporation (2012, p. 1260)[43], the ARH1 for the R_i matrix is

$$R_i = Var(\varepsilon_i) = \begin{pmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \dots & \sigma_{n_i}\sigma_1\rho^{n_i-1} \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \dots & \sigma_{n_i}\sigma_2\rho^{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_i}\sigma_1\rho^{n_i-1} & \sigma_{n_i}\sigma_2\rho^{n_i-2} & \dots & \sigma_{n_i}^2 \end{pmatrix}. \quad (2.5)$$

Expanding the approach with AR1, all adjacent measurement time points are assumed to have heterogeneous variances and correlations, thus it consists of $t+1$ parameters σ_i, σ_j and $\rho^{|i-j|}$ with the $\{i,j\}$th element being $\sigma_{ij} = \sigma_i\sigma_j\rho^{|i-j|}$. This increases the number of parameters to be estimated by one for every measurement as well as the time to calculate the models. The variability in the measurements of dV_VAL, dV_ARO and dV_SCR is assumed not to be constant at different points in time, i.e. taking up the above example, the residuals of arousal rating values of subjects $1\dots24$ at second 45 in the piece are assumed not to be correlated. ARH1 was used for all following calculations. All HLMs were run using residual maximum likelihood estimation (REML).

Finally taking up the beginning of the chapter, the relation from the introduced LMMs to HLMs used in this work is established as follows: HLMs consist of LMMs that are specified "in terms of an explicitly defined hierarchy of simpler models, which correspond to the levels of a clustered or longitudinal data set" (West et al., 2007, p. 22).

2.2.3 The Top-Down Strategy

The primary goal of the model-building process is to develop a parsimonious model structure in terms of the number of parameters that offers the best fit to the data. During this iterative process, numerous steps and tests have to be performed against the background of a useful balance between statistical and scientific considerations (West et al., 2007; Cheng et al., 2009).

Different procedures are recommended to fit LMMs. A top-down strategy[44] was selected as generally recommended by Cheng et al. (2009) and also presented

[43] Modified to provide comparability to equation 2.4.
[44] Besides the top-down approach, a step-up strategy is also proposed by West et al. (2007) starting with an initial minimum model with a fixed intercept as the only fixed-effect parameter.

by West et al. (2007, chap. 5) as a good approach for an exemplary repeated-measures data set. Following West et al. (2007), this strategy involves the steps (a) start with a well-specified mean structure for the model, (b) select a structure for the random effects in the model, (c) select a covariance structure for the residuals in the model and finally (d) reduce the model. However, Cheng et al. (2009) recommend the steps (a) specify the maximum model to be considered, (b) specify a criterion of goodness-of-fit of a model, (c) specify a predictor selection strategy, (d) conduct the analysis and (e) evaluate the reliability of the model chosen. The top-down approach assumes two nested models implying that the best reduced desired model is supposed to be 'hidden' inside the larger model. The final model can be uncovered by 'stripping the bark off' the larger model meaning removing redundant items. This uncovering can be realised by applying limitations to the parameters of the maximum model to be considered step by step.

First modelling attempts with LMMs were made considering individual differences of participants as random effects including gender, musical level [non-musician, amateur musician, professional musician], instrumental performance [years], ear training/music theory [years] and half of the big five questions asked on the questionnaires by means of seven-step bipolar rating scales [1: Extraversion, 3: Conscientiousness, 5: Openness, 7: Agreeableness, 9: Emotional stability]. As these attempts were not successful, individual differences were neglected. Although it was desirable to consider them, it is possible that there was too little variation in the data. As described by Grace-Martin (2012), this could have been the reason for the error.

The second top-down LMMs approach was made as follows. To choose random effects, all fixed effects and all random effects (slopes and intercepts) were included with an unstructured (UN) D-matrix and without R-matrix. Then, the same models were run with a variance components (VC) D-matrix. After that, the same was run again just with intercepts without D-matrix and without R-matrix. To choose the R-matrix, all fixed effects were included with (a) no R-matrix, (b) AR1, (c) ARH1, (d) TP and (e) TPH. Unfortunately, the first results of this second approach were not reliable due to error messages occurring that contained problems with the calculation of positive definite parameters for the D-matrix. Different recommendations for troubleshooting were followed as suggested by West et al. (2007, p. 32):

1. choose alternative starting values for covariance parameter estimates,
2. rescale the covariates,
3. based on the design of the study, simplify the model by removing random effects that may not be necessary,
4. fit the implied marginal model,
5. fit the marginal model with an unstructured covariance matrix.

Whereas step 3 made the error disappear, the setting was reduced to a HLM structure without any random effects. This third final, now HLM version resulted in the following model equation (exemplary for S1):

$$\left.\begin{aligned} &Y_{ti} = \\ &\beta_0 + \beta_1 \times PC\_RGN\_BRN\_RMS_{ti} \\ &+ \beta_2 \times PC\_TMP\_MEL_{ti} \\ &+ \beta_3 \times PC\_SPC_{ti} + \beta_4 \times PC\_UEH_{ti} \\ &+ \beta_5 \times PC\_UEL_{ti} + \beta_6 \times PC\_ICH_{ti} \\ &+ \beta_7 \times PC\_ICL_{ti} \end{aligned}\right\} \text{fixed effects} \\ + \varepsilon_{ti} \quad \} \text{residuals.} \tag{2.6}$$

Here, Y_{ti} represents the measure of the continuous response variable to be estimated (dV_ARO, dV_VAL, dV_SCR) and taken on the t-th occasion for the i-th subject. $t(t = 1, \ldots, n_i)$ indexes the n_i longitudinal observations on the dependent variable for a given subject, i.e. the samples of the dVs in the dataset at 1, 3, 5 etc. seconds ($f_s = .5Hz$). For each subject i, it is $t = 1, \ldots, \mathrm{SSN}_{Sj}$ ($j = 1, \ldots, 6$). $i(i = 1, \ldots, m)$ indicates the i-th subject (listener). It is $i = 1, \ldots, 24$ for dV_ARO and dV_VAL resp. $i = 1, \ldots, 48$ for dV_SCR. In the example case of S1, the 7 time-varying second order predictors are included as covariates associated with the fixed effects $\beta_1, \ldots, \beta_7$ for each measurement. The β-coefficients represent the fixed effects of a one-unit change in the corresponding covariates on the mean value of the dependent variable Y assuming that the other covariates remain constant at some value and ε_{ti} represents the error term (R-matrix) associated with the t-th observation on the i-th subject. An SPSS® syntax example is given in figure 2.11[45]. To clarify the analysis part up to here, a graphical summary is given

[45] See appendix E for detailed SPSS® syntax.

in figure 2.12.

```
MIXED arousal_d_z WITH FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) ...
      SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1 | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /REPEATED = time_ArVA | SUBJECT(Subj_ID_ArVa) COVTYPE(ARH1)
  /SAVE=PRED (PRED_2b) RESID (RESID_2b) FIXPRED (FIXPRED_2b).
```

Figure 2.11. Example of SPSS® model syntax.

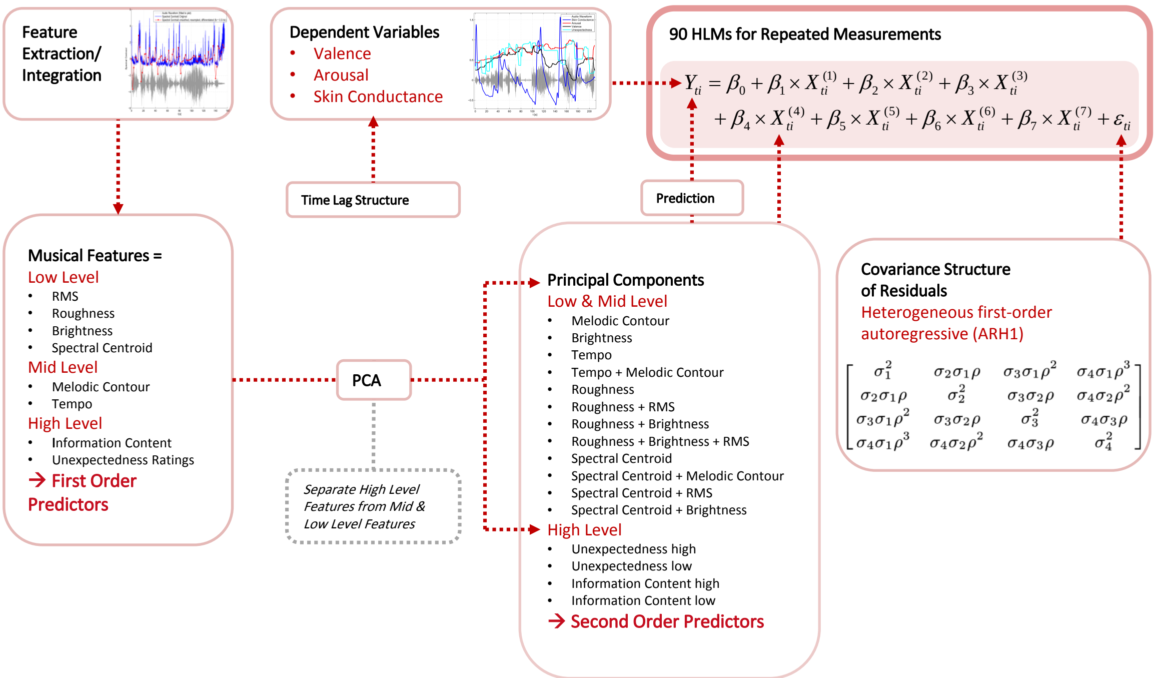
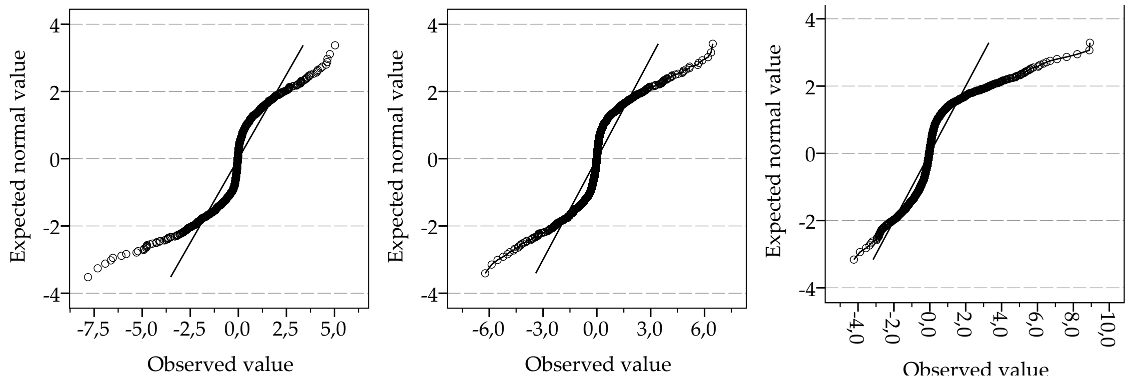


Figure 2.12. Graphical summary of the analysis part.

2.2.4 Diagnostics

The diagnostics were carried out during the whole model building process. The single steps will be described in the following.

Residual Diagnostics Ideally, the residuals associated with different subjects should be independent of each other. As shown in figure 2.13, the independence is rather limited. The attempt has been made to remove cases with residuals ≤ 2, but weighing the lost cases against the little improvement in independence of residuals lead to keeping the initial state. The distribution looks more or less similar for all other models.



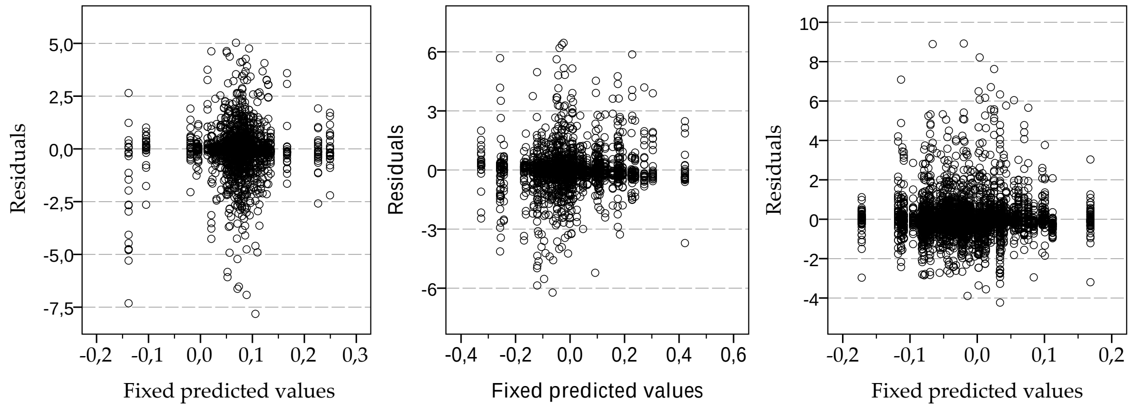
(a) (b) (c)

Figure 2.13. Distribution of conditional residuals for dV_VAL (a), dV_ARO (b) and dV_SCR (c), all exemplary for S3 at time lag 2.

The residual time series for the scatter plots (`RESID_2b`) and the predicted values (`FIXPRED_2b`) were written into the `.sav`-files via `/SAVE` commands within the `MIXED` procedure.

Homogeneity of variance The next step was using scatterplots to investigate the assumption of equal residual variance. Especially in figure 2.14(b) and 2.14(c), slight heteroskedasticity may be supposed. Plots for all the other models indicate similar results, but mostly show homoskedasticity of residuals. Like mentioned above, the attempt has been made to remove outliers removing all cases with residuals ≤ 2 which was not beneficial. The fixed predicted values (`FIXPRED_2b`) were also written into the `.sav`-files via `/SAVE` commands within the `MIXED` procedure.

Autocorrelation of the residuals It is important to check autocorrelation of residuals because (a) autocorrelation represents unmodelled information and (b) significance tests lose relevance (Dean and Bailes, 2010, p. 154). Due to the



(a) (b) (c)

Figure 2.14. Scatterplots to investigate the assumption of equal residual variance for dV_VAL (a), dV_ARO (b) and dV_SCR (c), all exemplary for S3 at time lag 2.

procedures described in section 2.1.4 (downsampling), section 2.2.2 (controlling autocorrelation via ARH1) and section 2.1.2 (differentiation), autocorrelation of residuals could be kept within limits and is high only on the first, sometimes on the second lag which is more or less the same for all models and dVs. Figure 2.15 shows ACFs and PACFs[46] for the first 16 seconds of all three dVs exemplary for S3 at time lag 2[47].

[46] With the PACF, it is possible to identifiy dependencies between two points of a time series (higher order autocorrelation) *excluding* influences from the values between them.

[47] All further plots of ACFs and PACFs are provided in the .spv-files in appendix G

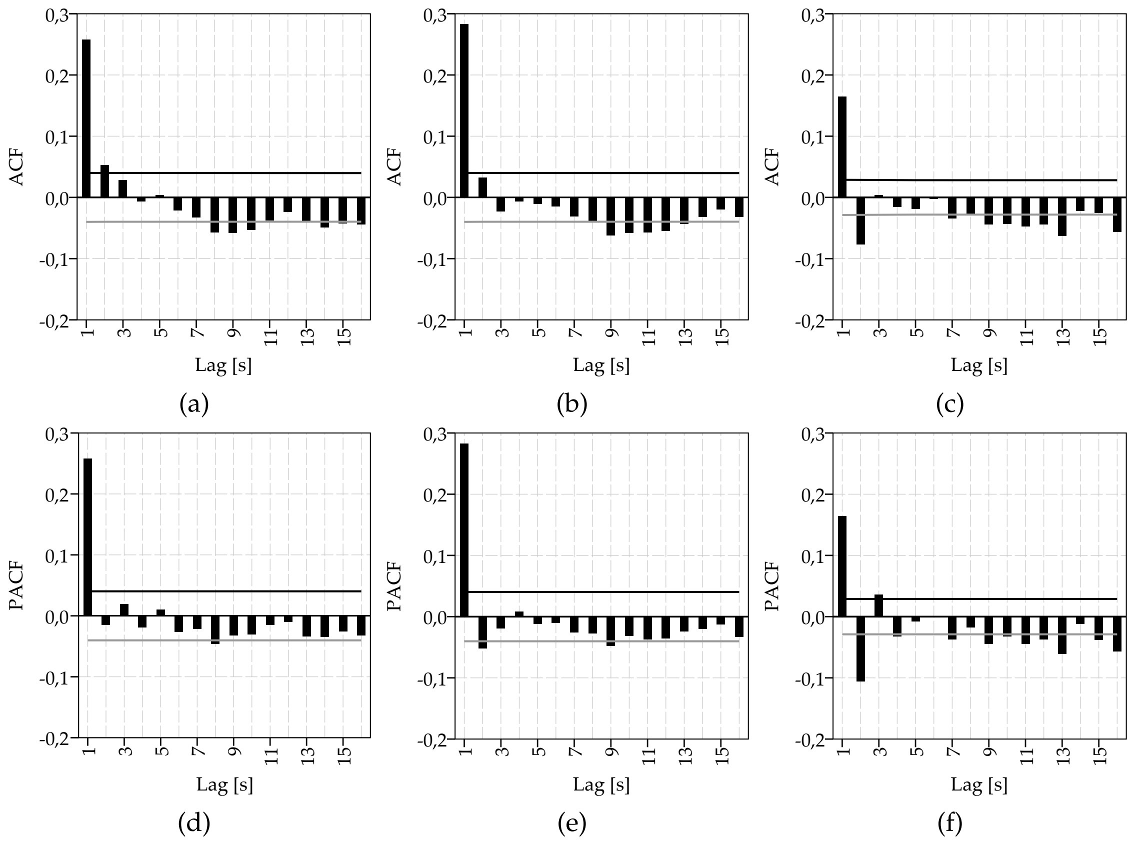


Figure 2.15. ACFs and partial autocorrelation functions (PACFs) of the residuals for the first 16 lags of dV_VAL (a,d), dV_ARO (b,e) and dV_SCR (c,f), all exemplary for S3 at time lag 2.

3 Results

In this section, the results of the 90 HLMs are presented. The induced emotion for each piece will be discussed based on the results data in table 7 (double page) with one example plot for the best model per dV. It should be mentioned again at this point that all results are *gradients*. The plots have been chosen based on the best R^2 values that have been calculated based on the fixed predicted values[48] and the time series of the dVs. In the case that these values were the same for all models, the plots have been chosen visually. In general, the R^2 values turned out very small which may be caused by the high variability due to individual differences in the listener ratings. Therefore, the models will mainly be evaluated based on the number of significant coefficients that indicate a clear relationship between predictors and dVs that was the aim of this work.

S1 - Acht Stücke für Flöte Allein: VI. Lied, leicht bewegt

dV_ARO

As a combination of both mid level features, PC_TMP_MEL shows a significant positive relation at lag 1 ($\beta = .16, p < .01$) which means that an increase in the values of this predictor causes increased arousal ratings with a delay of approximately 1 second. The same relationship exists between arousal ratings and PC_RGN_BRN_RMS, which is a combination of all four low level features, also at lag 1 ($\beta = .11, p < .01$). A rise in PC_ICH causes higher arousal ratings at lag 3 ($\beta = .12, p < .01$). The only negative relationship can be found for PC_UEL at lag 0 ($\beta = -.16, p < .01$) which means that decreasing values of low unexpectedness cause increasing arousal ratings. Figure 3.1 shows the model for time lag 1.

dV_VAL

Contrary to the arousal ratings, a negative relationship consists between dV_VAL and PC_RGN_BRN_RMS ($\beta = -.15, p < .01$). A decrease in PC_ICH causes lower valence ratings ($\beta = -.12, p < .01$). PC_UEH had a negative effect on valence ratings ($\beta = -.01, p < .01$) as expected. Consequently, PC_UEL shows the opposite

[48] Fixed predicted values: the regression means without the random effects.

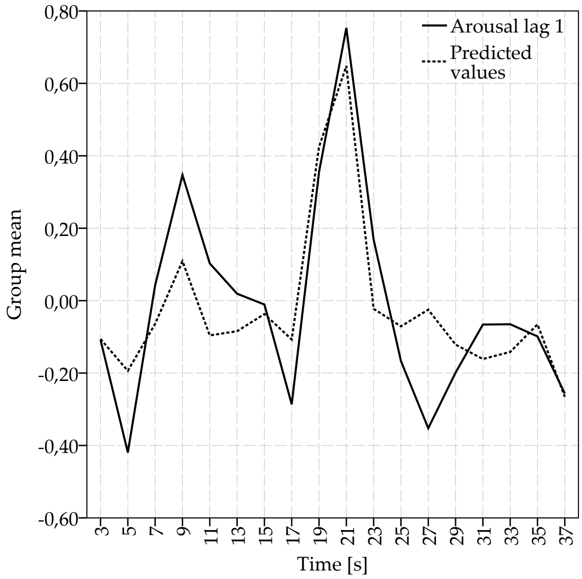


Figure 3.1. Plot of subjective arousal ratings (dV_ARO) and predicted values from HLMs for S1 as a function of time (time lag 1).

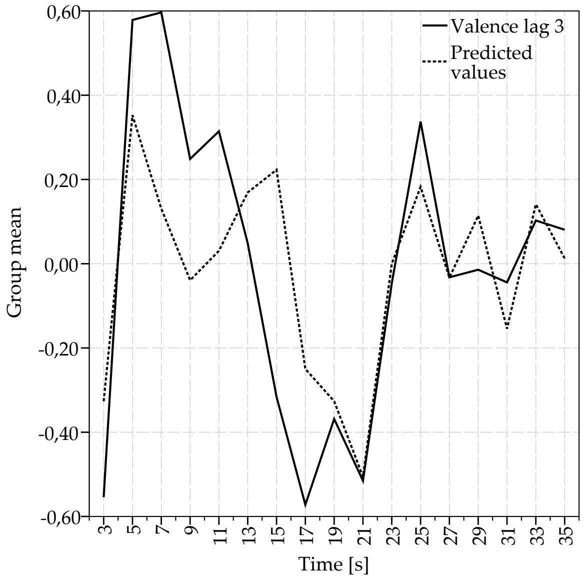


Figure 3.2. Plot of subjective valence ratings (dV_VAL) and predicted values from HLMs for S1 as a function of time (time lag 3).

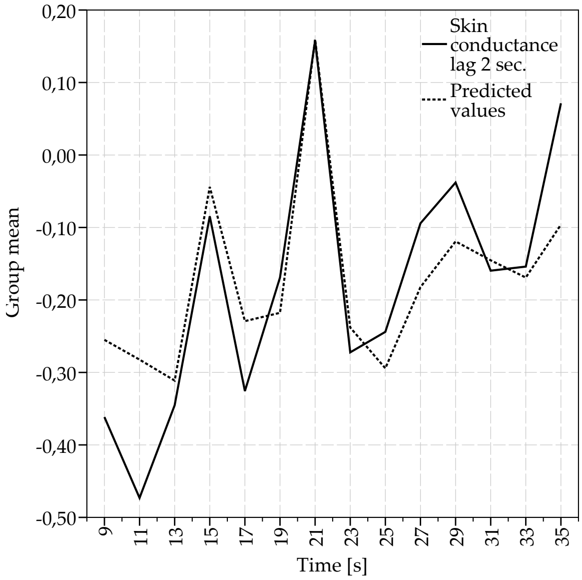


Figure 3.3. Plot of physiological measurements (dV_SCR) and predicted values from HLMs for S1 as a function of time (time lag 2).

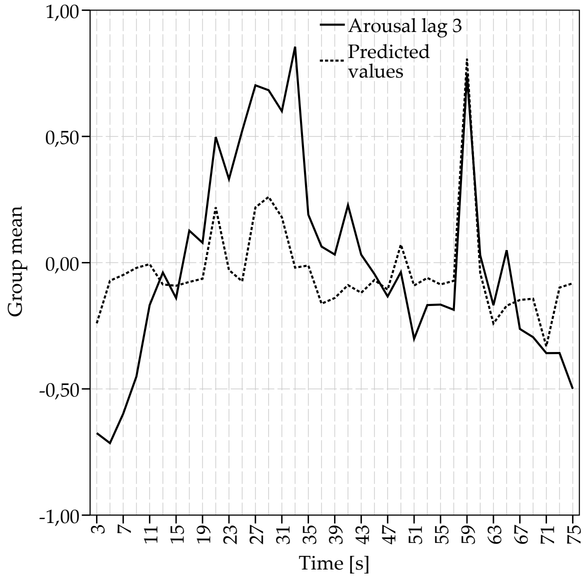


Figure 3.4. Plot of subjective arousal ratings (dV_ARO) and predicted values from HLMs for S2 as a function of time (time lag 3).

effect ($\beta = .11, p < .01$). Figure 3.2 shows the model for time lag 0.

dV_SCR

The most conspicuous and unexpected findings for SCR are the positive relations with PC_ICL on the time lags 0, 1 and 2 ($\beta = .09, p < .01$; $\beta = .12, p < .01$ and $\beta = .09, p < .01$). However, expected findings are the positive relation to PC_UEH ($\beta = .05, p < .01$) and the negative relation to PC_TMP_MEL ($\beta = -.06, p < .01$). Figure 3.3 shows the model for time lag 2.

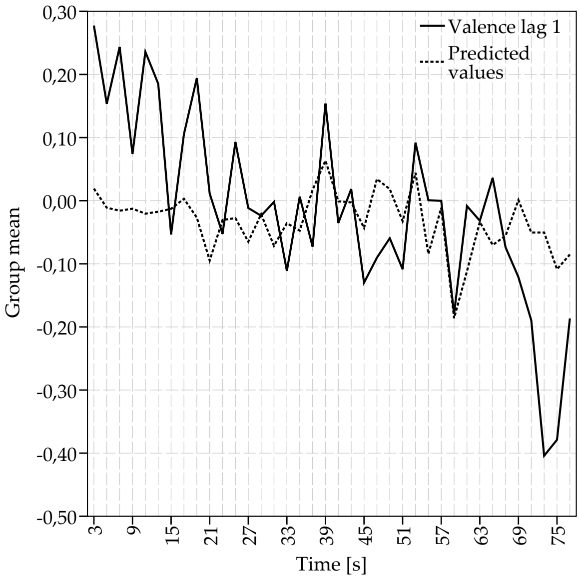


Figure 3.5. Plot of subjective arousal ratings (dV_VAL) and predicted values from HLMs for S2 as a function of time (time lag 1).

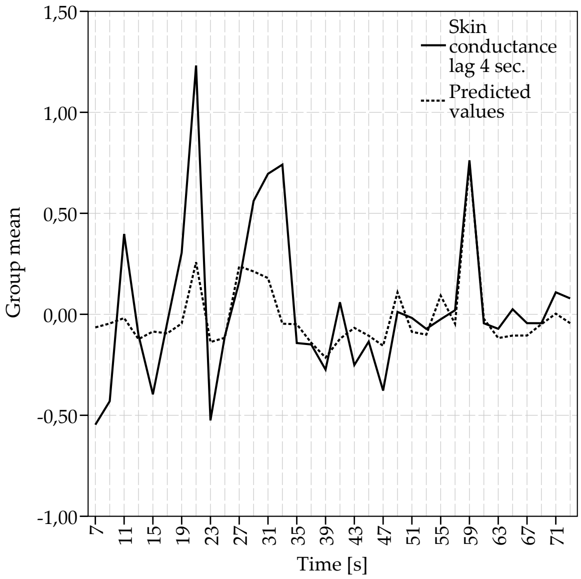


Figure 3.6. Plot of physiological measurements (dV_SCR) and predicted values from HLMs for S2 as a function of time (time lag 4).

S2 - Un Joueur de Flûte Berce les Ruines

dV_ARO

PC_UEH shows two highly significant positive coefficients at time lag 3 and 4 ($\beta = .14, p < .01$; $\beta = .08, p < .01$) which was to be expected. Three further positive significant coefficients appear with PC_RGN_RMS ($\beta = .09, p < .01$), PC_UEL ($\beta = .10, p < .01$) and PC_TMP_MEL ($\beta = .08, p < .01$). Figure 3.4 shows the model for time lag 2.

dV_VAL

dV_VAL does not show any highly significant coefficients for the $p < .01$ criterium. Nevertheless, Figure 3.5 shows the model for time lag 3.

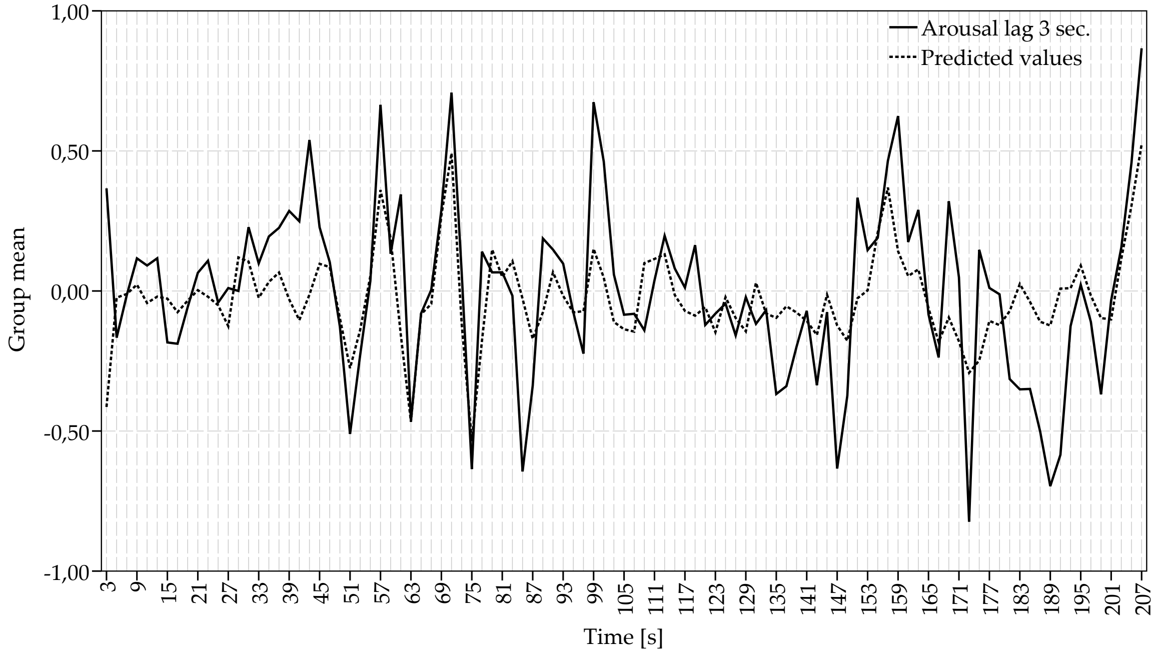


Figure 3.7. Plot of subjective arousal ratings (dV_ARO) and predicted values from HLMs for S3 as a function of time (time lag 3).

dV_SCR

However, dV_SCR shows significant coefficients especially for the low level feature combination PC_RGN_RMS at time lags 2, 3 and 4 ($\beta = .07, p < .01$; $\beta = .11, p < .01$; $\beta = .08, p < .01$). Just like for S1, both mid level features again share one component with PC_TMP_MEL showing a weak positive relationship ($\beta = .03, p < .01$). As expected, PC_UEL and PC_UEH show contrary behaviour ($\beta = -.05, p < .01$ vs. $\beta = .05, p < .01$ and $\beta = .14, p < .01$). A further positive influence comes from PC_SPC_BRN on lag 2 ($\beta = .11, p < .01$). Figure 3.6 shows the model for time lag 4.

