Non-verbal Signals in Oral History Archives

Francisca Pessanha
f.pessanha@uu.nl
Utrecht University
The Netherlands

ABSTRACT

Oral History Archives (OHA) are a rich source of emotional narratives, encapsulating the personal stories of people across different demographics, historical periods, and cultures. Computational technologies have transformed the oral history archival field by facilitating the transcription and verbal content analysis of interview collections where manual inspection is too time-consuming. However, these methods fail to include the subjective part of the archives. In this project, we explore the potential of automatic breathing patterns and non-verbal cues analysis applied to OHA interviews to gain new insights into the individual and collective emotional responses across different demographics. The proposed framework will investigate if automatic breathing signal prediction enhances the performance of speech emotion recognition models and if a cross-dataset learning approach for breathing signal prediction and paralinguistics analysis will work in OHA. Next, we will further use the emotional information gathered to study cultural differences when it comes to narrating traumatic experiences, focusing on different OHA collections. Lastly, to enhance our research and the literature, we will also design emotion elicitation experiments to create new emotional speech breathing datasets.

CCS CONCEPTS

• **Computing methodologies** → *Machine learning approaches*; *Feature selection.*

KEYWORDS

Affective Computing, Oral History, Paralinguistics, Breathing Analysis, Interpretability, Data Collection

ACM Reference Format:
Francisca Pessanha. 2022. Non-verbal Signals in Oral History Archives. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22), November 7–11, 2022, Bengaluru, India.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3536221.3557036

1 INTRODUCTION

Computational technologies have revolutionized the archival sciences field, prompting new approaches to process the extensive data in these collections. For example, automatic speech recognition (ASR) and natural language processing (NLP) create unique possibilities for analyzing oral history interviews where otherwise, the transcription and analysis of the full recording would be too time-consuming. However, many oral historians note the loss of aural information when converting the speech into text, pointing out the relevance of subjective non-verbal cues to understand the interviewee's narrative fully.

Hence, the ability to look at human emotions and somatic reactions during these interviews, both on an individual and on a collective level, would give scholars from many fields the means to focus on the feelings, mood, culture, and subjective experiences on a mass scale. Here, computational approaches are imperative for large-scale data analysis. By combining oral history with data sciences, I hope to provide meaningful tools to study emotional interviews and open new sources and challenges for data/AI scientists, as well as researchers from biomedical fields, such as psychiatry. In my PhD thesis, I will focus on audio features since this is common to both audio recordings (the most frequent way of interview collection) and video recordings. Under this domain, I will explore acoustic features and breathing signals during speech.

Although we do not use breathing to communicate explicitly, it signals affective cues, as it is an anatomical action that changes with different emotions [28]. Consequently, narrating painful events can cause a potential change in breathing patterns since, in a sense, remembering a traumatic event may lead to "re-experiencing" the old associations again [9]. This idea was previously explored by Akdag Salah et al., showing how breathing patterns change in survivor testimonies, and illustrated how these mark emotional moments during recordings [3]. However, unlike speech and paralinguistic speech analysis, for which extensive literature and automatic tools have been introduced, automated breathing patterns analysis in nonverbal communication is a relatively new research line, particularly challenging due to the lack of ground truth of breathing signals in most speech datasets and real-world data, like oral history collections. During this project, I want to tackle the challenges and opportunities of this new research field towards answering the research question "Can we use computational methods to gain new insights into the emotional response after stressful events across various cultures, geographies, and timespans by analyzing breathing patterns and non-verbal cues of thousands?". I divided this goal into three main challenges, defined below. For simplicity, paralinguistics and breathing signal features are defined as "PBS features".

- *Can automatic breathing signal prediction enhance the performance of speech emotion recognition models?* - In this project, we aim to use breathing features in combination with paralinguistics to improve the state-of-art for speech emotion recognition. At the time of writing, there is no published work on the automatic assessment of affect states using breathing signal features; therefore, our initial goal is to test this hypothesis.

