INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Prosodic Cues and Answer Type Detection for the Deception Sub-Challenge

Claude Montacié[1], *Marie-José Caraty*[1-2]

[1] STIH Laboratory, Paris Sorbonne University, 28 rue Serpente, 75006, Paris, France
[2] Paris Descartes University, 45 rue des Saints-Pères, 75006, Paris, France
Claude.Montacie@paris-sorbonne.fr, Marie-Jose.Caraty@ParisDescartes.fr

Abstract

Deception is a deliberate act to deceive interlocutor by transmitting a message containing false or misleading information. Detection of deception consists in the search for reliable differences between liars and truth-tellers. In this paper, we used the Deceptive Speech Database (DSD) provided for the Deception sub-challenge. DSD consists of deceptive and non-deceptive answers to a set of unknown questions. We have investigated linguistic cues: prosodic cues (pauses and phone duration, speech segmentation) and answer types (e.g., opinion, self-report, offense denial). These cues were automatically detected using the CMU-Sphinx toolkit for speech recognition (acoustic-phonetic decoding, isolated word recognition and keyword spotting). Two kinds of prosodic features were computed from the speech transcriptions (phoneme, silent pause, filled pause, and breathing): the usual speech rate measures and the audio feature based on the multi-resolution paradigm. The answer type features were introduced. A set of answer types was chosen from the transcription of the Training set and each answer type was modeled by a bag-of-words. Experiments have shown improvements of 13.0% and 3.8% on the Development and Test sets respectively, compared to the official baseline Unweighted Average Recall.

Index Terms: deception detection, linguistic cues, speech recognition, computational paralinguistics

1. Introduction

Deception is an integral part of human communication through spoken or written language. Many researchers have attempted to define the deception [1]. The definitions differ according to the presence of the three following statements: the objective falsity of the proposition, the senders's believe in this falsity and the intention of the sender to deceive the receiver [1].

Deceptive speech is deliberately elaborated and uttered by a speaker in the purpose to deceive an interlocutor. The message transmitted is generally designed in real-time causing cognitive load to the liar [2] and [3]. Moreover, inhibiting the truth is a more consuming cognitive task than telling the truth [4] and [5]. Nevertheless, an experimented liar can pre-build his message. In related works, it was observed that this cognitive load has consequences on the linguistic content [6] and often on the paralinguistic content [7]. Deception is a complex phenomenon for which many other cues were investigated in a large amount of research across psychophysiological (polygraph [8] and [9]), neurological (event related potential [10]) and behavioral (facial expressions [11], body gesture [12]) dimensions.

The detection of deception is of great interest for various human organizations such as forensic investigation for which the search for truth from witnesses and interrogations is crucial. Many interview strategies have been defined to take advantage of the different cognitive processes of truth-tellers and liars [13] such as unanticipated questions. In this context, trained and experienced interviewers are able to detect deception with an Unweighted Average Recall (UAR) of around 70% [14] and [15].

After the controversial Voice Stress Analyser [16] and [17], the majority of studies on the automatic detection of deceptive speech have been conducted over the past decade. Data-driven approaches were used for all this research. Corpora of deceptive and non-deceptive speech have been built, such as the Columbia/SRI/Colorado (CSC) corpus [18]. Four kinds of cues were used: three computed from the speech signal (prosodic [19], [18], and [20], spectral [21] and non-linear features [22]), and the fourth coming from the analysis of the lexical content [21], [23], and [24].

For the Interspeech'2016 Computational Paralinguistics Deception Sub-Challenge [25], we paid particular attention to linguistic cues that should impact classification performance according to related works on deception cues [11]. In particular, we investigated the prosodic cues [7] and the lexical content [26] well-known to be sensible cues to deception. The paper is organized as follows. The speech corpus of the Deceptive Sub-Challenge is described in the next section. The characteristics of two systems are given in Section 3: the Official Baseline System (OBS) and the Extended Baseline System (EBS) taking into account the unbalancing of deceptive and non-deceptive speech and the two kinds of speaker (guilty and non-guilty). The prosodic and temporal cues are studied in Section 4. First, the automatic segmentation (phonemes, silent and filled pauses) of the audio files is described. