Hidden in a breath: Tracing the breathing patterns of survivors of traumatic events

Introduction

Many people experience a traumatic event during their lifetime. In some extraordinary situations, such as natural disasters, war, massacres, terrorism or mass migration, the traumatic event is shared by a community and the effects go beyond those directly affected. Today, thanks to recorded interviews and testimonials, many archives and collections exist that are open to researchers of trauma studies, holocaust studies, historians among others. These archives act as vital testimonials for oral history, politics and human rights. As such, they are usually either transcribed, or meticulously indexed. In this project, we look at the nonverbal signals emitted by victims of various traumatic events and seek to render these for novel representations that are capable of representing the trauma without the explicit (and often highly politicized) content. In particular, we propose to detect breathing and silence patterns during the speeches of trauma patients for visualization and sonification. We are hoping to glean into cultural and contextual differences of bodily expression of trauma through automatic processing of thousands of hours of testimonials from all over the world.

Trauma Diagnosis

DSM-IV defines a traumatic event as an event that is "generally outside the range of usual human experience" and would "evoke significant symptoms of distress in almost everyone." In the diagnosis of Posttraumatic Stress Disorder (PTSD) and acute stress disorder (ASD), DSM symptoms are essential. If the duration of symptoms is less than 3 months, it is defined as Acute Stress Disorder, otherwise, it is defined as Post Traumatic Stress Disorder. There are scarcely any studies that detect PTSD and acute stress disorder without clinical data. Some of these studies include information about heart rate, pulse and breathing patterns of individuals (Davis et al, 1996; Shalev et al, 1998; Ogden and Minton, 2000). Although vocal parameters are used in order to detect clinical depression, no studies investigating the relationship between voice and trauma were found. Some recent work investigates PTSD symptom severity in a multimodal fashion, using questionnaires and skin conductance physiology (Mallol-Ragolta et al. 2018). In this work, we focus on the relationship between trauma, speech and breathing. It is important to note that we do not work with PTSD, as not all trauma survivors have PTSD.

General Approach

Our general approach is to use speech features to automatically segment trauma survivor testimonials into speaking, breathing, and silence classes. The duration of the silences, the frequency of breathing and its quality, the whole dynamics of the testimonial are investigated

through unsupervised learning and visualization. Statistical testing will be used to measure the effect of cultural and contextual factors.

Speech and Breath Features

The most commonly used speech processing techniques in the recognition of emotions and clinical depression in the literature are related to prosody (i.e. pitch, jitter, energy, pause time and speaking rate), as well as the spectral features (i.e. formants) and cepstral features (i.e. Mel frequency cepstral coefficients). Prosodic, source, and acoustic features, as well as vocal tract dynamics are speech-related features affected by depression. Researchers have found that depressed subjects are prone to possess a low dynamic range of the fundamental frequency, a slow speaking rate, a slightly shorter speaking duration, and a relatively monotone delivery (Low et al., 2010; Cummins et al., 2011; Williamson et al., 2013; Mundt et al., 2007; Alghowinem 2013).

Breathing consists of two phases called *inspiration* and *expiration*. During inspiration, the diaphragm is used to increase the volume of the chest cavity, causing the air to enter by mouth or nose to fill the low-pressure lungs. During a normal expiration, the diaphragm and external muscles are relaxed, the chest volume is lowered, the pressure increases, and breath is dispelled. During speech activity, some of the air stored in the lungs are spent to produce sounds. Subsequently, long bouts of speaking also require taking a breath.

Breathing is intimately connected to oxygen flow, and thus is involved in all internal and external systems, including circulation, hormone systems, and the nervous system. Deep breathing in general seems to allow emotional release and processing of the energy stuck in the body from the trauma. Additionally, Ogden and Minton (2000) observed that subjects can have breathing difficulty when they are talking and extemporizing about their trauma.

Many features can be extracted from the speech signal for the automatic detection of breathing, including MFCC parameters, short-time energy, zero crossing rate, spectral slope, and duration. However, this is a difficult problem to tackle "in-the-wild". The testimonials we are planning to use in this work have many sources of noise to confound the automatic algorithms, including background music, other persons talking (and breathing), and external noise sources. Sometimes the level of the sound is far from ideal, and the performance of the breath detection will not come close to the ideal situation where the speaker is isolated in a lab environment with high quality recording equipment.

Dataset & Annotations

In order to test the quality of various collections, as well as if the language, culture, and the traumatic events' nature and date have a measurable effect on breathing patterns, we have collected short clips from survivors' testimonials. We sampled different languages (English, Chinese, Japanese, Spanish, Kinyarwanda), and different mass-traumas (Holocaust, Nanjing Massacre, Tsunami, Guatamalan Genocide, Tutsi Massacre). In the first phase of the project, 20 clips from 10 survivors were manually annotated for "speaking", "silence", "breathing", "lip noise",

"other people speaking" classes. For each survivor, we selected a 'normal' speech and a more 'traumatic' speech segment. For some survivors, the date of the events were decades ago (i.e. Nanjing Massacre and Holocaust), for some, the memories were fresh (i.e. Tsunami). Each speech segment is annotated to extract the time span of speech, silence, and breathing periods.