ICMI '22, November 7–11, 2022, Bengaluru, India

ACM ISBN 978-1-4503-9390-4/22/11…$15.00
https://doi.org/10.1145/3536221.3557036

- *Is cross-dataset learning robust for PBS prediction in mismatched conditions?* - Cross-dataset prediction will be essential to make affect predictions in extensive collections of unannotated data, such as OHA. However, differences in audio characteristics such as background noise, lower recording quality or microphone position will constitute challenges when working with OHA data. Further, OHA are local recordings of culture, i.e. the language used and its dialect is often not well studied in the fields of ASR and paralinguistics. This phenomenon will be particularly challenging when performing cross-dataset learning across languages of different language families (for example, English to Arabic).
- *How can we use the information gathered to contribute to the study of OHA in the humanities field?* - After assessing the relevance of PBS features for emotion assessment and implementing robust cross-dataset learning approaches, we plan to design tools to assist the analysis of sets of interviews and provide new incites over extensive collections of data. Examples of possible tools are: assessing the emotional flow of records, identifying key emotional points and comparing reactions to stressful events across cultures. As stated by Michael Frisch, "The "deep dark secret" of oral history, is that "nobody spends much time listening to or watching recorded and collected interview documents" [11] so making archival collections more engaging and accessible for both scholars and general public would be a great contribution for the field.

While OHA collections include a variety of topics, from everyday life narrations to war stories, we intend to focus on instances when a traumatic experience is narrated. Hence I argue that these interviews can provide extensive datasets for sentiment studies and the psychiatric field in general. Further, although my research will focus on applications in OHA, the research questions proposed are relevant for breathing research across multiple disciplines: if looking into respiration as a continuous sentiment feature, automatic breathing signal prediction and analysis can be used, for example, to enhance already existent sentiment recognition approaches, modulate sentiment in artificial agents and as feedback for human-computer interactions. In this paper, I will first explore the different application areas, and scientific contributions of social signal processing applied to oral history archival collection. Then, I will summarize relevant literature for automatic paralinguistic and breathing analysis from audio, with a focus on OH. In Section 3, the research methodology is introduced, followed by the contributions until the date. Lastly, I will reflect on the challenges found and define future work.

2 BACKGROUND AND RELATED WORK

2.1 Research opportunities found in OHA

2.1.1 Humanities field. Oral history embraced the subjective side of the interviewing process, using the act of remembering not only as a tool to understand what happened but as an object that needs to be studied on its own, i.e. how what happened is remembered [33, 34]. For the oral historians who emphasize the subjective side of memory and storytelling not as shortcomings but as the strong points of oral history studies, the nonverbal cues of the interviewees, as well as the dialogue between the interviewee and the interviewer contain important information that needs to be included in the archive and should be analyzed further to complete the research. Therefore, the information is not solely on the narrative, but also in the breaks and gaps of it, gestures, facial expressions, periods of silence and other non-verbal interactions [17, 25].

OHA interviews related to stressful events may contain sensitive data and, in the case of testimonies from highly politicized environment settings, such as the Rwanda and Bosnia genocide [15], the informants can face life-threatening situations due to their narrations. Being able to access the archival material as an enriched transcription with paralinguistic details about the emotional state of the interviewee but with removed identifiable information, would contribute enormously to the humanities/social research. Moreover, we suggest using OHA to study the effects of large scale traumatic events in different demographics and extend the archives with non-transcribable information such as mood, gestures and breathing. Computational methods could provide a more structured analysis of large collections of data and identify patterns not seen when manually analysing the interviews.

2.1.2 Psychiatry field. Psychiatric disorder classification is inherently complex since mental health depends not only on biological functions but also on how those functions are related to individual environmental and experiential challenges [14]. Arguably psychiatry is the medical speciality most sensitive to cultural influences. In this field, cultural differences can influence the assessment of symptoms and the expression of the disorder itself. For instance, in the case of distress, it is important to include region-specific syndromes and the idea of cultural concepts of distress in diagnosis questionnaires, like suggested by DSM-5, the Diagnostic and Statistical Manual of Mental Disorders [4]. However, it is impossible to manually include all world's cultural concepts of diseases into a single document, showcasing a limitation of this method. This issue is particularly important now with the globalization of clinical research when medication is tested in different populations to access the efficiency of psychiatric treatment. Differences in symptoms expression, evaluation and interpretation can have an impact on the outcome of the study and should be taken into account.