S3 - Density 21.5

dV_ARO

Each with three positive significant coefficients, PC_RGN_BRN_RMS (β = .08, p < .01; β = .12, p < .01; β = .09, p < .01) and PC_UEH (β = .06, p < .01; β = .06, p < .01; β = .06, p < .01) seem to have a strong influence on arousal ratings. In addition,

Figure 3.8. Plot of subjective valence ratings (dV_VAL) and predicted values from HLMs for S3 as a function of time (time lag 4).

with two significant coefficients each, PC_SPC (β = .04, p < .01; β = .06, p < .01), PC_ICL (β = −.04, p < .01; β = −.04, p < .01), PC_TMP (β = .03, p < .01; β = .03, p < .01) and PC_UEL (β = .05, p < .01; β = .06, p < .01) also have a clear influence on emotional reactions. Figure 3.7 shows the model for time lag 3.

dV_VAL

The only negative significant coefficient for the valence ratings can be found at PC_RGN_BRN_RMS at time lag 3 (β = −.06, p < .01). Figure 3.8 shows the model for time lag 3.

dV_SCR

In contrast to S1, SCR is negatively related to PC_ICL at time lags 1, 2 and 3 ($\beta = -.04, p < .01$; $\beta = -.04, p < .01$; $\beta = -.03, p < .01$). PC_RGN_BRN_RMS has a positive influence on SCR at three time lags ($\beta = .03, p < .01$; $\beta = .05, p < .01$; $\beta = .06, p < .01$) that seems to increase with the time lags getting higher. Up to now, PC_SPC shows the strongest influence with four significant coefficients at all time lags except lag 3. Here, it is important to note the different directions of relationships on different time lags: on lag 0 and 1, the relationship is negative ($\beta = -.04, p < .01$; $\beta = -.03, p < .01$) but turns positive on lag 3 and 4 ($\beta = .04, p < .01$; $\beta = .05, p < .01$). Another weak negative relationship can be seen for HL_ICH ($\beta = -.02, p < .01$) at lag 3. Figure 3.9 shows the model for time lag 3.

dV_ARO

PCs	Lag 0 β	Sig.	R²	Lag 1 β	Sig.	R²	Lag 2 β	Sig.	R²	Lag 3 β	Sig.	R²	Lag 4 β	Sig.	R²
S1 – Hindemith															
1: PC_RGN_BRN_RMS	,00	,93	,04	,11**	,01	,06	,07	,17	,04	,11*	,01	,05	,07	,14	,02
2: PC_TMP_MEL	,06	,06		,16**	,01		,19	,24		-,06	,67		-,14	,30	
3: PC_SPC	,04	,17		,07*	,04		,11	,08		,13	,12		,09	,18	
4: PC_UEH	-,01	,78		-,02	,76		,05	,22		,06	,06		,02	,66	
5: PC_UEL	-,16**	,00		-,10*	,03		-,03	,58		,03	,54		,03	,49	
6: PC_ICH	,02	,73		,06	,16		,12*	,01		,12**	,01		,01	,80	
7: PC_ICL	,01	,67		,04	,28		,02	,62		-,02	,62		,03	,55	
S2 – Poulenc															
1: PC_RGN_RMS	,01	,67	,01	-,02	,55	,01	,06*	,02	,03	,09**	,00	,06	,07*	,01	,05
2: PC_SPC_BRN	,02	,30		,02	,37		-,02	,52		,03	,31		,02	,43	
3: PC_ICH	,00	,96		,00	,85		,03	,19		,02	,43		-,01	,74	
4: PC_ICL	,02	,40		,01	,52		,03	,11		,01	,78		,01	,53	
5: PC_UEL	,04	,12		-,02	,46		,02	,40		,05*	,03		,10**	,00	
6: PC_TMP_MEL	,04	,08		,06	,06		,08**	,01		,05	,07		,02	,50	
7: PC_UEH	-,04	,09		-,02	,11		,06*	,02		,14**	,00		,08**	,01	
S3 – Varèse															
1: PC_RGN_BRN_RMS	-,01	,31	,01	,03	,05	,02	,08**	,00	,03	,12**	,00	,04	,09**	,00	,04
2: PC_SPC	,04**	,01		,04*	,01		,04*	,01		,06**	,00		-,01	,55	
3: PC_ICL	-,04**	,00		-,04**	,00		-,03	,06		-,02	,12		,01	,73	
4: PC_UEH	,00	,95		,02	,19		,06**	,00		,06**	,00		,06**	,00	
5: PC_ICH	-,02	,19		-,01	,44		,02	,12		,02	,07		,02	,14	
6: PC_TMP	-,01	,40		,03**	,00		,03**	,01		,02	,12		-,01	,63	
7: PC_UEL	-,04*	,02		-,02	,24		,01	,66		,05**	,00		,06**	,00	
8: PC_MEL	,00	,88		,00	,83		,02	,32		-,03	,29		,02	,35	
S4 – Debussy															
1: PC_SPC_RMS	,01	,70	,01	,01	,50	,01	,02	,19	,01	-,01	,70	,01	,01	,72	,02
2: PC_RGN	-,05*	,02		-,02	,42		,02	,23		,03	,11		,01	,35	
3: PC_TMP	,01	,64		,00	,79		,01	,39		,01	,50		,00	,95	
4: PC_BRN	,02	,20		,02	,34		,01	,59		-,03	,10		-,02	,24	
5: PC_ICH	,01	,67		,00	,90		,01	,41		,05*	,01		,04*	,02	
6: PC_ICL	-,02	,34		-,02	,37		,01	,53		,02	,13		,01	,44	
7: PC_UEH	-,01	,25		,00	,73		,00	,74		,05**	,00		,03*	,02	
8: PC_UEL	-,01	,50		-,04	,05		-,02	,32		-,02	,24		,00	,87	
9: PC_MEL	,02	,33		,03*	,03		,02	,26		,02	,45		,02	,43	
S5 – Bach 1															
1: PC_SPC_MEL	,03	,15	,01	,09*	,04	,04	,16**	,00	,07	,20**	,00	,04	,12	,06	,03
2: PC_RGN_RMS	,00	,89		,00	,82		,02	,27		,04*	,01		,03	,06	
3: PC_ICL	,03	,07		,00	,83		-,01	,49		,03	,15		,01	,64	
4: PC_TMP	,01	,65		,03*	,02		,02	,09		,00	,88		-,02	,32	
5: PC_UEH	-,02	,34		,00	,85		,01	,47		,02	,19		,01	,25	
6: PC_UEL	-,02	,05		,00	,99		,02	,36		,01	,72		,00	,89	
7: PC_BRN	,00	,98		-,01	,73		-,03	,11		-,05*	,03		-,01	,66	
8: PC_ICH	,02	,22		,04*	,02		,04**	,01		,01	,81		,00	,95	
S6 – Bach 2															
1: PC_RGN_BRN	-,04	,06	,01	,00	,91	,01	,03	,19	,01	,05*	,01	,01	,03	,20	,01
2: PC_SPC_RMS	,01	,70		-,03	,30		-,02	,32		,01	,78		,01	,72	
4: PC_UEL	,01	,54		,00	,88		-,03	,12		-,02	,22		,01	,46	
3: PC_TMP	,00	,99		,02	,29		,01	,51		,01	,62		,00	1,00	
5: PC_ICH	,05**	,01		,04*	,01		,00	,93		,00	,93		,03	,14	
6: PC_MEL	,03	,27		,04	,23		-,01	,68		-,01	,67		,00	1,00	
7: PC_UEH	,01	,76		,01	,41		,00	,82		-,01	,53		-,03	,09	
8: PC_ICL	-,05*	,01		-,03	,06		-,02	,30		-,01	,74		-,01	,73	

dV_VAL

PCs	Lag 0 β	Sig.	R²	Lag 1 β	Sig.	R²	Lag 2 β	Sig.	R²	Lag 3 β	Sig.	R²	Lag 4 β	Sig.	R²
S1 – Hindemith															
1: PC_RGN_BRN_RMS	-,01	,85	,04	-,02	,74	,04	-,08	,06	,05	-,15**	,00	,07	-,08	,09	,04
2: PC_TMP_MEL	-,05	,27		-,06	,39		-,17	,28		,10	,54		,15	,43	
3: PC_SPC	-,01	,75		,01	,84		,02	,69		,01	,82		-,22*	,03	
4: PC_UEH	-,09*	,01		-,07	,10		-,10**	,01		-,08	,06		-,01	,70	
5: PC_UEL	,12*	,01		,08	,09		,07	,13		,11**	,01		-,04	,33	
6: PC_ICH	-,12**	,00		-,06	,21		,02	,68		-,01	,78		-,08	,07	
7: PC_ICL	,03	,40		-,01	,87		-,07	,13		-,02	,63		,03	,49	
S2 – Poulenc															
1: PC_RGN_RMS	-,01	,62	,01	-,03	,21	,01	,00	,84	,00	,02	,54	,00	,00	,90	,00
2: PC_SPC_BRN	-,02	,30		-,05*	,05		-,02	,49		,01	,78		,01	,84	
3: PC_ICH	,00	,84		,01	,76		-,01	,48		-,01	,73		,01	,43	
4: PC_ICL	-,01	,45		-,03	,12		,00	,87		,01	,52		,02	,21	
5: PC_UEL	,05*	,02		,02	,23		-,02	,33		-,02	,38		,04	,09	
6: PC_TMP_MEL	,02	,47		-,02	,52		,01	,80		,03	,27		,03	,38	
7: PC_UEH	,00	,97		-,03*	,04		-,02	,32		,01	,72		,00	,85	
S3 – Varèse															
1: PC_RGN_BRN_RMS	,00	,71	,00	,01	,55	,00	-,02	,18	,01	-,06**	,00	,02	-,02	,20	,03
2: PC_SPC	,02	,09		,02	,17		-,01	,24		-,02	,19		,01	,62	
3: PC_ICL	,01	,57		,01	,67		,00	,82		,01	,48		-,02	,29	
4: PC_UEH	,00	,99		,00	,73		-,01	,51		,00	,77		-,02	,45	
5: PC_ICH	-,01	,53		-,01	,53		-,01	,16		,00	,97		,00	,69	
6: PC_TMP	-,01	,53		-,01	,27		-,01	,31		,00	,90		,00	,74	
7: PC_UEL	,02	,09		,02	,13		-,01	,34		,00	,71		,01	,65	
8: PC_MEL	-,02	,23		-,01	,63		,03	,18		,01	,49		-,03*	,05	
S4 – Debussy															
1: PC_SPC_RMS	,01	,65	,01	,00	,79	,01	,01	,56	,01	,03	,07	,01	,01	,74	,00
2: PC_RGN	,01	,68		,01	,65		,00	,87		-,01	,64		,02	,21	
3: PC_TMP	,02	,27		,02	,28		,02	,11		,01	,57		-,02	,10	
4: PC_BRN	-,01	,56		,00	,95		-,03	,17		-,04*	,02		-,03	,08	
5: PC_ICH	,00	,81		,03	,12		,06**	,01		,06**	,00		,00	,91	
6: PC_ICL	,01	,61		,02	,30		,02	,28		,02	,19		-,01	,71	
7: PC_UEH	,01	,58		,00	,78		,00	,98		,01	,37		,03	,15	
8: PC_UEL	-,03	,07		-,03	,09		-,04**	,01		,00	,81		,01	,42	
9: PC_MEL	,00	,84		,00	,84		,04	,16		,04	,18		,02	,48	
S5 – Bach 1															
1: PC_SPC_MEL	-,01	,83	,00	,02	,51	,00	,11**	,01	,03	,21**	,00	,06	,12	,06	,05
2: PC_RGN_RMS	,02	,22		,00	,94		-,01	,73		,01	,40		,03	,09	
3: PC_ICL	,00	,91		,00	,83		-,01	,43		,03	,10		-,01	,78	
4: PC_TMP	,01	,24		,02	,17		,00	,74		,02	,17		,00	,78	
5: PC_UEH	,00	,96		,00	,86		-,01	,59		-,02	,10		-,03*	,04	
6: PC_UEL	,02	,10		,01	,27		,01	,66		,01	,71		-,01	,58	
7: PC_BRN	-,01	,37		-,01	,54		-,06*	,01		-,06**	,01		-,04	,06	
8: PC_ICH	-,01	,42		,01	,77		-,01	,75		,04*	,04		,03	,17	
S6 – Bach 2															
1: PC_RGN_BRN	,00	,82	,00	-,02	,18	,00	-,02	,24	,00	-,02	,16	,00	,01	,53	,00
2: PC_SPC_RMS	-,02	,33		-,03	,27		,01	,75		,00	,87		,05	,05	
4: PC_UEL	,01	,61		,00	,81		-,01	,62		,00	,82		,00	,95	
3: PC_TMP	,00	,89		-,01	,41		-,02	,29		,00	,77		,03	,07	
5: PC_ICH	,02	,26		-,02	,27		,00	,77		,02	,36		-,01	,45	
6: PC_MEL	-,01	,72		,01	,79		,01	,68		,03	,53		,02	,56	
7: PC_UEH	,00	,82		,01	,50		,00	,84		,02	,36		,03	,06	
8: PC_ICL	,00	,79		,00	,92		-,01	,63		,01	,60		,01	,62	

dV_SCR

PCs	Lag 0 β	Sig.	R²	Lag 1 β	Sig.	R²	Lag 2 β	Sig.	R²	Lag 3 β	Sig.	R²	Lag 4 β	Sig.	R²
S1 – Hindemith															
1: PC_RGN_BRN_RMS	-,05*	,02	,06	-,02	,50	,06	,02	,57	,06	,07*	,03	,02	,08*	,01	,01
2: PC_TMP_MEL	-,06**	,00		-,02	,54		,03	,73		-,02	,73		-,06	,45	
3: PC_SPC	,09	,07		,15*	,04		,12	,20		,02	,83		-,13	,20	
4: PC_UEH	,05**	,00		,03	,09		,00	,80		-,01	,49		,00	,81	
5: PC_UEL	-,03	,19		-,01	,67		-,01	,65		,00	,94		,02	,49	
6: PC_ICH	,03	,22		-,01	,76		-,04	,08		-,01	,72		,01	,67	
7: PC_ICL	,09**	,00		,12**	,00		,09**	,00		,02	,60		-,08*	,02	
S2 – Poulenc															
1: PC_RGN_RMS	,02	,27	,00	,01	,61	,02	,07**	,00	,03	,11**	,00	,04	,08**	,00	,06
2: PC_SPC_BRN	,03	,15		,04	,11		,11**	,00		,06*	,05		,05	,07	
3: PC_ICH	-,02	,33		-,01	,62		,00	,83		,02	,18		,02	,18	
4: PC_ICL	,00	,91		,01	,64		,01	,68		,01	,42		,02	,17	
5: PC_UEL	-,03	,07		-,05**	,00		-,02	,13		,00	,93		,01	,50	
6: PC_TMP_MEL	,02	,12		,03**	,01		,03*	,01		,03*	,03		-,01	,42	
7: PC_UEH	,01	,68		-,01	,46		,01	,51		,05**	,00		,14**	,00	
S3 – Varèse															
1: PC_RGN_BRN_RMS	,00	,94	,00	-,01	,44	,01	,03**	,00	,01	,05**	,00	,01	,06**	,00	,01
2: PC_SPC	-,04**	,00		-,03**	,00		,00	,75		,04**	,00		,05**	,00	
3: PC_ICL	-,02	,05		-,04**	,00		-,04**	,00		-,03**	,00		-,01	,28	
4: PC_UEH	,00	,56		-,01	,05		-,01	,20		-,01	,22		,01	,53	
5: PC_ICH	,00	,67		,00	,91		-,01	,14		-,02**	,00		-,01	,39	
6: PC_TMP	-,01	,28		-,01	,19		,01	,29		,01	,15		,00	,99	
7: PC_UEL	,01	,20		-,01	,33		,00	,92		,02*	,01		,02*	,03	
8: PC_MEL	,02*	,03		,02*	,02		,01	,63		-,01	,71		,00	,83	
S4 – Debussy															
1: PC_SPC_RMS	,01	,28	,01	,02	,10	,01	,00	,99	,01	-,02	,24	,01	,00	,81	,01
2: PC_RGN	,00	,74		,00	,72		,01	,68		,01	,71		,03*	,02	
3: PC_TMP	,00	,79		,02*	,03		,04**	,00		,02*	,02		,01	,62	
4: PC_BRN	-,03**	,01		-,02	,16		,00	,89		,03*	,01		,02	,06	
5: PC_ICH	,02	,20		,01	,65		-,04**	,01		-,08**	,00		-,04*	,02	
6: PC_ICL	,04**	,01		,01	,50		-,04**	,01		-,02	,14		-,02	,10	
7: PC_UEH	,02*	,05		,03**	,00		,03**	,00		,05**	,00		,03**	,00	
8: PC_UEL	-,04**	,00		-,02	,20		,01	,68		-,01	,48		-,03**	,01	
9: PC_MEL	,01	,81		,01	,56		-,03	,30		-,07*	,02		-,06	,08	
S5 – Bach 1															
1: PC_SPC_MEL	-,03	,20	,00	,01	,76	,00	-,06	,11	,00	-,01	,91	,00	,13	,10	,00
2: PC_RGN_RMS	,00	,83		,00	,79		-,01	,52		-,01	,69		,01	,69	
3: PC_ICL	,01	,52		,02	,23		,00	,76		-,01	,42		,00	,84	
4: PC_TMP	,00	,91		-,01	,32		,00	,78		,02	,09		,01	,29	
5: PC_UEH	-,02	,10		,00	,77		,02	,13		,02	,10		,01	,28	
6: PC_UEL	,01	,65		,00	,85		,00	,98		-,01	,49		-,02	,07	
7: PC_BRN	,01	,70		-,02	,34		,00	,97		,00	,83		-,01	,51	
8: PC_ICH	,00	,90		,00	,84		-,03*	,04		,00	,83		,00	,88	
S6 – Bach 2															
1: PC_RGN_BRN	-,02	,12	,01	,02	,24	,00	,03*	,02	,00	,02	,07	,01	,03*	,02	,01
2: PC_SPC_RMS	-,09**	,00		-,07**	,00		-,04*	,04		-,03	,25		,02	,48	
4: PC_UEL	,01	,56		,00	,64		,00	,89		,02	,13		-,01	,64	
3: PC_TMP	,03**	,00		,00	,80		-,02	,13		,00	,79		,01	,32	
5: PC_ICH	-,01	,56		-,01	,24		,00	,95		,01	,63		,00	,77	
6: PC_MEL	,00	,99		,00	,99		-,02	,39		-,10**	,00		-,14**	,00	
7: PC_UEH	,01	,39		,01	,49		,00	,82		-,02	,12		,01	,46	
8: PC_ICL	,00	,68		,01	,25		,03**	,01		,02	,06		,01	,35	

Table 7

*Results of the HLMs (*p<.05, **p<.01). Underlined R^2 values mark plotted models.*

S4 - Syrinx

dV_ARO

For the arousal ratings, only PC_UEH shows a positive significant coefficient at lag 3 ($\beta = .05, p < .01$). Figure 3.10 shows the model for time lag 3.

dV_VAL

Here, a double positive relation occurs between dV_VAL and PC_ICH at time lags 2 and 3 ($\beta = .06, p < .01$; $\beta = .06, p < .01$). The negative effect of PC_UEH on valence ratings ($\beta = -.04, p < .01$) is as before at S1. Figure 3.11 shows the model for time lag 2.

dV_SCR

The strongest influence on dV_SCR comes from PC_UEH on all lags except lag 0 ($\beta = .03, p < .01$; $\beta = .03, p < .01$; $\beta = .05, p < .01$; $\beta = .03, p < .01$). As expected, PC_UEL shows negative effects on lag 0 and 4 ($\beta = -.04, p < .01$; $\beta = -.03, p < .01$). Concerning IC, it is interesting to note that PC_ICH shows two negative coefficients at lag 2 and 3 ($\beta = -.04, p < .01$; $\beta = -.08, p < .01$) and that PC_ICL shows the same negative coefficient at lag 3 ($\beta = -.08, p < .01$) but a positive one at lag 0 ($\beta = .04, p < .01$). PC_BRN again has a negative coefficient at lag 0 ($\beta = -.03, p < .01$), and finally PC_TMP shows a positive influence on dV_SCR as the only component member (like the influence on dV_ARO with S3) which is as expected. Figure 3.12 shows the model for time lag 3.

S5 - Solo Partitas for Flute, A-minor: "Corrente"

dV_ARO

It is the only time with PC_SPC_MEL that a low and a mid level feature share one component with positive coefficients at time lags 2 and 3 ($\beta = .16, p < .01$; $\beta = .20, p < .01$). One further positive influence comes from PC_ICH at lag 2 ($\beta = .04, p < .01$). Figure 3.13 shows the model for time lag 3.

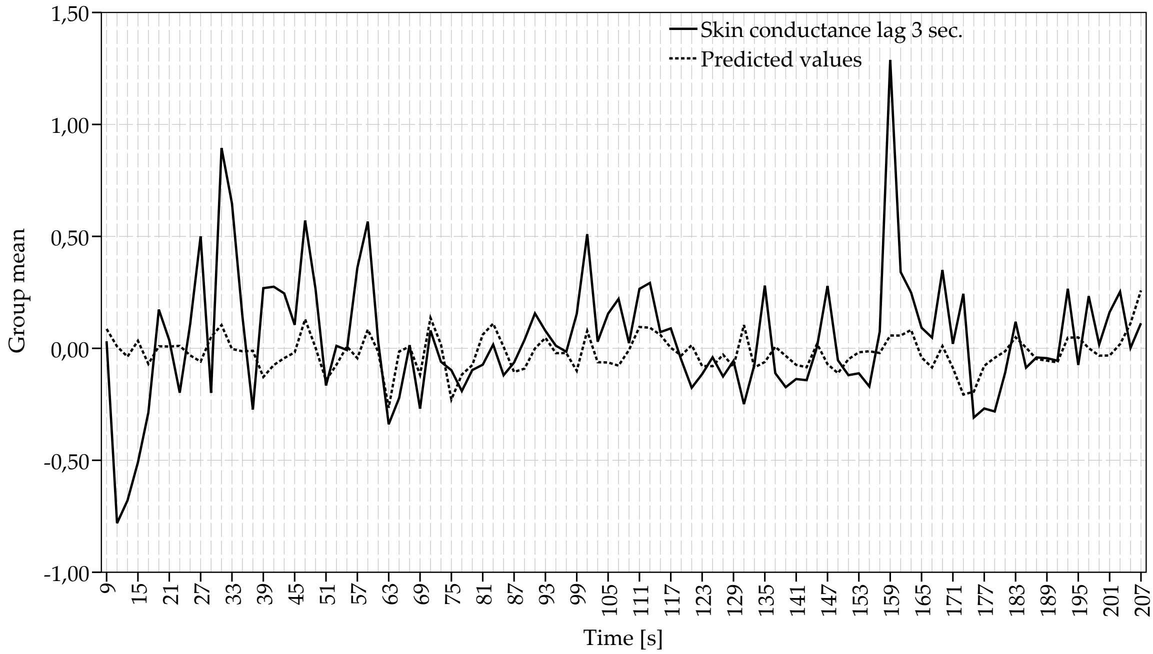


Figure 3.9. Plot of physiological measurements (dV_SCR) and predicted values from HLMs for S3 as a function of time (time lag 3).

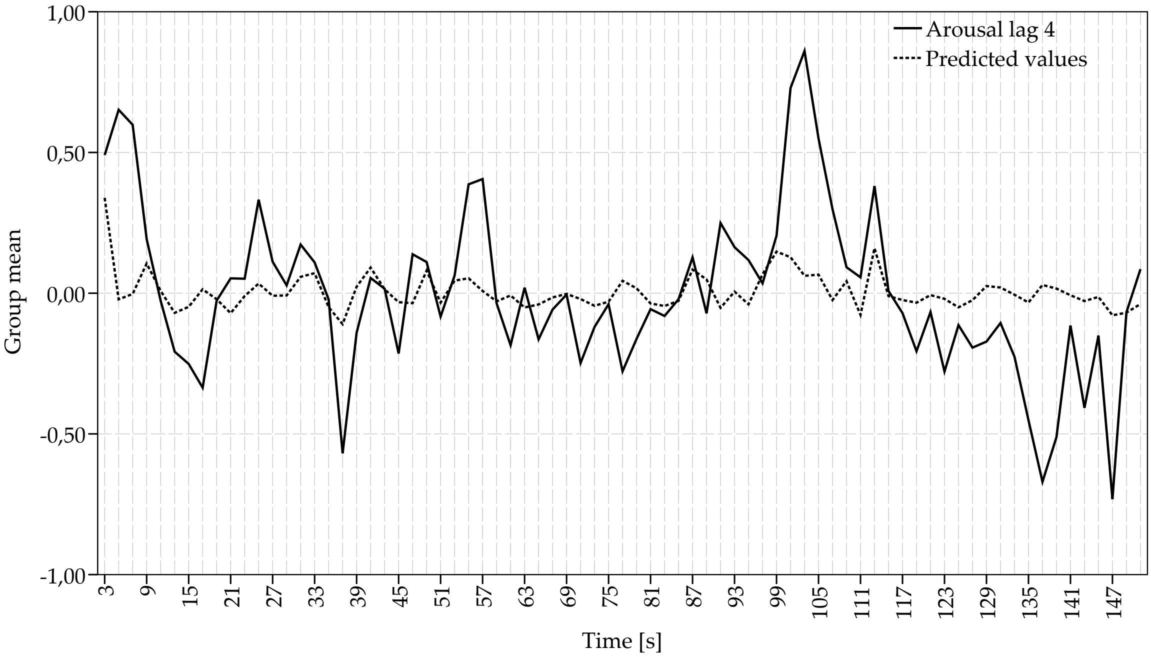


Figure 3.10. Plot of subjective arousal ratings (dV_ARO) and predicted values from HLMs for S4 as a function of time (time lag 4).

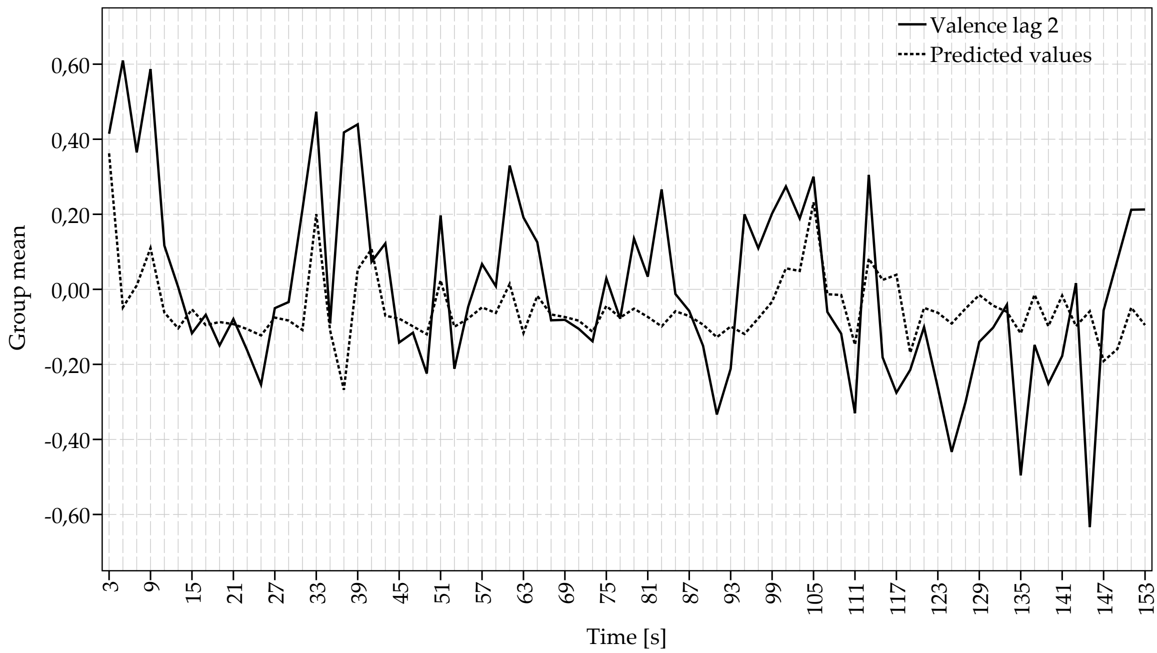


Figure 3.11. Plot of subjective valence ratings (dV_VAL) and predicted values from HLMs for S4 as a function of time (time lag 2).

