An Improved Dissonance Measure Based on Auditory Memory

Jensen, Kristoffer; Hjortkjær, Jens

Published in:
Journal of the Audio Engineering Society

Publication date:
2012

Document Version
Også kaldet Forlagets PDF

An Improved Dissonance Measure Based on Auditory Memory

KRISTOFFER JENSEN, *AES Member,* **AND JENS HJORTKJÆR**
(krist@create.aau.dk) (jensh@hum.ku.dk)

University of Aalborg, Denmark and University of Copenhagen, Denmark

Dissonance is an important feature in music audio analysis. We present here a dissonance model that accounts for the temporal integration of dissonant events in auditory short term memory. We compare the memory-based dissonance extracted from musical audio sequences to the response of human listeners. In a number of tests, the memory model predicts listener's response better than traditional dissonance measures.

0 INTRODUCTION

Dissonance is a fundamental perceptual attribute of harmonic tones that have particular relevance to musical sounds. The concept of dissonance is used in a number of different aspects. It may refer to tone combinations within a musical system that are perceived as tense or unstable (tonal dissonance). Movement from dissonance to consonance gives rise to the perception of tension and release, which is an essential part of tonal music. Dissonance may also refer to the psychoacoustic basis of this (sensory dissonance), as the amount of beating or roughness produced by simultaneous partials within an auditory filter [1, 2, 3]. Behavioral listening studies confirm that the experience of musical tension is correlated with dissonance [4, 5].

High-level models use dissonance measures extracted from pitch intervals, e.g., in music theory [6, 7]. Dissonance may also be used as a perceptually motivated audio feature in, e.g., music information retrieval. In audio signals, dissonance may be estimated by summing the dissonance between each peak in the short-term spectrum of the continuous audio signal. This audio feature is a powerful descriptor that may be used in many music applications.

However, measures of sensory dissonance do not take temporal integration or memory into account. The perception of dissonance is, nonetheless, likely to be affected by the local temporal context as other psychophysical sound properties are. The temporal integration of loudness, for instance, is well-known [8, 9]. Dissonant sounds in music give rise to a physiological arousal response in the listener [10]. Huron suggests that dissonance is associated with a startle response lasting around 3 to 4 seconds with physiological markers such as increased heart rate [11]. This means that the dissonance of identical sound events in close temporal sequence will not have identical impact on the perceiver.

We present a dissonance model that takes the temporal integration of dissonance in short-term memory into account. This model can be used to extract a dissonance measure from continuous musical audio. We show that this measure predicts listener's experience of tension in music better than dissonance measures without memory. The memory model is tested for different kinds of music and is found superior in all tested cases.

1 AUDITORY MEMORY MODEL

Short-term memory (STM) or working memory refers to the process of retaining a limited number of perceived elements at a relatively short time scale. STM is distinct from long term remembering at both a functional [12] and neural [13] level. Limitations on information processing may define STM in terms of either *time* or *content* of the memory process. Some models assume that representations in STM decay as a function of time unless actively maintained [14]. The duration of STM are estimated in the range of 3 to 5 seconds [15, 16]. Others focus on the limited number of elements or chunks that can be retained. Miller proposed a capacity of 7 ± 2 elements [17] while Cowan argued for a capacity around 4 elements [18].

Anderson and Lebiere [19] suggest that the *activation strength* of STM decays exponentially as a function of time:

$$A_t = 1 - 0.5 \ln(t + 1). \quad (1)$$

An element is held in STM as long as the activation strength is positive. We now include a term A_N to account for the way in which stored elements are displaced by new ones because of the limitation on the number of elements in STM. We assume that the current number of elements N_c in STM influences the total activation strength in a similar

