STREAM-BASED SONIFICATION DIAGRAMS

Stephen Barrass
Sonic Communications Research Group
University of Canberra
stephen.barrass@canberra.edu.au

Virginia Best
School of Medical Sciences
University of Sydney
ginbest@physiol.usyd.edu.au

ABSTRACT

The van Noorden Diagram describes the auditory streaming of two tones with changes in pitch difference and intertone onset interval (IOI). There are regions where the listener hears one stream or two streams, and an ambiguous region between where listening attention affects what is heard. The ambiguous region dominates at IOI > 200ms which is where many sonifications are designed. We propose a Stream-Based Sonification region at IOI < 200ms to control streaming and reduce the ambiguous effects of attention. In this paper we generalise this region in a series of four Stream-Based Sonification Diagrams. The first is a repetition of the original van Noorden Diagram at higher temporal resolution in the SBS region. The other three show the same general pattern of regions for new mappings of the brightness, amplitude and pan of a noise. The results show that streaming by brightness and pitch are closely related. They also show a new coherence boundary for streaming by amplitude, and that streaming by spatial panning is relatively unaffected by IOI. The palette of Stream-Based Sonification Diagrams developed here provides a foundation for the design of sonifications that control streaming and take listening attention into account.

1. INTRODUCTION

At the first ICAD in 1992, Sheila Williams observed the difficulty of interpreting an auditory display when changes in acoustic parameters do not necessarily lead to corresponding changes in perception. Changes in amplitude may lead to changes in rhythm, changes in frequency may be obscured by harmonic grouping, steps in the data may be masked by perceptual category, and temporal relations may be obscured by the grouping of data items in different streams [1]. She proposed that the theory of Auditory Scene Analysis could be used to design sonifications that were more perceptually meaningful and predictable, and presented an experiment where she scaled the metric of the strength of the streaming of a tone against captor tones. This experiment involved the use of a sequence of tones known as the van Noorden gallop that the listener can hear in two different ways depending on the perceptual organization by streaming. She concluded with the suggestion that an auditory display could take advantage of streaming effects in situations where the actual value of a parameter does not matter as much as the deviation from a general trend.

Although Williams shows the early recognition of the value of the Auditory Scene Analysis theory for designing sonifications, there has been very little research on streaming in the context of auditory display since. In the next section we present an overview of streaming theory and especially the van Noorden effect. Next we review research on streaming in the context of auditory display. We then develop the hypothesis that the van Noorden Diagram can be used to design sonifications that can take into account the effects of streaming and attention. The section on experiments develops a series of Stream-Based Sonification Diagrams modelled on the van Noorden Diagram. We analyse the results from these experiments and conclude with an agenda for further research into Stream-Based Sonification.

2. BACKGROUND

People can switch their attention between the sounds around them, such as a motorbike going by, the call of a bird outside and the tap of the keys on the keyboard. Auditory streams usually correspond with physical objects and events in the environment. The theory of Auditory Scene Analysis describes heuristics that organise the acoustic array into streams that can be selected by listening attention. However there are circumstances when what is heard does not necessarily correspond with a physical source. In one of his earliest experiments, Albert Bregman spliced together 26 short recordings of everyday sounds in a series. When he listened to the edited tape he heard unexpected sounds that he had not recorded, and found that it was difficult to tell the order of the sounds in the series [2]. The psychoacoustic effects of auditory stream formation can be demonstrated by the van Noorden gallop produced by arranging tones in a triplet sequence ABA-ABA- where A are reference tones, the B tones are variable, and the – are silences [3]. When the pitch difference between A and B is small the listener hears a 'galloping' rhythm, as illustrated in Figure 1a. As the pitch difference, Δf, is increased the sequence splits into two distinctly different rhythms, A-A-A-A- and –B—B-, with regular beats at different timings, as illustrated in Figure 1b. The sequence splits in a similar manner when it is sped up by reducing the intertone onset interval (IOI)..

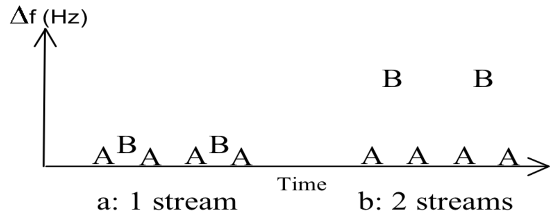

Figure 1. The van Noorden gallop.