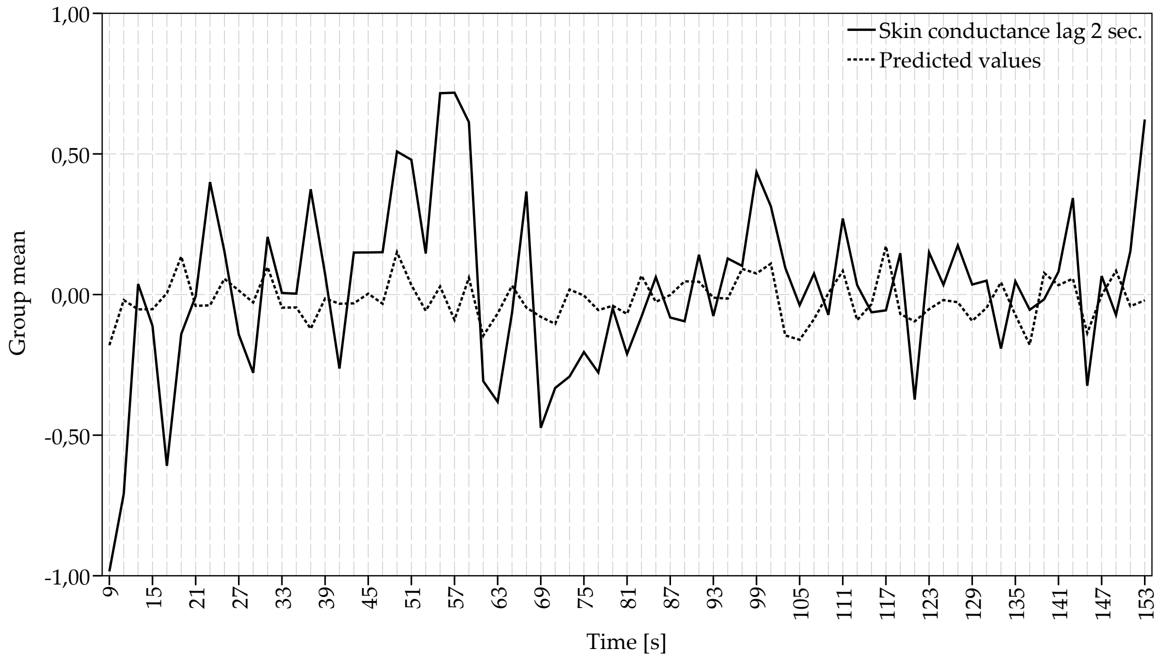


Figure 3.12. Plot of physiological measurements (dV_SCR) and predicted values from HLMs for S4 as a function of time (time lag 2).

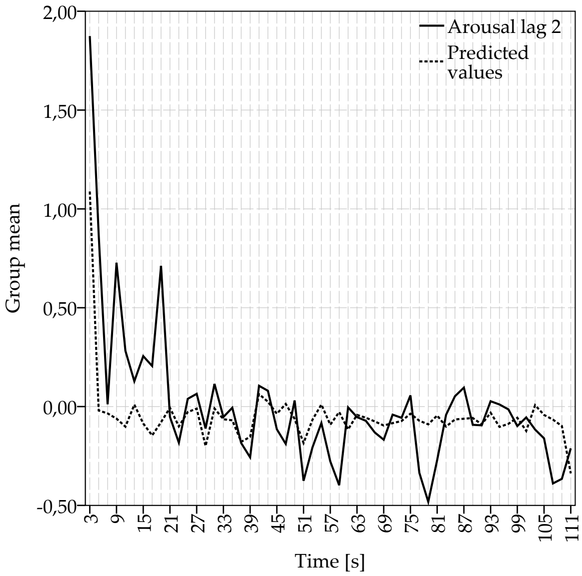


Figure 3.13. Plot of subjective arousal ratings (dV_ARO) and predicted values from HLMs for S5 as a function of time (time lag 2).

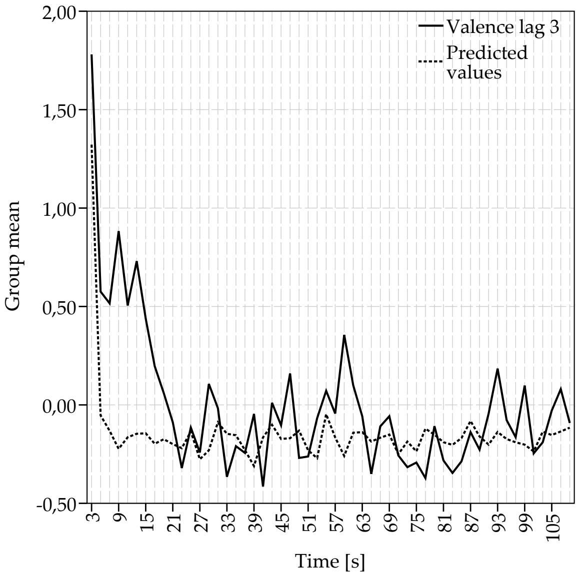


Figure 3.14. Plot of subjective valence ratings (dV_VAL) and predicted values from HLMs for S5 as a function of time (time lag 3).

dV_VAL

Like with arousal, PC_SPC_MEL also shows a positive influence on valence ratings at the same time lags 2 and 3 ($\beta = .11, p < .01$; $\beta = .21, p < .01$). One further negative influence can be observed from PC_BRN at time lag 2 ($\beta = -.06, p < .01$). Figure 3.14 shows the model for time lag 2.

dV_SCR

dV_SCR of S5 does not show any highly significant coefficients for the $p < .01$ criterium. Nevertheless, figure 3.15 shows the model for time lag 4.

S6 - Solo Partitas for Flute, A-minor: "Sarabande"

dV_ARO With PC_ICH ($\beta = .05, p < .01$), only one significant coefficient can be observed. Figure 3.16 shows the model for time lag 0.

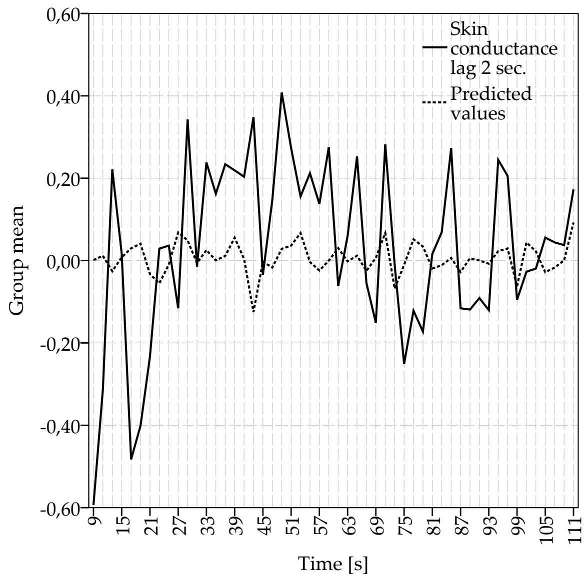


Figure 3.15. Plot of physiological measurements (dV_SCR) and predicted values from HLMs for S5 as a function of time (time lag 2).

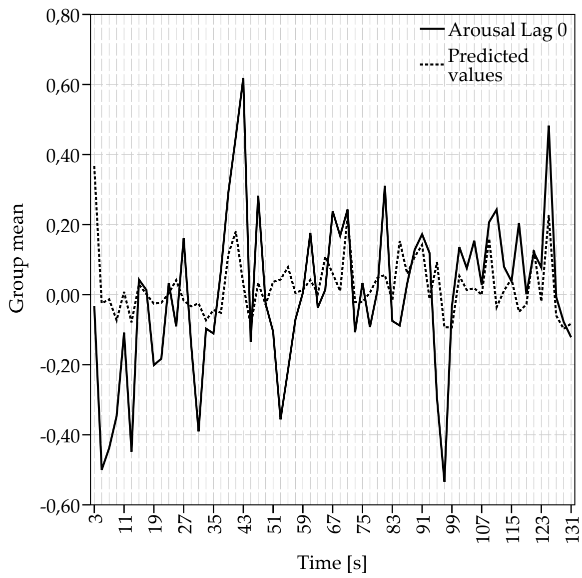


Figure 3.16. Plot of subjective arousal ratings (dV_ARO) and predicted values from HLMs for S6 as a function of time (time lag 0).

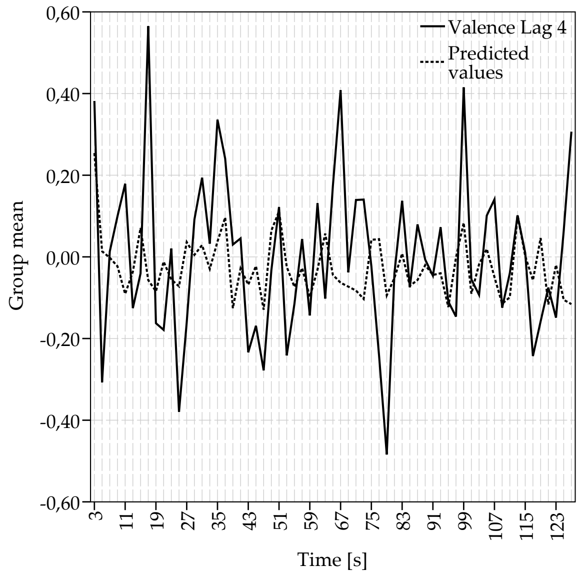


Figure 3.17. Plot of subjective valence ratings (dV_VAL) and predicted values from HLMs for S6 as a function of time (time lag 4).

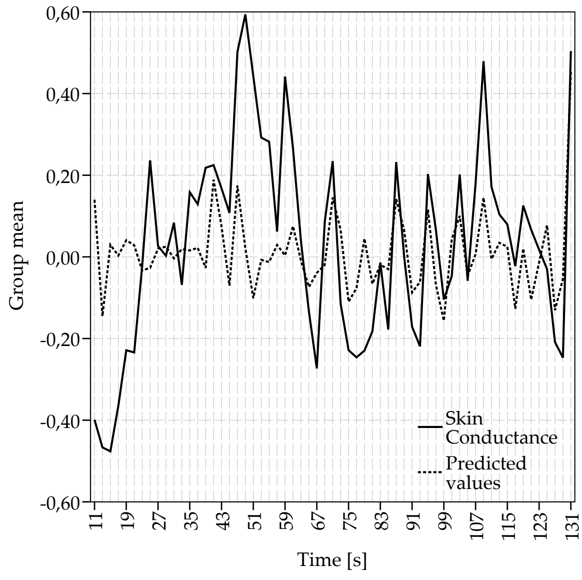


Figure 3.18. Plot of physiological measurements (dV_SCR) and predicted values from HLMs for S6 as a function of time (time lag 0).

dV_VAL

dV_VAL of S6 does not show any significant coefficients for the $p < .01$ criterium. Figure 3.17 shows the model for time lag 4.

dV_SCR

A strong negative influence on dV_SCR comes from PC_MEL as the only high level component part on lags 3 and 4 ($\beta = -.10, p < .01$; $\beta = -.14, p < .01$). A double negative influence comes from PC_SPC_RMS at time lags 0 and 1 ($\beta = -.09, p < .01$; $\beta = -.07, p < .01$), followed by weak positive influences from PC_TMP at lag 0 ($\beta = .03, p < .01$) and by PC_ICL at lag 2 ($\beta = .03, p < .01$). For IC, PC_ICL shows a positive significant coefficient at time lag 2 ($\beta = .03, p < .01$). Figure 3.18 shows the model for time lag 0.

4 Discussion

In the previous chapters, the secondary analysis of a dataset collected during a live concert experiment with solo flute performance was presented. Autoregressive hierarchical linear modelling was used to investigate if time series from continuous ratings of the subjective components valence (hypothesis H_1) and arousal (hypothesis H_2) as well as continuous measurements of skin conductance (hypothesis H_3) could each be predicted by a group of different low, mid and high level musical features as predictors. To enable the examination of H_4, PCAs were applied to all first order predictors to get a group of second order predictors and to reach a separation of firstly, the low and mid level features and secondly, the high level features. In the following, the hypotheses are checked and possible explanations for the influences of the predictors on the three dVs will be derived from the mechanisms of the BRECVEM framework introduced in section 1.1.1. For this purpose, a closer look is taken at the significant coefficients and especially their kind of influence (positive or negative). Table 8 shows an overview of the kind of influence of the coefficients for each PC and the three dVs. The connection between the counts in this table and the four psychological mechanisms is finally established in table 9. Both tables are the basis for the following considerations.

4.1 Checking The Hypotheses

H_1 – H_3: Predictability of Arousal, Valence and SCR For dV_ARO, table 9 shows exclusively positive influences from the predictors for brain stem reflex, rhythmic entrainment and emotional contagion. Musical expectancy is the mechanism that also two negative influences appear at. However, for dV_VAL, table 9 shows a balance of positive and negative influences from the predictors for all mechanisms except rhythmic entrainment where no influences were found. Balanced positive and negative influences from all predictors except PC_SPC_MEL for all four mechanisms appear for dV_SCR.

Brain stem reflex Triggered by acoustic signals that are loud, dissonant, sudden and fast in tempo, the brain stem reflex is assumed to be closely linked to second order predictors that involve the first order predictors LL_RMS, LL_RGN,

	positive						negative						+/-
PC label / piece number (S1-S6)	1	2	3	4	5	6	1	2	3	4	5	6	
dV_ARO													
PC_TMP			2										+
PC_TMP_MEL	1	1											+
PC_RGN_RMS		1											+
PC_RGN_BRN_RMS	1		3										+
PC_SPC			2										+
PC_SPC_MEL					2								+
PC_UEH		2	3	1									+
PC_UEL		1	2				1						(+)
PC_ICH	1				1	1							+
PC_ICL									2				-
dV_SCR													
PC_MEL												2	-
PC_BRN										1			-
PC_TMP				1		1							+
PC_TMP_MEL		1					1						0
PC_RGN_RMS		3											+
PC_RGN_BRN_RMS			3										+
PC_SPC			2						2				0
PC_SPC_RMS												2	-
PC_SPC_BRN		1											+
PC_UEH	1	2		4									+
PC_UEL								1		2			-
PC_ICH									1	2			-
PC_ICL	3			1		1			3	1			(+)
dV_VAL													
PC_BRN											1		-
PC_RGN_BRN_RMS							1		1				-
PC_SPC_MEL					2								+
PC_UEH							1						-
PC_UEL	1									1			0
PC_ICH				2			1						(+)

Table 8

Overview of the kind of influence (positive or negative) of the coefficients. The last column indicates the main influences of each PC across all stimuli for the three dVs. Ambiguous values are set in parentheses, balanced values are marked by 0.

LL_BRN, LL_SPC and ML_MEL. For LL_RMS and LL_RGN, this may be obvious because they represent these properties by definition. For LL_BRN and LL_SPC, it is supposed that fast changes in the high frequency spectrum may trigger the mechanism. Related to the performance, these changes could arise from transient respiratory sounds in the pauses and strong blowing sounds emitted from the mouthpiece as already mentioned in section 1.2.3. Additionally, very high tones from the flute could contribute here. For ML_MEL, fast tonal jumps could be the reason for brain stem reflex to be activated. The second order predictors assumed to be involved in activating this mechanism are coloured in table 9, first column.

Regarding this mechanism, the A4+ result for the predictor PC_RGN_BRN_-RMS is conspicuous and as expected: it combines portions of roughness, brightness and RMS changes and through an interaction that the lens model (fig. 1.7) might explain, seems to have a strong positive influence on the listeners' emotional judgements of arousal. As stated at the end of section 1.2.5, it is assumed that the boundaries between structural features and performance features are blurred. With regard to figure 1.7, both composer matching and performer matching are represented in the second order predictor that can be seen to embody one of the interactions shown in the middle of the lens model structure. This should be kept in mind for the results of all following second order predictors including more than one first order predictor. The influences of this predictor to the remaining three mechanisms may be explained similarly. Thus, PC_RGN_RMS may be classified as a weaker version of the aforementioned with a weaker result A1+ which also was to be expected. For the remaining three predictors that include portions of changes in tempo, melodic contour and spectral centroid, the A2+ result is also as expected because all three are assumed to have positive influences on arousal ratings.

Now crossing over to the results for valence, PC_BRN shows a negative influence on valence ratings with a result of V1-. If the relative amount of energy above $1000Hz$ as a percentage of the total signal energy increases, listeners judged only S5 as more unpleasant (cf. table 8). As the brain stem reflex is supposed to be activated when the signal entering the auditory systems is intended to warn from dangers as described in section 1.1.1, this finding is to be expected. The result for PC_RGN_BRN_RMS (V2-) expands this finding because with roughness and RMS, two further signal properties assumed to activate the brain stem reflex and to cause unpleasantness (especially roughness) are included. A change of sign takes place with PC_SPC_MEL (V2+): an increase of the spectral centroid in combination with an increase in melodic contour seems to be comfortable for listeners. This may be a small confirmation of the aforementioned findings of Gerardi and Gerken (1995) that melodic contour is positively related to valence.

However, the findings for influences on skin conductance measurements are rather ambiguous. While the results for PC_RGN_RMS (S3+), PC_RGN_BRN_RMS (S3+) and PC_SPC_BRN (S1+) are plausible (also regarding the positive influences on arousal ratings), the remaining predictors considered here show mainly negative influences (PC_MEL with S2-, PC_BRN with S1- and PC_SPC_RMS with S2-) or both negative and positive influences (PC_TMP_MEL with S1+/S1- and PC_SPC with S2+/S2-). These results have to be interpreted considering the different time lags: An example of this can be found by looking at the PC_SPC row in the main results table 7 for S3/dV_SCR. It is apparent that the direction of the influences on skin conductance for S3 change from time lag 0 and 1 (negative) to time lag 2 and 3 (positive) (as described in the results section of S3). The change of directions in the influences of the spectral centroid may happen due to the typical behaviour of SC described in section 2.1.3: while the listener is resting, SC gradually decreases and as soon as the listener is exposed to a new stimulus, changes of SC are reversed thus it increases quickly and will gradually drop again when the same stimulus is repeated along with another change of direction. Another change of influence direction appears at the PC_ICL row in the main results table 7 for S4/dV_SCR that could be explained similarly.

Rhythmic entrainment The average age of participants was 23 years. Following Topend Sports - The Sport + Science Resource (2013), the average resting heart rate for persons aged between 18 and 25 years is 74 – 78 bpm (females) respective 70 – 73 bpm (males). The stimuli with an average absolute tempo value[49] being closest to these average resting heart rate values are:

[49] Calculated in Matlab® based on the raw tempo time series imported from Sonic Visualiser 2.3.

Second Order Predictors	Brain Stem Reflex	Rhythmic Entrainment	Emotional Contagion	Musical Expectancy
PC_MEL	S2-	S2-	S2-	S2-
PC_BRN	V1- S1-	V1- S1-	V1- S1-	V1- S1-
PC_TMP	A2+ S2+	A2+ S2+	A2+ S2+	A2+ S2+
PC_TMP_MEL	A2+ S1+,S1-	A2+ S1+,S1-	A2+ S1+,S1-	A2+ S1+,S1-
PC_RGN_RMS	A1+ S3+	A1+ S3+	A1+ S3+	A1+ S3+
PC_RGN_BRN_RMS	A4+ V2- S3+	A4+ V2- S3+	A4+ V2- S3+	A4+ V2- S3+
PC_SPC	A2+ S2+,S2-	A2+ S2+,S2-	A2+ S2+,S2-	A2+ S2+,S2-
PC_SPC_MEL	A2+ V2+	A2+ V2+	A2+ V2+	A2+ V2+
PC_SPC_RMS	S2-	S2-	S2-	S2-
PC_SPC_BRN	S1+	S1+	S1+	S1+
PC_ICH	A3+ V2+,V1- S3-	A3+ V2+,V1- S3-	A3+ V2+,V1- S3-	A3+ V2+,V1- S3-
PC_ICL	A2- S4+,S4-	A2- S4+,S4-	A2- S4+,S4-	A2- S4+,S4-
PC_UEH	A6+ V1- S7+	A6+ V1- S7+	A6+ V1- S7+	A6+ V1- S7+
PC_UEL	A3+,A1- V1+,V1- S3-	A3+,A1- V1+,V1- S3-	A3+,A1- V1+,V1- S3-	A3+,A1- V1+,V1- S3-

Table 9

Summary table of second order predictors as possible triggers for mechanisms from the BRECVEM framework (coloured entries imply an influence according to the considerations on the mechanisms. Greyed out entries imply no relationship. A = dV_ARO, V = dV_VAL, S = dV_SCR (with numbers of significant coefficients), +/- = positive/negative influence).

1. S3 with $\mathfrak{T} = 74,1$,
2. S6 with $\mathfrak{T} = 68,5$,
3. S1 with $\mathfrak{T} = 90,0$,
4. S2, arguing that doubling the absolute tempo value of $\mathfrak{T} = 43,6$ to $\mathfrak{T} = 87,2$ causes a shift towards the relevant range of resting heart rate values and
5. S4, also arguing that doubling the absolute tempo value of $\mathfrak{T} = 48,5$ to $\mathfrak{T} = 97,0$ causes a shift towards the relevant range of resting heart rate values.

With a value of $\mathfrak{T} = 107,1$, S5 stays outside the relevant range. Assuming that rhythmic entrainment is closely related to tempo, a ranking may be extracted from the results data in table 7 (double page) for the tempo-related predictors PC_TMP and PC_TMP_MEL (see tab. 8):

1. S3: two significant positive effects of PC_TMP on dV_ARO (each $\beta = .03, p < .01$) at time lags 2 and 3,
2. S6: significant positive effect of PC_TMP on dV_SCR ($\beta = .03, p < .01$) at time lag 0,
3. S1: strong significant positive effect of PC_TMP_MEL on dV_ARO ($\beta = .16, p < .01$) at time lag 1, significant negative effect of PC_TMP_MEL on SCR ($\beta = -.06, p < .01$) at time lag 0,
4. S2: significant positive effect of PC_TMP_MEL on dV_ARO ($\beta = .08, p < .01$) at time lag 2, significant positive effect of PC_TMP_MEL on dV_SCR ($\beta = .03, p < .01$) at time lag 1 and
5. S4: significant positive effect of PC_TMP on dV_SCR ($\beta = .04, p < .01$) at time lag 2.

This could support the theory of oscillator synchronisation (see Large, 2002), explain the stimulus-dependencies of the influences and furthermore confirm Juslin et al. (2010) who claim that "The entrainment-inducing properties of music that produce affect presumably depend on the music having a marked pulse - and preferably one that is relatively close to the 'natural' heart rate or respiration of the listener." (p. 621). Both predictors related to this mechanism are coloured in table 9, second column.

Regarding this mechanism, two predictors containing tempo were suspected to have positive influences which is confirmed by two A2+ results.

No influences from predictors suspected to be in relationship with rhythmic entrainment were found on the subjective ratings of valence.

However, PC_TMP shows a clear positive influence on skin conductance measurements (S2+) which was to be expected. For PC_TMP_MEL, two results with opposite directions appear (S1+ in S2, S1- in S1).

Emotional contagion Taking up the example of emotional speech from the end of section 1.1.1 again, the influences of certain predictors on the mechanism of emotional contagion will now be elaborated. Juslin and Laukka (2003, p. 802, tab. 11) list acoustic cues that vocal expression and musical performance have in common, relating to the five basic emotions *anger, fear, happiness, sadness* and *tenderness*. These cues are:

- speech rate/tempo,
- intensity/sound level,
- intensity/sound level variability,
- high frequency energy,
- F0/pitch level,
- F0/pitch variability,
- F0/pitch contour,
- onsets/tone attacks and
- micro-structural irregularity.

For the purpose of this study, some of these cues can be connected to all predictors based on low and mid level features. The influences that may be explained by emotional contagion via similarities between vocal expression and music performance emerge from table 9, third column. The same predictors are coloured as for the brain stem reflex but with PC_TMP, one more is added. The attempt will be made to interpret the same results against a different theoretical background.

As shown in Juslin and Laukka (2003, p. 801, tab. 10), the *speech rate* has almost only positive effects on basic emotions which is clearly confirmed by the A2+ result for PC_TMP and also by PC_TMP_MEL where the positive influences of *F0 contour* can be seen as additional confirmation. Different acoustic cues of the theory by Juslin and Laukka (2003) show almost only positive results for anger, fear and happiness which can be confirmed by the overall positive influences of the remaining four second order predictors.

Turning to the influences on valence, only three predictors provide results: PC_BRN and PC_RGN_BRN_RMS show negative influences with V1- and V2-. Related to Juslin and Laukka (2003, p. 801, tab. 10), these findings cannot be confirmed. However, PC_SPC_MEL shows positive influences on valence ratings (V2+).

The influences on skin conductance measurements that are supposed to be explainable by emotional contagion are diverse. First, it seems unclear why SC decreases with rising PC_MEL (S2-) because a rising F0 contour of vocal expression would be expected to arouse a listener. Maybe a solution can be found considering the time lags: table 7 reveals that these two influences happen at S6 on time lags 3 and 4. Here, the first increase of SC could already be over and a first decrease might have started. However, the positive influences of PC_TMP with S2+ are as expected because an increasing speech rate would cause the same effect in a listener. For PC_TMP_MEL and PC_SPC, it has to be admitted that the concurrence of first order predictors within the second order predictors and the ambiguity do not allow clear interpretation. Even the negative effects from PC_BRN (S1-) and PC_SPC_RMS (S2-) are hard to interpret: increasing high frequency energy and loudness in a human voice would normally cause physiological arousal in a listener. Hence, the clear positive effects from PC_RGN_BRN_RMS with S3+ are as expected. Maybe the roughness information included in the predictor contributes to increasing SC due to higher unpleasantness.

Musical expectancy Musical expectancy is assumed to be closely related to PC_ICH/PC_ICL and PC_UEH/PC_UEL (Egermann et al., 2013). Furthermore, as indicated in the introduction, the mechanism may also be linked to musical thrills[50]. Guhn et al. (2007) looked for "unique musical and/or acoustic characteristics" of chill passages characterised by a significant increase of SC. These chill passages were accompanied by "a sudden or gradual volume increase from *piano* [. . .] to *forte*" and by "an expansion in its frequency range in the high or low register" (p. 480). This leads to the predictors coloured in the fourth column of table 9.

For this mechanism, the predictor PC_RGN_RMS shows a positive effect on

[50] The term *chills* is used equivalent to *thrills, frisson* and even *skin orgasm* (Huron and Margulis, 2010, p. 592).

arousal ratings with A1+. It may be presumed that this influence mainly originates from the RMS portion of the predictor which would be as expected. The predictor PC_RGN_BRN_RMS adds a further portion of brightness that seems to increase the impact on arousal ratings with a result of A4+. The predictor PC_SPC only representing spectral centroid seems to have a strong positive influence on arousal which may confirm findings of Guhn et al. (2007) describing chill passages to be accompanied by changes in the frequency range. This also applies for PC_SPC_MEL with a result of A2+. The results for PC_ICH with A3+ and PC_ICL with A2- confirm the findings of Egermann et al. (2013, p. 12, fig. 4, upper left) indicating event-related changes of arousal ratings into the same directions for high IC vs. low IC peak segments. Again, the results for PC_UEH with A6+ confirm the findings of Egermann et al. (2013, p. 15, fig. 6, upper left), whereas the results for PC_UEL with A3+ as well as A1- are ambiguous because the positive influences on arousal from PC_UEL, appearing exclusively for S2 and S3 at time lags 3 and 4, seem to be illogical, thus its origin cannot be clarified here.

The findings for the influences on valence cannot clearly confirm the findings of Egermann et al. (2013, p. 12, fig. 4, upper right) because PC_ICH shows ambiguous results (V2+/V1-) and PC_ICL shows no influences on valence ratings. The same applies to PC_UEH (V1-) and PC_UEL (V1+/V1-) not confirming the results showed in Egermann et al. (2013, p. 15, fig. 6, upper right).

Looking at the influences on measurements of skin conductance, the findings of Egermann et al. (2013, p. 13, fig. 5, upper left) could not be reproduced: PC_ICH (S3-) showed a contrary influence, and PC_ICL (S4+/S4-) shows ambiguous influences. However, the findings of Egermann et al. (2013, p. 16, fig. 7, upper left) could be partially confirmed by PC_UEH (S7+) and PC_UEL (S3-). Finally, the overall positive influences of tempo-related and loudness-related second order predictors on arousal successfully confirm findings of Schubert (2004) outlined in the introduction, even the difficulty to predict valence is confirmed.

To summarise, dV_ARO showed a total of 28 significant coefficients (see tab. 10). Because these numerous influences from second order predictors on arousal ratings were found, hypothesis H_1 is considered to be fully corroborated. In contrast, dV_VAL showed 11 significant coefficients, thus hypothesis H_2 is considered to be partially corroborated because far fewer influences from predictors were found. Even though no highly significant influences on valence ratings were found for

S2 and S6. Finally, dV_SCR quantitatively showed the best predictability with a total of 42 significant coefficients, thus hypothesis H_3 is considered to be fully corroborated. At the stimuli level, S3 (Density 21.5 by Edgar Varèse from the year 1936) seems to be the piece showing most emotional reactions with 26 significant coefficients.

Stimulus	dV_ARO	dV_VAL	dV_SCR	$\sum$
S1	4	4	5	13
S2	5	0	8	13
S3	14	1	11	26
S4	1	3	12	16
S5	3	3	0	6
S6	1	0	6	7
$\sum$	28	11	42	**81**

Table 10

Sums of significant coefficients $(p < .01)$ *in the results of the HLMs related to stimuli and dVs (summed from tab. 7).*

H_4 - Effects Of High Level Features vs. Low and Mid Level Features From the sums of significant coefficients of all HLMs listed in table 10, 38 are attributable to low and mid level features distributed over 12 predictors. However, 43 significant coefficients are attributable to the high level features (20 for PC_ICH and PC_ICL, 23 for PC_UEH and PC_UEL) distributed over only 4 predictors. Thus, H_4 is considered to be corroborated because independent predictions by means of both groups seem to be possible, whereas the group of high level features is suspected to make stronger predictions than the group of low and mid level features.

