GUIDING AUDITORY ATTENTION TOWARD THE SUBTLE COMPONENTS IN ELECTROCARDIOGRAPHY SONIFICATION

Hiroko Terasawa[1,2], Yota Morimoto[3], Masaki Matsubara[1], Akira Sato[1], Makoto Ohara[1], and Masatoshi Kawarasaki[1]

[1] University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki 305-8550 Japan

[2] JST, PRESTO
7, Gobancho, Chiyoda, Tokyo 102-0076 Japan

[3] Royal Conservatory of The Hague
Juliana van Stolberglaan 1, 2595 CA The Hague, The Netherlands

terasawa@slis.tsukuba.ac.jp

ABSTRACT

ECG (electrocardiography) consists of a few components and features that indicate the details of electrical conduction in each part of the heart. The simple pitch-mapping sonification of ECG cannot indicate these components to listeners, limiting its potential for diagnostic usages. We present an improved method to emphasize such key features in ECG sonification, using the principle that "we are attentive to loud, high-pitched, and rapidly fluctuating sounds." With this principle, we can guide the listener's attention to subtle yet critical elements of the ECG waveform by controlling the degree of auditory saliency for each component. In this report, we describe ECG sonification step by step and discuss the cognitive and attentive issues as well as the emotional responses of the listeners.

1. INTRODUCTION

In the diagnostic use of ECG (electrocardiography), medical doctors watch for key features that indicate the details of electrical conduction in each part of the heart. This paper describes a sound design methodology to emphasize such key features in ECG with the objective of guiding the listener's attention to the important information.

1.1. ECG components

ECG measures the electrical activity in the heart using electrodes placed externally on the chest or limbs. The electrical charging and releasing of the heart muscle cells control heart beating. ECG signals reflect this charging and releasing pattern that occurs at different parts in the heart.

A normal ECG signal shows a pattern within each cycle. Figure 1 shows one cycle of a sinus rhythm (a normal heartbeat pattern). A cycle is defined as a combination of five waves, specifically, the P, Q, R, S, and T waves. The Q, R, and S waves occur very rapidly and are often denoted as the QRS complex. The P and R waves occur with the atrial and ventricular discharges, respectively. The atria charge during the QS interval, and the ventricles charge during the T wave.

The R wave is the most notable wave, and the cardiac monitor most commonly used in hospitals (the type which generates periodic beeps) utilizes the occurrence of the R wave as a vital sign. The interval between consecutive R waves is called the RR interval and is the most critical information to be considered. Therefore, previous works on ECG sonification have mainly focused on displaying information related to the R wave [1, 2].

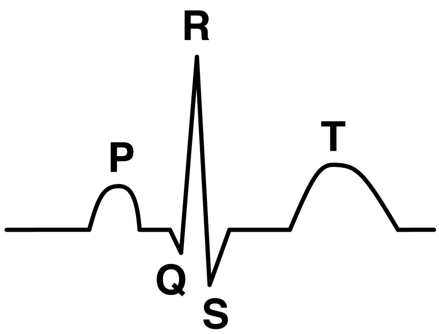

Figure 1: One normal ECG cycle, consisting of a P wave, QRS complex, and T wave. Usually, the U wave is hidden by the T wave and is thus invisible.

However, the waves other than the R wave also carry critical information [3]. For example, the shape of the P wave reflects the functional balance between the left and right atria. The T wave is generated when the heart restores in the resting stage, and deformation of the T wave may indicate conditions such as heart muscle problems or cardiomegaly (heart enlargement). There are many other important features in the P and T waves and in the intervals between waves as well. Cardiologists read such information from ECGs in order to perform diagnoses.

1.2. Goal and expected usage

Our objective was ECG sonification such that the sound would provide details of the P and T waves that would yield useful information for arrhythmia diagnosis.

This project was a collaboration between cardiologists and sound engineers. We decided that descriptive sonification of the P and T waves would be a research question of interest to both cardiologists and sound engineers. Cardiologists still examine these intricate waves visually, seeking critical signs for arrhythmia without assistance from information technology. For sound engineers, creating a design beyond monotonous beeps that can describe subtle changes in the data poses a significant research challenge.

Our expected users are cardiologists and clinical staff. The system is currently working offline (not in real time, but on stored data), but our approach is simple enough to be applied to real-time systems. We do not expect that the users will make diagnoses only based on sonifications, but rather that sonifications will contribute informative supplementary displays to be used along with conventional ECG graphs. In other words, we hope this kind of sonification will be useful as part of multi-modal (i.e., audiovisual) displays, increasing the robustness and decreasing the cognitive stress in the diagnostic process.