From a computational psychiatry point-of-view, Oral History Archives offer us a large dataset of interviews of people exposed to traumatic experiences. Although these circumstances are not enough for a diagnosis, studies point to a correlation between psychopatologies and traumatic events such as natural disasters [5, 24]. Additionally, interviews offer an extensive collection of emotions across different cultures and languages. Core emotions have been repeatedly validated in emotion-recognition studies across cultures showing that there are some universal, biologically determined, emotions for which core expression and identification are similar for all human beings. On this note, we suggest using OHA to go a step further and try to identify if other biological signals can be used to assist diagnosis across cultures [12].

2.2 Social Signal Processing

2.2.1 Paralinguistics. Paralinguistic features commonly used in emotion recognition tasks, such as emotional stability evaluation and personality prediction [23]. For example, voices from subjects with PTSD or depression were found to have significantly tenser voice quality [29]. Moreover, 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. Computational paralinguistics offers different feature sets and classification frameworks [31]. Several toolboxes exist, such as the openSMILE toolkit [10], or the COVAREP repository [8], which provide feature extraction capabilities for the analysis of paralinguistic features in speech.

At the moment, there are very few existing scientific works on processing OH archives with the tools of paralinguistics. Working on an OH dataset documenting war and violence in Croatia, Truong et al. [7] modeled the verbal and prosodic features of emotional expression in narrative data. Under the hypothesis that the interviewee would be more emotional after open-ended questions, the authors analyzed the correlation between nonverbal emotional expression in the voice (pitch, vocal effort, and pauses) and changes in the intensity of emotional expressions during the interviews. A correlation between the intensity of the verbal expression and the pitch and pause duration was observed, but there was no correlation for vocal effort. Studies like this offer a new way of looking into OH research.

2.2.2 Breathing. Breathing is a constant in a human's life and happens naturally and effortlessly; this process is adjusted continuously to the individual's needs. For example, we coordinate breathing with eating or speaking. Emotions happen with physiological changes within the entire body, a critical instance of which is the changes in breathing. An example of this is the relationship between anxiety and breathing: studies show an increase in the respiratory rate with anticipation anxiety, which is unrelated to a higher demand for oxygen; so the resulting uncomfortable urge to breathe ("air hunger") depends on the affective state of the subject [13].

Analyzing breathing patterns via machine learning approaches in nonverbal communication is a relatively new research line. Most approaches focus on automatic breathing detection, which is especially challenging for uncontrolled recording conditions. Because of the temporal nature of the problem, approaches that model the sound dynamics are preferred, such as Long Short Term Memory (LSTM) models [19–21]. LSTM networks contain memory units to preserve and propagate information over time, rendering them especially useful for audiovisual datasets. Recent studies relate breathing patterns to emotional states and mental health. For example, using breath signal information, Cho et al. [6] proposed a neural network-based classifier to discriminate between levels of stress while performing a task. In addition, in [16], signals from nonverbal parts of the recordings, such as breathing and silences, are combined with linguistic information for automatic depression detection.

3 METHODOLOGY

3.1 Data collection and acquisition

At the time of writing, only two speech datasets with a breathing signal ground truth are available [22, 32]. Both studies measure the breathing signal using two respiratory belts, one over the ribcage and the other over the abdomen. The speech presented in these datasets is either spontaneous speech or vocal exercises, with neither of them exploring the influences of emotionality in the breathing signal. For the present research, we propose the collection of an emotional speech breathing dataset; we believe this project would benefit our research long term and enrich the fields of paralinguistics and affective computing. Here we define two main paths regarding the emotional component on the dataset:

- *Focus on mild emotional elicitation.* For this purpose, we will conduct experiments at science festivals and other public settings. This approach would lead to a dataset of considerable size with ground truths for multiple modalities. The ethical implications of collecting such data are low compared to distress-related interview collection. The produced subsets would likely include children and the elderly, which would be a meaningful contribution to the field of paralinguistics. A drawback of this approach is the lack of control over the diversity of the dataset - gender, age, ethnicity, or personality-wise. Further, emotional elicitation is a complex experience and may not transfer well when applying the findings in OHA data.
- *Focus on distress and traumatic events recollection.* Here we will seek meaningful collaborations with scholars from oral history or other relevant fields already working on collecting distress-related interviews for their line of research. Collecting distress-related interviews raises ethical concerns; thus, collaborating with specialists is fundamental. By centering the data collection on narrations of real-life traumatic experiences, we expect to find a closer connection to OHA topic with the advantage of having ground truths for different modalities such as breathing signals. Further, we hope to complement our collaborators' research with computational insights regarding the non-verbal cues present in the interaction, particularly the variations in breathing patterns. This approach would allow a focus on a particular topic from the start, such as gender-based violence. It would also arguably give me more control over the demographics presented, allowing for a more balanced dataset.

Considering the dimension of this task we decided that it would be prudent to start by projecting a pilot study to prove of concept, allowing us to test different biometric collection material and emotion inducing tasks. Additionally, to evaluate the performance of the proposed models on non-annotated oral history interview collections, we will establish collaborations with existing archives.

3.2 Paralingustics

In [30], Schuller offers an excellent overview for voice and speech analysis. The author divides the analysis into two subgroups: analysis of voice and speech. Voice analysis will deal with the acoustic properties of the sound (how things are said). In contrast, speech analysis will focus on the linguistic content and non-linguistic vocal outbursts.

We aim to do additional research on voice/audio, centered on explainability, to combine the voice and breathing components of the project. In this context, the use of spectrum and cepstrum representations to extract audio-related features motivated me to research more into signal processing and what can be done from a spectral perspective to improve future breathing detection models. In the context of speech analysis, we want to collaborate with natural language processing researchers to create multimodal approaches for affect prediction and analysis. Textual information can be very informative of the emotionality across the interview; thus, combining verbal cues with non-verbal cues would be beneficial to assess affect in OHA interviews.

3.3 Breathing analysis

The lack of a breathing ground truth is a recurrent challenge, particularly when transitioning to the OHA. Cross-dataset learning provides a good solution for this problem; however, since the breathing signal describes the changes in diameter in the thorax/abdomen and not necessarily a sound, it's not intuitive to evaluate its accuracy by, for instance, comparing the audio and signal plots manually. For this reason, we want to propose an approach to assess the accuracy of the breathing prediction based on the audio recording. With this goal in mind, we will use breath speech corpus to define the correlation between the audio and breathing signal and then extrapolate this knowledge to breathing signal predictions in emotional/distress datasets. Additionally, we will study the correlation between annotated breathing events, such as "laughter", "filler", "breathing" and "speech" with the signal prediction.

Moreover, we will research multimodal approaches using linguistic, paralinguistic and breathing features to assess psychopathologies. In this project, we want to understand how breathing during speech relates to emotionality in the text and voice features (for instance, pitch and voice quality). We expect to find variations in the breathing pattern in emotional moments due to, for example, faster breathing or holding the breath.

4 CONTRIBUTIONS TO DATE

To better understand the research opportunities for computational approaches in the field of Oral History, I wrote an extensive survey on the topic [26]. In this survey, I looked into the field through different application lenses and explored computational techniques for breathing and paralinguistic analysis and the challenges/opportunities they find in OHA. After publication, I was invited to present this work during the UCL Multimodal Digital Oral History Seminar Series [1].

