An annotation scheme for sighs in spontaneous dialogue

Khiet P. Truong[1], *Gerben J. Westerhof*[2], *Franciska de Jong*[1,3], *and Dirk Heylen*[1]

[1]University of Twente, Human Media Interaction Group, Enschede, The Netherlands
[2]University of Twente, Dept. of Psychology, Health and Technology, Enschede, The Netherlands
[3]Erasmus University Rotterdam, Erasmus Studio, Rotterdam, The Netherlands
{k.p.truong, g.j.westerhof, f.m.g.dejong, d.k.j.heylen}@utwente.nl

Abstract

Sighs are non-verbal vocalisations that can carry important information about a speaker's emotional (and psychological) state. Although sighs are commonly associated with negative emotions (e.g. giving up on something, 'a sigh of despair', sadness), sighs can also be associated with positive emotions such as relief. In order to gain a better understanding of sighing as a social and affective signal in dialogue, and to advance towards an automatic classification and interpretation of the emotional content of sighs, it is necessary to learn more about the various phonetic characteristics of sighs. To that end, we developed an annotation scheme for sighs that takes the variation in phonetic form into account. Using this scheme, an oral history corpus containing emotionally-coloured dialogues was annotated for sighs. Results show that sighs can be annotated with a sufficient level of reliability (Cohen's Kappa of 0.713), and that indeed, various types of sighs can be identified as well (Cohen's Kappa between 0.637 and 0.805). Through a preliminary analysis of emotional content words, indications were found that certain types of sighs can be associated with specific emotional contexts.

Index Terms: sigh, non-verbal vocalisation, dialogue, social signal processing, annotation, oral history

1. Introduction

A sigh can be characterised as a non-verbal vocalisation that has physiological (respiratory) and psychological (emotional) functions. From a physiological point of view, sighing can be described as a respiratory phenomenon, consisting of deep inspiration and expiration phases. Usually, these deep inspiration and expiration phases are audible and may vary on certain phonetic features. For example, some sighs may have strong audible inspiration and expiration phases, while others sound more subdued; some sighs may be interspersed with speech while others are not. In other words, there might be various forms of sighs that may or may not serve specific functions.

With respect to the physiological function of sighing, the prevalent view seems to be that sighing functions as a resetter to restore the balance in the respiratory system [1]. Adopting a physical sense of 'relief', Wilhelm et al. [2] theoretize that sighing, i.e., expanding the lungs, may relieve chest tightness and as a result, causes temporary relaxation. Vlemincx et al. [1] argue that it is reasonable to assume that if sighing causes tension relief in a physiological manner, then sighing will also cause psychological comfort. Hence, it is probable to believe that in times of stress, sighing may induce or reduce psychological states such as relaxation or tension respectively.

In fact, "a sigh of relief" is a common expression in English that illustrates one of the possible psychological and emotional functions of sighing: the relief of tension. Paradoxically, sighing cannot only be associated with positive emotional states such as relief [1], it is also often associated with negative emotional states such as panic [3], depression [4], and sadness [5]. In general, it is known that emotions are linked to respiratory patterns [6] but it is still largely unknown how exactly emotions affect and/or cause certain sighing behaviour.

Linking non-verbal behaviour in dialogue and in particular non-verbal vocalisations, such as laughter and filled pauses, to emotional, social, and psychological states in an automated way has been one of the key challenges in social signal processing (SSP) and affective computing for the past few years, e.g. [7, 8]. In contrast to previous sigh studies [5, 1, 9, 3, 2], the focus in SSP and the current study is on sighing in spontaneous dialogue with a narrative structure, such as interviews, where subjects are talking to another interaction partner and where sighing may serve (interpersonal) emotional and/or social functions. Although sighing can be seen as a salient non-verbal vocal marker of emotional states, attempts to automatically detect and interpret sighs in dialogue have been scarce. To the best of our knowledge, the automatic classification of sighs in spontaneous dialogue has only been attempted in one study by Gupta et al. [10]. If we can increase our knowledge and understanding of how certain sighing behaviour and possible forms of sighs can be linked to specific emotional or psychological states in dialogue, we can advance the automatic detection and interpretation of sighs and its emotional contents. Consequently, as a first step, one needs to learn more about what phonetic features characterise sighs and what variation there exists in their possible phonetic forms.

