KRISTOFFER JENSEN

Sensory Dissonance Using Memory Model

Introduction

In its traditional form, music is a sequence of notes and other sounds. As such, it has rhythm, i.e. the notes have onset times at regular intervals, and it has harmony, meaning that the notes played have discrete pitch values. While notated notes exist in discrete time locations, performed notes may be played early or late, or with longer or shorter duration. This performance may accentuate the perceived expression of the music. In a similar manner, the note pitch may also be expressed with intonation, vibrato, glides, etc. that also accentuate the expression of the music.

The rhythm and in particular the harmony contribute in some music to the build-up of tension and release. This tension/release scheme is highly dependent on the consonance and dissonance of the music. Thus, the study of dissonance and consonance is important in music theory and related areas. In theory, the consonance and dissonance can be identified by the note intervals directly, although timbre and loudness, as well as intonation or vibrato may have a further influence of the perceived dissonance. Often, however, the dissonance is calculated based on the position of the notes in the music, and this can be determined for instance using the GTTM model.[1] However, such an approach is only possible if the music follows the underlying structure, and it may only be valid if the listener is sufficiently trained in the music style. In both these cases, a machine approach to the calculation of dissonance may prove useful; such a method would in an initial phase calculate the local dissonance only, which would incorporate all expressive influences, and be useful for atonal and other music that does not fit classical music theory models. In addition, such a machine measure would be closer related to non-informed listeners. Machine approaches are also useful when making comparative studies, as will be shown here.

While sensory dissonance has been known for a long time, it does not allow the dissonance measure of notes played in sequence instead of in parallel. This non-withstanding, most people would hear approximately the same consonance or dissonance in both situations. For this reason, the sensory dissonance has been improved with a model inspired by the working memory. In this model, the previous acoustic elements

---

[1] Fred Lerdahl and Ray Jackendoff, *A generative theory of tonal music* (Cambridge, MA: MIT Press, 1983).

(corresponding to the previous notes), are retained in the working memory, and fading. While it is not yet purged from the memory, each element is participating in the calculation of the sensory dissonance, which can thus be applied to calculate the dissonance of notes in sequence too. Further details of the model are presented below, together with how and where the working memory is undertaken in the human brain.

Sensory Dissonance

Sensory dissonance is the term for dissonance calculated using knowledge about the auditory system in relationship with the beats of tones. Of particular interest is the tonotopic (close frequencies are located at proximity) organisation in the auditory system and the related perceptual difference between closely related partial frequencies and dissonance. The key term in auditory research is critical band, and pure tones (for instance overtones in harmonic sounds) within one critical band participate in the total dissonance. Plomp and Levelt[2] investigated this using psycho-acoustic experiments and determined the standard curve of dissonance that indicates the maximum dissonance for two pure tones are found at approximately one quarter of the critical band, and they also demonstrated that dissonance is additive. The calculations can be made using a mathematical expression for the dissonance, and using the amplitudes multiplied together as the loudness weight for each two-tone pair. The total calculation of dissonance of two notes is thus done by identifying all partial tone pairs and summing the dissonance weighted with the amplitudes multiplied together. As a sidenote, it can be shown[3] that the musical scale commonly used is closely linked with the harmonic sound, and other scales are appropriate if non-harmonic sounds are used.

The sensory dissonance is thus dependent on the spectrum of the sounds. For the sake of demonstration, the discrete spectrum of three musical instruments sounds (Piano, Trumpet and Viola) and of one synthetic tone, with fundamental frequency approximately 260 Hz for all four tones, are shown in figure 1.

These spectrums have been used in order to calculate the sensory dissonance along one octave, and how this dissonance varies with the spectrum. This has been obtained by calculating the dissonance between one sound and the same sound transposed across one octave (figure 2).

It is clear that the spectrum influences greatly how the sensory dissonance evolves across one octave. The more spectrally rich sounds, such as the trumpet, render a higher dissonance across the octave, while the less spectrally rich piano sound has very low sensory dissonance, except at the proximity of the fundamental. An informal listening of the tone pairs confirms this finding. The trumpet is harsher in itself, but it also seems to give more dissonance because of the richness of the tone. A similar conclusion can be made with the synthetic tone. All of the tones have decreasing disso-

---

[2] Reinier Plomp R. and Willem. J. M. Levelt, "Tonal Consonance and Critical Bandwidth," *J. Acoust. Soc. Am.* 38(4) (1965): 548-560.

[3] Sound Examples for Tuning Timbre Spectrum Scale, http://sethares.engr.wisc.edu/html/soundexamples.html. Visited 26/3-2015.

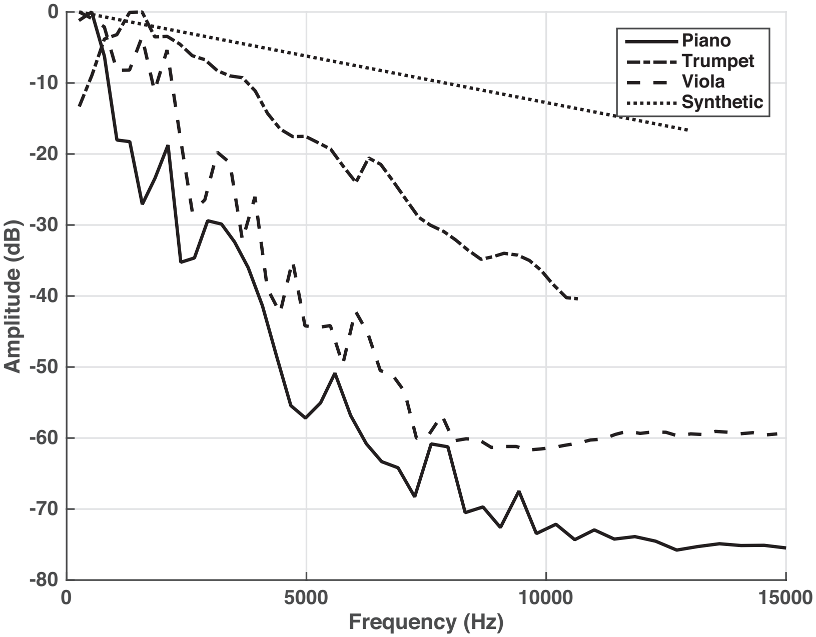

Figure 1. Spectrum of three acoustic instruments, Piano, Trumpet and Viola and of one synthetic tone

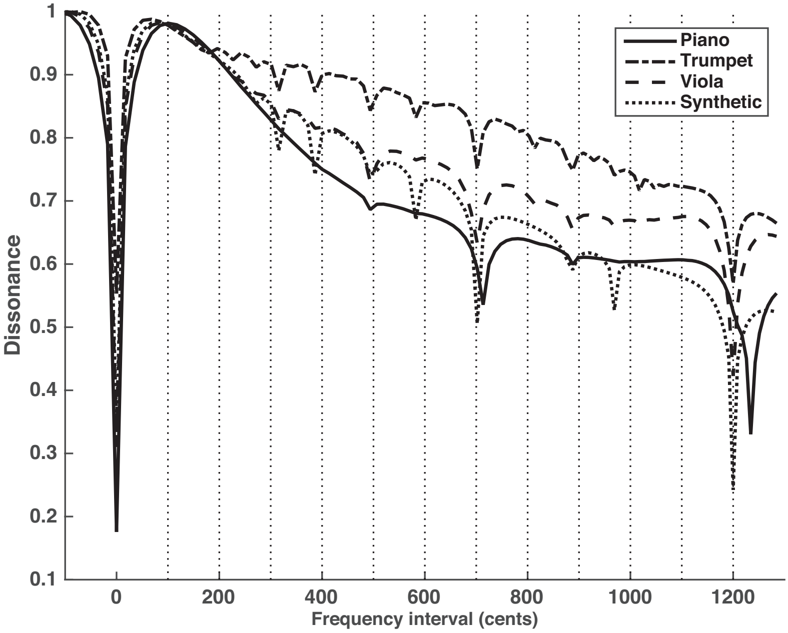

Figure 2. Dissonance curves for one octave for three instrument sounds and one synthetic sound. The equal temperament pitch values are shown with vertical lines.

nance along the note distance. This may be true in a non-musical setting, and it can be confirmed by playing the twelve intervals on a piano, for instance. Krumhansl and Shepard[4] tested how well one tone fitted in a major scale context, and they found that the fit (which should be related to consonance) were dependent on the individual assessment probably related to musical training. Some subjects made greater distinction between scale and non-scale notes, while others made very little such distinction. In addition, the test was performed by first playing an ascending or descending seven note major scale and then one probe tone for all subjects, though in particular the less musical subjects, the notes farther away from the last note of the context have much higher judgment of fit. The last notes of the context should have higher activation strength in the working memory, and thus contribute more to the total dissonance, and when the last tone is farther away in pitch, the fit is higher, and thus the dissonance is lower, which confirms the findings of figure 2 above.

The sensory dissonances across one octave for different sounds are shown with the equal temperament pitch values in figure 2. It is clear that some intervals fit well with the equal temperament, while others have minima in the dissonance (maxima in the consonance then) not on the equal temperament pitch values. Some of these consonance maxima fall on natural scale pitch values, while other maxima are related to the particular spectrum of the analysed sounds. For instance, all sounds have consonance maxima at the equal-temperament fifth, while the consonance maxima for the fourth is located lower than the equal-temperament fourth. As the maximum is rather pointed, even a slight misplacement of the maxima renders a much higher sensory dissonance.

Memory and Dissonance

Before recent time, most cognitive studies were concerned with either psycho-physical studies, that could show the duration of the short-term memory, or cognitive models, that would give the building blocks of part of the cognitive system. Today, recordings of brain activity, such as the Electroencephalography (EEG), which records the electric activity of the brain, may give more detailed information about the activity of the brain in certain conditions, and brain scan methods, such as the Functional magnetic resonance imaging (fMRI), makes it possible to see in more detail the location in the brain where the activity takes place. Because of different time resolutions, the EEG, in particular the time synchronous event-related potential (ERP) is better suited for fast activities (in the milliseconds) while fMRI are better suited for slower activities (seconds to minutes).

The EEG data are often analysed using the MisMatch Negativity (MMN). The MMN is obtained by subtracting the ERP of a standard stimulus from that of a deviant stimulus that occurs infrequently and randomly. The MMN is a reaction to the unexpectedness of the deviant stimulus, and it has been used mainly in connection with audi-

---

[4] Carol Krumhansl L, and Roger N. Shepard, "Quantification of the hierarchy of tonal functions within a diatonic context," *J Exp Psychol Hum Percept Perform* 5(4) (1979): 579-94.

tory stimuli. Alain, Woods and Knight[5] (1998) found evidence that the auditory sensory memory is located in the auditory cortex, and that the dorsolateral prefrontal cortices are also facilitating auditory sensory memory.

Using different approaches, in particular the study of humans with anatomical defects in the brain, significant discoveries about the brain have been in the last 150 years. Memory, in particular the working memory is modality dependent. As such, the auditory memory is located at the auditory centre in the superior temporal cortex[6]. The auditory information is then consolidated[7] in the hippocampus. This is also where stimuli that occur together but in different modalities are merged. After the system consolidation, the information trace is no longer bound to the main memory part, the hippocampus. How the information is subsequently distributed is not known in full details. Apparently, some highly connected regions, called rich club organization[8], exist in the human brain. The rich club consists of, in the cortical level, the precuneus (connected with working memory, involved with episodic memory), the superior frontal cortex (behavior and personality), the superior parietal cortex (spatial orientation, visual and hand input), and in the subcortical level, the hippocampus (consolidation of information), the putamen (motor control and learning) and thalamus (sensory switchboard, including auditory nucleus). One of the strongly interconnected regions is the hippocampus, and it is connected to the precuneus, involved with episodic memory, and the putamen, involved with learning. It is also interesting that the thalamus is part of the rich club. The thalamus contains the auditory nucleus, which is the auditory pathway before it reaches the auditory cortex. As such, it seems that the auditory information reaches large parts of the brain independently of the auditory cortex. It may be hypothesized that this enables the time synchronicity between modalities, but it is probably not strongly related to the working memory.

It is the rich club of the central nervous system (CNS) that enables the communication between the different parts of the brain in an optimized manner. The emplacement of the memory is dependent on the consolidation; if the memory is relatively new it is placed at the hippocampus vicinity in the limbic system, while if enough associations have been made, the memory is located in the cortical area. This enables the brain to consolidate new information in three stages, the first is attention-driven (the working memory), the second is placed in the limbic system, and it has the advantage that it does not alter the LTM significantly, while still enabling rapid learning, and possible readjustment, while the third is the LTM placed in the cortical parts of the brain, at or around Wernicke's area for audio.

---

[5] Claude Alain, David L. Woods, Robert T. Knight, "A distributed cortical network for auditory sensory memory in humans," *Brain Res* 812 (1998): 23–37.

[6] Marek-Marsel Mesulam M., "From sensation to cognition," *Brain* 121 (1998): 1013–1052.

[7] Yadin Dudai. "The neurobiology of consolidations, or, how stable is the engram?" *Annu. Rev. Psychol.* 55 (2004): 51-86.

[8] Martijn P. van den Heuvel and Olaf Sporn, "Rich-Club Organization of the Human Connectome," *The Journal of Neuroscience* 31(44) (2011):15775–15786.

Sensory dissonance using memory model

A two-step model has been designed to take into account the temporal context of tones when calculating the sensory dissonance. In the first step, new tones are identified by peaks in the perceptual spectral flux. The perceptual spectral flux is the spectral difference over time weighted by a model of the human frequency characteristics. Once a tone is identified, it is inserted into the memory store and it is retained in there as long as the activation strength for the tone is positive. The activation strength is exponentially decreasing for time and number of elements, and therefore, the tone is retained unless too many tones occur, or too much time has passed since it was inserted into the memory. Finally, for each step in time, the sensory dissonance is calculated for the current sound isolated and the dissonance between the current sound and all elements in the memory is added to this. All sensory dissonance involving tones in the memory are weighted with the activation strength of the tones. Jensen and Hjortkjær[9] give more details of the model and show that this model gives better sensory dissonance estimates when compared to human dissonance assessments.

Comparisons

In order to show the validity of this approach, the sensory dissonance measure using memory model is compared to the results of Krumhansl and Shepard.[10] They asked subjects to assess how well a final tone completed a sequence of seven major scale tones starting with middle C. The tones were approximately 0.75 seconds long each, and the final tone was chosen from each of the 13 chromatic tones from middle C to C′. Two conditions were tested, the ascending and descending, where each sequence starts with the C or C′, and ends with the tone preceding C′ and C, respectively. 24 subjects participated in this study, and the assessments were clustered in four groups, dependent on the similarity of the assessments. For our purpose, the first group assessment will be used as comparison to the sensory dissonance obtained with the memory model. The human goodness-of-fit can be seen in figure 3. The second group of this study only has high goodness-of-fit for the fundamental, while the third group seems to have goodness-of-fit related to the distance to the last note in the scale context. One isolated subject was singled out for having absolute pitch, and she gave high goodness-of-fit to C, E, G and C′.

---

[9] Kristoffer Jensen and Jens Hjortkjær, "An Improved Dissonance Measure Based on Auditory Memory," *Journal of the Audio engineering Society* 60(5) (2012): 350-354.

[10] Carol Krumhansl and Roger N. Shepard, "Quantification of the hierarchy of tonal functions within a diatonic context."

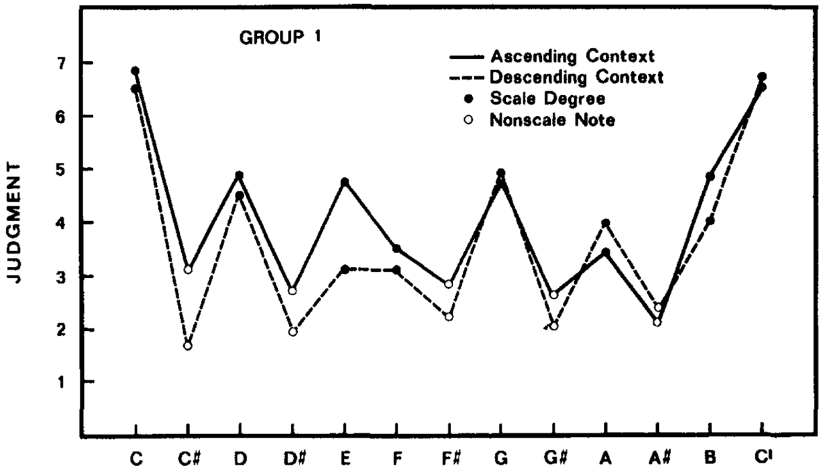

Figure 3. Goodness-of-Fit of final tone after seven-tone sequence.[11]

