Falling silent, lost for words ...
Tracing personal involvement in interviews with Dutch war veterans

Henk van den Heuvel, Nelleke Oostdijk
CLS/CLST, Radboud University
Erasmusplein 1
Nijmegen, the Netherlands
E-mail: {H.vandenHeuvel|N.Oostdijk}@let.ru.nl

Abstract

In sources used in oral history research (such as interviews with eye witnesses), passages where the degree of personal emotional involvement is found to be high can be of particular interest, as these may give insight into how historical events were experienced, and what moral dilemmas and psychological or religious struggles were encountered. In a pilot study involving a large corpus of interview recordings with Dutch war veterans, we have investigated if it is possible to develop a method for automatically identifying those passages where the degree of personal emotional involvement is high. The method is based on the automatic detection of exceptionally large silences and filled pause segments (using Automatic Speech Recognition), and cues taken from specific n-grams. The first results appear to be encouraging enough for further elaboration of the method.
Keywords: discourse annotation, emotion recognition, speech databases, automatic speech recognition, oral history

1. Introduction

Eye witness reports that relate first hand observations and personal experiences of events happening do not only include factual statements about 'who did what to whom' or 'what happened when', but also the witness' opinion or attitude towards what he or she has witnessed or experienced. Witnesses in their accounts of one and the same event will differ in a variety ways, for example in their choice of words or the attention they give to certain details by elaborating upon them. Combining what information is contained in multiple reports, can help the reconstruction of what (may have) happened and how witnesses say what they think of it. Where a witness report has been captured and kept as an audio recording, there is important additional information that can be learned from the way that things were said, that are not conveyed through the wording, but by making use of various paralinguistic means (prosody, pitch, volume, intonation). Thus, in spontaneous speech, silences and filled pauses (like 'eh') are important cues to a speaker's emotional state and involvement (Tisljar-Szabó & Pléh, 2014).

In sources used in oral history research (such as interviews with eye witnesses), passages where apparently speakers personally are very much emotionally involved are of particular interest, as these may give insight into how historical events were experienced, and what moral dilemmas and psychological or religious struggles were involved (Van den Berg et al., 2010). So far, however, studies in oral history have mostly been based on written versions (verbatim transcriptions), rather than on the primary audio data, as exploration of the audio is not trivial.

In the present paper we report on a pilot study we conducted. Using a large corpus of interview recordings with Dutch veterans (IPNV; Van den Heuvel et al., 2012) we investigated if it is possible to develop a method for automatically identifying passages where the degree of personal emotional involvement is high.[1] The data originally was collected for the purpose of historical documentation and research. As such the present study exemplifies the re-use of data that was made accessible to the research community after curation (Van den Heuvel, et al. 2012, 2015).

The remainder of this paper is structured as follows. Section 2 describes the method. It also includes details about the data and tools that we used for our pilot study. In Section 3 we present our first results and discuss our findings in some detail. Section 4 concludes this paper with a summary of the research presented and an assessment of its value as a first pilot.

2. Method

The method is directed at mining the audio and transcriptions for cues leading to passages in the recording where apparently the speaker is personally (emotionally) involved. As we know from previous research that silences and filled pauses are important cues, we base our approach on the automatic detection of exceptionally large silences and filled pause segments (using Automatic Speech Recognition (ASR)). In addition, we investigate what cues may be taken from specific n-grams.

In this section we first introduce our data and tools, before we go on to explain how ASR was used to detect silences and filled pauses and what possible role there is for n-grams.

2.1 Data and tools

For our pilot study, we selected 9 interviews from the IPNV collection of about 1,000 interviews with Dutch war veterans.[2] About 250 of these interviews are made

[1] As passages of personal involvement we consider those fragments where a speaker describes a situation or an event with a heavy emotional impact on his/her state of mind.

[2] The metadata for this collection are accessible through

audible and searchable through a web interface called OHAT (Oral History Annotation Tool) as described in Van den Heuvel et al. (2012).[3] These 250 interviews were disclosed by Automatic Speech Recognition (ASR) for keyword spotting purposes in OHAT. The ASR output contains the duration of the identified words, filled pauses and silences. For this study we concentrated on silence and filled pause durations which we shall refer to collectively as *pause durations*. By including both silences and filled pauses, various pausing strategies that speakers employ are taken into account (Van Donzel & Van Beinum, 1996).