To that end, we introduce an annotation scheme for sighs that takes these phonetic features and possible forms into account. The scheme was developed iteratively and applied to a corpus consisting of emotionally-coloured interviews with war trauma victims. This type of data was considered suited for the purpose of this study because of a) the fact that the narratives in this type of dialogue are emotionally coloured and b) the availability of the collection for research purposes. We discuss how we developed the scheme, how reliable the scheme is, and we present a preliminary analysis in which we explore emotional word usage in the vicinity of various forms of sighs.

The paper is structured as follows. We review previous studies on sighs in Section 2 and describe the corpus used in Section 3. The annotation scheme and a preliminary analysis on sighs are presented in Section 4 and 5 respectively. Finally, a conclusion and discussion are given in Section 6.

Figure 1: *Two still images from the CroMe corpus.*

2. Related work

In this section, we will introduce the relevant concepts and identify the gaps in existing coding schemes (Section 2.1), and provide a brief description of the available body of literature on the emotional and psychological states that have been associated with sighs (Section 2.2).

2.1. Annotations and definitions

Although there are some definitions of sighs given in previous studies, almost all of these definitions were not specifically developed for the use of annotation with the exception of the study by Gupta et al. [10]. In their study, sighs were annotated with the goal to automatically classify them. The annotation protocol used was relatively simple and based on audible information only: "a deep intake and release of breath that was audible in the audio channel." No information about the reliability of this annotation protocol was provided in their paper.

Sighs have been mentioned in other annotation protocols of large publicly available speech corpora as well. In a study by Trouvain & Truong [11], it was found that sighs are relatively often part of various annotation protocols for speech corpora, but that in general, the number of annotated sighs found in the corpora studied are very low. Given that breath sounds *are* frequently annotated in these corpora, a possible explanation for the low number of annotated sighs could be that breath sounds and sighs are being mixed up with each other due to the lack of clear and distinctive descriptions of the two events. In fact, when we study the annotation protocols of the corpora used in [11], we indeed find that breath sounds (and even inbreaths and outbreaths) are annotated but that additionally, sighs are also annotated as a separate category [12, 13, 14]. Clear and distinctive descriptions of the two events are lacking in these annotation protocols. Another possible explanation could be that the annotators were instructed to focus on speech events rather than non-speech events.

In case when coding instructions for sighs *are* available for other purposes than classification, sighs have been described as "a single breath characterized by a relatively long duration, deep in- and expiration, and a particular sound ..." [6]. A more elaborate description of sighs can be found in Teigen et al. [5] who coded "...clear (genuine) sighs if they consisted of clearly audible deep in- and outbreaths, or as uncertain (doubtful) sighs if they were weak, incomplete, voiced or in other ways departing from an unequivocal sigh." These descriptions leave some room for the observer's own subjective perception of a sigh, but more objective assessments of sighs through physiological measurements are also possible.

Based on physiological measurements, sighs have been defined as "breaths with an inspiratory volume of at least two times the mean inspiratory volume..." [1]. Abelson et al. [9] defines a sigh as "any breath that was at least 500 mL larger than the mean of the prior three breaths and at least 400 mL larger than the following breath...". Although these type of definitions are useful for an objective assessment of sighing, our focus here is to assess sighing from the visual and/or acoustic signal recorded in a natural, social setting.

In conclusion, the main characteristics of a sigh are that it has clearly audible deep inbreaths and outbreaths. Although all of the studies described include these characteristics in their sigh definitions, none of these studies have actually evaluated the reliability of such a definition for annotation purposes or paid attention to the phonetic variation in sighs.