4.2 Limitations And Suggestions For Future Research

The limitations of this work and proposals for future improvements and extensions will be pointed out in the following.

Concerning the live concert experiment, see also Egermann et al. (2013, p. 18) who list limitations like, among others, a lack of control over the participants actual behaviour and possible distractions by other participants. Moreover, recent critics of conventional dimensional models of emotion have led to new developments: Schubert et al. (2012) designed and tested the *emotion-face-clock* (fig. F.2, right). Their motivation for the new model arises from the drawbacks of the rating response formats of the first self-reported continuous response methods developed during the 1980s and 1990s that claimed to be very reductive. What they criticise most is that the one-word-labelling of main dimensions disguises underlying 'micro-dimensions'. Their example is the arousal dimension that, following Schimmack and Rainer (2002), is a combination of *energetic arousal* and *tense arousal*. Therefore, participants rate sadness as arousing because of sadness partially acting with high tension arousal (Schubert et al., 2012, p. 2). The proposed new system consists of six face symbols considering the position of every dimension following Hevner's adjective circle (fig. F.2, left) and could provide a more differentiated alternative to the 2DES used to record the data analysed in this work.

Crossing over to the analysis part, the selection of musical features could be further refined and fitted to the acoustic properties of the stimuli better. Additionally, even the features that seem to be robust and meaningful should be observed exactly in relation to the signal they are applied to: what is, for example, the significance of the tempo feature extracted from *Density 21.5* by Edgar Varèse where the tempo dimension seems to be completely dissolved and no more recognisable at some points in the piece?

Concerning the time lag structure, an improvement could be an automated approach of cross correlation calculations to find out the optimal time lags where the maximum emotional reactions take place. This was omitted due to the large number of 90 models that the cross correlation had to be calculated for manually. A further improvement and alternative approach for the time lag structure could be to introduce a time delay for the *predictors* instead of the *dVs* and to include all stimuli into one single model where they are represented by a piece-variable. Finally, this new attempt could include random factors and consider individual differences of listeners as well as a more detailed troubleshooting.

Unfortunately, it was only possible to consider the second order predictors regarding the interpretation of results with the BRECVEM framework. Of course, it would have been interesting to do the same analyses again with the first order predictors, but this was omitted due to the desired separation of high level features from mid and low level features that was of higher priority, especially to make the results comparable to those of Egermann et al. (2013). Furthermore, the suspected connections between HR and tempo, explained by rhythmic entrainment, could be further studied including the measurements of HR also recorded during the live concert. In general, it is unclear how negative effects of predictors on skin conductance measurements may be interpreted at all because an active reduction of this physiological reaction is questionable.

For the interpretation of musical features as predictors for measurements regarding musical emotions, approaches of music semiotics could possibly be helpful in the future as outlined by Juslin and Sloboda (2012, p. 599): they refer to the terms *index, icon* and *symbol* tracing back to Charles Sanders Pierce and suggest the following iconic relationships to be the most promising principle to explain: not only *how* mechanisms of musical emotions work, but even the *origin* of relationships between perceived emotions and musical features. Even Scherer and Zentner (2001, p. 364) describe structural and performance features to have iconic and symbolic characteristics.

Furthermore, all interpretations have to be made with the individual properties of the stimuli in mind. The broad variety of styles suggests that completely different mechanisms work inside the music. The approach made here with four mechanisms from BRECVEM could be expanded and the remaining three mechanisms (evaluative conditioning, visual imagery and episodic memory) could yet be considered.

4.3 Conclusions

This study tested the predictability of emotional reactions that were continuously captured during a live concert experiment with six pieces of classical solo flute music. For the purpose of finding factors from musical structure as well as from (psycho-)acoustics influencing these emotional reactions, resulting time series were used in a hierarchical linear modelling approach. Some promising results could be found, whereas findings for the predictability of arousal and skin conductance were far more meaningful than for valence. The derivation of explanations for the influences from a set of psychological mechanisms was partially successful, but could not be precisely completed due to the properties of the predictors: too many ambiguities appeared because it was hard to specify the information contained in the second order predictors that resulted from the first order predictors entering principal component analyses. In summary, the findings of this work could successfully contribute to and solidify previous findings of for example Egermann et al. (2013) and Schubert (2004) and seem to be a good basis for further analyses of music and emotion.

Bibliography

Aramaki, Mitsuko; Mathieu Barthet; Richard Kronland-Martinet; and Sølvi Ystad (Eds.) (2012): *From Sounds to Music and Emotions: LNCS 7900*. Springer. URL http://link.springer.com/content/pdf/10.1007%2F978-3-642-41248-6.pdf.

Baayen, R.H.; D.J. Davidson; and D.M. Bates (2008): "Mixed-effects modeling with crossed random effects for subjects and items." In: *Journal of Memory and Language*, **59**(4), pp. 390–412. doi:10.1016/j.jml.2007.12.005.

Bach, Johann Sebastian (1917a): *Partita in A minor, BWV 1013 (1722-23): II. Corrente*. URL http://javanese.imslp.info/files/imglnks/usimg/0/03/IMSLP05674-Bach-partita-flute-courante.pdf.

Bach, Johann Sebastian (1917b): *Partita in A minor, BWV 1013 (1722-23): III. Sarabande*. URL http://petrucci.mus.auth.gr/imglnks/usimg/8/87/IMSLP05675-Bach-partita-flute-sarabande.pdf.

Bellebaum, Christian; Irene Daum; and Patrizia Thoma (2012): *Neuropsychologie*. Basiswissen Psychologie, 1. auflage. Wiesbaden: VS Verlag.

Bortz, Jürgen (2010): *Statistik für Human- und Sozialwissenschaftler*. 6. auflage. Heidelberg: Springer.

Bullock, Jamie (2007): "libxtract: a Lightweight Library for Audio Feature Extraction." In: *Proceedings of the 2007 International Computer Music Conference*, vol. 2. ICMA, pp. 25–28.

Cabrera, Densil (1999): "'PSYSOUND': A COMPUTER PROGRAM FOR PSYCHOACOUSTICAL ANALYSIS: User's Manual." URL http://densil.tripod.com/software/PsySound.PDF.

Cannam, Chris; Christian Landone; and Mark Sandler (2010): "Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files." In: *Proceedings of the ACM Multimedia 2010 International Conference*. pp. 1467–1468.

Cheng, Jing; Lloyd J. Edwards; Mildred M. Maldonado-Molina; Kelli A. Komro; and Keith E. Muller (2009): "Real longitudinal data analysis for real people: Building a good enough mixed model." In: *Statistics in Medicine*, pp. 504–520. doi:10.1002/sim.3775.

Cook, Nicholas and Daniel Leech-Wilkinson (2008): "A musicologist's guide to Sonic Visualiser." URL http://charm.cchcdn.net/redist/pdf/analysing_recordings.pdf.

Coutinho, Eduardo and Angelo Cangelosi (2009): "The Use of Spatio-Temporal Connectionist Models in Psychological Studies of Musical Emotions." In: *Music Perception*, **27**(1), pp. 1–15. doi:10.1525/mp.2009.27.1.1.

Coutinho, Eduardo and Angelo Cangelosi (2011): "Musical emotions: Predicting second-by-second subjective feelings of emotion from low-level psychoacoustic features and physiological measurements." In: *Emotion*, **11**(4), pp. 921–937. doi:10.1037/a0024700.

Darwin, C. R. (1871): *The descent of man, and selection in relation to sex*, vol. Volume 2. 1st edition. London: John Murray. URL http://darwin-online.org.uk/content/frameset?itemID=F937.2&viewtype=text&pageseq=1.

Dawson, Michael E.; Anne M. Shell; and Diane L. Filion (2007): "The Electrodermal System." In: John T. Cacioppo; Louis G. Tassinary; and Gary Berntson (Eds.) *Handbook of Psychophysiology*. Cambridge University Press, pp. 159–181.

Dean, Roger T. and Freya Bailes (2010): "Time Series Analysis as a Method to Examine Acoustical Influences on Real-time Perception of Music." In: *Empirical Musicology Review*, **5**(4), pp. 152–175.

Dean, Roger T.; Freya Bailes; Emery Schubert; and Mark W. Greenlee (2011): "Acoustic Intensity Causes Perceived Changes in Arousal Levels in Music: An Experimental Investigation." In: *PLoS ONE*, **6**(4), p. e18591. doi:10.1371/journal.pone.0018591.

Debussy, Claude (1927): *Syrinx (1913)*. Paris: Jobert. URL http://imslp.eu/files/imglnks/euimg/2/2d/IMSLP12780-Debussy_-_Syrinx__solo_flute_.pdf.

Desain, Peter and Henkjan Honing (1993): "Tempo curves considered harmful." In: *Contemporary Music Review*, **7**(2), pp. 123–138. doi:10.1080/07494469300640081.

Di Pellegrino, G.; L. Fadiga; L. Fogassi; V. Gallese; and G. Rizzolatti (1992): "Understanding motor events: a neurophysiological study." In: *Experimental Brain Research*, **91**(1). doi:10.1007/BF00230027.

Eerola, Tuomas; Olivier Lartillot; and Petri Toiviainen (2009): "PREDICTION OF MULTIDIMENSIONAL EMOTIONAL RATINGS IN MUSIC FROM AUDIO USING MULTIVARIATE REGRESSION MODELS." In: *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*. URL http://ismir2009.ismir.net/proceedings/PS4-8.pdf.

Eerola, Tuomas and Petri Toiviainen (2004a): *MIDI Toolbox: MATLAB Tools for Music Research*. Department of Music, University of Jyväskylä. URL https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/Manual.

Eerola, Tuomas and Petri Toiviainen (2004b): "MIR IN MATLAB: THE MIDI TOOLBOX." In: *Proceedings of the 05th International Conference on Music Information Retrieval 2004*. pp. 22–27. URL http://ismir2004.ismir.net/proceedings/p004-page-22-paper193.pdf.

Egermann, Hauke; Marcus T. Pearce; Geraint A. Wiggins; and Stephen McAdams (2013): "Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music." In: *Cognitive, Affective, & Behavioral Neuroscience*. doi:10.3758/s13415-013-0161-y.

Ellsworth, P. C. and Klaus R. Scherer (2003): "Appraisal processes in emotion." In: R. Davidson; Klaus R. Scherer; and H. Goldsmith (Eds.) *Handbook of Affective Sciences*. New York: Oxford University Press, pp. 572–595. URL http://people.ict.usc.edu/~gratch/CSCI534/EllsworthScherer03.PDF.

Eyben, Florian; Felix Weninger; Florian Gross; and Björn Schuller (2013): "Recent developments in openSMILE, the munich open-source multimedia feature extractor." In: A. Jaimes (Ed.) *Proceedings of the 21st ACM international conference on Multimedia*. ACM, pp. 835–838. doi:10.1145/2502081.2502224.

Fastl, Hugo and Eberhard Zwicker (2007): *Psychoacoustics - Facts and Models*. Third edition. Springer.

Féré, C. (1888): "Note sur des Modifications de la Résistance électrique sous l'Influence des Excitations sensorielles et des Émotions." In: *Comptes rendus des séances de la Société de biologie et de ses filiales*, **40th, ser. 8, t. 5**, pp. 217–219. URL http://citebank.org/uid.php?id=123148.

Ferguson, Sam; Densil Cabrera; Emery Schubert; and Farhan Rizwi (2008): "PsySound3 User Manual: Draft (included in Psysound3 software package)." URL https://code.google.com/p/psysound3/downloads/detail?name=PsySound3Beta-2011-01-20.zip&can=2&q=.

Field, Andy P. (2009): *Discovering statistics using SPSS*. Introducing Statistical Methods Series, 3rd edition. Los Angeles and London: SAGE.

Frijda, N. H. and Klaus R. Scherer (2009): "Affect (psychological perspectives)." In: D. Sander and Klaus R. Scherer (Eds.) *Oxford Companion to Emotion and the Affective Sciences*. OUP Oxford, p. 10.

Garzorz, Natalie (2009): *Basics Neuroanatomie*. Basics. München and Jena: Elsevier, Urban & Fischer.

Gelfand, Stanley A. (2010): *Hearing: An introduction to psychological and physiological acoustics*. 5. London: Informa Healthcare.

Gerardi, Gina M. and Louann Gerken (1995): "The Development of Affective Responses to Modality and Melodic Contour." In: *Music Perception: An Interdisciplinary Journal*, **12**(3), pp. 279–290. URL http://www.jstor.org/stable/40286184.

Gjerdingen, Robert O. (2012): "Psychologists and Musicians: Then and Now." In: Diana Deutsch (Ed.) *The psychology of music*. Academic Press, pp. 683–707.

Gomez, P. and B. Danuser (2007): "Relationships between musical structure and psychophysiological measures of emotion." In: *Emotion*, **7**(2), pp. 377–387. doi:10.1037/1528-3542.7.2.377.

Grace-Martin, Karen (2012): "When the Hessian Matrix goes Wacky." URL http://www.theanalysisfactor.com/wacky-hessian-matrix/.

Guhn, Martin; Alfons Hamm; and Marcel R. Zentner (2007): "Physiological and Musico-Acoustic Correlates of the Chill Response." In: *Music Perception*, **24**(5), pp. 473–484. doi:10.1525/MP.2007.24.5.473.

Hedblad, Anton (2011): "Evaluation of Musical Feature Extraction Tools using Perpetual Ratings: Master's Thesis in Music Acoustics." URL http://www.speech.kth.se/prod/publications/files/3620.pdf.

Hevner, K. (1935): "The Affective Character of the Major and Minor Modes in Music." In: *The American Journal of Psychology*, **47**(1), pp. p 103–118. URL http://www.jstor.org/stable/1416710.

Hevner, K. (1936): "Experimental Studies of the Elements of Expression in Music." In: *The American Journal of Psychology*, **48**(2), pp. p 246–268. URL http://www.jstor.org/stable/1415746.

Hindemith, Paul (1958): *Acht Stücke für Flöte allein (1927)*. Mainz: Schott. URL http://petruccilibrary.ca/files/imglnks/caimg/3/3b/IMSLP310400-PMLP501598-Hindemith_-_8_St__cke_f__r_Fl__te_allein.pdf.

Huron, David (2006): *Sweet anticipation: Music and the psychology of expectation*. Cambridge and Mass: MIT Press.

Huron, David and Elizabeth Hellmuth Margulis (2010): "Musical expectancy and thrills." In: Patrik N. Juslin and John A. Sloboda (Eds.) *Handbook of Music and Emotion (reprinted 2012)*, Series in affective science. Oxford: Oxford University Press, pp. 575–604.

Husain, Gabriela; William Forde Thompson; and E. Glenn Schellenberg (2002): "Effects of Musical Tempo and Mode on Arousal, Mood, and Spatial Abilities." In: *Music Perception*, **20**(2), pp. 151–171. doi:10.1525/mp.2002.20.2.151.

IBM Corporation (2012): "IBM SPSS Statistics 21 Command Syntax Reference." URL ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/21.0/en/client/Manuals/IBM_SPSS_Statistics_Command_Syntax_Reference.pdf.

Ilie, Gabriella and William Forde Thompson (2006): "A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect." In: *Music Perception*, **23**(4), pp. 319–330. doi:10.1525/mp.2006.23.4.319.

Izard, Carroll E. (2009): "Emotion theory and research: highlights, unanswered questions, and emerging issues." In: *Annual review of psychology*, **60**, pp. 1–25. doi:10.1146/annurev.psych.60.110707.163539.

Juslin, Patrik N. (2000): "Cue utilization in communication of emotion in music performance: Relating performance to perception." In: *Journal of Experimental Psychology: Human Perception and Performance*, **26**(6), pp. 1797–1813. doi:10.1037//0096-1523.26.6.1797.

Juslin, Patrik N. and Petri Laukka (2003): "Communication of emotions in vocal expression and music performance: different channels, same code?" In: *Psychological bulletin*, **129**(5), pp. 770–814. doi:10.1037/0033-2909.129.5.770.

Juslin, Patrik N.; Simon Liljeström; Daniel Västfjäll; and Lars-Olov Lundqvist (2010): "How Does Music Evoke Emotions? Exploring The Underlying Mechanisms." In: Patrik N. Juslin and John A. Sloboda (Eds.) *Handbook of Music and Emotion (reprinted 2012)*, Series in affective science. Oxford: Oxford University Press, pp. 605–642.

Juslin, Patrik N. and John A. Sloboda (2012): "Music and Emotion." In: Diana Deutsch (Ed.) *The psychology of music*. Academic Press, pp. 583–645.

Juslin, Patrik N. and Daniel Västfjäll (2008): "Emotional responses to music: The need to consider underlying mechanisms." In: *Behavioral and Brain Sciences*, **31**(5). doi:10.1017/S0140525X08005293.

Kincaid, Chuck (2005): "Guidelines for Selecting the Covariance Structure in Mixed Model Analysis." In: SAS Institute Inc. (Ed.) *Proceedings of the Thirtieth Annual SAS Users Group International Conference*. Cary, NC. URL http://www2.sas.com/proceedings/sugi30/198-30.pdf.

Korhonen, M.D; D. A. Clausi; and M.E Jernigan (2006): "Modeling emotional content of music using system identification." In: *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, **36**(3), pp. 588–599. doi:10.1109/TSMCB.2005.862491.

Kramer, Adam D. I.; Jamie E. Guillory; and Jeffrey T. Hancock (2014): "Experimental evidence of massive-scale emotional contagion through social networks." In: *Proceedings of the National Academy of Sciences of the United States of America*, **111**(24), pp. 8788–8790. doi:10.1073/pnas.1320040111.

Krimphoff, Jochen; Stephen McAdams; and Suzanne Winsberg (1994): "Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique." URL http://hal.archives-ouvertes.fr/docs/00/25/28/11/PDF/ajp-jp4199404C5134.pdf.

Krumhansl, Carol L. (1996): "A Perceptual Analysis of Mozart's Piano Sonata K. 282: Segmentation, Tension, and Musical Ideas." In: *Music Perception: An Interdisciplinary Journal*, **13**(3), pp. p 401–432. URL http://www.jstor.org/stable/40286177.

Large, E. (2002): "Perceiving temporal regularity in music." In: *Cognitive Science*, **26**(1), pp. 1–37. doi:10.1016/S0364-0213(01)00057-X.

Lartillot, Olivier (2013): "MIRtoolbox 1.5: User's Manual." URL https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/Download/resolveuid/eaaa9c0f-97c9-41ab-92d3-da9886fd14b8.

Lartillot, Olivier and Petri Toiviainen (2007): "A MATLAB TOOLBOX FOR MUSICAL FEATURE EXTRACTION FROM AUDIO." In: *Proceedings of the 10th International Conference on Digital Audio Effects (DAFX)*.

Laukka, Petri; Patrik Juslin; and Roberto Bresin (2005): "A dimensional approach to vocal expression of emotion." In: *Cognition & Emotion*, **19**(5), pp. 633–653. doi:10.1080/02699930441000445.

Lerch, Alexander (2012): *An introduction to audio content analysis: Applications in signal processing and music informatics*. Hoboken and N.J: Wiley.

Lerch, Alexander; Gunnar Eisenberg; and Koen Tanghe (2005): "FEAPI: A Low Level Feature Extraction Plugin API." In: *Proceedings of the 8th International Conference on Digital Audio Effects (DAFX)*. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.59.9079.

Levitin, Daniel J. (2006): *This is your brain on music: The science of a human obsession*. New York: Plume.

Lippert, Herbert (2006): *Lehrbuch Anatomie: 184 Tabellen*. 7., erw. aufl. München [u.a.]: Elsevier, Urban & Fischer.

Mathieu, Benoît; Slim Essid; Thomas Fillon; Jacques Prado; and Gaël Richard (2010): "YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software." In: *Proceedings of the 11th International Conference on Music Information Retrieval 2010*. pp. 441–446. URL http://dblp.uni-trier.de/db/conf/ismir/ismir2010.html#MathieuEFPR10.

McEnnis, Daniel; Cory McKay; and Ichiro Fujinaga (2005): "jAudio: A feature extraction library." In: *Proceedings of the 06th International Conference on Music Information Retrieval 2005*.

McEnnis, Daniel; Cory McKay; and Ichiro Fujinaga (2006): "jAudio: Additions and improvements." In: *Proceedings of the International Conference on Music Information Retrieval*. pp. 385–386. URL http://jmir.sourceforge.net/publications/ISMIR_2006_jAudio_Improvements.pdf.

MediTECH Electronic GmbH (2004): "ProComp Infiniti Gebrauchsanleitung." URL http://www.meditech.de/fileadmin/download/anleitungen/ProComp%20Infiniti%20Anleitung.pdf.

Nagel, F.; Reinhard Kopiez; Oliver Grewe; and Eckart Altenmüller (2008): "Psychoacoustical correlates of musically induced chills." In: *Musicae Scientiae*, **12**(1), pp. 101–113. doi:10.1177/102986490801200106.

Pearce, Marcus T. (2005): "The Construction and Evaluation of statistical Models of melodic Structure in Music Perception and Composition: PhD Thesis." URL http://doc.gold.ac.uk/~mas01mtp/papers/Pearce2005.pdf.

Pearce, Marcus T.; María Herrojo Ruiz; Selina Kapasi; Geraint A. Wiggins; and Joydeep Bhattacharya (2010): "Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation." In: *NeuroImage*, **50**(1), pp. 302–313. doi:10.1016/j.neuroimage.2009.12.019. URL http://www.sciencedirect.com/science/article/pii/S1053811909013068.

Peeters, Geoffroy; Bruno L. Giordano; Patrick Susini; Nicolas Misdariis; and Stephen McAdams (2011): "The Timbre Toolbox: Extracting audio descriptors from musical signals." In: *The Journal of the Acoustical Society of America*, **130**(5), p. 2902. doi:10.1121/1.3642604.

Peretz, Isabelle (2010): "Towards a Neurobiology of Musical Emotions." In: Patrik N. Juslin and John A. Sloboda (Eds.) *Handbook of Music and Emotion (reprinted 2012)*, Series in affective science. Oxford: Oxford University Press, pp. 99–126.

Plomp, R. (1965): "Tonal Consonance and Critical Bandwidth." In: *The Journal of the Acoustical Society of America*, **38**(4), p. 548. doi:10.1121/1.1909741.

Poulenc, Francis (2000): *Un joueur de flûte berce les ruines (1942): For solo flute*. London: Chester Music.

Russell, James A. (1980): "A circumplex model of affect." In: *Journal of Personality and Social Psychology*, **39**(6), pp. 1161–1178. doi:10.1037/h0077714.

Salimpoor, Valorie N.; Mitchel Benovoy; Gregory Longo; Jeremy R. Cooperstock; and Robert J. Zatorre (2009): "The rewarding aspects of music listening are related to degree of emotional arousal." In: *PloS one*, **4**(10), p. e7487. doi:10.1371/journal.pone.0007487.

Sanden, Chris; Chad R. Befus; and John Z. Zhang (2010): "Camel: a lightweight framework for content-based audio and music analysis." In: *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound*, AM '10. ACM, pp. 22:1–22:4. doi:10.1145/1859799.1859821. URL http://doi.acm.org/10.1145/1859799.1859821.

Scherer, Klaus R. (1987): "Toward a dynamic theory of emotion: The component process model of affective states." In: *Geneva Studies in Emotion and Communication*, **1**(1), pp. 1–98. URL http://www.affective-sciences.org/system/files/biblio/1987_Scherer_Genstudies.pdf.

Scherer, Klaus R. (2000): "Psychological models of emotion." In: Joan C. Borod (Ed.) *The neuropsychology of emotion*, Series in affective science. New York: Oxford University Press, pp. 137–162.

Scherer, Klaus R. (2004): "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?" In: *Journal of New Music Research*, **33**(3), pp. 239–251. doi:10.1080/0929821042000317822.

Scherer, Klaus R. (2005): "What are emotions? And how can they be measured?" In: *Social Science Information*, **44**(4), pp. 695–729. doi:10.1177/0539018405058216.

Scherer, Klaus R. and Marcel R. Zentner (2001): "Emotional Effects of Music: Production Rules." In: Juslin, Patrik N. , Sloboda, John A. (Ed.) *Music and emotion*, Series in affective science. Oxford and New York: Oxford University Press, pp. 361–392.

Schimmack, Ulrich and Reisenzein Rainer (2002): "Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation." In: *Emotion*, **2**(4), pp. 412–417. doi:10.1037/1528-3542.2.4.412.

Schubert, Emery (1999): *Measurement and time series analysis of emotion in music - Volume 1*. Ph.D. thesis, University of New South Wales.

Schubert, Emery (2004): "Modeling Perceived Emotion With Continuous Musical Features." In: *Music Perception*, **21**(4), pp. 561–585.

Schubert, Emery (2010): "Continuous Self-Report Methods." In: Patrik N. Juslin and John A. Sloboda (Eds.) *Handbook of Music and Emotion (reprinted 2012)*, Series in affective science. Oxford: Oxford University Press, pp. 223–253. doi:10.1093/acprof:oso/9780199230143.003.0009.

Schubert, Emery; Sam Ferguson; Natasha Farrar; and Taylor, David, McPherson, Gary E. (2012): "The Six Emotion-Face Clock as a Tool for Continuously Rating Discrete Emotional Responses to Music." In: Mitsuko Aramaki; Mathieu Barthet; Richard Kronland-Martinet; and Sølvi Ystad (Eds.) *From Sounds to Music and Emotions: LNCS 7900*. Springer, pp. 1–18.

Schubert, Emery; Joe Wolfe; and Alex Tarnopolsky (2004): "Spectral centroid and timbre in complex, multiple instrumental textures." In: *Proceedings of the 8th International Conference on Music Perception & Cognition 2004*.

Sethares, William A. (2005): *Tuning, Timbre, Spectrum, Scale*. 2nd edition. London: Springer.

Singer, Tania; et al. (2004): "Empathy for pain involves the affective but not sensory components of pain." In: *Science (New York)*, **303**(5661), pp. 1157–1162. doi:10.1126/science.1093535.

Slaney, Malcolm (1998): "Auditory Toolbox: Version 2, Technical Report #1998-010." URL http://www.tka4.org/materials/lib/Articles-Books/Speech%20Recognition/AuditoryToolboxTechReport.pdf.

Sloboda, John A. and Andreas C. Lehmann (2001): "Tracking Performance Correlates of Changes in Perceived Intensity of Emotion During Different Interpretations of a Chopin Piano Prelude." In: *Music Perception*, **19**(1), pp. 87–120. doi:10.1525/mp.2001.19.1.87.

Synak, Piotr and Alicja Wieczorkowska (2005): "Some issues on detecting emotions in music." In: Dominik Ślęzak; JingTao Yao; James F. Peters; Wojciech Ziarko; and Xiaohua Hu (Eds.) *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, vol. 3642. Berlin and Heidelberg: Springer, pp. 314–322. doi:10.1007/11548706\textunderscore33. URL http://dx.doi.org/10.1007/11548706_33.

Tarchanoff, J. (1890): "Ueber die galvanischen Erscheinungen in der Haut des Menschen bei Reizungen der Sinnesorgane und bei verschiedenen Formen der psychischen Thätigkeit." In: *Archiv für die gesamte Physiologie des Menschen und der Tiere*, **46**(1), pp. 46–55. doi:10.1007/BF01789520. URL http://dx.doi.org/10.1007/BF01789520.

Thayer, Robert E. (1989): *The biopsychology of mood and arousal*. New York: Oxford University Press.

Topend Sports - The Sport + Science Resource (2013): "Resting Heart Rate Table." URL http://www.topendsports.com/testing/heart-rate-resting-chart.htm.

Tzanetakis, George and Perry Cook (2000): "MARSYAS: A framework for audio analysis." URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.4572&rep=rep1&type=pdf.

Varèse, Edgar (1946): *Density 21.5 (1936)*. Milan: Casa Ricordi.

Vuoskoski, Jonna K. and Tuomas Eerola (2011): "Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences." In: *Musicae Scientiae*, **15**(2), pp. 159–173. doi:10.1177/1029864911403367.

Watson, David and Lee Anna Clark (1999): "The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form: Updated Version 08/1999." URL http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1011&context=psychology_pubs.

Watson, David; Lee Anna Clark; and Auke Tellegen (1988): "Development and validation of brief measures of positive and negative affect: The PANAS scales." In: *Journal of Personality and Social Psychology*, **54**(6), pp. 1063–1070. doi:10.1037/0022-3514.54.6.1063.

Watson, David and R. L. Mandryk (2012): "Modeling Musical Mood from Audio Features and Listening Context on an IN-SITU Data Set." In: *Proceedings of the 13th International Conference on Music Information Retrieval 2012*.

Weinzierl, Stefan (Ed.) (2008): *Handbuch der Audiotechnik*. Berlin and Heidelberg, New York: Springer. doi:10.1007/978-3-540-34301-1.

West, Brady T.; Brenda Gillespie; and Kathleen B. Welch (2007): *Linear mixed models: A practical guide using statistical software*. London: Chapman & Hall/CRC.

Yeomans, John S. and Paul W. Frankland (1996): "The acoustic startle reflex: neurons and connections." In: *Brain Research Reviews*, **21**(3), pp. 301–314. doi:10.1016/0165-0173(96)00004-5.

Zentner, Marcel R. and Tuomas Eerola (2010): "Self-Report Measures and Models." In: Patrik N. Juslin and John A. Sloboda (Eds.) *Handbook of Music and Emotion (reprinted 2012)*, Series in affective science. Oxford: Oxford University Press, pp. 187–221.

List of Figures

List of Tables

Appendix A Scores

Figure A.1. Hindemith - Acht Stücke für Flöte allein: VI Lied, leicht bewegt (Hindemith, 1958, p. 5)

Figure A.2. Poulenc - Un joueur de flûte berce les ruines (Poulenc, 2000).