The selection of the interviews contained the overlap of interviews for which the ASR output was available and the interviews for which a complete hand-crafted transcription was available. The guidelines for these transcriptions are in Van den Berg et al. (2010). They include markers for silences and filled pauses (such as 'eh'). The full transcriptions allowed us to have a close look at the precise linguistic content associated with the pauses, which we considered essential at this stage of our study.

Thus we limited our material to the 9 interviews labeled in the database with tape numbers 208, 230, 348, 374, 488, 490, 508, 519, 549.

2.2 Locating silences and filled pauses

For each interview, the ASR-output file was used to compute the durations of the silences and filled pauses. These durations were displayed as percentages for each consecutive window of 30 seconds and transformed into z-scores. Next, we sorted the 30-second fragments for each interview by their z-scores. Table 1 gives some statistics on the pause durations for the selected interviews.

Interview	Average overall in interview	Max. pause duration (%)	Corresp. max. z-score
208	44.22	68.53	2.53
230	42.46	94.93	5.14
348	29.47	84.53	5.94
374	37.20	82.40	5.04
488	43.29	65.77	2.93
490	41.02	61.00	2.44
508	50.86	88.63	4.62
519	33.23	59.70	2.75
549	45.54	86.17	4.44

Table 1: Pause statistics over 9 interviews over 30-second fragments

For the interviews with the largest pauses, we inspected the audio, the transcriptions and the general context by a lookup in OHAT. As a rule of thumb, we inspected per interview the fragments for which the z-score is higher than 3.

We annotated in which of these fragments there were passages that were characterized by a high degree of personal involvement. Moreover, we inspected whether such passages were characterized by specific linguistic cues.

2.3 N-grams

N-grams were considered as a possible means to help identify relevant passages. As we expected to see personal involvement reflected in the extent to which attention is given to careful phrasing, we derived an n-gram list on the basis of the transcriptions that were available for the nine interviews, excluding the contributions of the interviewers. We specifically were interested in n-grams pointing towards recurrent patterns involving filled pauses, particularly as research in the field of discourse analysis has shown that speakers typically pause strategically.

The n-gram list was inspected, while distinguishing between unigrams on the one hand, and multigrams on the other hand:

Unigrams: single words

The 20 most frequent words used by the respondents in the 9 interviews were *eh* ('eh'), *en* ('and'), *die* ('that'; demonstrative and relative pronoun), *dat* ('that'; demonstrative and relative pronoun or subordinating conjunction), *de* ('the'), *ik* ('I'), *een* ('a'), *ja* ('yes'), *was* ('was'), *h*et ('it'), *je* ('you'), *in* ('in'), *van* ('of'), *dan* ('then', 'than'), *we* ('we'), *daa*r ('there'), *maar* ('but'), *niet* ('not'), and *toen* ('then').[4]

Multigrams: two or more consecutive words

Among the 20 most frequent multigrams used by the respondents, there are three bigrams in which the hestitation marker *eh* ('eh') appears. In descending order of frequency these are *en eh* ('and eh'), *die eh* ('that eh'), and *dat eh* ('that eh').

These findings are consistent with the observation that in spoken discourse speakers often pause after having made a start on the next clause, taking time to collect their thoughts and to plan what to say next.[5] Within a phrase, speakers are known to pause before important words. None of the multigrams in our list led to the identification of any such words.[6] This was presumably due to data sparseness, as on inspection of the transcriptions we did find occurrences such as *Ze hebben wel wat mensen natuurlijk eh … neergelegd* ('Of course they eh put some people down').

https://easy.dans.knaw.nl/ui/?wicket:bookmarkablePage=:nl.knaw.dans.easy.web.search.pages.PublicSearchResultPage&q=IPNV

[3] The tool is accessible via http://wwwlands2.let.ru.nl/spex/annotationtool/ and as demo version via http://wwwlands2.let.ru.nl/spex/annotationtooldemo//

[4] The words are listed in descending order of frequency.

[5] Frequently, upon continuation words uttered before the pause are repeated (*daar was een eh een* 'there was a eh a').

[6] We also looked at skipgrams. This did not lead to any tangible results.

