POLYNOMIAL EXTRAPOLATION FOR PREDICTION OF SURPRISE BASED ON LOUDNESS - A PRELIMINARY STUDY

Hendrik Purwins[*+]**,Piotr Holonowicz**[*]**,Perfecto Herrera**[*]
[*]Music Technology Group,Universitat Pompeu Fabra, Barcelona
[+]Neural Information Processing Group, Berlin Institute of Technology
hendrik.purwins,piotr.holonowicz,perfecto.herrera@upf.edu

ABSTRACT

The phenomenon of music surprise can be evoked by various musical features, such as intensity, melody, harmony, and rhythm. In this preliminary study we concentrate on the aspect of intensity. We formulate surprise as a critical derivation from the predicted next intensity value, based on the "immediate" past ($\sim$ 7 s), slightly longer than the short-term memory. Higher level cognition, processing the long range structure of the piece and general stylistic knowledge, is not considered by the model. The model consists of a intensity calculation step and a prediction function. As a preprocessing method we compare instantaneous energy (root mean square), loudness, and relative specific loudness. This processing stage is followed by a prediction function for which the following alternative implementations are compared with each other: 1) discrete temporal difference of intensity functions, 2) FIR filter, and 3) polynomial extrapolation. In addition, we experimented with different analysis window length, sampling rate and hop size of the intensity curve. Good results are obtained for loudness and polynomial extrapolation based on an analysis frame of 7 s, a sampling rate of the loudness measures of 1.2 s, and a hop size of 0.6 s. In the polynomial extrapolation a polynomial of degree 2 is fitted to the loudness curve in the analysis window. The absolute difference between the extrapolated next loudness value and the actual value is then calculated and divided by the standard deviation within the analysis window. If the result is above a threshold value we predict surprise. The method is preliminarily evaluated with a few classical music examples.

SMC 2009, July 23-25, Porto, Portugal

1 INTENSITY CALCULATION AND SURPRISE MEASURES

We compare three representations for the intensity: 1) the instantaneous energy, i.e. the root mean square of the amplitude, 2) the specific loudness in 24 bark bands (Zwicker and Fastl [1999] p.225) and 3) loudness as the sum of the specific loudness across the bark bands. We use the implementation within the IRCAM descriptor (Peeters [2004]).

After experimenting with different frame and hop sizes we chose a hop size of 0.6 s and a frame size of 1.2 s (rectangular window), calculating the intensity or (specific) loudness respectively.

The intensity is fed into a surprise prediction function. The output of such a model yields a curve of "surprisingness". Each surprise curve is normalized with respect to its maximal value. Applying a threshold leads to the binary decision surprise point/ no surprise. We use a threshold of 0.95. We compare four different models. The first two methods (Δ energy and Δ loudness) are defined by taking the differences of consecutive samples from energy or loudness. The third method is an FIR filter across the sampled loudness within an analysis frame of 7 s. The forth method is based on polynomial regression of the sampled loudness. Across a time frame of 7 seconds a regression polynomial is calculated and used for extrapolating the subsequent loudness value. In addition, the root mean square approximation error σ is calculated across this window. The value of "surprisingness" is calculated by dividing the deviation of the extrapolated and the actual next loudness value by σ.

2 DATA SET AND EVALUATION METHOD

We use a small set of 8 excerpts of classical music by Haydn (*Symphony No. 94 mit dem Paukenschlag*) Beethoven (*Symphonies 5 & 6*), Strauss (*Tod und Verklärung*), and Rossini (*Guillaume Tell*). The surprise points have been manually annotated by a subject

Method	F-Measure
Δ Energy	0.71
Δ Loudness	0.75
Filter	0.25
Polynomial Regression	0.83

Table 1: Comparison of surprise measures.

who had 13 years of violin instruction. The subject listened to the excerpts several times to determine the point in time when he was maximally surprised, expressed by high attention that this moment called and/or a gooseflesh shortly after, due to the dynamics of the piece when listening to it for the first time.

