EXTENDED ABSTRACT: USING VOCAL-BASED SOUNDS TO REPRESENT SENTIMENT IN COMPLEX EVENT PROCESSING

Jeff Rimland

College of Information Sciences and Technology
Penn State University
University Park, PA 16802
jrimland@ist.psu.edu

Mark Ballora

School of Music/School of Theater
Penn State University
University Park, PA 16802
ballora@psu.edu

ABSTRACT

There is an intricate and evolving relationship between sonification and Complex Event Processing (CEP) for improved situational awareness. In a paper presented at ICAD 2013 [1], we introduced a series of techniques using CEP for simultaneous sonification of both quantitative "hard" data and human-derived "soft" data in the context of assistive technology. The connection of CEP and sonification was explored further in the context of a severe weather tracker that relies on fusion of quantitative (sensor-based) weather data along with human observations about storms and related conditions [2]. An area of shortcoming in both of these earlier works was the difficulty in creating sounds that represented human sentiment about observed conditions (e.g. unanticipated obstacles for a blind person crossing a busy street, or impending dangerous weather conditions) in a format that enabled intuitive listening for improved situational awareness. This extended abstract provides an update on that continuing research by representing human sentiment data, via the use of vocal synthesis that is driven by Complex Event Processing.

1. INTRODUCTION – COMPLEX EVENT PROCESSING (CEP)

The Complex Event Processing (CEP) paradigm provides capabilities for rapid processing of multiple distributed heterogeneous streams of data in a manner that allows high-level semantic information to be extracted from low-level data by applying relatively simple rules, filters, and aggregations to assist in recognition of increasingly higher-level events within each level below it [3]. For example, the low-level events of:

1) people throwing rice
2) tin cans tied behind a limousine
3) women dressed in bridesmaid gowns

can be combined into the higher-level event called "wedding." It could also be interesting (perhaps to a bridal magazine) to calculate that several weddings within a short period of time comprise an event called a "wedding frenzy" [4].

Additionally, CEP is well suited to performing real-time processing from a direct input stream as opposed to forensic analysis of past events that are recorded in a database. This approach is much different from conventional relational database querying, which relies on the event being completed before analysis begins. In fact, raw data processed by CEP is often never stored in a persistent database [5].

There is the potential for a symbiotic relationship between sonification and CEP because each can help to deal with key shortcomings of the other. One of the main challenges of CEP is the difficulty of determining event thresholds (such as how many "wedding" events per month makes a "wedding frenzy"). Listening to auditory displays comprised of key indicators of performance in the CEP process (e.g. ratios of low-level to high-level events, aggregation statistics, etc.) could potentially provide an intuitive human interface for tuning and optimizing these thresholds. Additionally, since CEP can assist in abstracting low-level events into higher-order inferences, complex sonification challenges can be simplified and improved through the utilization of CEP techniques. For example, datasets conventionally considered too large or complex to sonify could be converted into a more tractable problem by working with manageable numbers of high-level events instead of being hindered by huge numbers of low-level events [6]. Even with this synergistic relationship, certain types of data are difficult to integrate into a CEP framework. The remainder of this extended abstract describes such an example and proposes the use of vocal synthesis for sonification.

2. WORKING WITH SENTIMENT DATA

2.1. Measuring sentiment

Past work [2] leveraged human sentiment to analyze weather conditions in order to help corroborate data derived from physical sensors such as weather satellites and radar. Current technology exists to analyze unstructured text and assign numerical metrics of sentiment [7], [8]. We relied on the AlchemyAPI web service to algorithmically assign values to short excerpts of text describing current weather conditions. These values were within the range of ± 1, with lower values representing danger or unhappiness, and higher values representing safety or happiness. While many algorithms and services are available to perform such a conversion, it is very difficult to provide an intuitive human interface to this data.

2.2. Sonifying sentiment

We propose that using vocal synthesis to create sounds representing sentiment data can provide a superior capability for comprehension by leveraging the innate human ability to associate vocal sounds with human emotions. Although it is trivial to map the normalized sentiment values (ranging from -1 to +1) to sound attributes such as pitch and volume, there is no intuitive and innate association between sentiment and pitch (for example). By synthesizing vocal sounds, we hope to leverage millions of years of evolution that have resulted in the human association of certain vowel transitions with corresponding feelings or emotions. This is related to the human capacity for "everyday listening" that is often cited as an advantage in the creation of effective auditory displays [9].

3. DEMONSTRATION

As a demonstration and assessment tool, we have assembled a simulation consisting of several datasets each comprised of one hundred weather-related sentiment values with a range of -1 (for extremely negative sentiment) to +1 (for extremely positive sentiment). This simulation data is derived from the tornado observation data used in [2], where textual data was converted to numerical sentiment assessments via AlchemyAPI [7].