3. Results and discussion

As pointed out in Section 2.2 we inspected passages with pause durations with a z-score higher than 3. Table 2 below shows, per interview, how many of the pause fragments with z>3 displayed personal involvement.

Interview	#Fragments z-score >3	#Fragments z-score>3 & showing personal involvement
208	0	0
230	3	3
348	2	2
374	2	0
488	0	0
490	0	0
508	2	2
519	0	0
549	1	0

Table 2: Pause statistics over fragments in 9 interviews with a z-score >3 displaying personal involvement

We observed that for interviews 230, 348, 488 and 508 (which is 4 out of 9) the top passages indeed displayed genuine personal involvement. For interviews 230, 348 and 508 this holds for all passages with z>3. Remarkably these are the interviews with the largest max z-score in Table 1. Below are three striking examples from interviews 230, 348 and 508.

Context for the passage taken from Interview 230:
Upon returning home from a mission, the veteran is confronted with questions by his family and fiancee about his having shot some enemies. The relatives simply do not understand how this must have been. There is a very long silence due to heavy emotions.

Resp:	*nou het, ja, dus he m'n broers, zussen, ouders en nou eh laat ze nou weer eventjes alleen (lacht) god ja**……..** Sorry*
EN:	'well the, yes, so … my brother, sisters, parents and well eh leave them just a moment alone (laughs) god yes … Sorry'

Fig. 1. Excerpt from Interview 230 (01:22:30 to 01:23:00; z=5.14). Resp=respondent; EN=English translation

Context for the passage taken from Interview 348:
During the night of 5-6 April 1944 there was a riot of Russian captives against the Germans on the Dutch island Texel. The veteran saw that 20 German boys were shot in a line-up. This event evokes a long silence.

Resp:	*En eh.. op op 100 meter afstand dan schoten die Georgiers met hun mitrailleur op die eh op die rij jongens, ik heb eh... ik heb t voor m'n ogen gezien, die lui sprongen soms 2 meter hoog waren erbij, ja getroffen. Op 1 na dood, alles dood. Ja.*
EN:	'And eh. On on 100 meters distance then those Georgains with their machine gun shot that eh that row of boys, I have eh … I have seen it before my eyes, those guys sometimes jumped 2 metres high were there, yes hit. All but 1 dead, all dead. Yes.'
Int:	*Dat tekent uw leven.*
EN:	'that has marked your life'
Resp:	*ja natuurlijk. [even stil] tuurlijk. [Even stil]*
EN:	'yes of course. [silent for a while] of course [silent for a while]'

Fig. 2. Excerpt from Interview 348 (00:08:30 to 00:09:00; z=3.54). Resp=respondent; Int=interviewer; EN=English translation

Context fragment from Interview 508:
This veteran served in a submarine unit during WW II. Soldiers like him were hardly honoured for their services as opposed to people from the resistance. Reflecting on this perceived injustice and the loss of his friends who served with him, he turns emotional.

Resp:	*weet u dat klinkt allemaal zo theatraal maar, ik had teveel vriendjes verloren. **…….** neem me niet kwalijk eh.. dit ophalen is niet goed voor mij maar goed.*
EN:	'you know that all sounds so theatrical but, I had lost too many friends. ……. pardon me … eh. This memory is not good for me but anyway.'

Fig. 3. Excerpt from Interview 508 (2:49:00 to 2:49:30; z=4.62). Resp=respondent; EN=English translation

When we look for linguistic cues in these passages, we find that there are many false starts and repetitions, indications that the speaker is struggling, trying to keep control over his emotions and attempting to find the right words. Strikingly, in two passages, as the respondent is overcome with emotions, we find him apologizing (*Sorry, neem me niet kwalijk,* 'sorry', 'excuse me').

In two of the selected interviews (208, 519), the longest pause passages displayed possible personal involvement. In Table 1 their max. z-scores are 2.53 and 2.75. In 208 the veteran is telling about a patrol with a low voice indicating suspense. Between the lines one senses a deep personal involvement, but this is not obvious at all. Interestingly, there is one passage of personal involvement in 208 which is cited often in Van den Berg et al. (2010) and which appears in our sorted list at position 4 with a z-score of 2.

For three of the interviews (374, 490, 549), the longest pause passages did not display personal involvement. In 374 the silence originated from the interviewer when she was looking up her next question.