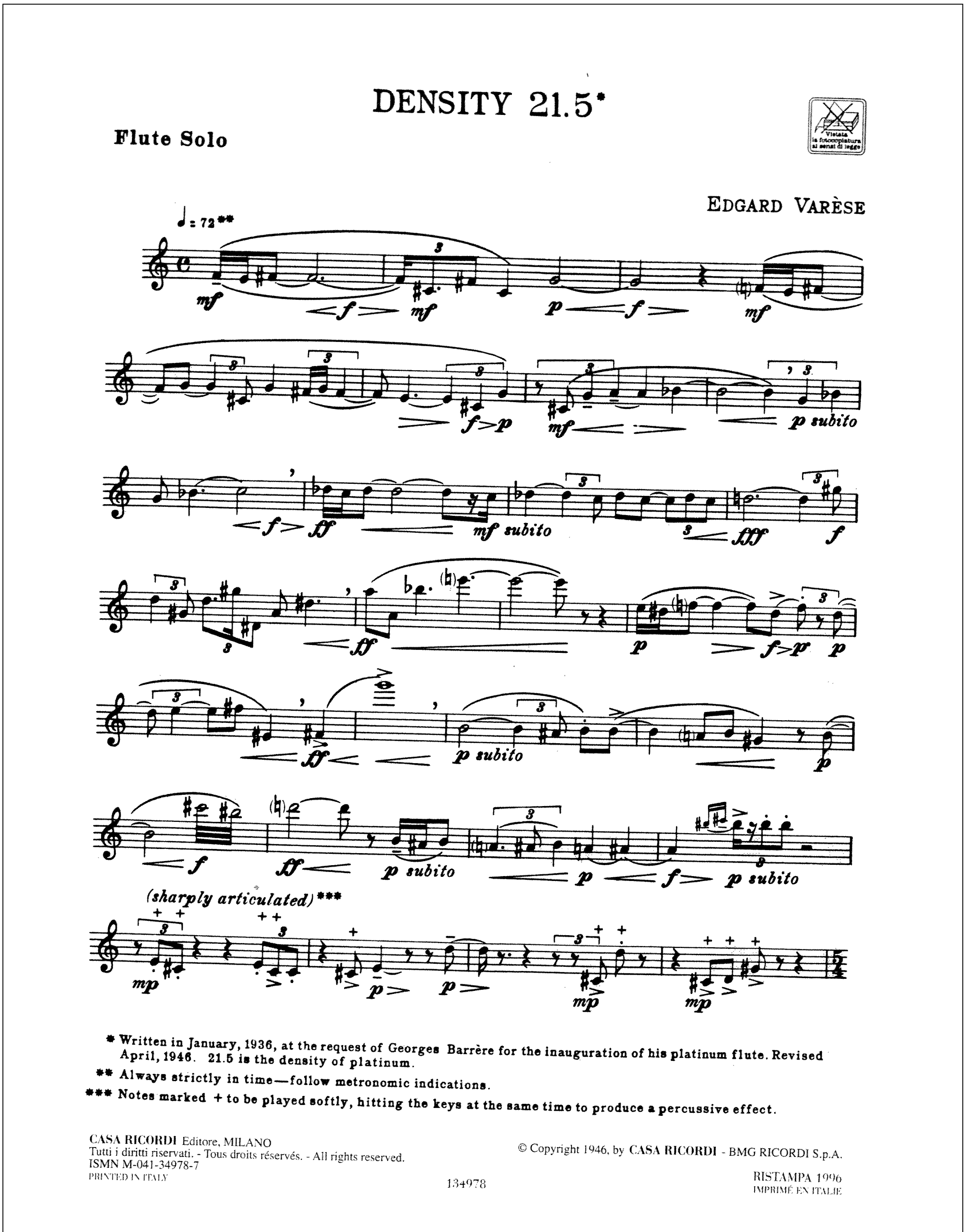


Figure A.3. Varèse - Density 21.5 (Varèse, 1946, p. 1/2).

Figure A.4. Varèse - Density 21.5 (Varèse, 1946, p. 2/2).

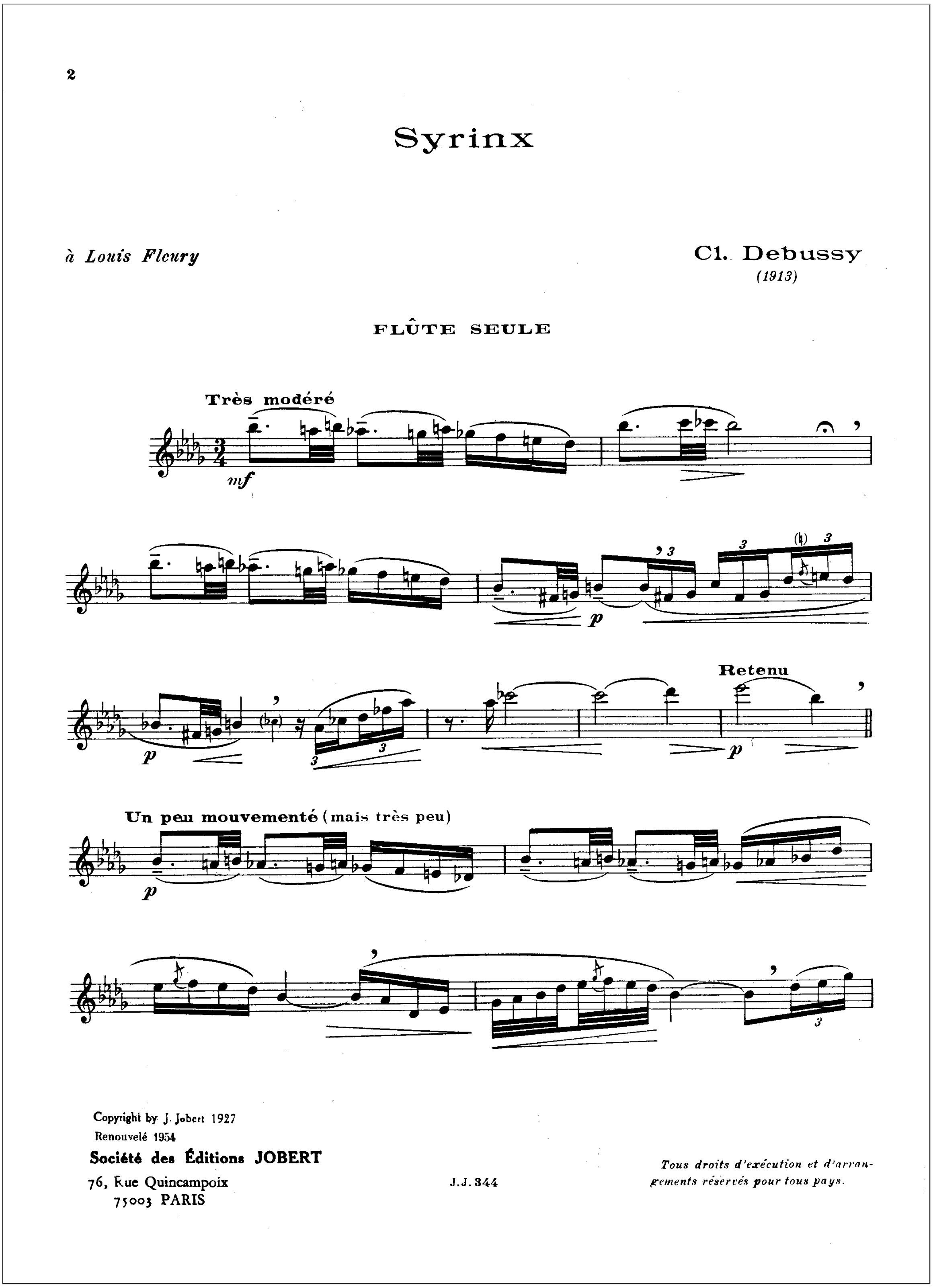


Figure A.5. Debussy - Syrinx (Debussy, 1927, p. 1/2).

Figure A.6. Debussy - Syrinx (Debussy, 1927, p. 2/2).

Figure A.7. Bach - Partita in A minor, BWV 1013: II. Corrente (Bach, 1917a, p. 1/2).

2 flute solo

Music engraving by LilyPond 2.10.20—www.lilypond.org

Figure A.8. Bach - Partita in A minor, BWV 1013: II. Corrente (Bach, 1917a, p. 2/2).

Figure A.9. Bach - Partita in A minor, BWV 1013: III. Sarabande (Bach, 1917b).

Appendix B Plots Information Content/Unexpectedness

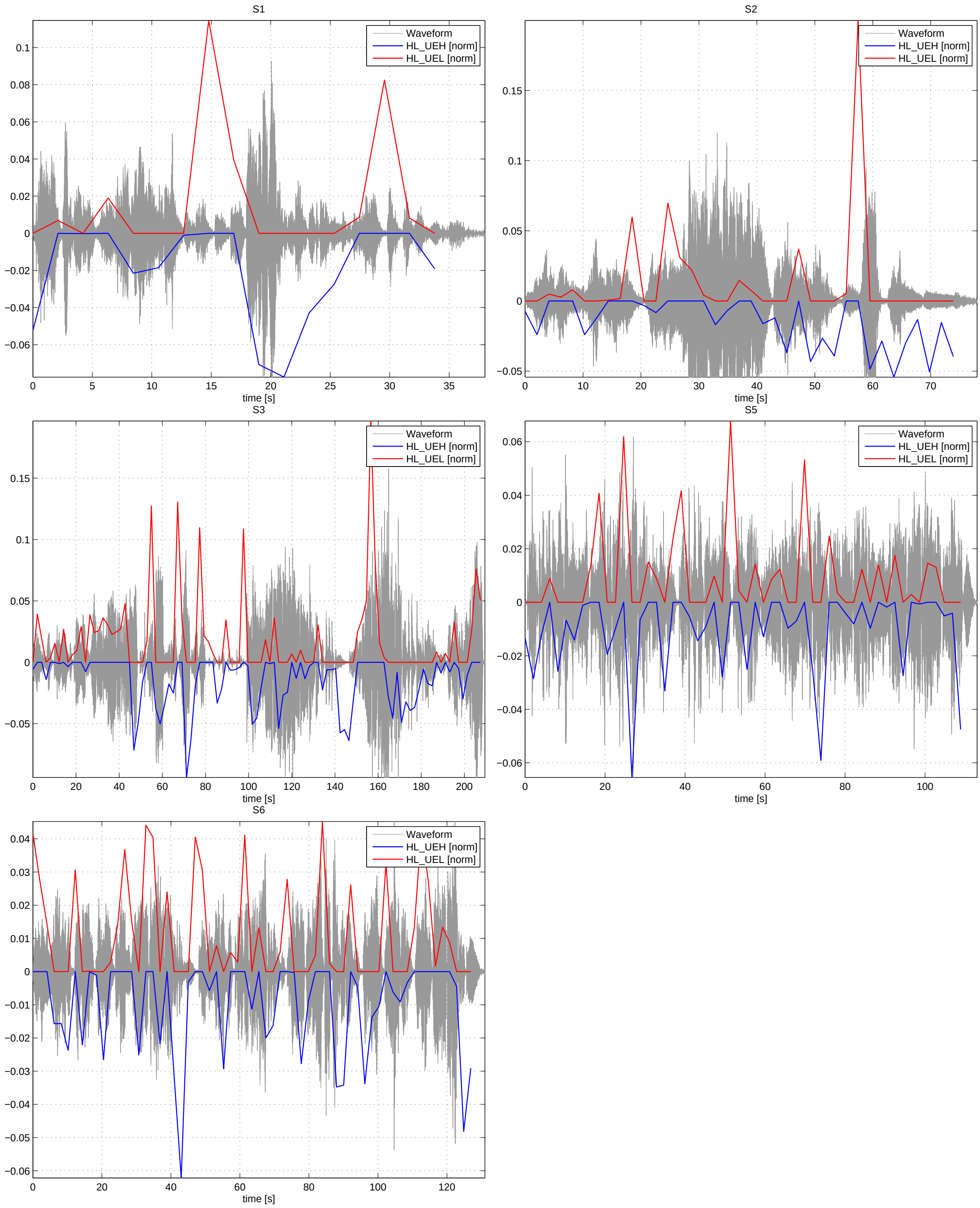


Figure B.1. UE high/low S1-S3 and S5-S6.

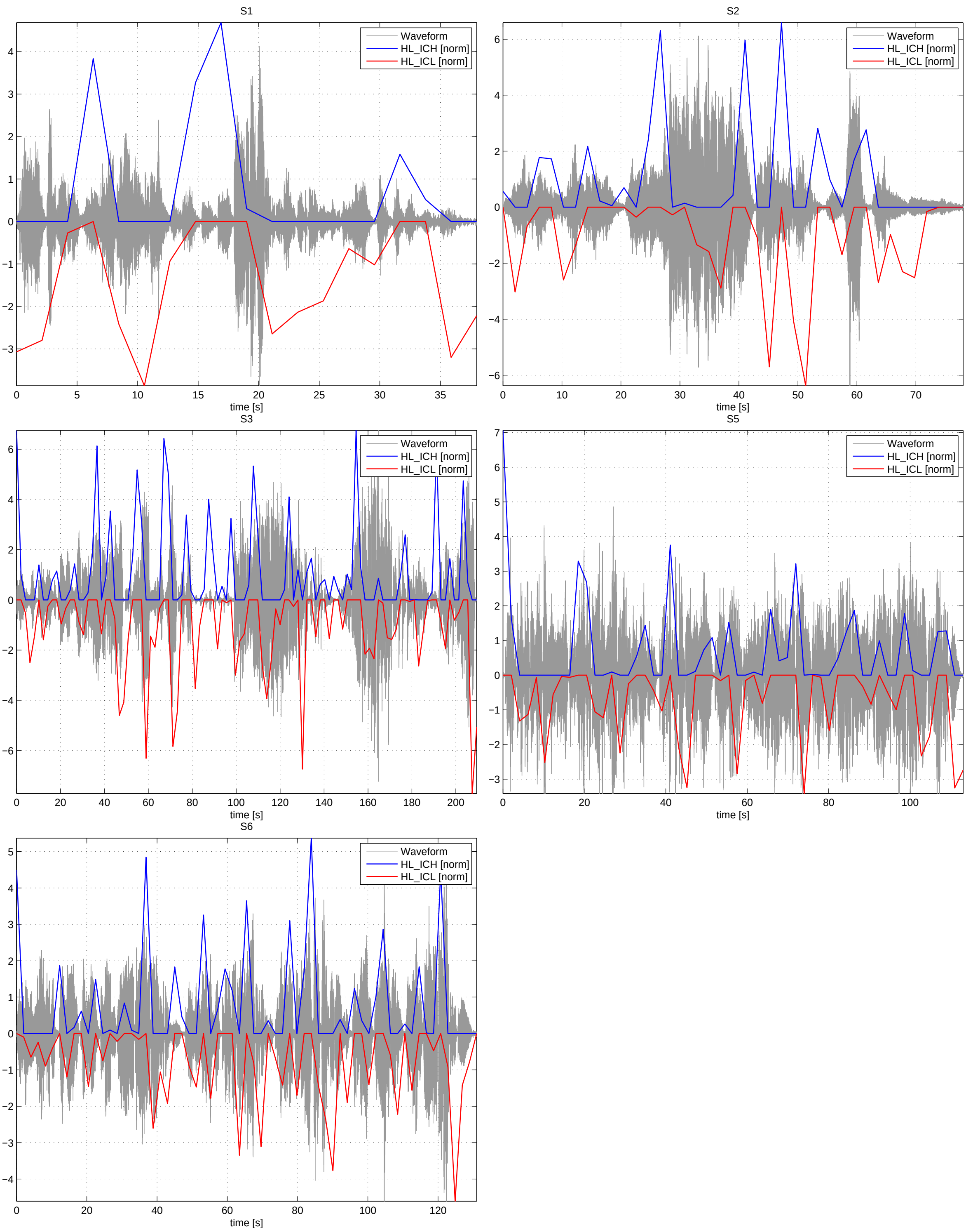


Figure B.2. IC high/low S1-S3 and S5-S6.

Appendix C Plots Second Order Predictors vs. dVs

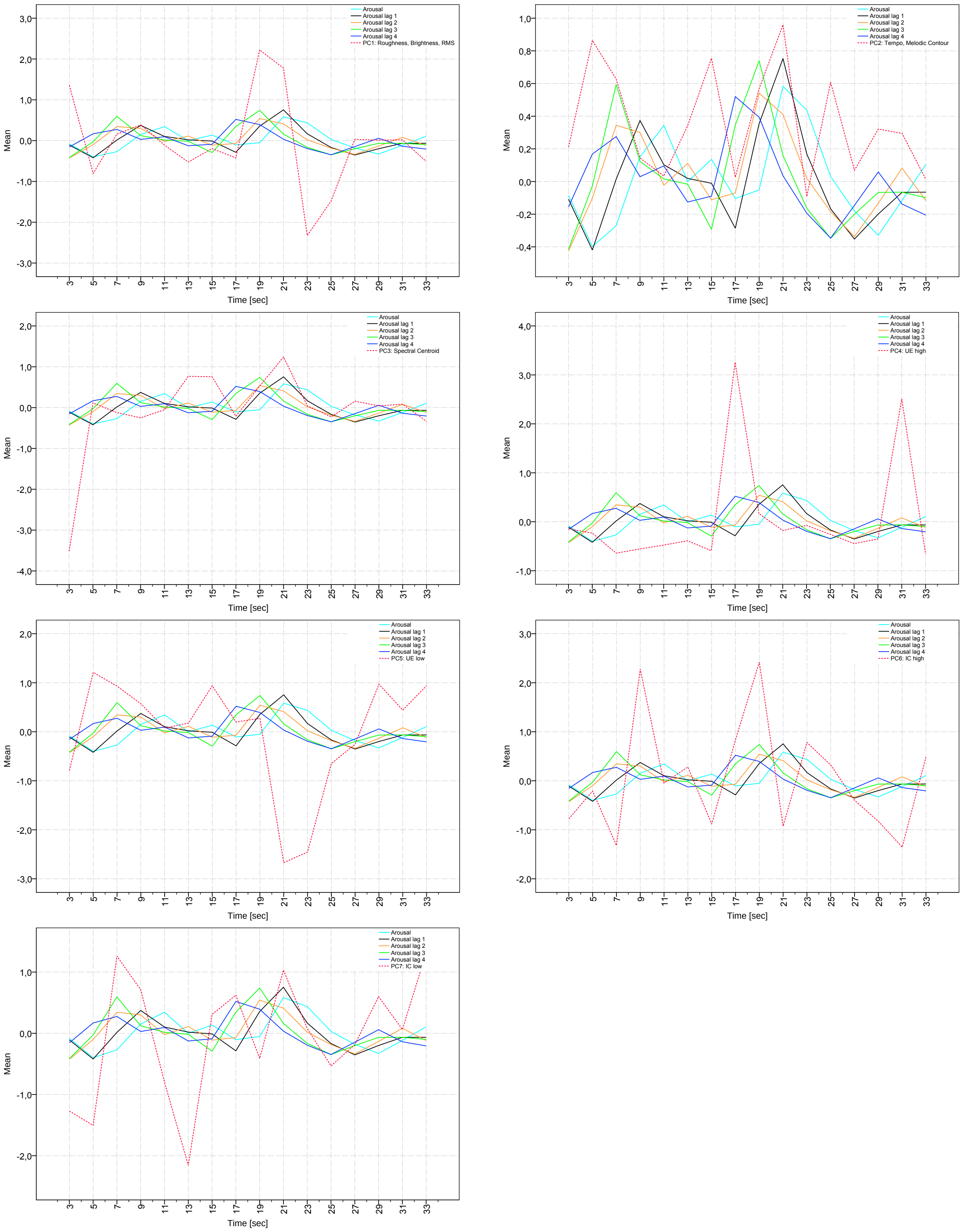


Figure C.1. S1, Arousal.

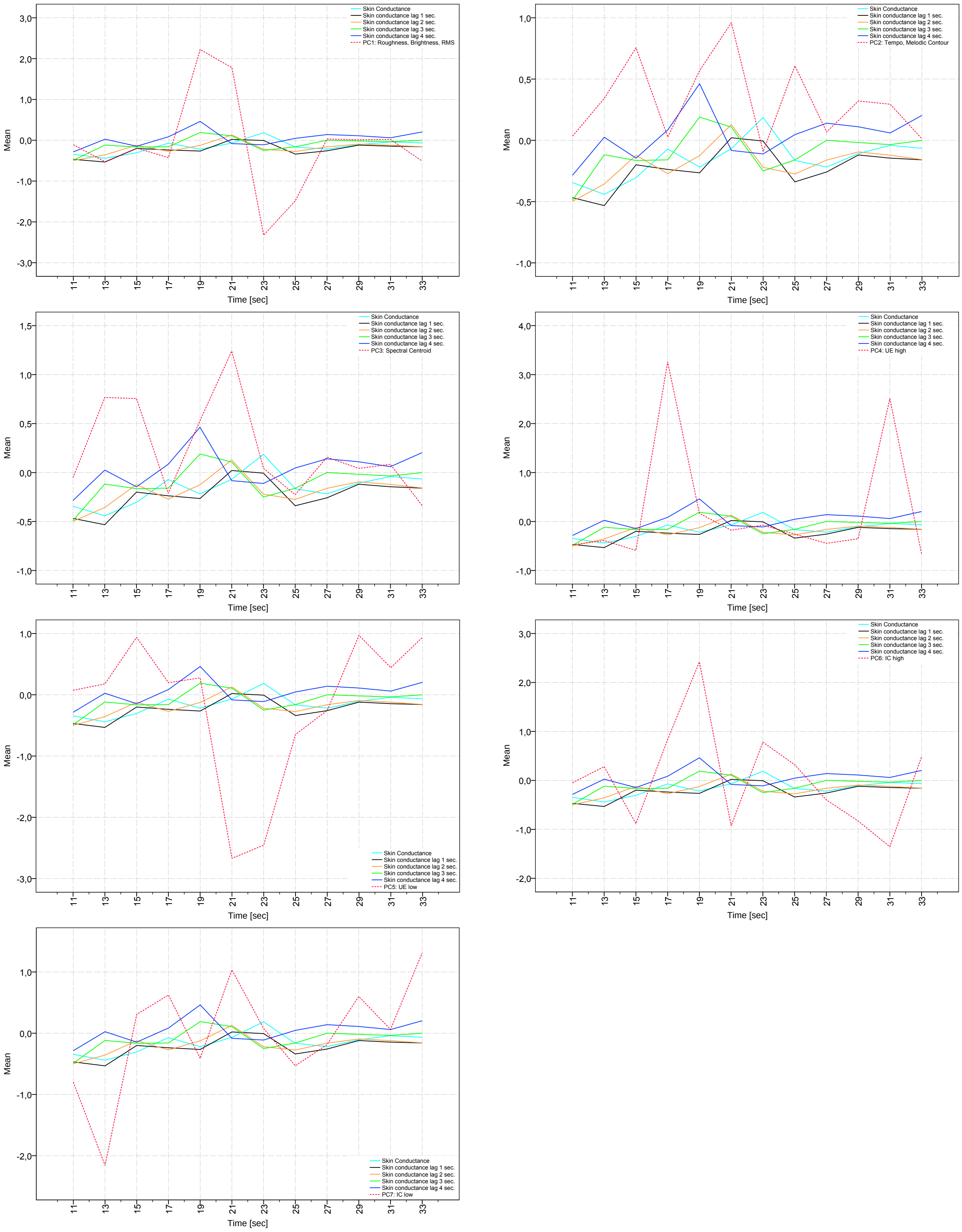


Figure C.2. S1, Skin Conductance.

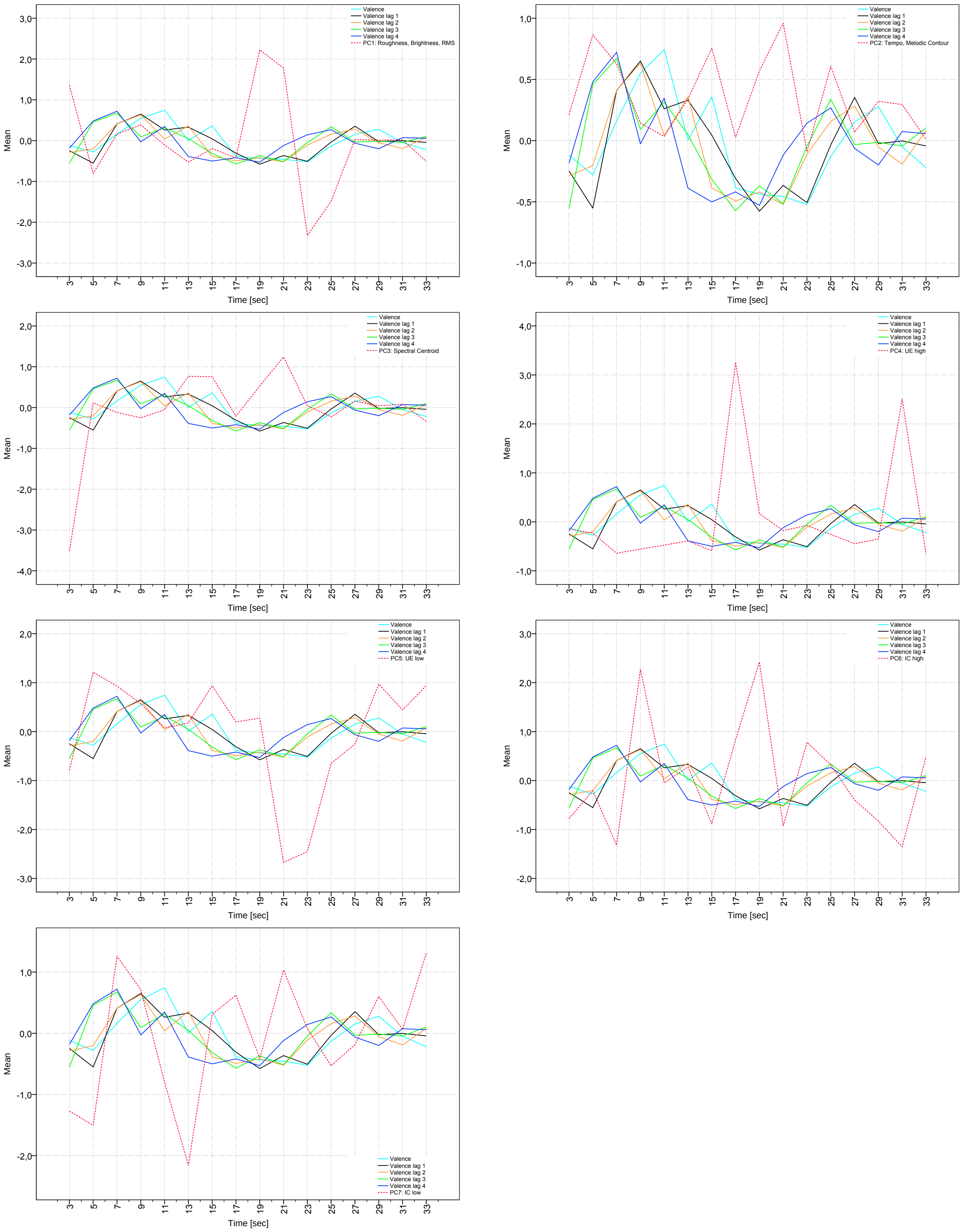


Figure C.3. S1, Valence.

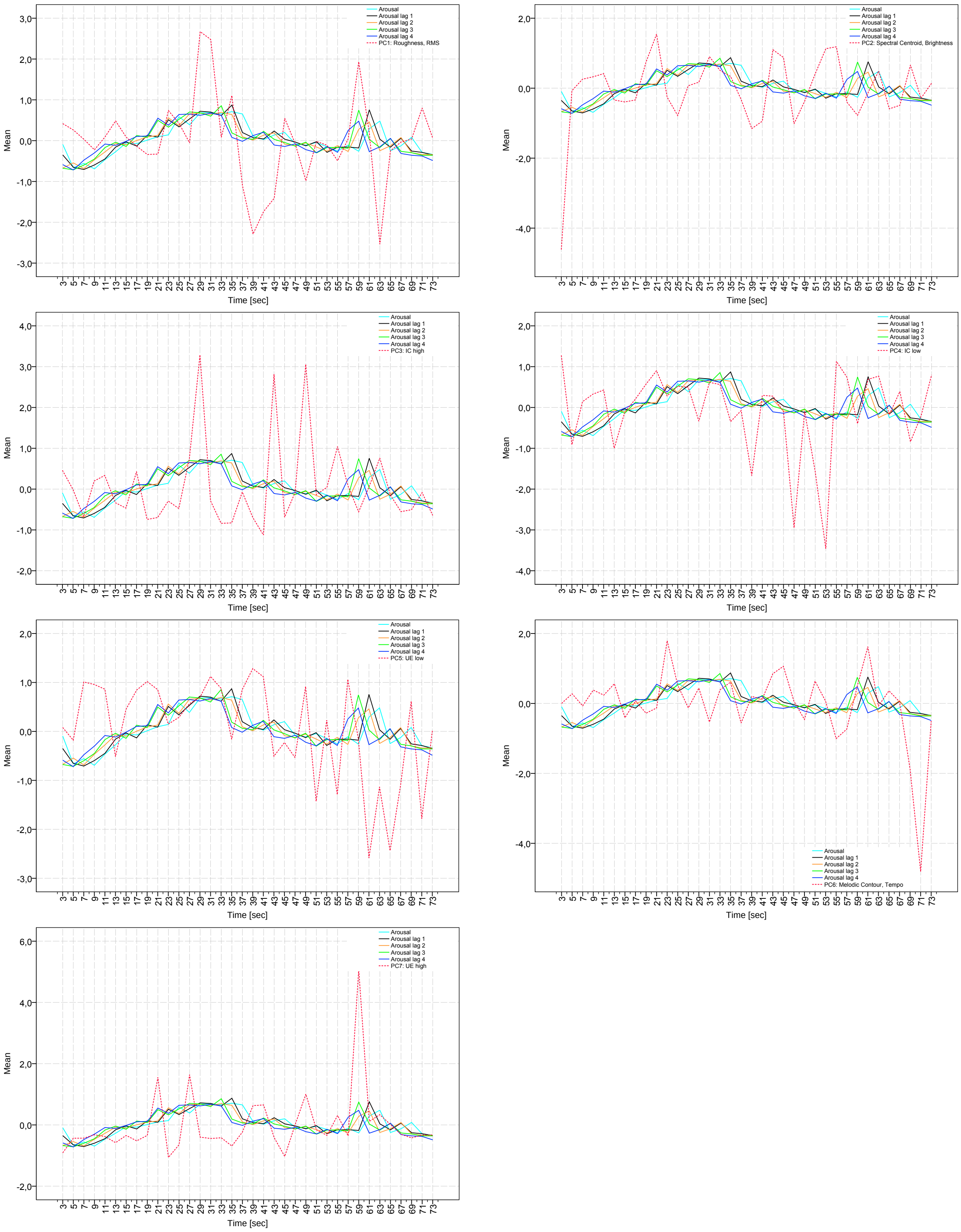


Figure C.4. S2, Arousal.