I dedicated my first year to learning about paralinguistics and signal processing theory to access the full potential of the archives.In this context, I worked on subprojects on paralinguistics and how to apply these methods for emotion detection. As such, I participated in the ChaLearn 2021 to detect personality traits based on audio-video interviews. Personality trait prediction during dyadic interactions offers meaningful tools to analyze OHA. While working on this project, we noticed the disproportional influence of the metadata features in the prediction task. For this reason, we shifted our focus to propose an informed baseline to help disentangle the various contextual factors of influence in this type of case study. With this work, we received an honorable mention and presentation in ICCV Understanding Social Behavior in Dyadic and Small Group Interactions (DYAD) workshop [27].

Additionally, I conducted an exploratory study focused on the cross-dataset prediction challenge; this work was submitted as a long paper to ICMI 2022 and is currently under review. This study has two distinct parts: (1) cross-dataset breathing signal prediction under mismatched conditions and evaluation of the breathing signal predictions accuracy on the unannotated target dataset and (2) speech breath analysis in a distress corpus and subsequent depression severity assessment. We used the UCL Speech Breath Monitoring (UCL-SBM) corpus, containing continuous spontaneous speech, and is introduced for the Breathing Sub-Challenge of the INTERSPEECH 2020 Computational Paralinguistics Challenge [32] to design the breathing signal prediction model. For depression severity assessment and evaluation of the cross-dataset breathing signal predictions, we used the Distress Analysis Interview Corpus-Wizard-of-Oz dataset (DAIC-WOZ). To validate the breathing signal prediction in the target dataset, we manually annotated audible breathing events in both UCL-SBM and DAIC-WOZ corpus and compared the predicted breath event characteristics between the two sets. We observed a negative correlation between depression and arousal, valence and dominance, pointing to sentiments such as sadness and depression, and correlations for breathing features suggested slow and shallow breaths, consistent with the detected mood. Lastly, the simple regression models proposed for depression assessment surpassed the AVEC2017 Real-life Depression Challenge baseline, performed on the same dataset, illustrating the potential of breathing features for mood disorder analysis. The insights of this exploratory study will be the base for further experiments on the relevance of breathing features for affect state analysis.

In parallel, we have designed experimental scenarios for emotional elicitation and performed a pilot study on the influence of emotional music on breathing patterns. The dataset collected during the referred pilot study will be submitted to the GENEA (Generation and Evaluation of Non-verbal Behaviour for Embodied Agents) Workshop 2022 at ICMI 2022. Research has shown that people recognize emotionality in virtual agents better when specific respiratory patterns occur [18]; hence, the produced speech breath dataset will be a relevant data source for behavior synthesis in this area. At the time of writing, we have two data collection events scheduled during science festivals for the general public. With this and other small datasets, we aim to study the breathing/emotion correlation on unannotated datasets, such as multimedia archives and create tools to explore the emotional content of large sets of interviews.

5 NEXT STEPS

This section will discuss the specific research plans envisioned to answer the proposed research subquestions. The next steps to tackle each subquestion are:

- *Can automatic breathing signal prediction enhance the performance of speech emotion recognition models?* Preliminary results showed the potential of breathing in the context of emotion analysis. In future work, we want to focus on the relationship between breathing and paralinguistics and how their combined use can enhance the performance of already existing speech emotion recognition models. With this goal in mind, we intend to 1) make further improvements on respiratory features extraction based on clinical insights and 2) define interpretable models to combine speech features, such as voice quality and pitch, with the breathing episodes happening simultaneously. The respiratory features proposed depend on accurate peak detection to define the inhalation and exhalation periods; hence, improving this step is critical for more accurate respiratory features. Therefore, we want to enhance the signal preprocessing pipeline to better account for the signal variability observed during speech and adapt this process to the acoustic characteristics of the recording. For breathing feature extraction, we will first focus on frequency-based features, such as continuous wavelet transformation to bring valuable insights into respiratory patterns. Moreover, to tackle the limitations imposed by the DAIC-WOZ dataset, we will expand the research proposed in our exploratory study to suitable emotional speech sources to evaluate further the impact of breathing in affect state analysis tasks.