Analysis

In Figure 1, each line represents one survivor's speech segments, where light blue circles are speech, purple circles are breathing, and green circles are silences. The horizontal dimension shows the time (the longest segment is one and a half minutes), and the circle radii are proportional to duration. On the left side, the normal speech segments are aligned, whereas the ones on the right side are emotionally charged. Typical features in the latter include long silences, pierced by deep breath -especially before telling about the most traumatic event-, sometimes frequent and sharp breathing. As expected, even in such a small sample size, certain characteristics prevail: for instance, personal speech patterns are different, and set the tempo of the speech as well as breathing, but within the personal tempo, deep breathing emerges. Some of these patterns are not heard while listening to the video/audio recordings themselves, but become visible only after the annotated portions of breathing patterns are visualized. Cultural approaches to trauma and disasters might dictate the way events are described, but this small sample suggests the possibility that the breathing and silence patterns that occur while telling a traumatic event are shared across cultures. A much larger scale empirical investigation will follow.
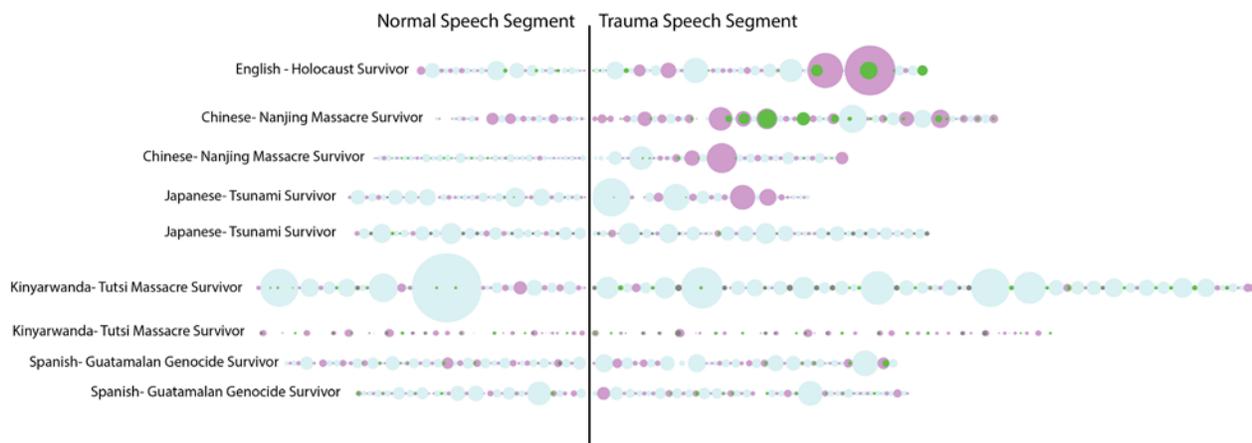

Figure 1: Visualization of Breath (purple), Silence (Green), Speech (Light Blue) Patterns

References

Alghowinem, S., "From joyous to clinically depressed: Mood detection using multimodal analysis of a person's appearance and speech", Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on, pp. 648– 654, IEEE, 2013.

Cahill, S. P. and K. Pontoski, "Post-traumatic stress disorder and acute stress disorder I: their nature and assessment considerations", Psychiatry (Edgmont), Vol. 2, No. 4, p. 14, 2005.

Cummins, N., J. Epps, M. Breakspear and R. Goecke, "An investigation of depressed speech detection: Features and normalization", Twelfth Annual Conference of the International Speech Communication Association, 2011.

Davis, J. M., H. E. Adams, M. Uddo, J. J. Vasterling and P. B. Sutker, "Physiological arousal and attention in veterans with posttraumatic stress disorder", Journal of Psychopathology and Behavioral Assessment, Vol. 18, No. 1, pp. 1–20, Mar 1996.

Low, L.-S. A., N. C. Maddage, M. Lech, L. Sheeber and N. Allen, "Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents", Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pp. 5154–5157, IEEE, 2010.

Mallol-Ragolta, A., S. Dhamija, and T.E. Boult, "A Multimodal Approach for Predicting Changes in PTSD Symptom Severity," Proceedings of the 2018 on International Conference on Multimodal Interaction (pp. 324-333). ACM, 2018.

Mundt, J. C., P. J. Snyder, M. S. Cannizzaro, K. Chappie and D. S. Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology", Journal of neurolinguistics, Vol. 20, No. 1, pp. 50–64, 2007.

Ogden, P. and K. Minton, "Sensorimotor psychotherapy: One method for processing traumatic memory.", Traumatology, Vol. 6, No. 3, p. 149, 2000.

Shalev, A. Y., T. Sahar, S. Freedman, T. Peri, N. Glick, D. Brandes, S. P. Orr and R. K. Pitman, "A prospective study of heart rate response following trauma and the subsequent development of posttraumatic stress disorder", Archives of general psychiatry, Vol. 55, No. 6, pp. 553–559, 1998.

Williamson, J. R., T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination", Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, pp. 41–48, ACM, 2013.