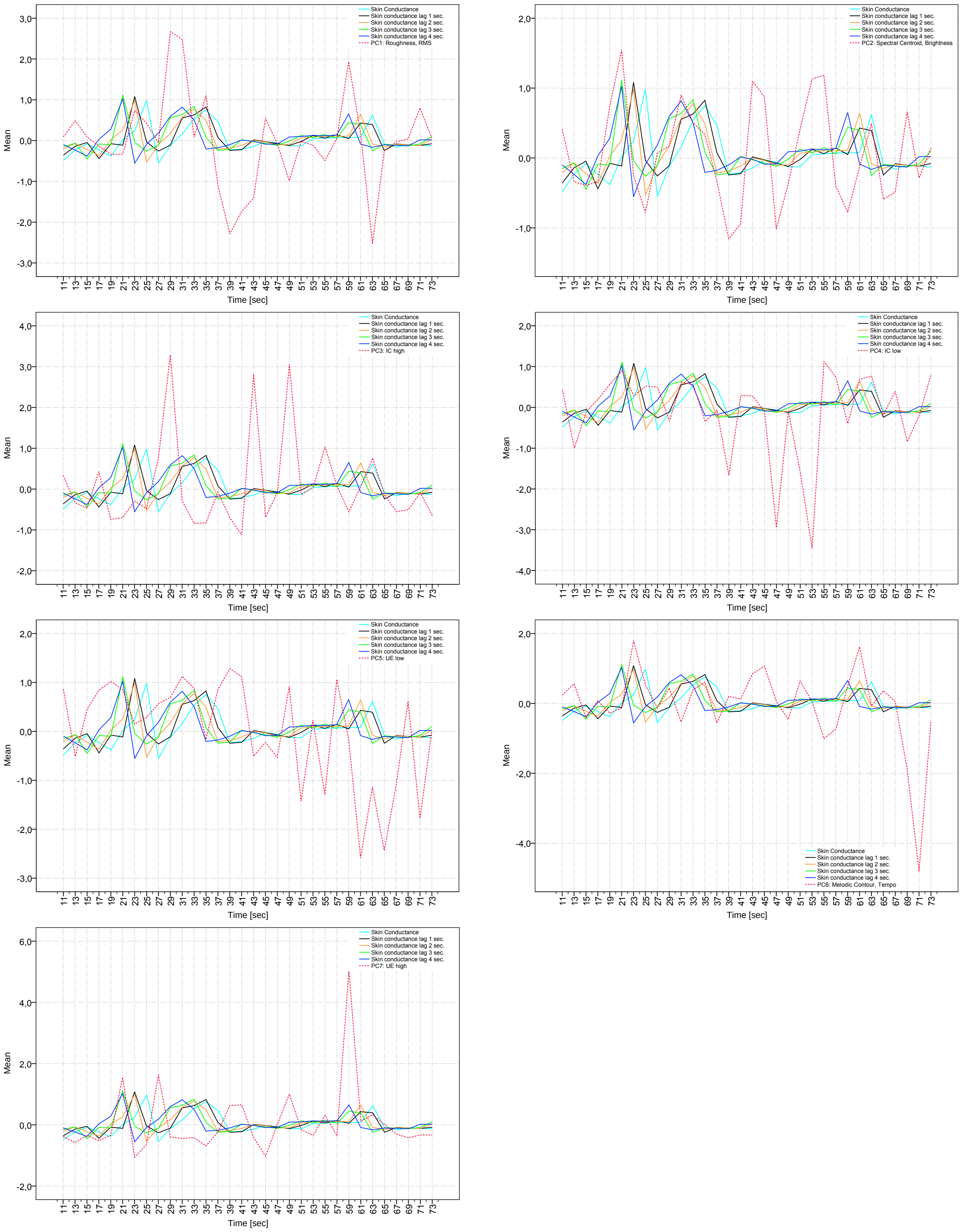


Figure C.5. S2, Skin Conductance.

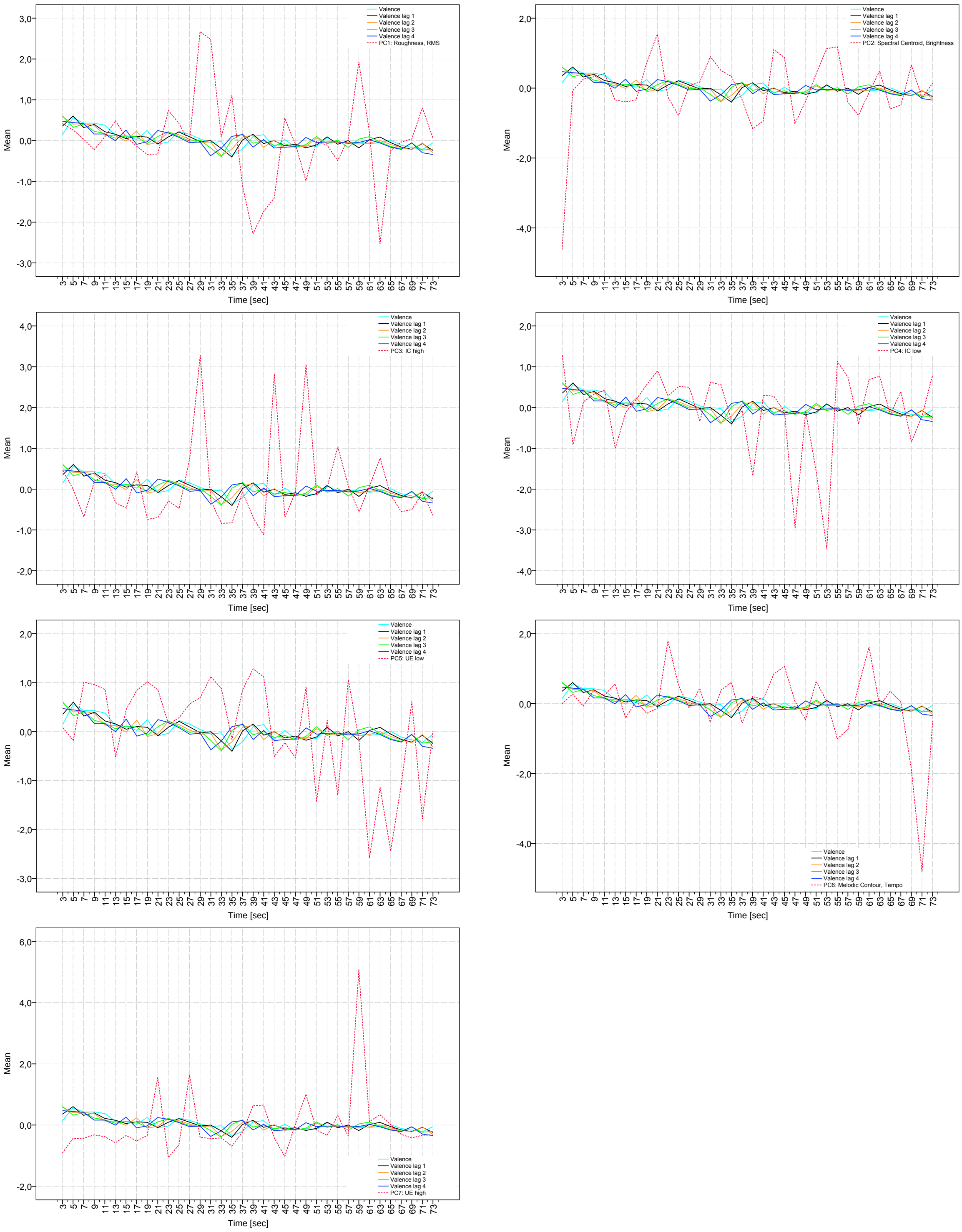


Figure C.6. S2, Valence.

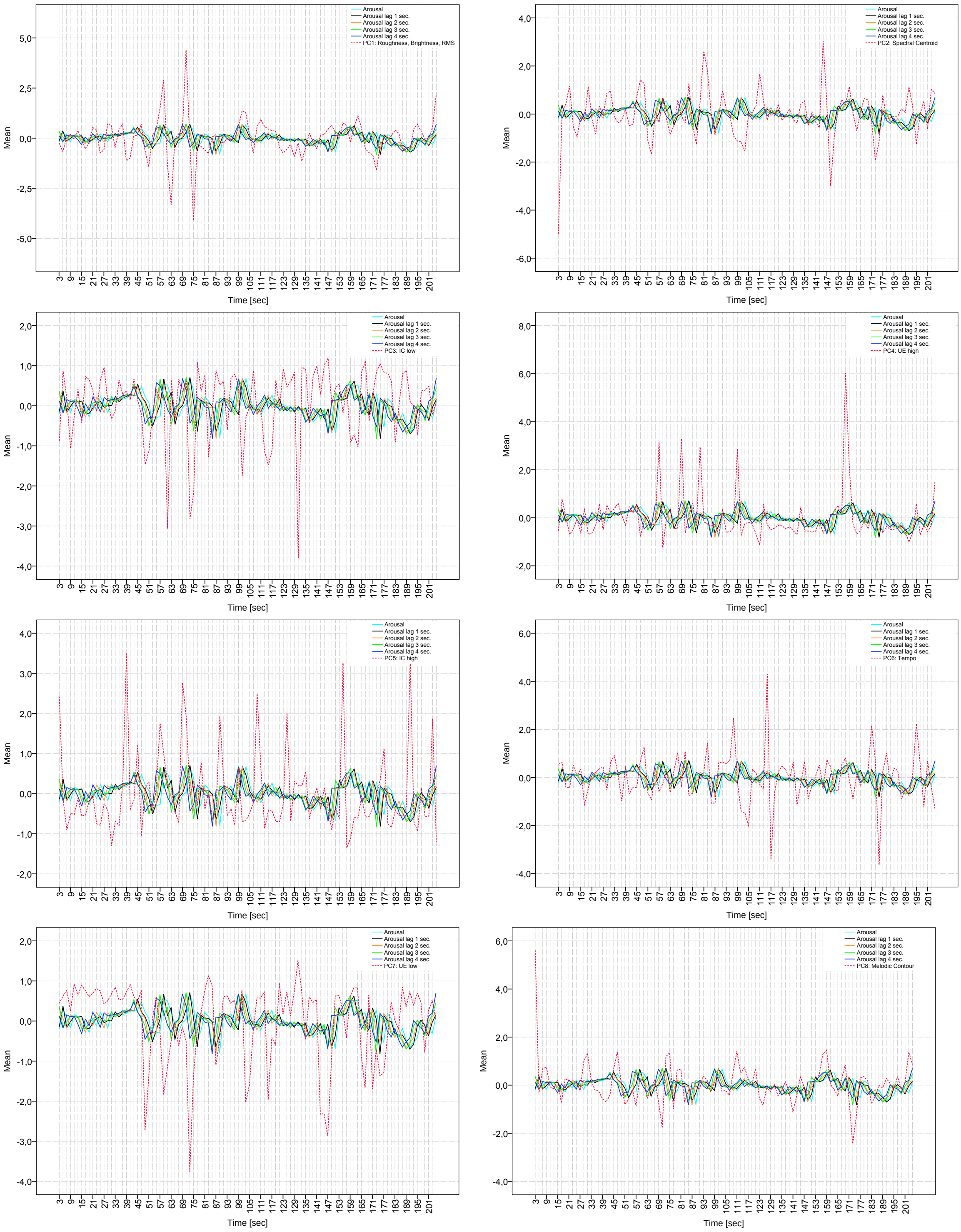


Figure C.7. S3, Arousal.

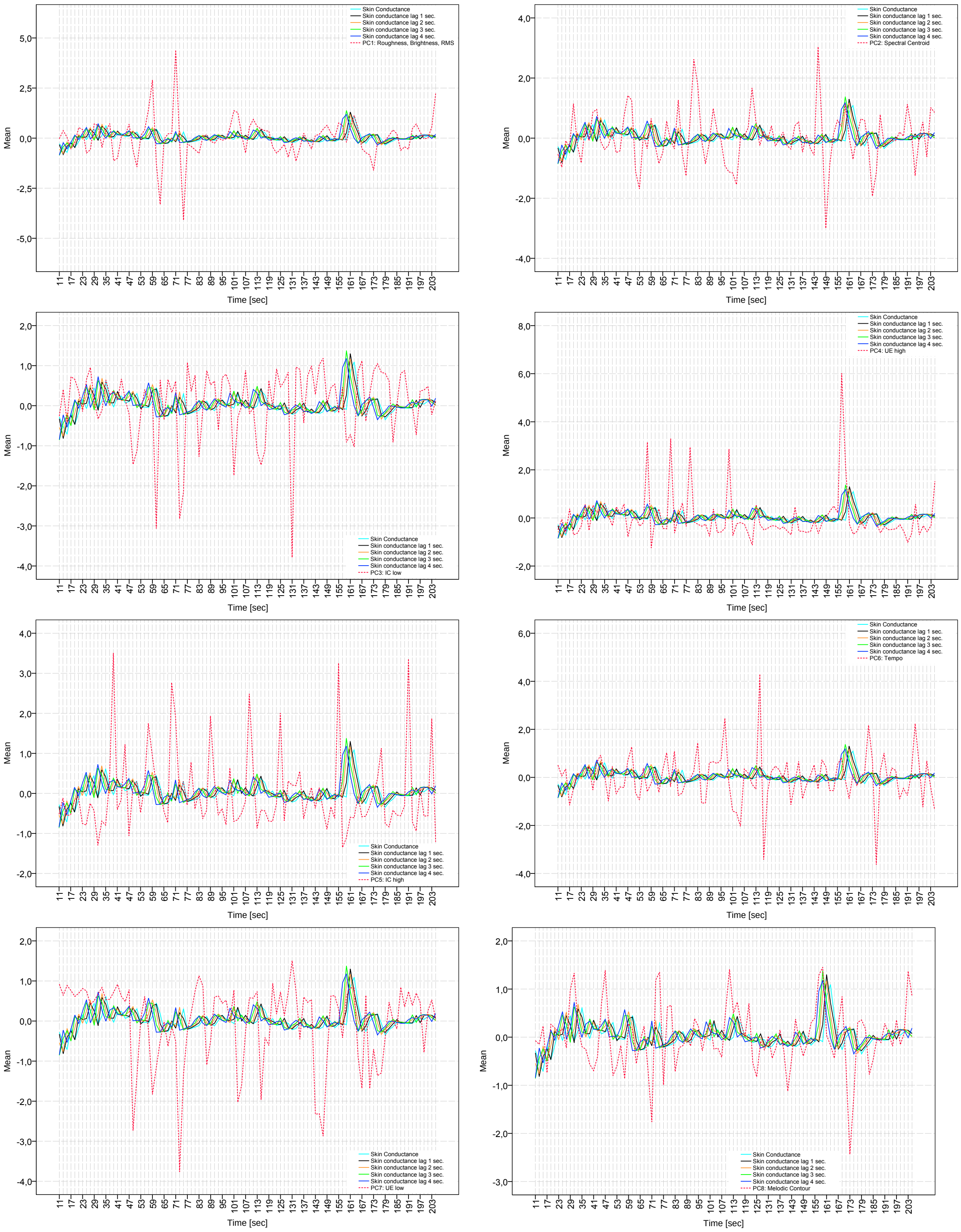


Figure C.8. S3, Skin Conductance.

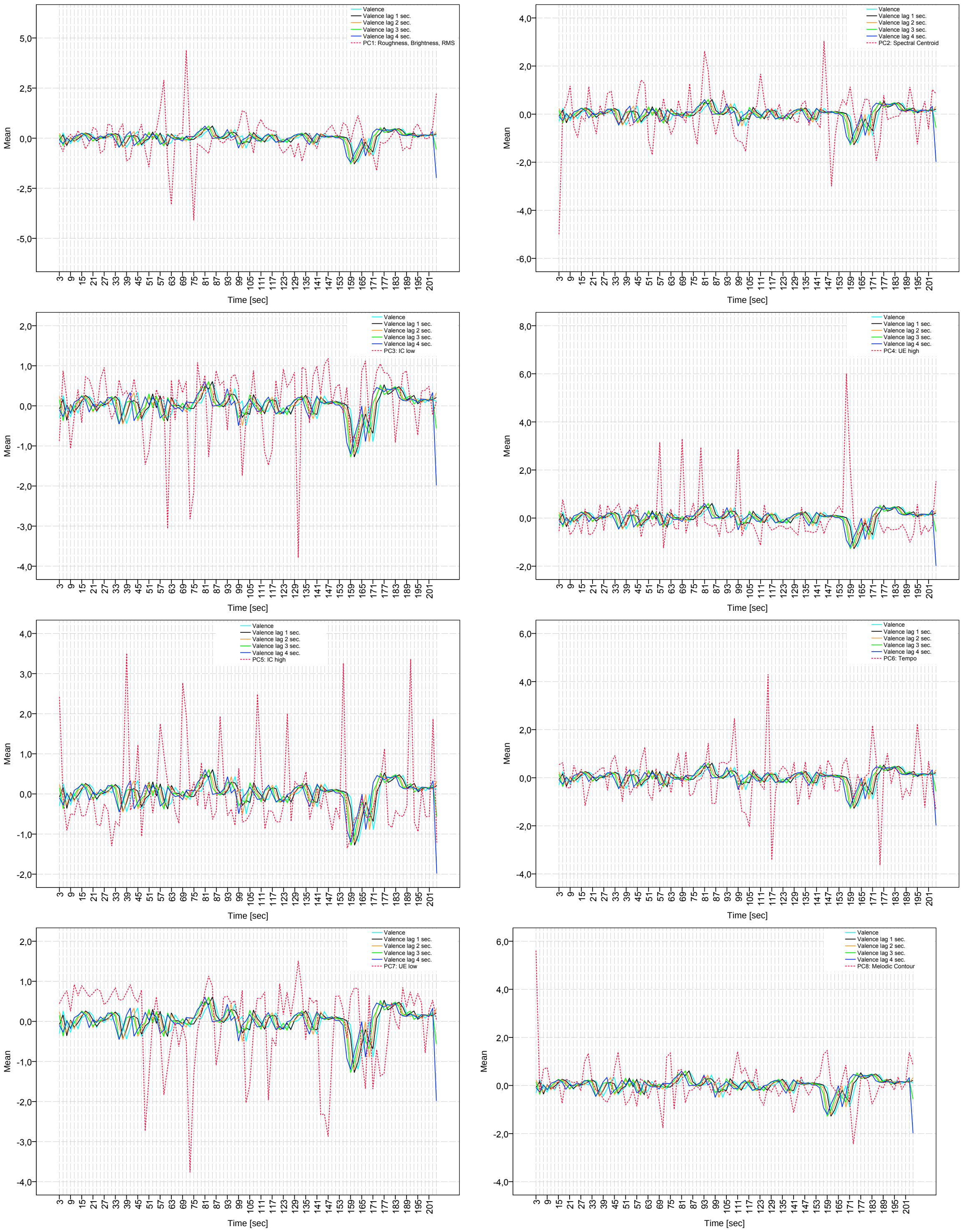


Figure C.9. S3, Valence.

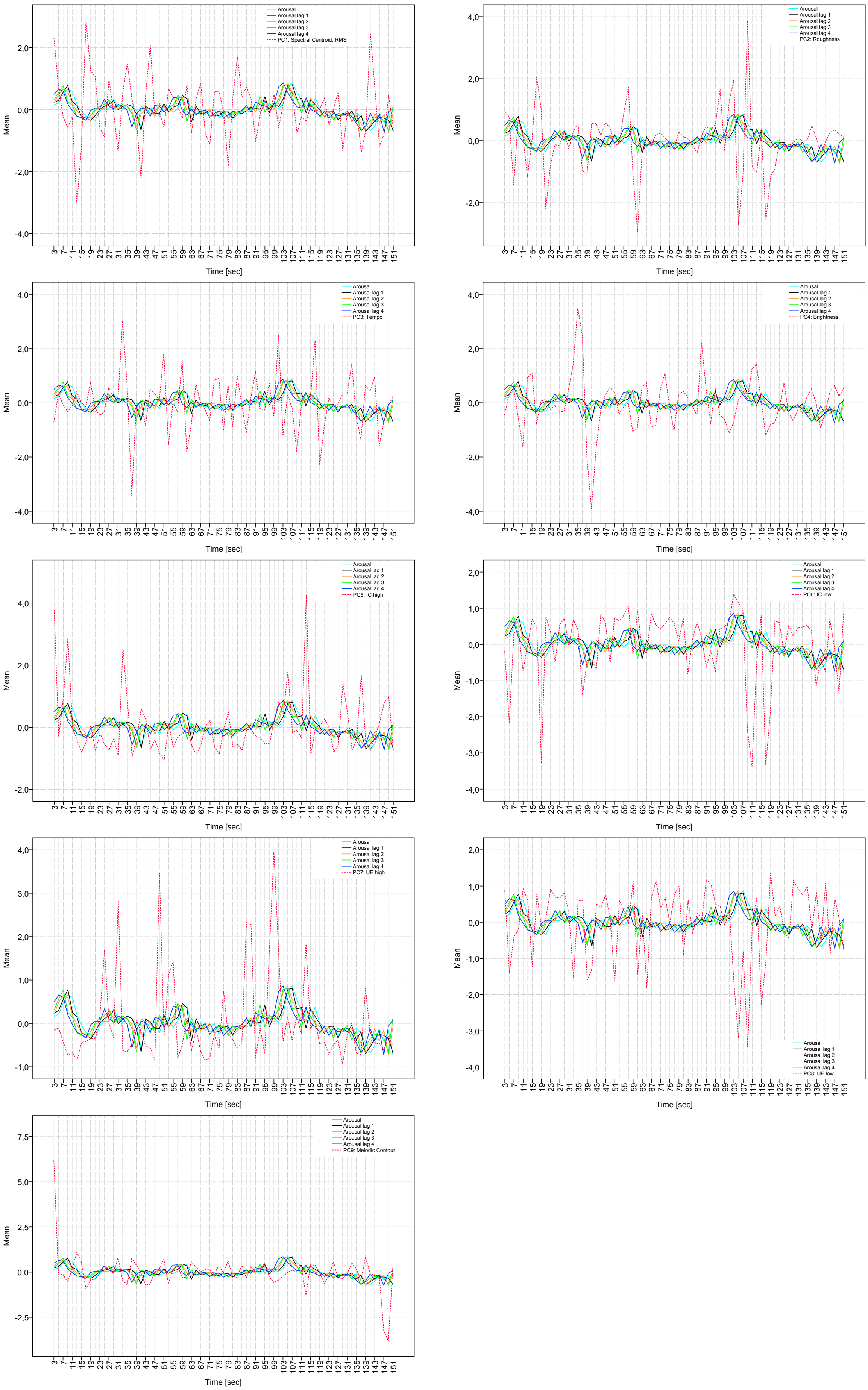


Figure C.10. S4, Arousal.

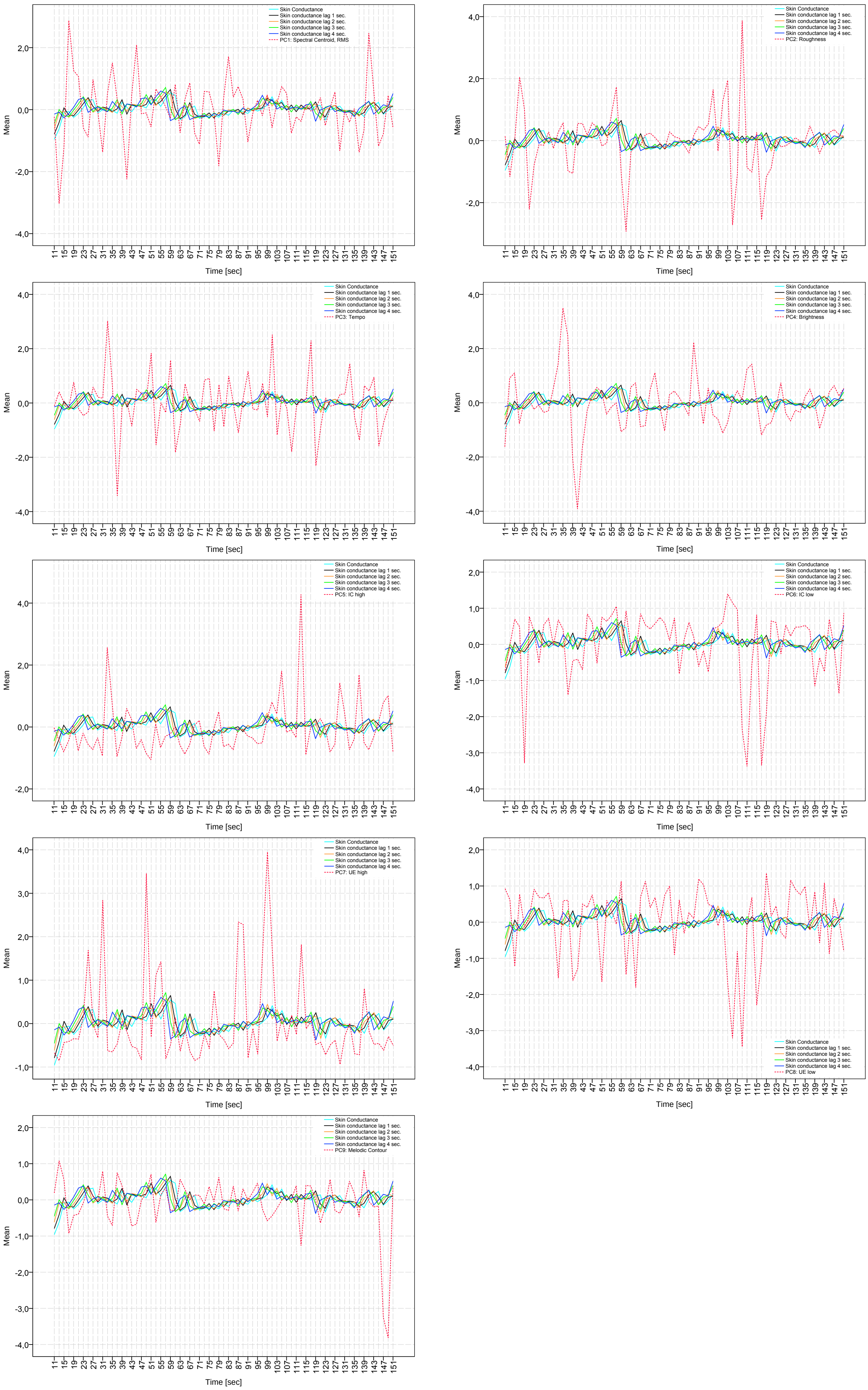


Figure C.11. S4, Skin Conductance.

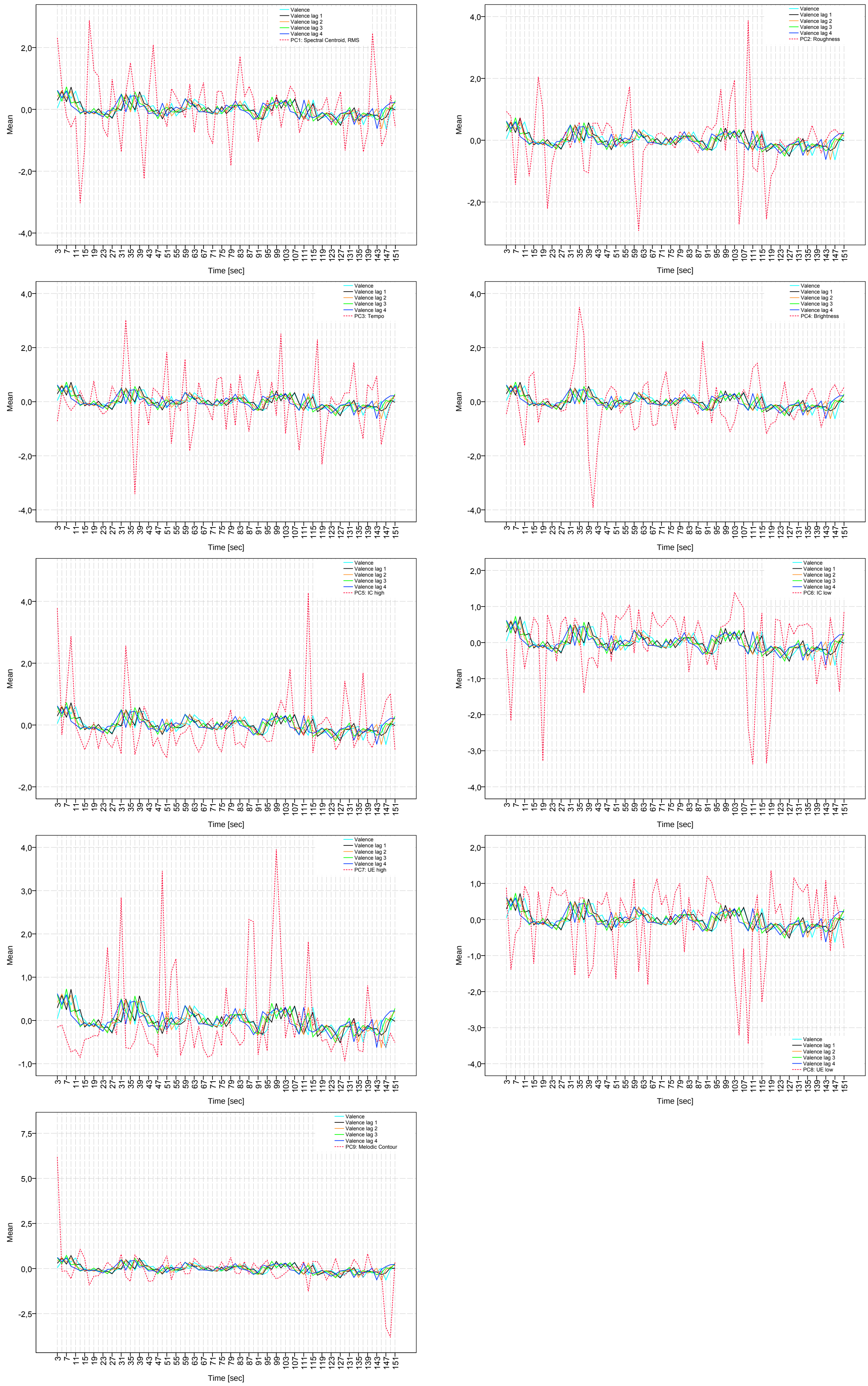


Figure C.12. S4, Valence.