For each of these datasets, the CEP process is used to generate a summary dataset by detecting the following patterns: Positive Isolated Spike, Positive Long-term Surge, Positive-to-Negative Sentiment Reversal, Negative-to-Positive Sentiment Reversal, Negative Isolated Spike, and Negative Long-term Surge. Details on this summarization process are available in [10].

Sonification of each dataset is accomplished by first playing a representation of the raw dataset, which is represented by sine wave grains; and then playing a representation of the CEP-derived summary of that data, which is represented via vocal synthesis. The sine wave grains representing the raw dataset are interpolated over a one-octave range with the low extreme mapping to a -1 sentiment value and the high extreme mapping to a sentiment of +1. The grains are played over a ten second period, resulting in hearing approximately 10 grains per second. This playback method allows trends in the data to readily be detected, but requires a relatively extended listening period for large datasets.

The summary data, which is represented via synthesis of sung vowel sounds, provides similar ability to recognize trends but requires a much shorter listening period. In general, negative summaries were represented as an "OH" vowel sound with high pitch, vibrato rate, and amplitude. Positive summaries were represented as an "OO" vowel with lower pitch, vibrato rate, and amplitude. Reversals patterns are represented as smooth transitions between these sounds. It should be noted that these mappings of sentiment values to sound attributes are only a starting point, and future human subject experiments are likely to reveal even more effective usage of innate response (e.g. the use of "growl" sounds to portend danger).

Each dataset of 100 sentiment values was typically reduced to three or four summary values via the CEP process, and each sung vowel representing these values requires approximately a half-second for playback. For larger datasets, multi-tier CEP processing could result in an even greater reduction in time required.

In future work, auditory and visual displays will be synchronized to allow selective playback of all permutations of data type and display type (see Figure 1). This will facilitate future human subjects experiments to further quantify the efficacy of our approach. Additional experimentation could compare vocal synthesis with a more conventional sonification approach mapping sentiment values to pitch or volume.

Data Type / Display Type	Sensor Data	Sentiment Data	Both
Auditory Display	☑	☑	☑
Visual Display	☑	☑	☑
Both	☑	☑	☑

Figure 1: Experimental grid showing data type / display type permutations.

4. CONCLUSION

Sentiment analysis is an evolving technique that is becoming mandatory in any domain that leverages human observations that are recorded in an unstructured or semi-structured manner. It is vital that the sonification community develops methods to integrate such data into auditory displays in a manner that is compatible with innate human modes of interpretation and understanding. Complex Event Processing facilitates this by reducing massive amounts of data into smaller numbers of relevant summary events that are well suited to representation via vocal synthesis. We hope that the application of vocal synthesis presented in this paper will provide useful guidance in that endeavor.

5. ACKNOWLEDGMENT

We gratefully acknowledge that this research activity has been supported in part by a Multidisciplinary University Research Initiative (MURI) grant (Number W911NF-09-1-0392) for "Unified Research on Network-based Hard/Soft Information Fusion", issued by the US Army Research Office (ARO) under the program management of Dr. John Lavery.

6. REFERENCES

[1] J. Rimland, M. Ballora, and D. Hall, "Hard and soft information fusion in sonification for assistive mobile device technology." In Proceedings of the 14th International Conference on Auditory Display, pp. 19–24, 2013

[2] J. Rimland, "Hybrid Human-Computing Distributed Sense-making: Extending the SOA paradigm for dynamic adjudication and optimization of human and computer

roles.", A Dissertation in Information Sceinces and Technology, The Pennsylvania State University, 2013.

[3] G. Cugola and A. Margara. 2012. "Processing flows of information: From data stream to complex event processing." ACM Comput. Surv. 44, 3, Article 15 (June 2012), 62 pages. DOI=10.1145/2187671.2187677 http://doi.acm.org/10.1145/2187671.2187677

[4] The term "wedding frenzy" is credited to Julie Coughlin, Penn State University.

[5] J. Rimland, M. McNeese, and D. Hall, "Conserving analyst attention units: Use of multi-agent software and CEP methods to assist information analysis." *Proceedings of SPIE*, 2013.

[6] J. Rimland, M. Ballora, and W. Shumaker, "Beyond visualization of big data: a multi-stage data exploration approach using visualization, sonification, and storification," In Proceedings of SPIE, 2013.

[7] http://alchemyapi.com

[8] http://semantria.com

[9] T. Hermann, *Sonification for Exploratory Data Analysis*. Unpublished doctoral dissertation, Bielefeld University, Germany, 2002.

[10] J. Rimland and M. Ballora, "Using Complex Event Processing (CEP) and vocal synthesis techniques to improve comprehension of sonified human-centric data." *Proceedings of SPIE*, 2014.