The transition from the unitary galloping rhythm to two distinctly different rhythms at different rates passes through an ambiguous stage where listeners can hear it either way, depending on how they focus their attention. The point at which a change in attention can cause a change in the galloping rhythm is called the 'fission' boundary, and the point at which attention can cause a change in the perception of two rhythms is called the 'coherence' boundary. These boundaries define the regions where the listener hears a single stream, two streams, and the ambiguous region between. These boundaries and regions were mapped for changes in pitch separation and IOI in the van Noorden Diagram shown in Figure 2.

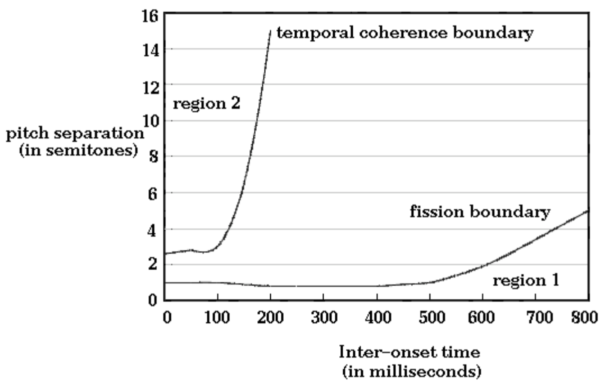

Figure 2. The van Noorden Diagram.

The fission boundary is flat for IOI < 500ms, so the gallop effect depends only on pitch difference there. Likewise the coherence boundary is flat for IOI < 100ms so the two-stream effect is mostly dependent on pitch difference in this region. This means that for IOI < 100 the perception of one or two streams is highly predictable and dependent on a simple pitch difference between the tones. From 100-200ms there is a rapid change in the coherence boundary, and above 200ms the two-stream region diminishes so that the ambiguous region dominates the Diagram.

3. STREAMS IN SONIFICATION

Fitch and Kramer presented a sonification of 8 physiological variables mapped to two auditory channels – one for heart variables and one for breath [4]. The sonification design, modeled on the pulse-oximeter used in operating theatres, maps pulse rate to a beeping tone that varies in pitch with blood oxygen level. The events in this channel have an IOI of 500-1000ms. A CO_2 variable was piggybacked onto this channel by altering the timbre of the tone using frequency modulation (FM). The breath channel maps respiration amplitude to the amplitude of a synthesised breath-like noise. Changes in body temperature were piggybacked onto this channel by modifying the centre frequency of a band-pass filter applied to the noise. In an evaluation, the subjects performed a simulated monitoring task faster and more accurately with this sonification than with a visual graph of the same information. The theory of ASA suggests that the choice of tones for one channel, and noise for the other, may have been an important factor in the perceptual segregation of the information in these auditory channels. Dannenbring and Bregman found that the "segregation effects seen with these [tone and noise] stimuli are the strongest we have seen" [5].

The extension of the pulse-oximeter design pattern to pulse and breath channels was reused by Seagull, Wickens and Loeb in a further study of visual, auditory and redundant displays for physiological monitoring [6]. The subjects monitored and adjusted the state of a simulated patient using three heart parameters and three breath parameters, and at the same time performed a manual tracking task that required hand-eye coordination. As previously, heart rate and breath rate were sonified by the rate of heart and breath channels. Oxygen saturation was mapped to pitch of the heart channel, and CO_2 level mapped to the pitch of the breath channel. Blood pressure modified the timbre of the blood channel, and breath volume modified the timbre of the breath channel. Performance with the sonification in this experiment was significantly slower than with the visual display. The sonification produced the best performance in the distracter task, by freeing visual attention from the monitoring task. Performance on the distracter task was poorest with the redundant audiovisual display, perhaps due to conflicting information or attention overload. Although the sonification is modeled on the previous experiment it is not clear from the paper whether the heart and breath channels were a tone and a noise. If tones were used for both then streaming between the auditory channels may have caused perceptual and cognitive interference.