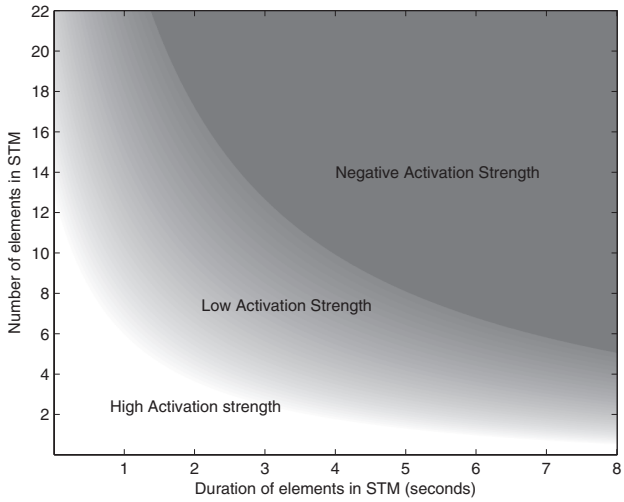

Fig. 1. Memory activation strength as a function of duration and number of elements (Eq. 2). An auditory component is erased from STM when the activation strength becomes negative.

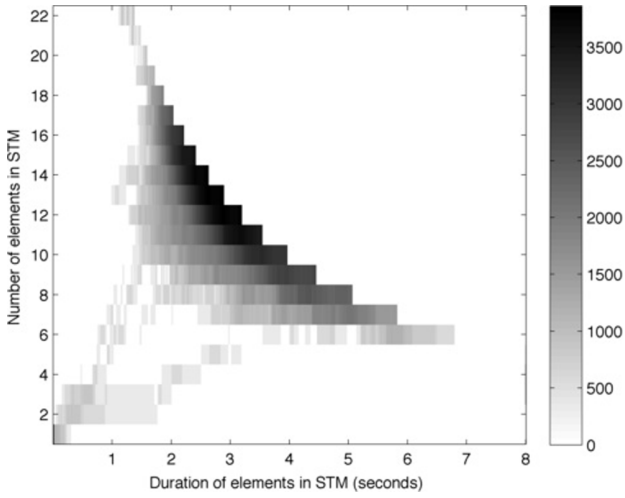

Fig. 2. Density of frames extracted from music audio in STM showing the typical area of activation.

decaying manner. The total activation strength becomes [20]:

$$A = A_t + A_N = 2 - 0.5\ln(t+1) - 0.5\ln(N_c+1). \quad (2)$$

In consequence, there is a trade-off between temporal decay and the number of elements in STM. This is illustrated in Fig. 1. If there is a high load on STM by a large number of elements then these can only be held in STM for a short period of time. Conversely, a small number of elements can be held in STM for a longer period of time. Fig. 2 shows the density of events for a given number of elements and duration in frames of musical audio as extracted by the model (explained below). This illustrates the typical area of content in STM in musical sequences.

Fig. 3 illustrates the activation strengths of a sequence of events at a constant inter-onset interval (IOI) of 500 ms. As can be seen, the number of elements in STM stabilizes around a given level. In musical sequences, the faster rate of events typically cause the STM model to settle around 12 to 18 events as can be seen in Fig. 2. This is more than is typically reported in the literature, but many elements have only weak activation strength and are likely not to be detected in experiments assessing STM function.

The general memory model (Eq. 2) can be used to account for the way in which auditory events are integrated over time, and we will suggest how to calculate dissonance within STM using the model. In the audio domain, we must first extract perceptually salient events from the musical signal to determine the temporal location of events that enter into STM. Here, we use a perceptually motivated measure of spectral flux to detect the onsets of musical events for the dissonance calculation. The *perceptual spectral flux* is calculated as the sum of rising magnitudes from an N-point FFT scaled by equal loudness contours:

$$psf = \sum_{a_k^t - a_k^{t-1} > 0} a_k^t - a_k^{t-1} \quad (3)$$

where a_k is the magnitude of the k^{th} frequency bin, normalized with an equal loudness contour weighting [21]. A peak detector identifies onsets as the instantaneous value of *psf* above 10% of the mean over 1.5 seconds plus 90% of the maximal *psf* over 0.9 seconds. This detection procedure captures many of the features used by humans to separate auditory streams [22].