Thirteen piano sounds are used in this experiment. Seven scale tones are created either ascending or descending, and, one by one, each of the chromatic tones is appended to it. Each time, the sensory dissonance is calculated using the memory model, and the maximum sensory dissonance of the last tone is retained. As such, thirteen sensory dissonances are saved for the ascending, and thirteen for the descending sequence. The result is shown in figure 4.

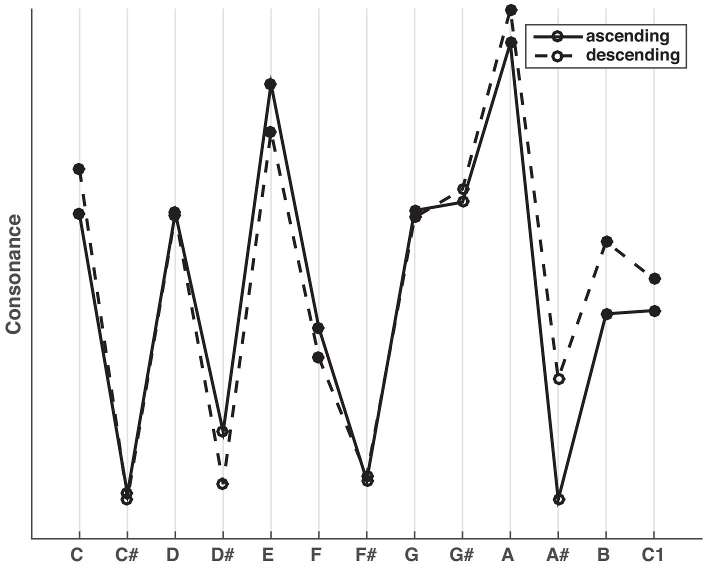

Figure 4. Sensory Dissonance using memory model of last note after seven note sequence. The major scale notes are denoted with black circle. Dissonance goes downwards, therefore the result is labelled consonance.

---

[11] *Ibid.*

While the results obtained through the sensory dissonance with memory model has many similarities to the goodness-of-fit of human subjects, such as higher consonance for scale notes, relatively low fifth consonance, and no large difference between ascending and descending scale context, some differences also exists, in particular the lack of relatively high fundamental consonance in the sensory dissonance with memory model results. Some of the discrepancies may be explained by the use of prior knowledge by the human subjects. It is still clear that the inclusion of a memory model in the calculation of the sensory dissonance enables the calculation of sensory dissonance of sequences of notes with good results.

Conclusions

Sensory dissonance has been known since the sixties, but its use has been hampered by the inability to take in account temporal context, i.e. prior tones.

A better measure of sensory dissonance is obtained if it includes a memory model. A review of the cortical and sub-cortical parts that participate in the forming and use of memory is presented here. A secondary episodic memory in the limbic system enhanced the robustness of the long-term memory, while retaining the flexibility necessary to retain new information. Information rich clubs enable the transmission of information between the different parts of the brain efficiently.

A novel improvement to the calculation of sensory dissonance, consisting of the identification and retention of prior tones, and the calculation of the sensory dissonance between the current tone and the previous tones retained in the memory model. As such, it has been shown that the sensory dissonance using memory model improves the estimation of sensory dissonance when comparing it with human subjects.

An experiment with the sensory dissonance using memory model confirms the validity of this approach by giving comparable results to a study using human subjects. While human subjects, of course, may take into account other knowledge, for instance by storing the context sequence in the temporary episodic memory, and identifying the important tones, a high degree of similarity was still found between the goodness-of-fit of a final tone in a seven tone major scale sequence context and the consonance obtained using the sensory dissonance using memory model.