The heart and breath pattern was reused again in a further study on physiological monitoring by Watson and Sanderson that additionally included the effects of expertise [7]. They found that expert anesthetists performed equally well with the auditory, visual, and combined audiovisual displays. In this experiment the experts performed better in a distracter task while monitoring an auditory display than with the other displays. Non-experts showed lower performance levels overall, and the auditory display resulted in worse performance in both the monitoring and distracter tasks. They observed that the attention required to track and integrate information across two auditory streams may have had an effect on cognitive load. This observation distinguishes between a sonification mapping of data to an auditory stream, from a mapping to an acoustic channel. The mapping to streams addresses the problem of perceptual crosstalk between information in auditory channels, and raises the higher-level issues of attention and cognition of information in auditory displays.

Anderson and Sanderson further investigated whether monitoring performance would be different if the six heart and breath variables were mapped to one, two or three auditory streams [8]. As well as pulse speed, frequency and spectral shape, the single stream configuration could vary in amplitude, pulse width, and tremolo (a regular amplitude variation within a pulse). In the two stream configuration each base stream was a tone, but one varied in pulse speed, frequency, and tremolo, while the other varied in amplitude, spectral shape, and pulse width. In the three stream case, the base streams were all tones, but one varied in pulse speed and frequency, one in spectral shape and tremolo, and one in amplitude and pulse width. The task was to detect changes in one variable when no other variables were changing, when one other variable was changing and when five other variables were changing. The results showed best performance with the single stream sonification and one changing variable. Performance with all configurations was similar in the presence of distracter variables. Performance was poorer when there was a distracter task, and significantly worse in the three stream configuration.

In another study of streaming in the context of auditory displays, Rivenez and colleagues [9] investigated the question of whether streams are formed by attention, or are formed pre-attentively. Subjects performed a tone detection task in the presence of silence, an environmental recording, and a white noise stream. They found that stream segregation requires cognitive effort, that it seems to occur in non-focused streams, and that the perceptual stability of the focused stream is highly dependent on attention.

4. PROPOSAL

The research on streaming in auditory display shows there are significant implications for :

- the order of sequential data.
- the grouping relations between data.
- interference across data channels.
- attention and focus.
- cognitive load and interpretation.

We propose that the van Noorden Dagram can be used to design sonifications that can control the effects of streaming to achieve better channel separation and reduce cognitive workload. The Stream-Based Sonification (SBS) region, shown in Figure 3, from IOI 0-200ms, reduces the ambiguous effects of attention, and enables both region 1 and region 2 to be usefully employed.

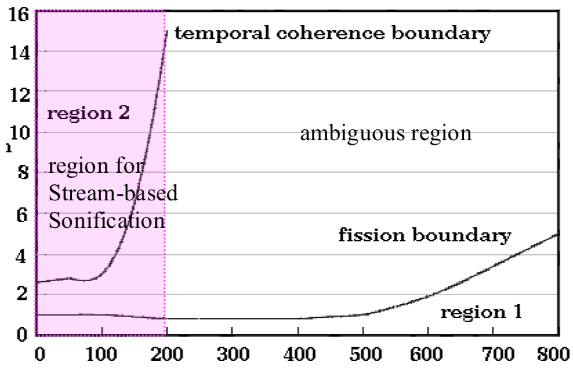

Figure 3. Region for *Stream-based Sonification*.

It is interesting to note that Williams found that the grouping of a tone followed a metric close to the log scale for IOI from 50-70ms, and that there was a categorically different mechanism above 100ms, which is where the coherence boundary turns to slope upward.

Up until now the van Noorden Diagram has only been used to describe the streaming of tones by pitch difference. However there are many other auditory parameters that cause streaming, such as spectral centroid, amplitude, inter-aural level differences (ILDs), envelope shape, spectral shape, phase difference and others. Moore and Hockel provide an overview of a wide range of experiments that use the van Noorden gallop, and conclude with the suggestion that any perceptible difference between two sounds may produce streaming effects [10]. However, although the gallop has been used extensively to study many different auditory parameters, the van Noorden Diagram seems only to have been used to describe the streaming by pitch of a tone. We propose that the van Noorden Diagram can be generalised to other auditory parameters to develop a palette of SBS Diagrams that can be used to design many different kinds of sonifications.

5. EXPERIMENTS

This section presents four experiments that develop the concept of SBS Diagrams based on the van Noorden Diagram. The first experiment repeats the original van Noorden experiment in one subject to establish a reference framework and methods. The three following experiments generalise the diagram in two subjects to the brightness of a noise grain, and then to the amplitude and ILD panning of the noise grain. The sounds were synthesised using the open source PD tool [11], and the synthesis patches referred to in these experiments can be downloaded from http://www.scrg.org.au.