1.3. Cognition and attention in auditory scenes

This goal cannot be accomplished by simple pitch mapping, in which the signal amplitude is mapped to the frequency of the sonified sound. This simple pitch-mapping approach causes cognitive issues in interpreting the P and T waves, as we describe in Section 2. In this paper, we describe the process of resolving this problem and how we guided the listener's attention to the ECG P and T waves by effective sound design.

Section 2 describes the cognitive problem with the simple pitch-mapping approach. Section 3 offers a solution based on the principle that "fast and high-pitched sounds grab auditory attention." Section 4 describes the technical framework of the system, and Section 5 is devoted to the additional discussion, conclusion, and future work.

2. GUIDING ATTENTION WITH SOUND DESIGN

2.1. Simple pitch-mapping approach

We first applied the simplest pitch-mapping approach to our ECG data. The amplitude of the ECG waveform was associated with the pitch of the sonification: a higher amplitude corresponded to a higher tone pitch. We employed granulated sinusoids in a musical scale, like that of the piano keyboard shown in Fig. 2, instead of a sweeping sine tone.

The advantage of using the musical scale is that changes in the waveform are more easily recognized as they form a kind of melody. It is very difficult to recognize the extent to which the frequency of a sweeping sine tone increases. Listeners can more easily recognize the pitch intervals of a musical scale. Thus, in order to sonify the form and amplitude of an ECG signal, the use of the musical scale has significant advantages over sweeping sinusoids.

The use of the musical scale also has aesthetic merits. By choosing a diatonic scale (or some other scale, such as pentatonic) the occurrence of beating and inharmonic intervals can be limited. For listeners who are accustomed to traditional Western music, this property is highly desirable, facilitating the recognition of sonification as a result of their already trained music recognition skills.

Another synthesis technique we employed was "thinning," that is, generating a sinusoidal grain only when the pitch changed, as illustrated in Fig. 3. This design does not generate sound constantly but rather only when a significant change is observed. It yields a kind of arpeggiated rhythm in the sonification with fluctuating micro-silences between sinusoidal grains. This technique produces a light texture in the sonification, rather than the flat and heavy texture of a continuous sound stream.

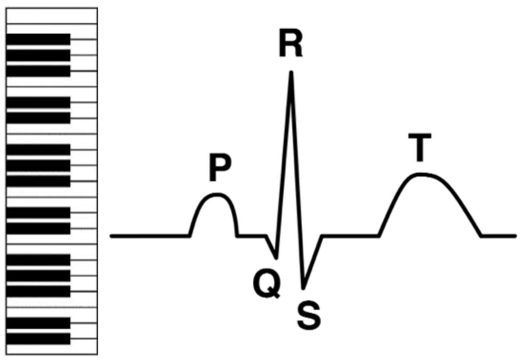

Figure 2: Simple pitch mapping in a piano-keyboard style. The amplitude of the ECG waveform is associated with the discrete pitch (note in a musical scale) of the sinusoidal grain.

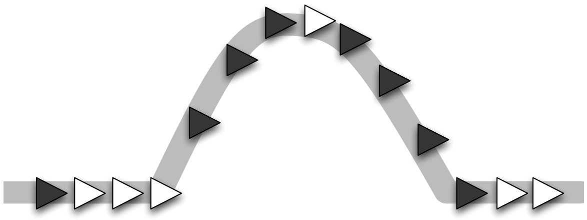

Figure 3: Illustration of sound thinning. The sinusoidal grains are produced at constant time intervals, but if the pitch of the subsequent sinusoidal grain is the same as that of the previous grain, only the first sound is played (dark triangles) and the later sound is omitted (white triangles). This process emphasizes the changes in the waveform and produces an irregular rhythmic texture in the sound.

2.2. The cognitive issue in P- and T-wave perception: "We are attentive to loud, high-pitched, and rapidly fluctuating sounds"

The next problem is the difficulty in differentiating the P and T waves with the simple pitch-mapping approach, because the sonified sound changes pitch very rapidly in the QRS complex, while the changes in the other waves are less rapid.