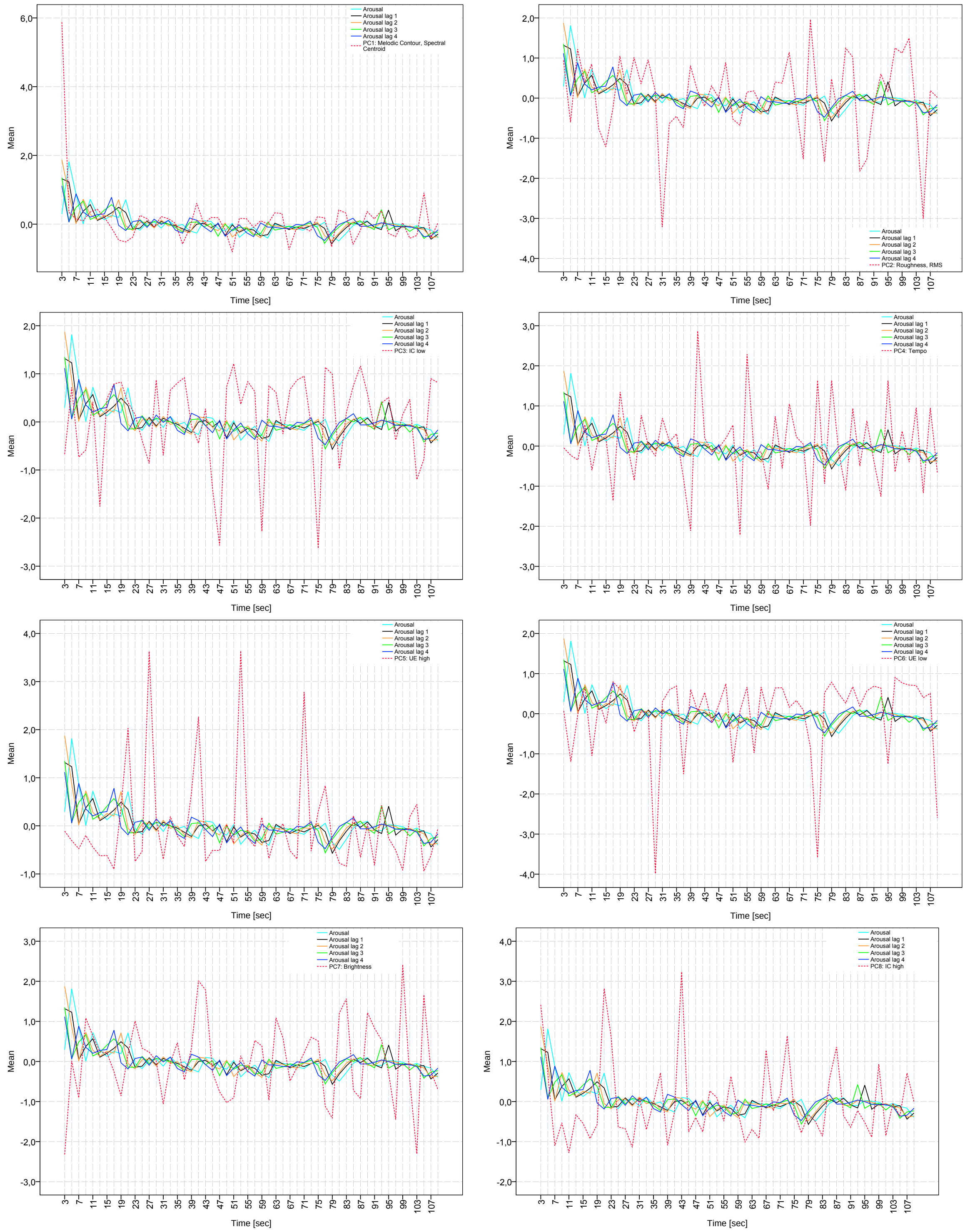


Figure C.13. S5, Arousal.

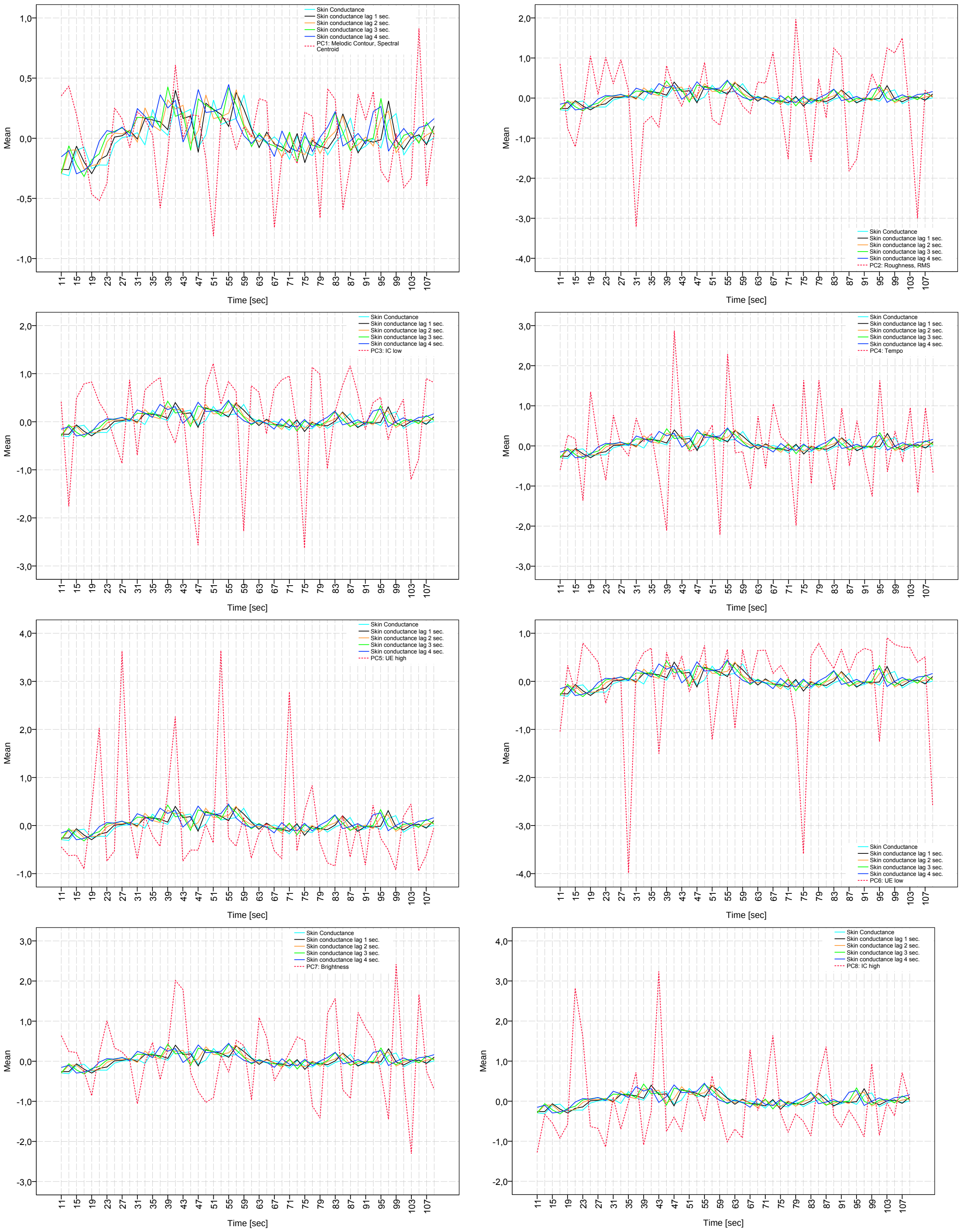


Figure C.14. S5, Skin Conductance.

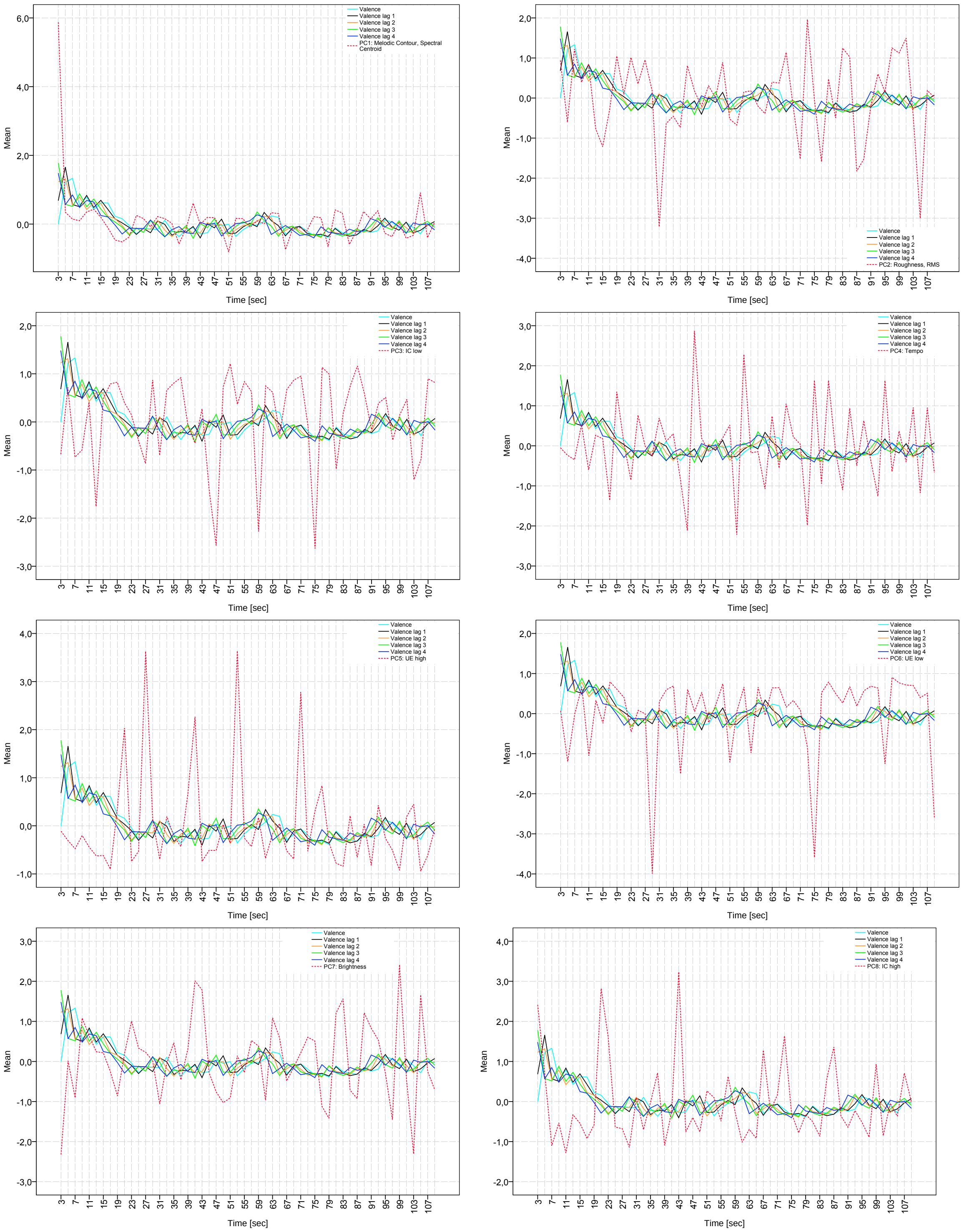


Figure C.15. S5, Valence.

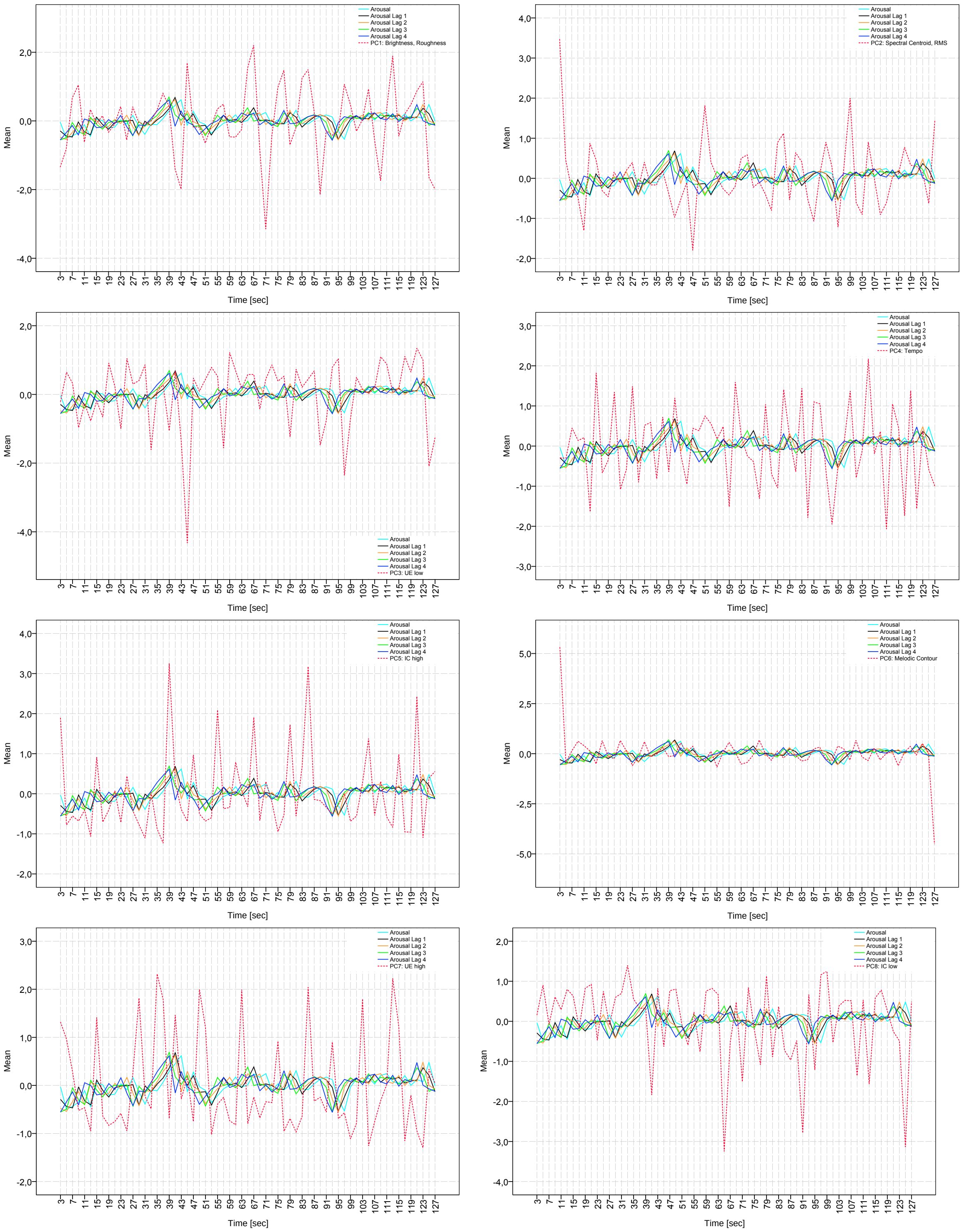


Figure C.16. S6, Arousal.

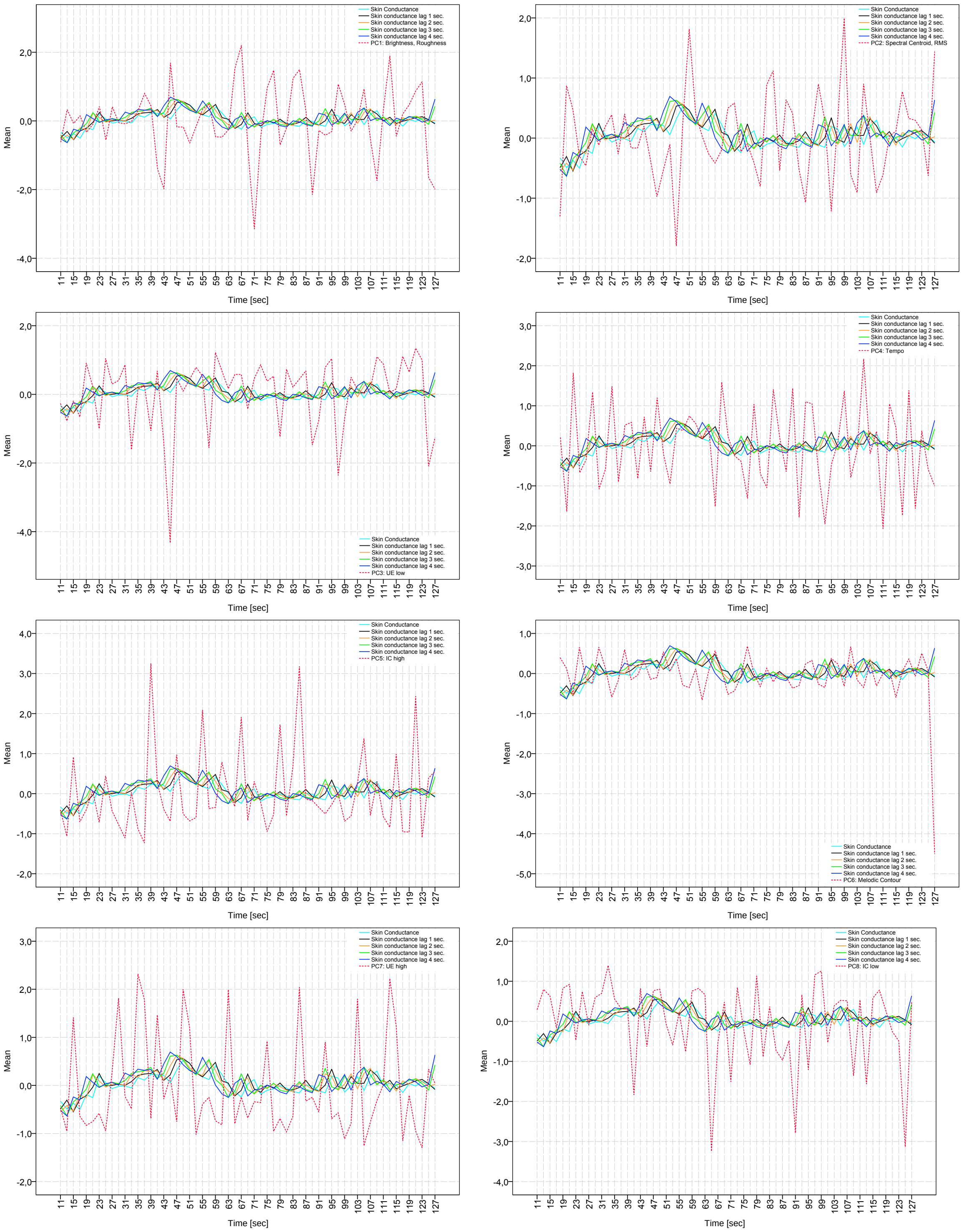


Figure C.17. S6, Skin Conductance.

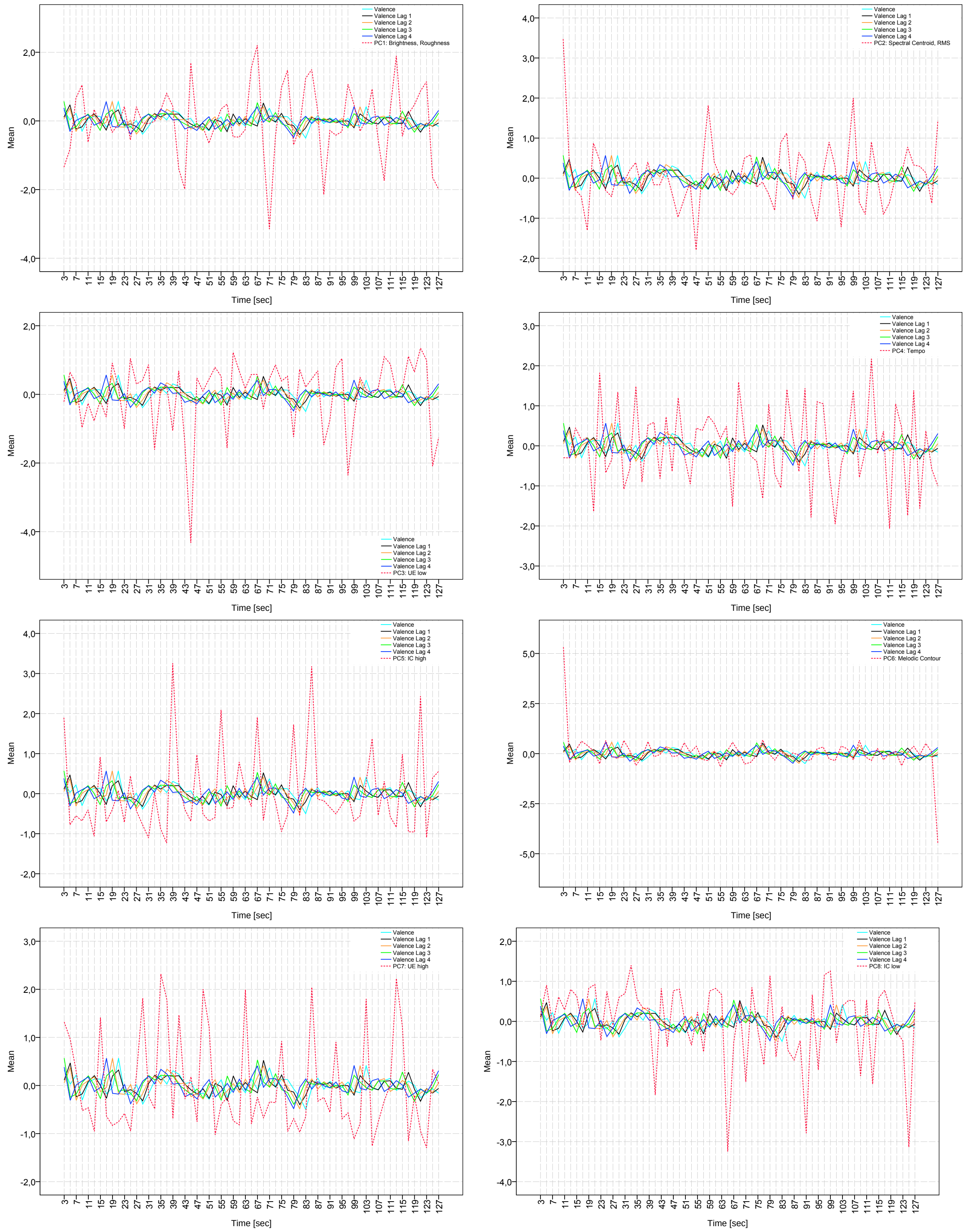


Figure C.18. S6, Valence.

Appendix D Statistics

Korrelationsmatrix

		Tempo	RMS	Brightness	Roughness	Spectral Centroid	Melodic Contour	IC high (basic)	IC low (basic)	Unexpectedness mean high lag 3 sec.	Unexpectedness mean low lag 3 sec.
Korrelation	Tempo	1.000	.067	.051	.069	-.275	.765	.071	.318	-.120	.032
	RMS	.067	1.000	.806	.919	-.615	.314	.496	.303	.184	.088
	Brightness	.051	.806	1.000	.880	-.263	.277	.474	.523	.159	.165
	Roughness	.069	.919	.880	1.000	-.351	.328	.541	.404	.204	.258
	Spectral Centroid	-.275	-.615	-.263	-.351	1.000	-.345	-.035	.067	.090	.267
	Melodic Contour	.765	.314	.277	.328	-.345	1.000	.471	.544	.082	.013
	IC high (basic)	.071	.496	.474	.541	-.035	.471	1.000	.568	.496	.371
	IC low (basic)	.318	.303	.523	.404	.067	.544	.568	1.000	.330	.355
	Unexpectedness mean high lag 3 sec.	-.120	.184	.159	.204	.090	.082	.496	.330	1.000	.392
	Unexpectedness mean low lag 3 sec.	.032	.088	.165	.258	.267	.013	.371	.355	.392	1.000

Korrelationsmatrix

		Tempo	RMS	Brightness	Roughness	Spectral Centroid	Melodic Contour	IC high (basic)	IC low (basic)	Unexpectedness mean high lag 3 sec.	Unexpectedness mean low lag 3 sec.
Korrelation	Tempo	1.000	.380	-.106	.389	-.416	.181	-.329	-.341	-.036	-.220
	RMS	.380	1.000	.159	.734	-.425	.268	-.086	.013	.462	.183
	Brightness	-.106	.159	1.000	.310	.345	.201	.383	.343	.018	.014
	Roughness	.389	.734	.310	1.000	-.206	.183	-.153	-.004	.431	.114
	Spectral Centroid	-.416	-.425	.345	-.206	1.000	-.173	.215	.102	-.127	-.094
	Melodic Contour	.181	.268	.201	.183	-.173	1.000	.269	.375	.217	.455
	IC high (basic)	-.329	-.086	.383	-.153	.215	.269	1.000	.363	.071	.164
	IC low (basic)	-.341	.013	.343	-.004	.102	.375	.363	1.000	.064	.446
	Unexpectedness mean high lag 3 sec.	-.036	.462	.018	.431	-.127	.217	.071	.064	1.000	.318
	Unexpectedness mean low lag 3 sec.	-.220	.183	.014	.114	-.094	.455	.164	.446	.318	1.000

Korrelationsmatrix

		Tempo	RMS	Brightness	Roughness	Spectral Centroid	Melodic Contour	IC high (basic)	IC low (basic)	Unexpectedness mean high lag 3 sec.	Unexpectedness mean low lag 3 sec.
Korrelation	Tempo	1.000	.003	.106	.075	.092	-.107	.015	-.087	.016	.027
	RMS	.003	1.000	.676	.745	-.176	.427	.416	.308	.311	.351
	Brightness	.106	.676	1.000	.663	.333	.268	.320	.276	.462	.323
	Roughness	.075	.745	.663	1.000	.044	.293	.354	.319	.358	.391
	Spectral Centroid	.092	-.176	.333	.044	1.000	-.392	-.096	-.086	.166	.017
	Melodic Contour	-.107	.427	.268	.293	-.392	1.000	.502	.525	.126	.153
	IC high (basic)	.015	.416	.320	.354	-.096	.502	1.000	.379	.377	.256
	IC low (basic)	-.087	.308	.276	.319	-.086	.525	.379	1.000	.143	.356
	Unexpectedness mean high lag 3 sec.	.016	.311	.462	.358	.166	.126	.377	.143	1.000	.368
	Unexpectedness mean low lag 3 sec.	.027	.351	.323	.391	.017	.153	.256	.356	.368	1.000

Figure D.1. PCA Correlation tables S1-S3.