A van Noorden sequence of sine tone triplets ABA-ABA- was synthesised with the PD patch sine-tone.pd that has inputs for a trigger, IOI (ms), duration (ms) frequency (Hz), and outputs for a trigger and an audio signal.

Six of these sine tones were chained together to build a van Noorden synthesiser, van-Noorden.pd. The pitch of the A tones was set to midi 83 = 987.3Hz, close to the 1000Hz used by van Noorden in the original experiment. The pitch difference Δf of the B tone from the A tone was adjusted by the subject using a yellow slider in the interface. The subject was also able to adjust the IOI in 10 ms steps from 10-300 ms using buttons on the left of the panel. The duration of each tone was 30ms with 5ms attack ramp, and a level of 70dB. The tones were synthesised in real-time on a MacBook Pro and delivered through Sennheiser HD490 headphones.

The experimental method is modelled on the method of adjustments used by van Noorden in his second series of experiments on pitch streaming [3]. The subject selects an IOI and then adjusts the frequency of the B tone to the point where the ABA sequence first divides at the fission boundary. The subject then enters the Δf at this point in a spreadsheet. The subject then adjusts the frequency of the B tone to the point where the ABA sequence can no longer be heard as a gallop on the coherence boundary. The subject then enters the Δf at this point in the spreadsheet.

The data points recorded in the first experiment by subject S are shown in Figure 4, together with the fission and coherence boundaries from van Noorden. The results are generally similar but there is a notable difference in the region from 40-160 ms where the ambiguous area in this pilot is up to 5 semitones wider than recorded by van Noorden. There is also a bump in the fission boundary between 30-50ms, and a slight increase from 210ms that is not visible in the original study. Referring back to van Noorden's results we see that there are individual differences of a similar order, and this plot is an aggregate from two quite different listeners. The judgements of the fission and coherence boundaries border on ambiguity, and subjects tend to become better at making this judgement with experience. Even explaining the concepts is difficult and this may be why the number of subjects in these experiments is small and the results show so much individual variation. Nevertheless, the general architecture of fission and coherence boundaries around an ambiguous region is confirmed using the current method.

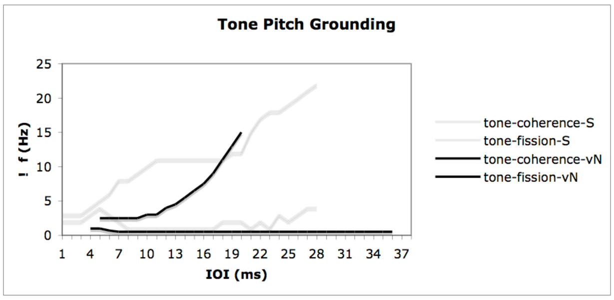

Figure 4: Tone Pitch Grounding

The following three experiments investigated whether the van Noorden Diagram could be extended to other aspects of sound that influence auditory streaming. In the first of this series the spectral centroid (related to brightness) of a noise was substituted for the pitch of a tone. The noise grain was synthesised by passing white noise through a 100Hz (6dB roll-off) band-pass filter with a variable centre frequency, in noise-grain.pd. The duration of the grain was enveloped to 30ms with 5ms attack ramp, with a level of 70dB as in the previous experiment. The user interface and procedure were the same as the previous experiment. The only difference was that the B slider controlled the midi frequency of the spectral centroid of the noise grain, rather than the frequency of a sine tone. The centroid of the A grain was set to 987.3 Hz to mirror the A tone in the previous experiment.

The noise grains produced streaming effects that were very similar to the classic sine tone study. The results in Figure 5 show the coherence and fission boundary, with an ambiguous region between.

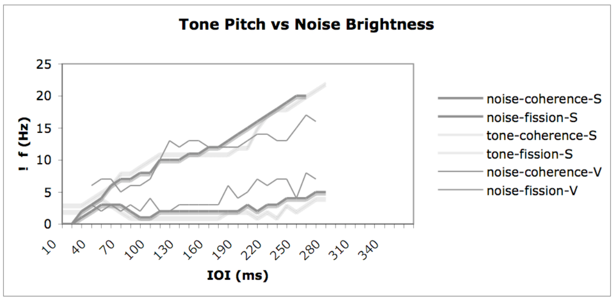

Figure 5: Tone Pitch vs Noise Brightness

The next experiment repeated van Noorden's mapping of the fission boundary for the amplitude level of a sine tone, but this time using a noise grain [12]. In his study van Noorden suggests that there may also be a coherence boundary where level differences cause two distinct streams to be heard. We followed this suggestion by also mapping the coherence boundary for a noise grain, using the same method as in the previous experiment on brightness. Although there are notable individual differences, the results for the two subjects, S and V in Figure 6 exhibit the form of a SBS Diagram, with a fission boundary and a coherence boundary each side of an ambiguous region. The fission boundary data from van Noorden's study is also shown.