2.2. Relation to emotional and psychological states

Although there have been some studies on the relation between sighing and certain emotional and psychological states, the amount of research remains relatively scarce and much of the exact relation is still unknown. In general, according to a study by Teigen [5], people associate sighing with negative, low-intensity and deactivated emotional states. In addition, it was found that during a puzzle task, people often sigh because they gave up and felt helpless. According to Teigen [5], a sigh mainly indicates a mismatch between, e.g., ideals and realities, and it indicates acceptance, i.e., 'letting go' and 'giving up'.

Sighing has been related to other negative psychological states as well such as depression. In a study by Robbins et al. [4], rheumatoid arthritis patients wore a voice recorder for whole weekends long that recorded the patients' voices. Sighs were coded in the recordings and it was found that the number of sighs per hour was strongly positively correlated to the patients' levels of depression.

Evidence that sighing can also be associated with positive feelings comes from a study by Vlemincx et al. [1] in which the expression 'sigh of relief' was put to the test. In their experiments, two conditions were created: one in which stress was induced by exposure to a loud noise stressor and one in which relief was induced by ending the stressor. Sighs were measured and determined in a physiological manner. The results showed that people sighed more often in the relief condition than in the stress condition.

In summary, sighs are non-verbal cues that can be important markers for a person's emotional and psychological state. However, much of the variance of the relation between sighing and emotional state remains unexplained. One of the steps we take in this paper in order to gain a better understanding of sighing in an emotional context, is identifying various types of sighs and comparing the emotional word usage in the vicinity of these sighs. For that purpose, we used a corpus containing emotionally-coloured dialogues: the CroMe corpus.

3. Material

3.1. The CroMe corpus

We used audiovisual recordings of interviews made in the Croatian Memories project (CroMe) [15], see Fig. 1 for some screen shots of the recordings. The goal of this oral history project was to collect personal testimonies on war and trauma in Croatia and to make these testimonies accessible. Each testimony consists of an interviewer and interviewee and has an interview-like structure determined by a semi-structured questionnaire. Approximately 400 video-recorded interviews are available. Partly for subtitling purposes, all interviews have been transcribed in Croatian and most of them have been translated to English as well. As the interviewees were interviewed about their expe-

SIGH1	**Clearly visible and/or audible inhalation and exhalation (through the nose or mouth) that are part of the same breathing cycle.** • The exhalation can be co-produced with voiced sounds. • The breathing cycle with clearly visible and/or audible inhalation and exhalation can be intermitted by speech sounds.
SIGH2	**Clearly visible and/or audible exhalation (through the nose or mouth).** • There is no clear inhalation visible and/or audible. • The exhalation can be somewhat articulated: some lip movement may be involved during the turbulent airflow of the exhalation which results in unvoiced /f/, /p/ or /h/-like sounds. • The exhalation can be co-produced with voiced sounds.

Table 1: *Annotation guidelines for sighs in dialogues.*

riences during war time, the tone of these interviews is both personal and emotional.

3.2. Data selected for analysis

From the set of interviews available, we selected 12 interviews (6 females and 6 males) for sigh annotation and analysis. The interviewees have a mean age of 62.3 (sd = 13.9) within a range of 45–92 years. The mean duration of each interview is 54.6 minutes (sd = 15.9 minutes). We manually segmented all the interviewer and interviewee speaking segments (since their speech was recorded on 1 audiochannel). Finally, we also used the time-aligned English subtitles for our analysis involving emotional word usage, see Section 5.

4. Development of annotation guidelines

A consensus on annotation guidelines for sighs in dialogues was reached after three phases of (iterative) annotation and discussion activities. Two annotators (LP and VV) performed the annotations and refined the guidelines after several iterations and discussions.