Korrelationsmatrix

		Tempo	RMS	Brightness	Roughness	Spectral Centroid	Melodic Contour	IC high (basic)	IC low (basic)	Unexpectedness mean high lag 3 sec.	Unexpectedness mean low lag 3 sec.
Korrelation	Tempo	1.000	.387	.009	.002	-.295	.121	.079	-.040	.089	.139
	RMS	.387	1.000	.266	.595	-.648	.498	.377	.013	.059	.087
	Brightness	.009	.266	1.000	.200	.202	.164	.296	.293	.157	.155
	Roughness	.002	.595	.200	1.000	-.319	.214	-.024	-.141	.154	.023
	Spectral Centroid	-.295	-.648	.202	-.319	1.000	-.383	-.203	.295	.042	-.084
	Melodic Contour	.121	.498	.164	.214	-.383	1.000	.660	.192	.181	.098
	IC high (basic)	.079	.377	.296	-.024	-.203	.660	1.000	.308	.157	.017
	IC low (basic)	-.040	.013	.293	-.141	.295	.192	.308	1.000	.222	.380
	Unexpectedness mean high lag 3 sec.	.089	.059	.157	.154	.042	.181	.157	.222	1.000	.451
	Unexpectedness mean low lag 3 sec.	.139	.087	.155	.023	-.084	.098	.017	.380	.451	1.000

Korrelationsmatrix

		Tempo	RMS	Brightness	Roughness	Spectral Centroid	Melodic Contour	IC high (basic)	IC low (basic)	Unexpectedness mean high lag 3 sec.	Unexpectedness mean low lag 3 sec.
Korrelation	Tempo	1.000	.202	.018	.089	-.166	.189	-.148	-.176	-.136	.067
	RMS	.202	1.000	.202	.785	-.542	.627	.537	.496	.022	.006
	Brightness	.018	.202	1.000	.353	.480	-.261	-.156	.191	.192	.130
	Roughness	.089	.785	.353	1.000	-.161	.370	.328	.363	-.119	-.074
	Spectral Centroid	-.166	-.542	.480	-.161	1.000	-.785	-.619	-.192	.101	.004
	Melodic Contour	.189	.627	-.261	.370	-.785	1.000	.641	.384	-.089	-.042
	IC high (basic)	-.148	.537	-.156	.328	-.619	.641	1.000	.376	-.069	-.122
	IC low (basic)	-.176	.496	.191	.363	-.192	.384	.376	1.000	.074	.036
	Unexpectedness mean high lag 3 sec.	-.136	.022	.192	-.119	.101	-.089	-.069	.074	1.000	.383
	Unexpectedness mean low lag 3 sec.	.067	.006	.130	-.074	.004	-.042	-.122	.036	.383	1.000

Korrelationsmatrix

		Tempo	RMS	Brightness	Roughness	Spectral Centroid	Melodic Contour	IC high (basic)	IC low (basic)	Unexpectedness mean high lag 3 sec.	Unexpectedness mean low lag 3 sec.
Korrelation	Tempo	1.000	.554	.094	.395	-.352	.233	-.057	.014	.046	.299
	RMS	.554	1.000	.422	.740	-.566	.517	.338	.475	.322	.427
	Brightness	.094	.422	1.000	.651	.204	.202	.421	.489	.208	.255
	Roughness	.395	.740	.651	1.000	-.092	.450	.277	.533	.317	.540
	Spectral Centroid	-.352	-.566	.204	-.092	1.000	-.496	-.198	-.004	-.298	-.112
	Melodic Contour	.233	.517	.202	.450	-.496	1.000	.407	.249	.284	.228
	IC high (basic)	-.057	.338	.421	.277	-.198	.407	1.000	.421	.189	.162
	IC low (basic)	.014	.475	.489	.533	-.004	.249	.421	1.000	.318	.503
	Unexpectedness mean high lag 3 sec.	.046	.322	.208	.317	-.298	.284	.189	.318	1.000	.494
	Unexpectedness mean low lag 3 sec.	.299	.427	.255	.540	-.112	.228	.162	.503	.494	1.000

Figure D.2. PCA Correlation tables S4-S6.

Rotierte Komponentenmatrix[a]

	Komponente						
	1	2	3	4	5	6	7
Roughness	.942	.087	-.119	.067	.143	.189	.049
Brightness	.928	.039	-.011	.033	.014	.061	.292
RMS	.864	.047	-.423	.086	.019	.194	.025
Tempo	-.005	.969	-.093	-.072	.063	-.089	.068
Melodic Contour	.163	.840	-.159	.038	-.079	.349	.245
Spectral Centroid	-.280	-.207	.916	.040	.161	.024	.084
Unexpectedness mean high lag 3 sec.	.087	-.051	.026	.954	.180	.181	.115
Unexpectedness mean low lag 3 sec.	.107	.009	.139	.182	.952	.112	.118
IC high (basic)	.356	.125	.022	.278	.172	.828	.215
IC low (basic)	.282	.296	.102	.160	.167	.209	.847

Extraktionsmethode: Analyse der Hauptkomponente.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Rotierte Komponentenmatrix[a]

	Komponente						
	1	2	3	4	5	6	7
Roughness	.907	.010	-.128	-.017	.029	.098	.187
RMS	.828	-.320	.011	-.011	.092	.103	.229
Spectral Centroid	-.215	.931	.076	.020	-.019	-.114	.000
Brightness	.505	.535	.415	.334	-.105	.079	-.195
IC high (basic)	-.098	.103	.950	.150	.069	.087	.055
IC low (basic)	-.003	.054	.162	.931	.221	.125	.011
Unxpectedness mean low lag 3 sec.	.077	-.046	.053	.213	.934	.163	.134
Melodic Contour	.101	-.054	.192	.233	.289	.852	.118
Tempo	.387	-.298	-.283	-.357	-.309	.550	-.178
Unxpectedness mean high lag 3 sec.	.308	-.032	.048	.015	.140	.056	.922

Extraktionsmethode: Analyse der Hauptkomponente.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 9 Iterationen.

Rotierte Komponentenmatrix[a]

	Komponente							
	1	2	3	4	5	6	7	8
Roughness	.890	.009	.190	.118	.135	.045	.135	-.071
RMS	.886	-.159	.043	.077	.158	-.024	.133	.199
Brightness	.746	.435	.050	.252	.033	.075	.102	.302
Spectral Centroid	-.002	.963	-.021	.077	-.032	.040	.002	-.183
IC low (basic)	.170	-.014	.928	.029	.145	-.050	.164	.214
Unexpectedness mean high lag 3 sec.	.220	.099	.029	.942	.162	.000	.159	.024
IC high (basic)	.220	-.031	.153	.173	.920	.013	.085	.197
Tempo	.039	.043	-.042	.001	.009	.996	.011	-.037
Unexpectedness mean low lag 3 sec.	.223	.011	.157	.158	.078	.013	.945	.019
Melodic Contour	.214	-.266	.310	.026	.268	-.064	.018	.815

Extraktionsmethode: Analyse der Hauptkomponente.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Figure D.3. PCA rotated component matrices S1-S3.

Rotierte Komponentenmatrix[a]

	Komponente								
	1	2	3	4	5	6	7	8	9
Spectral Centroid	-.909	-.147	-.125	.168	-.085	.198	.034	-.080	-.167
RMS	.637	.540	.302	.220	.218	.097	-.024	-.013	.190
Roughness	.172	.959	-.040	.096	-.057	-.086	.089	.004	.075
Tempo	.162	-.003	.981	.003	.020	-.023	.041	.062	.033
Brightness	-.076	.127	.007	.965	.131	.124	.057	.066	.051
IC high (basic)	.143	-.035	.027	.158	.903	.155	.074	-.033	.325
IC low (basic)	-.137	-.069	-.023	.134	.143	.937	.087	.198	.086
Unexpectedness mean high lag 3 sec.	-.036	.081	.042	.056	.065	.080	.961	.221	.066
Unexpectedness mean low lag 3 sec.	.067	.001	.068	.067	-.029	.191	.232	.945	.022
Melodic Contour	.232	.130	.046	.059	.347	.094	.081	.030	.887

Extraktionsmethode: Analyse der Hauptkomponente.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Rotierte Komponentenmatrix[a]

	Komponente							
	1	2	3	4	5	6	7	8
Melodic Contour	.889	.228	.200	.095	-.041	-.032	-.071	.150
Spectral Centroid	-.858	-.127	-.020	-.086	.027	-.037	.324	-.227
Roughness	.108	.949	.121	.011	-.094	-.040	.169	.069
RMS	.467	.762	.243	.147	.080	.008	.114	.189
IC low (basic)	.176	.223	.939	-.114	.039	.023	.094	.109
Tempo	.117	.072	-.098	.978	-.070	.040	.020	-.082
Unexpectedness mean high lag 3 sec.	-.039	-.042	.036	-.071	.971	.196	.086	-.017
Unexpectedness mean low lag 3 sec.	-.001	-.029	.020	.041	.192	.977	.053	-.044
Brightness	-.274	.248	.103	.028	.103	.068	.911	-.034
IC high (basic)	.484	.209	.158	-.141	-.028	-.072	-.052	.816

Extraktionsmethode: Analyse der Hauptkomponente.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Rotierte Komponentenmatrix[a]

	Komponente							
	1	2	3	4	5	6	7	8
Brightness	.871	-.187	-.016	.045	.292	.012	.134	.177
Roughness	.791	.152	.368	.183	-.009	.260	.052	.207
Spectral Centroid	.156	-.896	.002	-.175	-.109	-.241	-.177	.042
RMS	.535	.641	.193	.330	.090	.181	.059	.261
Unexpectedness mean low lag 3 sec.	.167	.042	.896	.161	.053	.058	.263	.220
Tempo	.145	.222	.142	.947	-.069	.080	-.023	-.031
IC high (basic)	.195	.125	.048	-.068	.931	.174	.046	.164
Melodic Contour	.164	.275	.066	.090	.189	.910	.109	.071
Unexpectedness mean high lag 3 sec.	.125	.164	.229	-.023	.048	.101	.934	.104
IC low (basic)	.301	.029	.238	-.035	.196	.073	.122	.889

Extraktionsmethode: Analyse der Hauptkomponente.
Rotationsmethode: Varimax mit Kaiser-Normalisierung.[a]

a. Rotation konvergierte in 7 Iterationen.

Figure D.4. PCA rotated component matrices S4-S6.

Appendix E Code

IBM SPSS® Modelling Syntax

```
*Example model syntax for dV arousal lag 0

set printback=OFF.
FILE HANDLE indir ...
    /NAME='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\ARH1_S1256\sav\'.
*FILE HANDLE outdir /NAME='D:\***\Output'.

FILE HANDLE input /NAME='\indir\01_ArVaExp_d_z.sav' .
GET FILE='input' .
OUTPUT CLOSE *.

INCLUDE FILE='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\PCA1.sps'.

*SAVE OUTFILE='outdir\sav\01_ARO_L0_AR1_131127.sav'
*/COMPRESSED.

MIXED arousal_d_z WITH FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) ...
      SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1 | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /REPEATED = time_ArVA | SUBJECT(Subj_ID_ArVa) COVTYPE(ARH1)
  /SAVE=PRED (PRED_2b) RESID (RESID_2b) FIXPRED (FIXPRED_2b).

*Step 2
*Vorraussetzungen und Vorhersageleistung testen:
*y= alle fixed effects + alle random effects

*PPLOT, SCATTERPLOT, LINE, ACF.
PPLOT
  /VARIABLES=RESID_2b
  /NOLOG
  /NOSTANDARDIZE
  /TYPE=Q-Q
  /FRACTION=BLOM
  /TIES=MEAN
  /DIST=NORMAL.

GRAPH
  /SCATTERPLOT(BIVAR)=FIXPRED_2b WITH RESID_2b
```

```
  /MISSING=LISTWISE
  /TEMPLATE='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\APA_Diag_Gitter_30.sgt'.

GRAPH
  /LINE(MULTIPLE) = MEAN(arousal_d_z) MEAN(PRED_2b)  BY time_ArVa
  /MISSING=LISTWISE
  /TEMPLATE='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\APA_Diag_Gitter_30.sgt'.
ACF VARIABLES=RESID_2b
  /NOLOG
  /MXAUTO 16
  /SERROR=IND
  /PACF.

*R^2.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS R
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT arousal_d_z
  /METHOD=ENTER FIXPRED_2b.

*Exclude cases.
COMPUTE filter=ABS(arousal_d_z)  ≤ 2.
EXECUTE.
USE ALL.
FILTER BY filter.
EXECUTE.

*2c.
MIXED arousal_d_z WITH FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1
  /CRITERIA=CIN(95) MXITER(200) MXSTEP(50) SCORING(1) ...
      SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1 | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /REPEATED = time_ArVA | SUBJECT(Subj_ID_ArVa) COVTYPE(ARH1)
  /SAVE=PRED (PRED_2c) RESID (RESID_2c) FIXPRED (FIXPRED_2c).

OUTPUT EXPORT
  /CONTENTS  EXPORT=VISIBLE  LAYERS=PRINTSETTING  ...
      MODELVIEWS=PRINTSETTING
  /XLSX  DOCUMENTFILE=
    'D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\output_xlsx\ARH1_S1256.xlsx'
     OPERATION=MODIFYSHEET  SHEET='Tabelle1'
     LOCATION=LASTCOLUMN  NOTESCAPTIONS=YES.
```

```
*PPLOT, SCATTERPLOT, LINE, ACF.
PPLOT
  /VARIABLES=RESID_2c
  /NOLOG
  /NOSTANDARDIZE
  /TYPE=Q-Q
  /FRACTION=BLOM
  /TIES=MEAN
  /DIST=NORMAL.

GRAPH
  /SCATTERPLOT(BIVAR)=FIXPRED_2c WITH RESID_2c
  /MISSING=LISTWISE
  /TEMPLATE='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\APA_Diag_Gitter_30.sgt'.

GRAPH
  /LINE(MULTIPLE) = MEAN(arousal_d_z) MEAN(PRED_2c)  BY time_ArVa
  /MISSING=LISTWISE
  /TEMPLATE='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\APA_Diag_Gitter_30.sgt'.
ACF VARIABLES=RESID_2c
  /NOLOG
  /MXAUTO 16
  /SERROR=IND
  /PACF.

*R^2.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS R
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT arousal_d_z
  /METHOD=ENTER FIXPRED_2c.

OUTPUT SAVE NAME=*
 OUTFILE='D:\Users\cgrafe\FINAL_MODELS_mitIC_ARH1\output_spv\01_ARO_L0.spv'
 LOCK=NO.
```

IBM SPSS® Assembling Syntax Excerpt

```
/* Christoph Graefe, Matr.-Nr. 215891
/* Master Thesis - August 2014

/* SUBJECTIVE MEASURES Arousal

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\1arousal_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  time_ArVa F1.0
  Subj_ID_ArVa F1.0
  arousal1_d_z COMMA6.5.
CACHE.
EXECUTE.
VARIABLE LABELS
time_ArVa 'Time [sec]' arousal1_d_z 'Arousal' Subj_ID_ArVa 'Subject ...
    ID Ar/Va'.
EXECUTE .
SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

[...]

/* SUBJECTIVE MEASURES Valence

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\1valence_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  valence1_d_z COMMA6.5.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'.
```

```
EXECUTE.
VARIABLE LABELS
    valence1_d_z 'Valence'.
EXECUTE.
SAVE ...
   OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

[...]

/* SUBJECTIVE MEASURES Unexpectedness

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\1exp_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  time_Exp F1.0
  Subj_ID_Exp F1.0
  exp1_d_z COMMA6.5.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\01_ArVaExp_d_z.sav'.
EXECUTE.
VARIABLE LABELS
    time_Exp 'Time [sec]' Subj_ID_Exp 'Subject ID Exp' exp1_d_z ...
        'Unexpectedness'.
EXECUTE.
SAVE ...
   OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

[...]

/* PHYSIOLOGICAL DATA skin conductance

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\1gsr_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
```

```
  /IMPORTCASE=ALL
  /VARIABLES=
  time_PhyMeas F1.0
  Subj_ID_PhyMeas F1.0
  gsr1_d_z COMMA6.5.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'.
EXECUTE.
VARIABLE LABELS
 time_PhyMeas 'Time [sec]' gsr1_d_z 'Skin Conductance' ...
     Subj_ID_PhyMeas 'Subject ID PhyMeas'.
EXECUTE.
SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

[...]

/* AUDIO FEATURES

[...]

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\1rghns_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  rghns_d_z COMMA6.3.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'.
EXECUTE.
VARIABLE LABELS
    rghns_d_z 'Roughness'.
EXECUTE .
SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

[...]
```

```
GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\1tmp_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  tmp_d_z COMMA6.3.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'.
EXECUTE.
VARIABLE LABELS
    tmp_d_z 'Tempo'.
EXECUTE .
SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\1icNEW_d_z.csv"
  /DELCASE=LINE
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  icNEW_d_z COMMA6.3.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'.
EXECUTE.
VARIABLE LABELS
    icNEW_d_z 'IC - d_z'.
EXECUTE .
SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

GET DATA
  /TYPE=TXT
  /FILE="d:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\1ic_z.csv"
  /DELCASE=LINE
```

```
  /DELIMITERS=","
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  ic_z COMMA6.3.
CACHE.
EXECUTE.
MATCH FILES /FILE=*
/FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'.
EXECUTE.
VARIABLE LABELS
    ic_z 'IC - z'.
EXECUTE .
SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'
/COMPRESSED.

INCLUDE ...
    FILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\07_ListenerData.sps'.

SAVE ...
    OUTFILE='D:\AUDIOKOMMUNIKATION\Masterarbeit\data\[...]\01_ArVaExp_d_z.sav'
  /COMPRESSED.
```

Appendix F Further Material

Study	Musical Style, notes	Features																	
Coutinho/Cangelosi (2011)	Classical	**Loudness**	**pitch level**	pitch contour	**tempo**	texture	**sharpness**												
		2 positive relation-ship between loudness, tempo, timbre, and arousal	2 pitch level relates positively with both arousal and valence	0	2 positive relation-ship between loudness, tempo, timbre, and arousal Tempo is another feature which relates to the valence ratings as well as to those of arousal	0	2 positive relation-ship between loudness, tempo, timbre, and arousal												
Coutinho/Cangelosi (2009)	Classical	**Dynamics (loudness)**		**Pitch variation (STFT Flux)**	**Tempo (bpm)**	**Texture (multiplicity)**	**Timbre (sharpness)**	**Mean Pitch (spectrum centroid)**											
	consistent relationships between music features and the emotional response, which support important empirical findings	2 Higher loudness relates with positive Arousal		2 The average spectral variations relate negatively with Valence and positively with Arousal, indicating that large pitch changes are accompanied by increased intensity and decreased hedonic value	2 Fast tempi are related to high Arousal (quadrants 1 and 2) and positive Valence (quadrants 1 and 4). Slow tempi exhibit the opposite pattern	2 Thicker textures have positive relationships with Valence and Arousal (quadrants 1, 2, and 4)	2 Sharpness showed positive asso-ciations with Arousal and Valence (especially the first)	2 The highest pitch passages relate with high Arousal and Valence (quad-rants 1, 2, and 4)											
Guhn, Hamm Zentner (2007)	Classical, 6 excerpts **>>> CHILLS**	**loudness**			**tempo**		**frequency range**								**harmonic progression**				
		2 sudden or gradual volume increase from piano (i.e., soft) to forte (i.e., loud)		**alternation/contrast of instruments (solo<>orchester)** > mehrere features auf einmal (anderung in loudness, frequency range etc.)	2 all slow (adagio-larghetto)		2 all chill passages are characterized by an expansion in its frequency range in the high or low reg-ister								2 all chill passages were characterized by harmon-ically peculiar progressions that potentially elicited a certain ambiguity in the listener [...] in all three chill passages, the music went through a harmonic progression that briefly deviated from a pattern that could have been expected based on the previous section [...] the chill passages are similar in that they are characterized by a specific interplay of the melodic and harmonic progression				
Gomez, Danuser (2007)	Manowar-Mahler	**loudness**			**tempo**						**mode (major/minor)**		**accentuation**	**rhythmic articulation**	**harmonic complexity**				
	> wide range of styles!	2 major predictor of arousal			2 most importantly to distinguish between low-arousal and high-arousal excerpts						2 most important contribution to distinguishing between negative and positive excerpts		2 most importantly to distinguish between low-arousal and high-arousal excerpts	2 most important contribution to distinguishing between negative and positive excerpts AND between low-arousal and high-arousal excerpts	2 most important contribution to distinguishing between negative and positive excerpts				

Schubert (1999) Diss	Western Classical Music	**Loudness**		**melodic contour**	**tempo**	**texture**	**timbre/centroid**			duration		rhythm	envelope	articulation	harmony	voicing	phrasing	form	style
		2 unambiguously positively correlated to arousal also effected valence, though not to the same degree		2 positively related to valence	2 Changing tempo had a positive relationship with both valence gradient and arousal Gradient	1 In Aranjuez, texture gradient is positively linked with the arousal dimension, while it is positively linked to the valence dimension in Morning texture is a variable that **may provide a rich area of future research** **Def.:** estimate of the number of tones simultaneously noticed in a sound	1 The findings suggest that centroid is not a useful variable in the investigation of music-emotion relationships > the frequency spectrum **centroid variable should not be abandoned** (more research needed) strong positive relationship between centroid for the arousal for Slavonic Dance. *WIDERSPRUCH* no significant effect on valence gradient for any of the four test pieces			1		1	1	1	1	1	1	1	1
Dean, Bailes, Schubert, Greenlee (2011)	Classical (Dvorak)	Loudness/Acoustical Intensity																	
		2 "...intensity is a powerful influence on perceived arousal in this case." "…intensity powerfully influences perceived arousal in a wide range of musical contexts." Perceptual loudness clearly mirrors acoustic intensity very closely... it is intensity rather than perceived loudness that seems to more Directly influence perceived arousal																	
Dean, Bailes, Schubert (2010)	Wishart – Red Bird (Electroacoustic)	**Intensity [SPL]**					**Spectral Flatness measured as Wiener Entropy**												
		2 "Intensity influences perceptions of change and expressed arousal substantially." "...strong positive correlations between temporal patterns of intensity and perceived change and arousal."					2 "Spectral flatness influences valence, while animate sounds influence the valence response and its variance."												
Schubert (2004)	Classical -Dvorak (Sl. Dance) -Grieg (Morning) -Rodrigo (Aranjuez) -Strauss (Pizzicato)	**Loudness**		Melodic Contour	**tempo**	texture	Centroid (related to perceived timbral sharpness												

	"more than 60% of variation in arousal response could be explained by musical features such as loudness and tempo"	2 Associated positively with changes in arousal, clearly dominant		0 varied posi-tively with valence, though this finding was not conclusive	2 Associated positively with changes in arousal	0 do not appear to have a clear relationship to perceived emotion did not produce consistent predictions	0 do not appear to have a clear relationship to perceived emotion did not produce consistent predictions												
Korhonen, Clausi, Jernigan (2005)	6 Classical **<<< Nach Schubert! > features unbewertet**	0 Loudness Level Short Term Max. Loudness **(Loudness)**	0 Power Spectrum Centroid Mean STFT Centroid **(Mean Pitch)**	0 Mean STFT Flux Std. Dev. STFT Flux Std. Dev. STFT Centroid **(Pitch Variation)**	0 Beats per Minute **(Tempo)**	0 Multiplicity **(Texture)**	0 Timbral Width Mean STFT Rolloff Std. Dev. STFT Rolloff Sharpness (Zwicker and Fastl) **(Timbre)**								Spectral Dissonance (Hutchinson and Knopoff) Spectral Dissonance (Sethares) Tonal Dissonance (Hutchinson and Knopoff) Tonal Dissonance (Sethares Complex Tonalness **(Harmony)**				
Nagel, Kopiez, Altenmüller (2008)		**Loudness**				**tone-to-noise ratio**	**Roughness**												
	Allgemein: Psychoacoustical features seem to play an important role in the emotional experiences of chills	2 Increasing and decreasing loudness between 8 and 18 Bark appears to influence the probability that chills will be experienced by participants				1	1												
Watson, Mandryk (2012)	Model of musical mood with in-situ data sets, **Features unbewertet, nur im Gesamtmodell erfolgreich**				**1** **Rhythm** tempo (bpm) pulse clarity		**1** **Timbre** brightness rolloff spectral flux spectral centroid MFCC low energy roughness				**1** **Mode** most probable key estimation major/minor			**1** **Articulation** attack slope (indicator of how agressive a note is played) Average Silence Ratio (ASR)					
Li (2003)	Ambient, Classical,Fusion, Jazz					texture	timbre					rhythmic							
						0	0					0							
Levitin (2006)	-	Loudness	pitch				timbre	spatial location	reverberant environment	tone durations	onset times								
Dubnov, McAdams, Reynolds (2006)	The Angel of Death by Roger Reynolds for piano, chamber orchestra, and computer-processed sound, in a live-concert setting						**cepstrum**												
							interested in a gross spectral envelope description of the sound, which captures sound properties that might be described as overall sound color or texture, without considering the more detailed features due to effects such as pitch, or notes and timbres of specific instruments												

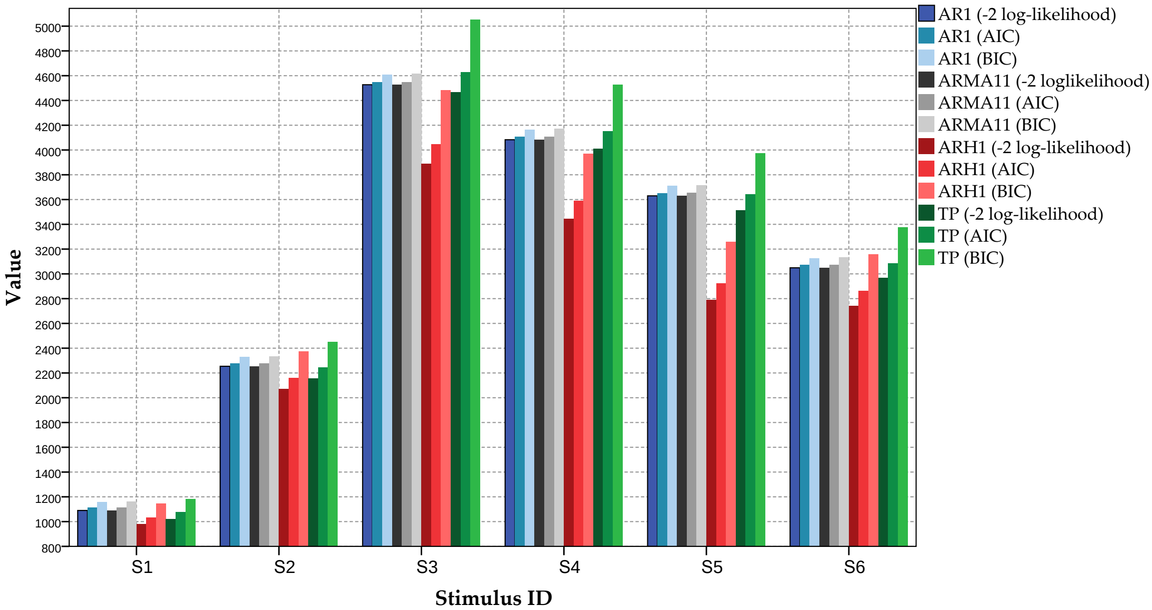


Figure F.1. Comparison of possible Covariance Structures for the R-Matrix.

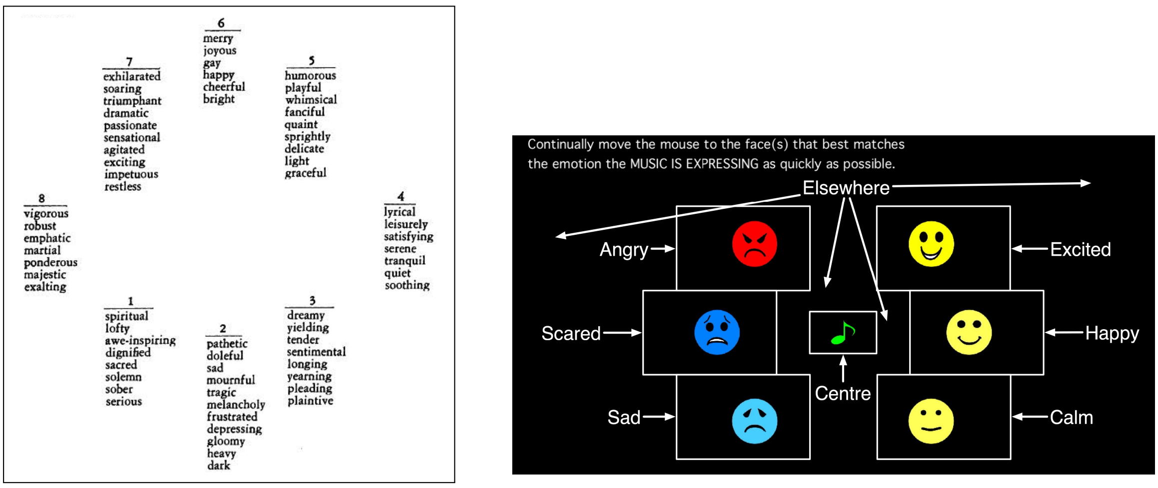
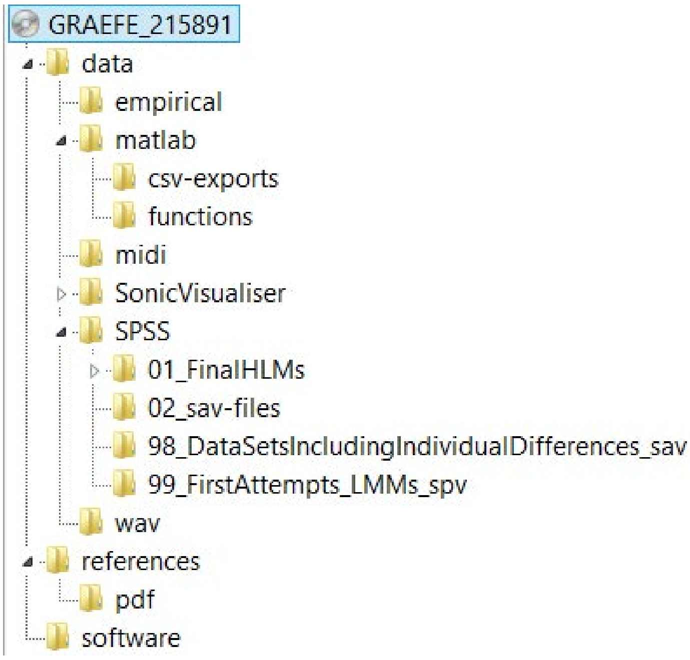


Figure F.2. Adjectives arranged in eight groups (Hevner, 1936, p. 249) (left), Structure of six emotion-face-clock graphic user interface (white boxes, arrows and labels not visible to participants) (Schubert et al., 2012, p. 5) (right).

Appendix G Electronic Documentation

The attached DVD-ROM contains all scripts, empirical data, audio files, scores and references used in this work stored in the following folder structure:



Folder structure of the attached DVD-ROM.

Notes to the single folders:

- **empirical** includes all forms, questionnaires and data collected during the live concert (from supervisor)
- **matlab** `main_05hz.m` is the main Matlab® script executing all feature extractions, analyses and plots (*MIRtoolbox 1.5* and *MIDI toolbox* from **toolboxes** folder must be installed)
- **SonicVisualiser** contains all projects for manually aligned tempo and manually aligned midi data
- **SPSS** includes the complete model syntax (`.sps`-files), all data tables (`.sav`-files) and outputs (`.spv`- and `.xlsx`-files)
- **wav** contains recordings of all six stimuli played back and performed during the live concert and for the analyses
- **references** contains all references in `.pdf`-format, a `.bib`-file listing all cited publications and an additional .cit-file (entries with red flags = `.pdf`-file included).