The onset detection identifies events for the dissonance calculation. For each detected event n, the corresponding spectrum with frequencies f^n and amplitudes a^n is inserted into STM.

Following Plomp and Levelt [2] and Sethares [23], the dissonance between two frequency components (where $f_1 > f_2$) are estimated as:

$$d = a_1 a_2 (e^{\frac{f_1 - f_2}{0.0245 f_1 + 22.57}} - e^{\frac{f_1 - f_2}{0.015 f_1 + 13.74}}) \quad (4)$$

where maximal dissonance occurs at a separation of roughly 1/4 of the critical bandwidth. Frequency components further apart than one critical band are not perceived as dissonant.

In complex sounds, the total dissonance can be estimated by assuming additivity between combinations of all frequency components [2]:

$$d_{tot} = \sum_k \sum_{l>k+1} d_0(f_k, a_k, f_l, a_l) \quad (5)$$

We now argue that dissonance is also affected by the local context, that is, by the dissonance of other events in STM. It is assumed that interference between spectral components of all events currently in STM contributes to the perceived dissonance of a given event. We do this by adding the dissonance between all spectral components within STM scaled by their activation strength to the dissonance of the current event d_{tot}:

$$d_{stm} = d_{tot} + \sum_n A^n \sum_k \sum_l d(f_k, a_k, f_l^n, a_l^n) \quad (6)$$

where f_k and a_k are the k^{th} frequency and amplitude of the current frame, while f_l^n and a_l^n are the frequencies and amplitudes of event n in STM. If the spectral components of a given frame interferes with those of other events in

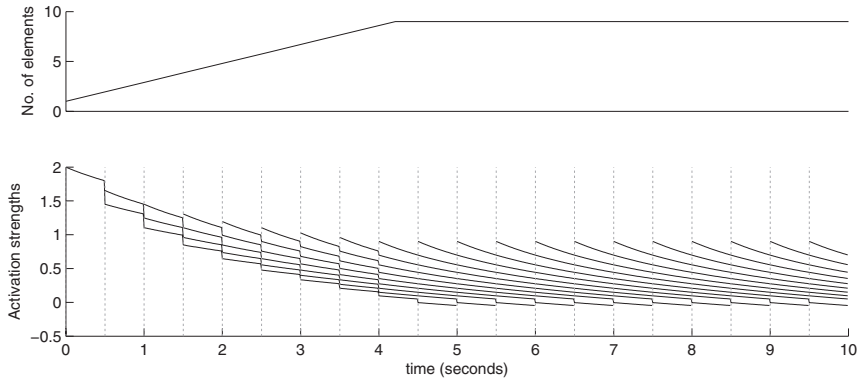

Fig. 3. Activation strengths for each element in an isochronous sequence of events (below) and the number of elements in STM (above).

STM, then this contributes to the dissonance of that frame. In the following, we indicate that this model does in fact predict listeners' response to music more accurately.

2 EXPERIMENT

It is well-known that dissonance in music correlates with listeners' experience of tension [4, 5]. Tension ratings given in music listening experiments may thus be used to assess the dissonance measure based on auditory memory described above. Previous studies have demonstrated a correlation between dissonance and tension in both longer [4] and shorter [24] musical sequences, and in tonal as well as atonal music [5]. The instruction to rate "tension" in music is frequently used to examine real-time music listening because the term seems to describe the way in which listeners attend to musical patterns in an essential way [25, 26]. Studies comparing alternative conceptualizations given in the experimental instructions have found ratings of "tension" to yield a more differentiated and reproducible response [27, 28]. It is thus preferable to examine ratings of "tension" rather than explicit ratings of "dissonance" or other sensory qualities because rating tension is a more meaningful task to music listeners. Dissonance is implicit in tension patterns in tonal music since musical harmony relies on tonal dissonance. It may, however, play a different role in atonal music and music based on timbre where sensory dissonance have a more structural impact [6]. For this reason, the musical stimuli in the current experiment comprised both classical pieces of tonal music as well as atonal music played on pitched instruments or abstract electronic compositions based on timbre.