This problem is actually one of auditory attention and perception. Auditory scene analysis is the process of recognizing the meaningful segments of an auditory scene, which consists of the overlap and connection of many sounds [4]. In this process, some sounds are more distinctive than others and attract more attention. This attention-grabbing characteristic is called saliency, and recent publications have further elaborated on its concept and models [5, 6, 7].

The most fundamental principle of auditory saliency is that loud and high-pitched sounds are very difficult to ignore [7]. Most auditory alarms are designed based on this principle; for example, fire alarms are extremely high-pitched and loud to attract attention and motivate evacuation. Furthermore, since our hearing is most sensitive to high frequencies, we tend to perceive high-pitched sounds as louder [8].

We propose another influence on saliency, roughness, which is the rapid fluctuation of sound. Alarms, bells, and sirens often are accompanied by rapid changes in amplitude and frequency, resulting in a rough texture. Roughness is generally regarded as the cause of annoyance [9, 10]. We do not yet understand the causal relationship between annoyance and attention, yet we can reasonably speculate that the emergent events, which should attract attention, are associated with annoying, negative emotions, so that people intuitively react repulsively to the events [11]. Another possibility, based on a Darwinian perspective, is that individuals who learned to interpret an emergency as emotional arousal have survived over the generations. Considering the alarming and annoying nature of rough sounds, we can conclude that rapidly fluctuating sounds draw attention in an auditory scene.

With our simple pitch-mapping approach, the QRS complex becomes a high-pitched, suddenly fluctuating sound, satisfying the above criteria that are required for auditory saliency. That is why the QRS complex is audibly distinctive and focusing on the P and T waves is extremely difficult. However, since our goal was to better represent the P and T waves, the QRS complex being salient was not desirable in this investigation. In order to accomplish our goal, we had to cause the P and T waves to be salient and to de-emphasize the QRS complex.

2.3. Attention-guiding strategy

The saliency principle of "we are attentive to loud, high-pitched, and rapidly fluctuating sounds" can be an influential tool for sound design. By inverting this principle, "quiet, low-pitched, and slowly fluctuating sounds" are less salient. Considering these principles enables effective sound design.

3. SONIFICATION DESIGN

The sonification design described in this section resolves this problem by inverting the order of saliency for the QRS complex and the other waves and by replacing less interesting information with a symbolic sound. We describe the sound design concept in this section and further technical details in the next section.

3.1. Control signals

In order to produce the control signal, we first conducted a simple QRS complex detection. We analyzed the signal with running windows and identified the segments showing the largest variance as the QRS complexes, since the rise-fall pattern of a QRS complex segment shows the greatest variation from the baseline signal. The segments between consecutive QRS complexes were simply divided into first and second halves. Figure 4 shows an example of the three control parameters in a sinus rhythm; we associated control parameters 1, 2, and 3 with the QRS complex, PR interval, and ST interval, respectively.

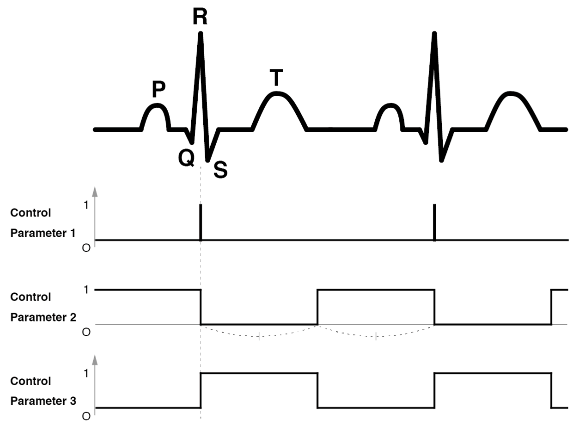

Figure 4: Control parameters for sonification. Parameters 1, 2, and 3 indicate the QRS complex, P-wave segment, and T-wave segment, respectively.

The detection of the QRS complex and allocation of PR and ST intervals can be further sophisticated, but that is beyond the scope of this paper.

3.2. Achieving a symbolic, less salient QRS-complex sound

In the pitch-mapping sonification, the QRS complex sounded too salient. Therefore, we replaced this segment with a bell-like, low-pitched, slowly decaying sine tone (which is, according to our principle, a less salient sound) at the time when control parameter 1 turned positive. This replacement step significantly facilitated the listener's ability to distinguish the P and T waves. Thus, listeners could perceive the details of the P and T waves yet simultaneously maintain awareness of the basic heartbeat rhythm, so the QRS complex part no longer interfered with attention to the P and T waves.