4.1. Main annotation guidelines

In the first phase, the annotators were asked to come up with a first description of sighs by listening and watching the audiovisual material. This phase functioned as a first exploration with the data and sighing behaviour for the annotators. The annotators first explored the data separately from each other, and subsequently met with each other to discuss discrepancies and to reach consensus over the annotation guidelines. This discussion resulted in a first version of annotation guidelines that already had the main criteria formulated as shown in Table 1: a sigh should have a clearly visible and/or audible inhalation and/or exhalation. A distinction was made between sighs that have clear inhalations and exhalations (this includes the stereotypical sigh), and sighs that only have clear exhalations. In this phase, a subset of 4 interviews (subset A) was used.

In the second phase, the annotators applied the annotation guidelines formulated in the previous phase to another subset of 4 interviews (subset B). The confusion matrix of both annotators is shown in Table 2a. Because we did not ask the annotators to label 'no sighs', the classification of 'no sighs' (the cell [0,0]) was calculated by considering the alternating pattern of 'sigh' and 'no sigh' which yielded a calculation of `number of sighs + 1` for each interview. Based on Table 2a, an agreement score of Cohen's $\kappa = .358$ was derived. Subsequently, both annotators compared and discussed their annotations with the aim to reach consensus on the discrepancies between them. An improved agreement score of Cohen's $\kappa = .524$ was achieved. As a result of this consensus discussion, several guidelines were added and refined. These additional guidelines mainly specified how speech sounds needed to be taken into account in annotating sighs, see Table 1.

		VV		
		0	1	2
LP	0	81	14	3
	1	15	21	11
	2	9	0	4

(a) *Second phase, set B*

		VV		
		0	1	2
LP	0	115	10	3
	1	7	68	3
	2	12	0	8

(b) *Third phase, set C*

Table 2: *Annotation results of LP and VV*

In the third and final phase, the annotation guidelines from the second phase as shown in Table 1 were applied by both annotators on a third separate subset of 4 interviews (subset C). The confusion matrix obtained is shown in Table 2b and yields an agreement score of Cohen's $\kappa = .713$. In a subsequent consensus meeting, both annotators met again to discuss their discrepancies and tried to reach consensus on the instances that they did not agree upon. This resulted in a small improvement of Cohen's $\kappa = .781$ which confirmed our suspicion that the annotation guidelines are of sufficient quality and that a ceiling was reached. Hence, we considered the guidelines presented in Table 1 as final and asked the annotators to re-annotate subset A in order to increase the amount of sigh data. Given the sufficient level of agreement obtained with the current guidelines, the annotators each annotated different interviews in subset A.

In total, we found 185 sighs in the 12 interviews studied, see Table 3. For subset B+C, only the sighs that were agreed upon by the two annotators, the 'consensus sighs' are included. In the remainder of this paper, we only consider these 'consensus sighs' and discard the sighs from subset A.

subset	SIGH1	SIGH2
A	38	25
B	34	4
C	74	10
All	146	39

Table 3: *Final number of sighs (= 185) found in 12 interviews. Sighs from subset B+C were obtained after reaching consensus.*

4.2. Additional guidelines: various types of sighs

Although the main annotation guidelines included phonetic aspects, the annotators were not asked to label these phonetic aspects explicitly. In the second task, we asked two new annotators, an expert annotator and a naive annotator, to carry out a sub-annotation on the sighs from set B and C (the 'consensus' sighs) that was more focused on labelling various types of sighs based on phonetic features. The sub-annotation guidelines are shown in Table 4.

SIGH1	1A	A 'typical' sigh: clear inhalation and exhalation of one breath cycle with no intermitted speech sounds. (21)
	1B	A sigh intermitted and/or interspersed with speech (or other voiced) sounds. A possible pattern could be inhalation some speech (or other voiced) sounds exhalation. Or the exhalation is interspersed with speech (or other voiced) sounds. (87)
	1C	Other. (0)
SIGH2	2A	An exhalation with some articulation resulting in sounds like 'pfff', 'puh' etc. (3)
	2B	The exhalation is interspersed with speech (or other voiced) sounds. (4)
	2C	The exhalation is through the nose. (4)
	2D	Other. (3)

Table 4: *Sub-annotation focused on form: types of sighs. The number in brackets indicate the number of sighs observed in that category.*

The agreement between the annotators yielded $\kappa = .805$ for the subcodes SIGH1X and $\kappa = .637$ for SIGH2X. The lower Kappa achieved for the subcodes SIGH2X confirmed our suspicions that SIGH2 consisted of the 'less clear' cases of sighs that are more susceptible to discussion. In subsequent analyses of types of sighs, we used the expert annotations as ground truth.