- *Is cross-dataset learning robust for PBS prediction in Oral History Archives (OHA)?* Collecting additional data with emotional breathing patterns will allow us to evaluate the cross-dataset prediction performance under mismatched conditions on datasets with a breathing ground truth. Although the current cross-dataset prediction evaluation strategy led to positive results, it can only evaluate the breathing prediction accuracy during annotated audible breath events. We are curious to see where the cross-dataset prediction approach fails over the entire breathing signal and how we can better tackle these problems in real-world data collections, such as OHA. Additionally, we will perform case studies of subsets of OHA collections pertinent for our research [2] to identify additional challenges and opportunities.
- *How can we use the information gathered to contribute to the study of OHA in the humanities field?* We intend to process tools to summarize the emotional component of interviews and highlight key points (for example, high emotionality and laughing) to help researchers navigate through a large number of interviews and explore patterns in specific topics, such as reactions to natural disasters . Here, establishing multidisciplinary collaboration with oral history scholars are extremely helpful in assessing the needs of the research field and their research and facilitating the exchange of interview data, often not open to the general public.

REFERENCES

[1] [n.d.]. Multimodal Digital Oral History - Seminar Series. https://multimodaldigitaloralhistory.omeka.net/events

[2] [n.d.]. *Nederlands Veteranen Instituut*. https://www.nlveteraneninstituut.nl

[3] Almila Akdag Salah, Albert Ali Salah, Heysem Kaya, Metehan Doyran, and Evrim Kavcar. 2020. The sound of silence : Breathing analysis for finding traces of trauma and depression in oral history archives. *Scholarship in the Digital Humanities* (2020).

[4] American Psychiatric Association, American Psychiatric Association, et al. 2013. Diagnostic and statistical manual of mental disorders: DSM-5. *Arlington, VA* (2013).

[5] Annetta Burger, Talha Oz, William G Kennedy, and Andrew T Crooks. 2019. Computational social science of disasters: Opportunities and challenges. *Future Internet* 11, 5 (2019), 103.

[6] Youngjun Cho, Nadia Bianchi-Berthouze, and Simon J Julier. 2017. DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 456–463.

[7] Franciska de Jong, Khiet Truong, Gerben Westerhof, Sanne Lamers, and Anneke Sools. 2013. Emotional expression in oral history narratives: Comparing results of automated verbal and nonverbal analyses. (2013).

[8] Gilles Degottex, John Kane, Thomas Drugman, Tuomo Raitio, and Stefan Scherer. 2014. COVAREP—A collaborative voice analysis repository for speech technologies. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 960–964.

[9] Kirsten Ekerholt and Astrid Bergland. 2008. Breathing: a sign of life and a unique area for reflection and action. *Physical therapy* 88, 7 (2008), 832–840.

[10] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*. 1459–1462.

[11] Michael Frisch. 2008. Three dimensions and more: Oral history beyond the paradoxes of method. *Handbook of emergent methods* (2008), 221–238.

[12] Daniel Guinart, John M Kane, and Christoph U Correll. 2019. Is transcultural psychiatry possible? *JAMA* 322, 22 (2019), 2167–2168.

[13] Ikuo Homma and Yuri Masaoka. 2008. Breathing rhythms and emotions. *Experimental Physiology* 93 (9 2008), 1011–1021. Issue 9. https://doi.org/10.1113/EXPPHYSIOL.2008.042424

[14] Quentin JM Huys, Tiago V Maia, and Michael J Frank. 2016. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience* 19, 3 (2016), 404–413.

[15] Erin Jessee. 2019. The limits of oral history: Ethics and methodology amid highly politicized research settings. *The Oral History Review* (2019).

[16] Heysem Kaya, Dmitrii Fedotov, Denis Dresvyanskiy, Metehan Doyran, Danila Mamontov, Maxim Markitantov, Alkim Almila Akdag Salah, Evrim Kavcar, Alexey Karpov, and Albert Ali Salah. 2019. Predicting Depression and Emotions in the Cross-roads of Cultures, Para-linguistics, and Non-linguistics. In *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*. 27–35.