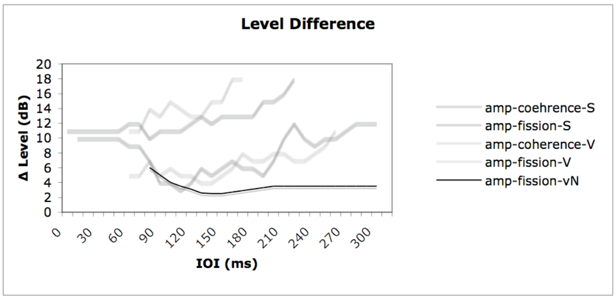

Figure 6: Level Difference

The final experiment extends the van Noorden Diagram to stereo panning of a noise grain by ILD. The vertical axis is the angle in degrees between the A and B elements, which were panned symmetrically from the midline. An angle of 0 degrees positions A and B together in the middle of the head. An angle of 40 degrees places A at -20 degrees to the left, and B at +20 degrees to the right. As in the previous experiments, the subject adjusted the variable parameter (in this case pan angle) to interactively find the coherence and fission boundaries. The results for subjects S and V in Figure 7 again show the form of the SBS Diagram with coherence and fission boundaries.

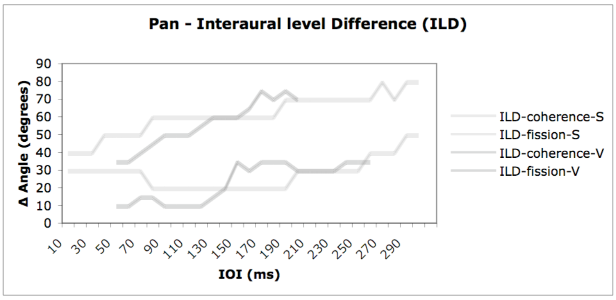

Figure 7: ILD Difference

6. DISCUSSION

The results of the first experiment confirm the narrow ambiguous region from 50-150ms, and the steep rise in the coherence boundary from 150-220ms that van Noorden found in his original experiment. However, there is a hump in the coherence boundary from 50-150ms that is not in the original results. A similar result was also found in the next experiment with the brightness of noise grains. At this stage it is not clear if this is due to a different interpretation of ‘two streams’, individual differences between the small number of subjects, or a difference in the resolution of data measured in this region.

The second experiment confirms that the method for measuring the van Noorden effect can be extended to other auditory parameters. The resulting van Noorden Diagram for the brightness of a noise grain mirrors that for the pitch of a sine tone with only minor differences at the coherence and fission boundaries. This similarity suggests that the same perceptual processes may be involved.

The third experiment further extends the mapping of coherence and fission boundaries to the loudness differences between noise grains. Although loudness is considered to have a weak effect on streaming, the one- and two-stream regions were distinct, and the coherence and fission boundaries were clearly perceptible. The coherence boundary for loudness exhibits a sharp upward slope around 150-200ms, similar to the coherence boundaries for pitch and brightness. The fission boundary also has a general similarity but is more complicated, starting with a downward slope from 50-100ms, then a hump at 150ms, an upward slope and another hump at 200ms. This boundary can accommodate the 8 data values from the 2 subjects in van Noorden’s study in the region 50-100ms, but is 5dB higher than the original data at 200ms. The individual differences between van Noorden’s subjects range up to 5 dB and could explain the variation from the current study.

The final experiment is a new extension of the van Noorden method to stereo panning that shows similar boundaries and regions. However the coherence boundary does not have the same steep rise from 150-200ms as in the other diagrams. From these results the grouping by stereo panning is largely independent of the IOI up to 300ms. Acoustic elements that are less than 30 degrees apart form a single stream, while those that are more than 60 degrees apart segregate into two streams. In the ambiguous region between the listener can use attention to hear the sequence in either way.