5. Analysis

A first exploration into the use of sighs as an emotional cue was carried out. We looked at the emotional context in which sighs are produced. In particular, we looked at word usage in the vicinity of sighs: can we say that sighs are produced in emotionally-laden contexts, and does type of sigh matter? For each sigh, 20s-long word segments from the English subtitles (10 s preceding and 10 s following the sigh) were extracted. 20s-Segments that do not contain a sigh were also extracted. The word count program Linguistic Inquiry and Word Count developed for measuring the emotional state as expressed in personal narratives (LIWC [16]) was used to calculate the percentages of words belonging to 'positive emotion' and 'negative emotion' categories for each segment of words (these percentages were normalised by speaker using z-scores). Subsequently, we made comparisons between the normalised LIWC scores of segments containing a sigh and segments not containing a sigh. Additionally, the subcodes were taken into account to see whether the type of sigh makes a difference in the emotionally-laden context in which it occurs.

Although statistical tests do not show significant differences among the LIWC values found for the different types of sighs (which can be partly due to the relatively small sample size), we do observe some interesting trends in Fig. 2 that are worthwhile to mention and to follow-up with future research. We observe that for most of the sighs, the amount of positivity decreases while the amount of negativity increases (SIGH, SIGH1, SIGH1A, SIGH1B) in comparison to when there are no sighs present (NO SIGH). This trend is in concordance with what we would expect given the setting of the interviews (i.e., war trauma). However, we did expect the amount of negativity to be more pronounced as in the case of SIGH1A. It is striking that the 'typical sigh' (SIGH1A) seems to show more pronounced positivity and negativity as expected, while the group of 'less clear' sighs (SIGH2) shows opposite behaviours. The ambiguity of sighs concerning their positive and negative connotations is also reflected in the trends observed in Fig. 2. In summary, these preliminary results warrant a more detailed investigation into the emotional values behind various types of sighs.

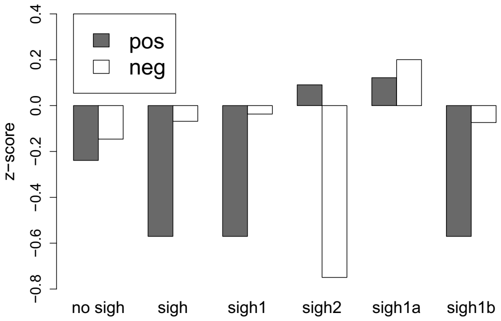

Figure 2: *Median LIWC values (z-scores) for sighs in general (SIGH), for different types of sighs (SIGH1, SIGH2, SIGH1A, SIGH1B) and for segments that do not contain sighs (NO SIGH). The other types of sighs are not shown here due to their small sample size.*

6. Conclusion and discussion

In order to gain a better understanding of what phonetic features characterise sighs and how sighs correlate with emotional word usage, we have annotated and identified various types of sighs in emotionally-coloured interviews. Through several phases, an annotation scheme for sighs was developed and assessed that resulted in acceptable reliability scores. A preliminary analysis of emotional word usage showed that there are signs that not all sighs are produced in the same emotional contexts. This observation should be investigated in more detail in future research through, for example, a manual annotation of the emotional value or emotional context of each sigh. For future work, we also recommend to investigate the feasibility of automatic detection of sighs since many sighs sound rather subdued and confusions with other breathy sounds are easily made. With the annotation scheme proposed, the identification of phonetic variation in sighs and our preliminary results on emotional word usage in the vicinity of sighs, we hope to have laid a firm basis for future research into spontaneous sighing behaviour in dialogues within the context of social signal processing.