3.3. Achieving a more salient P-wave sound

The P and T waves were then better recognized, yet the segmentation between the two consecutive waves was not evident. The cardiologist we consulted advised us to emphasize the P waves more because these include more critical information than do the T waves. Therefore, when control parameter 2 became positive, we raised the pitch of the P waves by a half octave, so that these could attract more attention. After doing so, the P waves were more salient, and listeners could focus on the P waves while recognizing the other parts in parallel.

4. TECHNICAL DETAILS

We implemented the sonification system in the SuperCollider programming environment combined with Ruby. As described in Section 2.1, we employed granular synthesis and mapped the amplitude of the ECG data onto the discrete steps of a musical scale. In order to enable sonic distinction of the PR interval, QRS complex, and ST interval, our system constructed three granular synthesis streams per ECG signal, each corresponding to one of the three waves. In this way, each sound stream could be adjusted independently (scale range, transposition, grain articulation, panning, etc.) according to its intended cognitive role.

4.1. Data analysis

First, we analyzed the original ECG data by using the statistical analysis class implemented in Ruby, since SuperCollider only minimally supports mathematical analysis. The variance of the windowed signal was calculated, and the sections with largest variance were identified as QRS complexes. The calculated control parameters 1, 2, and 3 were then used to control the associated granular synthesis streams. All three of the granular synthesis streams shared the original ECG data but were controlled (turned on or off) independently according to the three control parameters in Fig. 4.

4.2. Sound synthesis

We designed the internal sonic aspects of each of the three streams such that the sonic images of each stream would be distinct from those of the others: the P wave employed Hanning-windowed sine-tone grains and was rendered in a higher pitch range than that of the T wave (in order to emphasize the critical information); the T wave used a sine wave oscillator with phase-modulation feedback, adding clarity (harmonics) to the otherwise dull-sounding frequency range; the R wave was generated in the pitch range between those of the P and T waves and was distinctively articulated with a percussive envelope.

Additionally, the grain density was controlled for both the P and T waves by dynamically controlling the number of overlapping grains according to the amplitude of the ECG signal: a lower amplitude of the ECG (i.e., lower pitch of sonification) corresponded to fewer overlapping grains. This overlap was controlled to avoid the dull sonic image of the saturated lower frequency range and to smooth out the otherwise jagged articulation in the upper frequency range.

We decided to render the R wave weaker in amplitude than the P and T waves because of its simpler role; it only informs the regularity or irregularity of the beating intervals. However, we rendered it as a chord (notes: C, F, A, and B), and because of its slightly dissonant character (F major #4th) in the key of C major, it could be heard more clearly than if it were, for example, a single tone or a C major chord. As exemplified in this harmonic approach, by limiting the continuous pitch space to the discrete steps of a musical scale, we expanded our parameter space by also incorporating the musical parameter space (i.e., tonality, melody, harmony, etc.). The range of notes and chord used for sonification are depicted on a musical staff in Fig. 5.

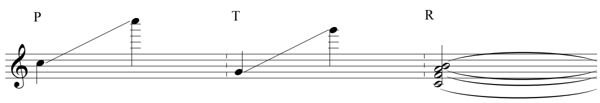

Figure 5: The range of notes used for sonification for the PR interval (left), ST interval (middle), and QRS complex (right).

5. DISCUSSION

5.1. Saliency, alarm, and emotion

The phenomenon that "loud, high-pitched, and rapidly fluctuating sounds are difficult to ignore" occurs quite often in sonification, even with data from sources other than ECG. Furthermore, most of the annoying sounds produced by sonification are of this nature. These sounds easily cause very negative emotional responses, and the Darwinian ideas described in Section 2.1 may explain why. For example, the smell of rotten food is perceived as repulsive and arouses a negative emotional response. This reaction is very natural because those who did not have negative responses to rotten food did not survive over the generations, so we are the descendants of those who experienced repulsion to rotten food. In the same way, considering that "loud, high-pitched, and rapidly fluctuating sounds" are most likely caused by potentially dangerous events with high energy and activity, it is also natural that many of us experience negative responses to these sounds.

In sonification, when the original data are "fast and strong," we may or may not wish to express the sound as such. If we do not wish to express the signals as originally generated, we can sonify them with various degrees of saliency by carefully reducing the relevant features in order to avoid unnecessary negative responses.