[17] Dominick LaCapra. 2016. Trauma, history, memory, identity: What remains? *History and Theory* 55, 3 (2016), 375–400.

[18] Benny Liebold and Peter Ohler. 2013. Multimodal emotion expressions of virtual agents, mimic and vocal emotion expressions and their effects on emotion recognition. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 405–410.

[19] Maxim Markitantov, Denis Dresvyanskiy, Danila Mamontov, Heysem Kaya, Wolfgang Minker, and Alexey Karpov. 2020. Ensembling end-to-end deep models for computational paralinguistics tasks: ComParE 2020 Mask and Breathing Sub-challenges. *INTERSPEECH, Shanghai, China* (2020).

[20] John Mendonça, Francisco Teixeira, Isabel Trancoso, and Alberto Abad. 2020. Analyzing Breath Signals for the Interspeech 2020 ComParE Challenge. *Proc. Interspeech 2020* (2020), 2077–2081.

[21] Venkata Srikanth Nallanthighal and H Strik. 2019. Deep sensing of breathing signal during conversational speech. (2019).

[22] Venkata Srikanth Nallanthighal and H Strik. 2019. Deep sensing of breathing signal during conversational speech. (2019).

[23] Michael Neff, Nicholas Toothman, Robeson Bowmani, Jean E Fox Tree, and Marilyn A Walker. 2011. Don't scratch! Self-adaptors reflect emotional stability. In *International Workshop on Intelligent Virtual Agents*. Springer, 398–411.

[24] Fran H Norris, Matthew J Friedman, Patricia J Watson, Christopher M Byrne, Eolia Diaz, and Krzysztof Kaniasty. 2002. 60,000 disaster victims speak: Part I. An empirical review of the empirical literature, 1981–2001. *Psychiatry: Interpersonal and biological processes* 65, 3 (2002), 207–239.

[25] Cord Pagenstecher and Stefan Pfänder. 2017. Hidden dialogues: Towards an interactional understanding of oral History interviews. *Oral History Meets Linguistics* (2017), 185–207.

[26] Francisca Pessanha and Almila Akdag Salah. 2021. A Computational Look at Oral History Archives. *Journal on Computing and Cultural Heritage* (2021).

[27] Francisca Pessanha and Gizem Sogancioglu. 2022. Fact sheet: Automatic Self-Reported Personality Recognition Track. https://doi.org/10.48550/ARXIV.2207.11012

[28] Pierre Philippot, Gaëtane Chapelle, and Sylvie Blairy. 2002. Respiratory feedback in the generation of emotion. *Cognition & Emotion* 16, 5 (2002), 605–627.

[29] Stefan Scherer, Giota Stratou, Gale Lucas, Marwa Mahmoud, Jill Boberg, Jonathan Gratch, Louis-Philippe Morency, et al. 2014. Automatic audiovisual behavior descriptors for psychological disorder analysis. *Image and Vision Computing* 32, 10 (2014), 648–658.

[30] Björn Schuller. 2011. Voice and speech analysis in search of states and traits. In *Computer Analysis of Human Behavior*. Springer, 227–253.

[31] Bjorn W Schuller. 2012. The computational paralinguistics challenge [social sciences]. *IEEE Signal Processing Magazine* 29, 4 (2012), 97–101.

[32] Björn W Schuller, Anton Batliner, Christian Bergler, Eva-Maria Messner, Antonia Hamilton, Shahin Amiriparian, Alice Baird, Georgios Rizos, Maximilian Schmitt, Lukas Stappen, et al. 2020. The interspeech 2020 computational paralinguistics challenge: Elderly emotion, breathing & masks. *Proceedings INTERSPEECH. Shanghai, China: ISCA* (2020).

[33] Linda Shopes. 2015. Community oral history: where we have been, where we are going. *Oral History* (2015), 97–106.

[34] Alistair Thomson. 2007. Four paradigm transformations in oral history. *The oral history review* 34, 1 (2007), 49–70.