7. CONCLUSION

The theory of Auditory Scene Analysis was taken up in the sonification research community at the very first International Conference on Auditory Display. Since then there have only been a handful of studies that have investigated this theory in the context of sonification. These studies have raised the issues of the effect of streaming on perceived temporal organisation of sonified data elements, and the effects of attention on stream formation and cognitive load. Sonifications that present data at rates lower than two elements per second are in the region of the van Noorden Diagram where the effects of attention dominate what is heard. We propose that a Stream-Based Sonification region, from 0-200ms in the van Noorden Diagram, can reduce the ambiguous effects of attention. In this paper we replicated the original van Noorden experiment to confirm the coherence and fission boundaries, and mapped these boundaries for the brightness, loudness and pan of a noise grain in the Stream-Based Sonification region. The extension to the brightness of noise grains showed a high correlation with pitch grouping. The amplitude experiment confirmed van Noorden’s results on the fission boundary, and extended them to include the coherence boundary. The final experiment extended the paradigm to stereo panning and identified similar regions and boundaries, although streaming was largely independent of the inter-onset interval in this case. These results show that the Stream-based Sonification region can be a general framework for designing auditory parameters using sequential stream segregation.

Further work will involve the development of a sonification design method that uses these Stream-Based Sonification Diagrams, some example designs for specific applications, and evaluations of these applied designs. By controlling streaming these sonifications will improve the perception of temporal order, reduce perceptual interference between data channels, and reduce cognitive load due to attention switching. Other further work will involve the verification of the diagrams obtained from the two subjects so far, the development of Stream-Based Sonification Diagrams for other auditory parameters, and the extension of the framework to simultaneous streaming.

8. REFERENCES

[1] S.M. Williams, "Perceptual Principles in Sound Grouping, " in G. Kramer (ed) *Auditory Display: Sonification, Audification and Auditory Interfaces*. SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII. Addison Wesley, Reading, Mass, pp. 95–126.

[2] A.S. Bregman, *Auditory Scene Analysis*. MIT Press, Cambridge, MA, 1990.

[3] L.P.A.S. van Noorden, "Temporal Coherence in the Perception of Tone Sequences," Unpublished Thesis, University of Technology, Eindhoven, 1975.

[4] W.T. Fitch and G. Kramer, "Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex Multivariate System." in: G. Kramer (ed) *Auditory Display: Sonification, Audification and Auditory Interfaces*. SFI Studies in the Sciences of Complexity, Proceedings Volume XVIII. Addison Wesley, Reading, Mass, pp. 307-325

[5] G.L. Dannenbring and A.S. Bregman,."Stream Segregation and the Illusion of Overlap", *Journal of Experimental Psychology: Human Perception and Performance*., vol. 2, no. 4, pp. 544-55, Nov. 1976.

[6] F.J. Seagull, C.D. Wickens and R.G. Loeb, "When is Less More? Attention and Workload in Auditory, Visual, and Redundant Patient-monitoring Conditions," in *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 2001.

[7] M.0. Watson and P.S. Sanderson, "Designing for Attention with Sound: Challenges and Extensions to Ecological Interface Design," *Human Factors*, vol. 49, no. 2, pp. 331-346, April 2007.

[8] J. Anderson and P.S. Sanderson, "Designing Sonifications for Effective Attentional Control in Complex Work Domains," in *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society (HFES),* Santa Monica, CA. pp. 1818-1822, 2004.

[9] M. Rivenez, C. Drake, A. Guillame, and D. Sebastein. "Listening to Environmental Scenes in Real Time, " in *Proceedings of the International Conference on Auditory Display (ICAD 2002),* Kyoto, July 2002.

[10] B.C.J. Moore and H. Gockel, "Factors Influencing Sequential Stream Segregation," Acta Acustica united with Acustica, vol. 88, no. 3, May/June 2002, pp. 320-333..

[11] M.S. Puckette, "Pure Data," in *Proceedings of the International Computer Music Conference. (ICMC 1996)* San Francisco, International Computer Music Association, 1996, pp. 269-272.

[12] L.P.A.S. van Noorden, "Minimum Differences of Level and Frequency for Perceptual Fission of Tone Sequences ABAB, *The Journal of the Acoustical Society of America* vol. 61, no. 44, pp. 1041-1045, 1977.