Our method to automatically detect in interviews passages that reflect personal involvement by measuring pause durations and analyzing linguistic cues shows some encouraging first results. Especially the fact that the interviews displaying the highest outliers (z-scores) for pause durations appear to entail high personal impact is reassuring. However, the method can be improved in a number of ways.

As we noted above, long pauses may originate from

the interviewer. It seems worthwhile to concentrate exclusively on the respondent's contributions, or at least to distinguish between contributions by the respondents on the one hand, and contributions by the interviewer on the other.. Note that the latter can be very relevant for evoking utterances with high personal involvement from the respondent, e.g. by employing long silences and back channels (like 'eh eh').

In this pilot we measured the durations of the filled pauses that could be identified on the basis of manual transcriptions of the interviews. However, these transcriptions are not always accurate. We may therefore benefit from using the ASR output proper to measure the filled pause durations.

Except for durations of pauses and silences also their numbers can be included. The method may be improved by weighing durations and counts of paralinguistic and linguistic cues. In addition, in future work we will also investigate how other verbal indicators (such as 'tsja', 'ja', i.e. 'well', 'yes') or non-verbal markers (such as speech tempo and pitch range) relate to and may help to finetune our method.

As a final note, we are aware that the method as developed so far aims at high *precision* detection of passages in interviews showing high personal involvement. For optimizing the *recall* of the retrieval of such passages one would need to identify first all passages with genuine personal impact and evaluate which of these appear in the high regions of our sorted lists. For this we would need to annotate a number of representative interviews, preferably by multiple annotators, regarding passages showing emotional involvement so as to obtain a golden standard. This annotation should also include a grading in the judgment of personal involvement, for instance on a 5-point scale.

4. Conclusion

While transcriptions provide a means to search for specific keywords, thus revealing the factual content of a story, important information as to subjective dimensions is contained in audio and (para)linguistic cues. Disclosing these dimensions, too, is important in oral history research in order to truly appreciate what occurred. In a pilot study involving a large corpus of interview recordings with Dutch war veterans passages, we have investigated if it is possible to develop a method for automatically identifying where the degree of personal emotional involvement is high. The method is based on the automatic detection of exceptionally large silences and filled pause segments (using Automatic Speech Recognition), in combination with cues taken from specific n-grams. Our first results suggest that for interviews with large outliers in terms of pause durations, fragments containing passages demonstrating high personal involvement emerge in the top of our rankings, whereas for interviews with less outspoken pause lengths the results are diffuse. Despite the fact that our study in its present setup is targeted at precision rather than at recall, it should be stressed that a quick way of reliably finding at least some passages reflecting personal involvement is in its own right a serious step forward in oral history research. Strategies to improve our findings are proposed as objects of further study.

5. References

Tisljar-Szabó, E., Pléh. C. (2014). Ascribing emotions depending on pause length in native and foreign language speech. *Speech Communication* Vol. 56, 35-48.

Van den Berg, H., Boeije, H., Scagliola, S., Wester, F., Woelders, S. (2010). Over het ontsluiten en gebruiken van kwalitatieve onderzoeksdata. In: Van den Berg, H., Scagliola, S., Wester, F. (eds.). (2010). *Wat veteranen vertellen. Verschillende perspectieven op biografische interviews over ervaringen tijdens militaire operaties*, pp. 297-312. Pallas Publications.

Van den Heuvel, H., Sanders, E., Rutten, R., Scagliola, S. (2012). An oral history annotation tool for INTER-VIEWs. In: *Proceedings of the Seventh International Conference of Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey.

Van den Heuvel, H., Oostdijk, N., Sanders, E., De Lint, V. (2015). Data curations by the Dutch Data Curation Service. Overview and future perspective. In: Odijk, J. (2015): *Selected Papers from the CLARIN 2014 Conference, October 24-25, 2014, Soesterberg, The Netherlands*. Linköping Electronic Conference Proceedings, 116, pp. 54-62. ISBN: 978-91-7685-954-4.
http://www.ep.liu.se/ecp/116/005/ecp15116005.pdf

Van Donzel, M., Koopmans-van Beinum, F. (1996). Pausing strategies in discourse in Dutch. In: *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP 96) October 3-6, Wyndham, Pennsylvania (1996)*. Retrieved from http://www.asel.udel.edu/icslp/cdrom/vol2/505/a505.pdf.