In a behavioral experiment, music listeners ($N = 21$) were asked to rate the experienced tension by adjusting a continuous slider while listening to nine different musical stimuli [26]. Stimuli comprised both shorter excerpts as well as entire pieces. Two excerpts were presented in both a recorded version and in a MIDI version with equalized onset velocity and timing in order to examine the tension response while minimizing the effect of loudness (another predictor of musical tension).

The mean sampled tension response by human subjects was then compared to the dissonance measures described above. The tension ratings were compared with both the dissonance measure based on STM (Eq. 6) and the immediate dissonance (Eq. 5). The two time-series were first cross-correlated to find the time lag of subjects response, and the correlation was calculated at this lag. The ratings and the smoothed dissonance measure for the first musical excerpt are shown in Fig. 4. However, the correlations are calculated using the unsmoothed dissonances in order to avoid introducing artificial serial correlation into the data (producing artificially high correlations with the continuous tension response, [29]).

The results for each excerpt are listed in Table 1. As can be seen, the dissonance measure with temporal integration are better correlated with tension ratings than the immediate dissonance in all excerpts. A nonparametric Friedman's test

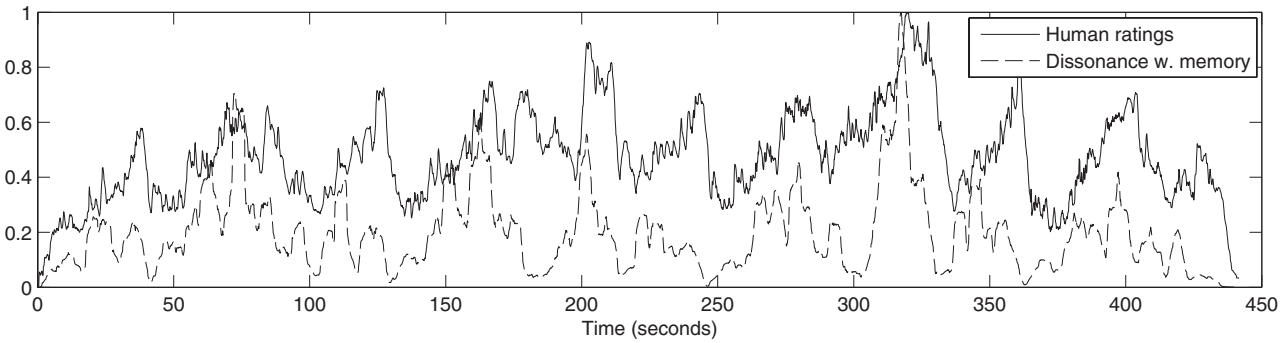

Fig. 4. Behavioral tension ratings and dissonance (w. memory) for the first musical excerpt.

Table 1. Results showing the correlations (ρ) between mean tension ratings and dissonance without and with the memory model for all 9 musical pieces

No.	Genre	Composer	Duration (sec.)	ρ w/o memory	ρ w memory	Difference
1	Classical tonal	Mozart	444	0.17	0.33	+94.1%
2	Classical tonal - MIDI	Mozart	355	0.17	0.25	+47.1%
3	Classical atonal	Webern	123	0.49	0.55	+12.5%
4	Classical atonal - MIDI	Webern	113	0.09	0.16	+77.8%
5	Electronic	Stockhausen	28	0.60	0.68	+13.3%
6	Electronic	Stockhausen	27	0.59	0.67	+13.6%
7	Electronic	Stockhausen	60	0.33	0.46	+39.4%
8	Electronic	Alva Noto	47	0.28	0.42	+50.0%
9	Classical tonal	Debussy	39	0.23	0.26	+13.0%

on the correlations found the memory model to perform significantly better ($p < .005$).