We adopt the usual evaluation method in music information retrieval for onset detection to surprise by transferring it to a larger time scale. The predicted surprise points are compared to the annotated surprise points. Within a tolerance range of $\pm 0.6s$, coinciding annotated and predicted surprise points are considered as correct hits. Multiple hits within the tolerance window are considered as simple hits. Then precision, recall, and f-measure are calculated.

3 EVALUATION AND EXAMPLES

Table 1 shows that the polynomial regression method works best. Due to its psychoacoustical justification, the Δ loudness method works better than the Δ energy method. As a demonstrate we show the analysis of two sound examples in Figures 1 and 2.

4 CONCLUSION AND DISCUSSION

The use of the polynomial regression model allows us to distinguish a continuous *crescendo* from an abrupt *subito forte*. Due to the rather short analysis window of 7 s (which is slightly longer than what is considered the short term memory) only surprise effects are considered that reflect a direct reaction to the sound. Surprise that is due to longer range structure or due to knowledge of stylistic particularities cannot be considered by the model suggested here. The optimal threshold of the polynomial regression model could be learned on pieces of a particular style and evaluated on a hold-out set of pieces of the same style. Other loudness models can be compared with the ones used here. It would help to use more sophisticated prediction models, e.g. such that consider periodic regularities in the loudness. Possible candidates would be autocorrelation or wavelets. Another effect to be taken into account is that the repetition of a surprising passage is less surprising. Therefore

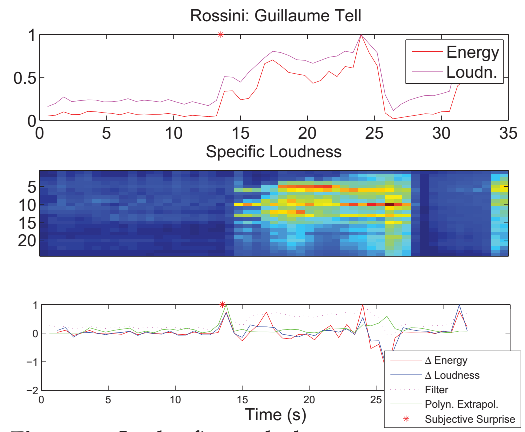

Figure 1: In the first subplot, root mean square energy and loudness is displayed. The specific loudness reveals the perceptually relevant energy distribution across the Bark bands. For surprise prediction, the loudness proofs more useful than the channel-wise specific loudness. The lowest subplot shows the various surprise measures. We observe a pronounced peak of the polynomial extrapolation method at the subjective surprise point with low values otherwise. On the contrary, the Δ methods yield high surprise values also at other points in time.

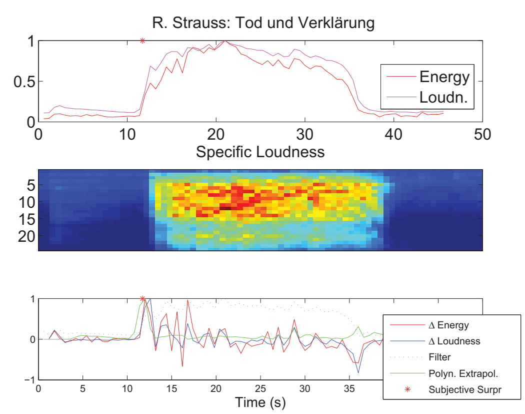

Figure 2: The analysis of a Richard Strauss excerpt (cf. Figure 1). Oscillating behavior of the Δ methods can be seen.

patterns of temporal development leading to a surprise point should be stored and time aligned to future surprise candidates. This could be performed by Dynamic Time Warping. For considering multiple context dependencies, Bayesian networks provide a useful methodology.

Acknowledgments

This research was supported by the European projects EmCAP (Emergent Cognition through Active Perception, FP6-IST, contract 013123) and CLOSED (Closing the Loop of Sound Evaluation and Design, FP6-NEST-PATH, project no. 29085). The first author received support from a Juan de la Cierva scholarship from the Spanish Ministry of Education and Science.

References

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, IRCAM, Analysis/Syntesis Team.

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics – Facts and Models*. Springer, 2nd edition.