7. Acknowledgements

We would like to thank Vivian Veldkamp and Loes Pepers for their annotation efforts, discussions, and help with this paper. Daphne Karreman and Jered Vroon are also thanked for their annotation efforts. The project Croatian Memories (CroMe) was funded by the Netherlands Ministry of Foreign Affairs.

8. References

[1] E. Vlemincx, I. Van Diest, S. De Peuter, J. Bresseleers, K. Bogaerts, S. Fannes, W. Li, and O. Van den Bergh, "Why do you sigh? Sigh rate during induced stress and relief," *Psychophysiology*, vol. 46, pp. 1005–1013, 2009.

[2] F. H. Wilhelm, R. Gevirtz, and W. T. Roth, "Respiratory dysregulation in anxiety, functional cardiac, and pain disorders: Assessment, phenomenology, and treatment," *Behavior Modification*, vol. 25, pp. 513–545, 2001.

[3] F. H. Wilhelm, W. Trabert, and W. T. Roth, "Physiologic instability in panic disorder and generalized anxiety disorder," *Biological Psychiatry*, vol. 49, pp. 596–605, 2001.

[4] M. L. Robbins, M. R. Mehl, S. E. Holleran, and S. Kasle, "Naturalistically observed sighing and depression in rheumatoid arthritis patients: A preliminary study," *Health Psychology*, vol. 30, pp. 129–133, 2011.

[5] K. H. Teigen, "Is a sigh just a "sigh"? Sighs as emotional signals and responses to a difficult task," *Scandinavian journal of Psychology*, vol. 49, pp. 49–57, 2008.

[6] F. A. Boiten, N. H. Frijda, and C. J. E. Wientjes, "Emotions and respiratory patterns: review and critical analysis," *International Journal of Psychophysiology*, vol. 17, pp. 103–128, 1994.

[7] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. R. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 Computational Paralinguistics challenge: social signals, conflict, emotion, autism," in *Proceedings of Interspeech*, 2013, pp. 148–152.

[8] T. F. Krikke and K. P. Truong, "Detection of nonverbal vocalisations using Gaussian mixture models: looking for fillers and laughter in conversational speech," in *Proceedings of Interspeech*, 2013, pp. 163–167.

[9] J. L. Abelson, J. G. Weg, R. M. Nesse, and G. C. Curtis, "Persistent respiratory irregularity in patients with panic disorder," *Biological Psychiatry*, vol. 49, pp. 588–595, 2001.

[10] R. Gupta, C.-C. Lee, and S. Naryananan, "Classification of emotional content of sighs in dyadic human interactions," in *Proceedings of ICASSP*, 2012, pp. 2265–2268.

[11] J. Trouvain and K. P. Truong, "Comparing non-verbal vocalisations in conversational speech corpora," in *Proceedings of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, 2012, pp. 36–39.

[12] A. Janin, D. Baron, D. Edwards, D. Ellis, D. Gelbart, and N. Morgan, "The ICSI meeting corpus," in *Proceedings of ICASSP*, 2003, pp. 364–367.

[13] J. Carletta, "Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus," *Language Resources and Evaluation*, vol. 41, pp. 181–190, 2007.

[14] A. H. Anderson, M. Bader, E. Gurman Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weintert, "The HCRC Map Task Corpus," *Language and Speech*, vol. 34, pp. 351–366, 1991.

[15] F. M. G. de Jong, A. J. van Hessen, T. Petrovic, and S. I. Scagliola, "Croatian memories: speech, meaning and emotions in a collection of interviews on experiences of war and trauma," in *Proceedings of LREC 2014*, to appear.

[16] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "Linguistic Inquiry and Word Count: LIWC," Austin, TX, 2007, [Computer software].