3 CONCLUSION

We have presented a dissonance measure based on auditory memory. New auditory events are identified and stored in a short term memory module. For each frame, the total dissonance is calculated as the sum of local dissonance and the interaction with elements in STM. We have shown that STM-based dissonance is significantly better correlated with human ratings than dissonance measures without memory integration.

It is noticeable that these relatively simply assumptions about human perception improves the traditional dissonance measures considerably. An audio measure of dissonance that accurately corresponds to listener's experience is valuable in many audio applications. Nonetheless, audio extracted features seldom include psychological assumptions or knowledge. There are a number of ways in which the memory model may improve audio features even more by including more psychological premises. In particular, chunking elements together may reduce the number of elements in STM, and the memory model may thus be coupled with a theory of chunking [30].

4 REFERENCES

[1] H. L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (New York: Dover Publications, 1885).

[2] R. Plomp and W. J. M. Levelt, "Tonal Consonance and Critical Bandwidth," *J. Acoust. Soc. Am.*, vol. 38, no. 4, pp. 548–560 (1965).

[3] E. Terhardt, "The Concept of Musical Consonance: A Link between Music and Psychoacoustics," *Music Perception*, vol. 1, no. 3, pp. 276–295 (1984).

[4] E. Bigand and R. Parncutt, "Perceiving Musical Tension in Long Chord Sequences," *Psychological Research*, vol. 62, no. 4, pp. 237–254 (1999).

[5] D. Pressnitzer, S. McAdams, S. Winsberg, and J. Fineberg, "Perception of Musical Tension in Nontonal Orchestral Timbres and its Relation to Psychoacoustic Roughness," *Perception & Psychophysics*, vol. 62, no. 1, pp. 66–80 (2000).

[6] F. Lerdahl, *Tonal Pitch Space* (New York: Oxford University Press, 2001).

[7] R. Parncutt, *Harmony* (Springer Series in Information Sciences, Berlin: Springer-Verlag, 1989).

[8] W. A. Munson, "The Growth of Auditory Sensation," *J. Acoust. Soc. Am.*, vol. 19, pp. 584–591 (1947).

[9] D. Algom and L. E. Marks, "Range and Regression, Loudness Scales, and Loudness Processing: Toward a Context-Bound Psychophysics," *J. Experimental Psychology: Human Perception and Performance*, vol. 16, no. 4, pp. 706–727 (1990).

[10] A. J. Blood, R. J. Zatorre, P. Bermudez, and A. C. Evans, "Emotional Responses to Pleasant and Unpleasant Music Correlate with Activity in Paralimbic Brain Regions," *Nature Neuroscience*, vol. 2, pp. 382–387 (1999 Apr.).

[11] D. Huron, *Sweet Anticipation* (Cambridge: MIT Press, 2006).

[12] A. Baddeley and G. Hitch, "Working Memory," in *The Psychology of Learning and Motivation: Advances in Research and Theory* (G. Bower ed.), vol. 8, pp. 47–89 (New York: Academic Press, 1974).

[13] E. E. Smith and J. Jonides, "Storage and Executive Processes in the Frontal Lobes," *Science*, vol. 283, pp. 1657–1661 (1999 Mar.).

[14] P. Barrouillet, S. Bernardin, and V. Camos, "Time Constraints and Resource Sharing in Adults' Working Memory Spans," *J. Experimental Psychology: General*, vol. 133, pp. 83–100 (2004 Mar.).

[15] G. Radvansky, *Human Memory* (Boston: Allyn and Bacon, 2005).

[16] B. Snyder, *Music and Memory. An Introduction* (Cambridge: MIT Press, 2000).

[17] G. Miller, "The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Rev.*, vol. 63, no. 2, pp. 81–97 (1956).

[18] N. Cowan, "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity," *Behavioral and Brain Sciences*, vol. 24, pp. 87–114 (2001 Feb.).