5.2. Compressing information based on expected function

The other interesting aspect in this sonification process is the mixture of symbolic and analogic representations, which were described by Kramer [12]. Symbolic representations display the processed information, while analogic representations directly display the original data or signal. In our case, we combined these two types of representations, representing the QRS complex as symbolic and the P and T waves as analogic.

A symbolic representation is equivalent to the compression of the signal information. In our case, the details of the QRS complex were omitted, and only the information about its timing was utilized, in order to mark the rhythmic intervals. Of course if we had wanted to utilize all of the original data, this kind of compression would have been impossible. However, knowing which aspects of the data to retain or discard enables the combined use of the symbolic and analogic representations.

Mixing the symbolic and analogic representations facilitates the parallel processing of multiple kinds of information. Symbolic representations tend to generate impressions similar to those of musical sounds, while analogic representations generate impressions similar to those of environmental sounds. Humans are able to distinguish musical and environmental sounds and process them in parallel. Perhaps because of this skill, it was easy to recognize and distinguish the two types of sound in our sonification.

As described in Section 4.2, by limiting the pitches to those of a musical scale, our degree of expressive freedom increased, enabling the use of musical parameters. Furthermore, the waveforms became more easily recognizable due to the melodic patterns. Flexible switching between symbolic and analogic representations, or superposing of those representations, provides a sonic resource that can be further exploited.

5.3. Collaborative process and evaluation

This project was conducted as a close collaboration between sound engineers and cardiologists. Although we have not conducted a formal evaluation, the frequent meetings of the two parties facilitated the estimation of the validity of the sound design. The cardiologists and engineers sometimes had different opinions (or complaints) about the sound design, and such conflicts usually developed into creating a better solution.

6. CONCLUSION AND FUTURE WORK

In this paper we described a design method to guide a listener's attention in sonification, using ECG sonification as an example, according to the principle that "we are attentive to loud, high-pitched, and rapidly fluctuating sounds." The QRS complex was replaced with a symbolic, less salient sound, and the critical P wave was emphasized with a distinctive, higher pitch. We also discussed the emotional response to and auditory saliency of the sonification.

Topics of future work include investigating the potential for arrhythmia diagnosis and sophisticating the data analysis process. A formal listening test is also desirable, but further close collaboration between domain scientists (cardiologists) and sound engineers could be equally advantageous.

7. ACKNOWLEDGMENT

This project is funded by PRESTO program of Japan Science and Technology Agency.

8. REFERENCES

[1] D. Worrall, B. Thoshkahna, and N. Degara, "Detecting Components of an ECG Signal For Sonification," in *Proc. of the 20th Int. Conf. on Auditory Display (ICAD-2014)*, New York, USA, 2014.

[2] M. Ballora, et al., "Detection of Obstructive Sleep Apnea through Auditory Display of Heart Rate Variability," in *Computers in Cardiology*, vol 27, pp. 739-740, 2000.

[3] S. Watanabe, and I. Yamaguchi, *ECG Perfect Manual - A Comprehensive Tutorial with Theory and Waveform patterns (in Japanese)*, Tokyom Japan: Yodosha Publishing, 2006.

[4] A. Bregman, *Auditory Scene Analysis*, Cambridge: MIT Press, 1990.

[5] M. Elhilali, et al., "Interaction between Attention and Bottom-Up Saliency Mediates the Representation of Foreground and Background in an Auditory Scene," *PLoS Biology*, Volume 7, Issue 6, e1000129, pp. 1-14, June 2009.

[6] V. D. Delmotte, *Computational Auditory Saliency*, Ph.D. Thesis, Georgia Institute of Technology, Atlanta, 2012.

[7] S. H. Chon, *Timbre Saliency, The Attention-Capturing Quality*, Ph.D. Thesis, McGill University, Montreal, 2013.

[8] B. Moore, *An Introduction to the Psychology of Hearing: Sixth Edition*, Leiden, Netherlands: Brill, 2013.

[9] W. A. Sethares, *Tuning Timbre Spectrum Scale: Second Edition*, London, UK: Springer-Verlag, 2005.

[10] I. Peretz et al., "Cortical Deafness to Dissonance," *Brain*, vol 124, pp. 928-940, 2001.

[11] D. Evans, *Emotion: A Very Short Introduction*, Oxford, UK: Oxford University Press, 2003.

[12] G. Kramer, *Auditory Display*, Reading: Addison-Wesley, 1994.