[19] J. Anderson and C. Lebiere, *Atomic Components of Thought* (Hillsdale: LEA, 1998).

[20] K. Jensen, "On the Use of Memory Models in Audio Features," in *Symposium of Frontiers of Research on Speech and Computer Music Modeling and Retrieval (FRSM/CMMR - 2011)* (Bhubaneswar, India), pp. 100–107 (2011 Mar.).

[21] ISO-226, "Acoustics—Normal Equal-Loudness-Level Contours," Tech. Rep., International Organization for Standardization, Geneva (2003).

[22] K. Jensen, "Multiple Scale Music Segmentation Using Rhythm, Timbre and Harmony," *EURASIP Journal on Applied Signal Processing*, vol. Special issue (2007).

[23] W. A. Sethares, "Local Consonance and the Relationship between Timbre and Scale," *J. Acoust. Soc. Am.*, vol. 94, no. 3, p. 1218–1228 (2003).

[24] E. Bigand, R. Parncutt, and F. Lerdahl, "Perception of Musical Tension in Short Chord Sequences: The Influence of Harmonic Function, Sensory Dissonance, Horizontal Motion, and Musical Training," *Perception & Psychophysics*, vol. 58, no. 1, pp. 125–141 (1996).

[25] F. V. Nielsen, *Oplevelsen af musikalsk spænding [The Experience of Musical tension]* (Copenhagen: Akademisk Forlag, 1983).

[26] J. Hjortkjær, "Toward a Cognitive Theory of Musical Tension," *Ph.D. thesis*, University of Copenhagen, 2011.

[27] J. A. Lychner, "An Empirical Study Concerning Terminology Relating to Aesthetic Response to Music," *J. Research in Music Education*, vol. 46, no. 2, pp. 303–319 (1998).

[28] W. E. Fredrickson, "A Comparison of Perceived Musical Tension and Aesthetic Response," *Psychology of Music*, vol. 23, no. 1, pp. 81–87 (1995).

[29] E. Schubert and W. Dunsmuir, "Regression Modelling Continuous Data in Music Psychology," in *Music, Mind, and Science* (S. W. Yi, ed.), pp. 298–352 (Seoul, Korea: Seoul National University, 1999).

[30] K. Jensen, "Retrieving Musical Chunks," in *Proceedings of the 2008 Computer in Music Modeling and Retrieval*, pp. 339–347, Re: New – Digital Arts Forum (2008).

THE AUTHORS

Kristoffer Jensen

Jens Hjortkjær

Kristoffer Jensen obtained his Masters degree in 1988 in computer science at the Technical University of Lund, Sweden, and a D.E.A in signal processing in 1989 at the ENSEEIHT, Toulouse, France. His Ph.D. was delivered and defended in 1999 at the Department of Computer Science, University of Copenhagen, Denmark, treating signal processing applied to music with a physical and perceptual point-of-view. This mainly involved classification and modeling of musical sounds. Dr. Jensen has been involved in synthesizers for children, state of the art next generation effect processors, and signal processing in music informatics. His current research topic is signal processing with musical applications and related fields, including perception, psychoacoustics, physical models, and expression of music. Dr. Jensen has chaired more than 8 major conferences, been the editor of many books and conference proceedings, and published more than 200 papers. He currently holds a position at the Institute of Architecture, Design and Media Technology, Aalborg University Esbjerg as associate professor.

●

Jens Hjortkjær is a research assistant at the Department of Arts and Cultural Studies at the University of Copenhagen, Denmark. He received his Masters degree in music and psychology in 2007 at the University of Copenhagen, Denmark and Paris Sorbonne University, France, with supplementary studies in medical engineering at the Technical University of Denmark (DTU). His doctoral dissertation was defended in 2011 and dealt with modeling the perception of musical tension based on behavioral experiments with continuous response interfaces. He is currently conducting research on the neural basis of auditory categorization at the Danish Research Centre for Magnetic Resonance (DRCMR), Denmark.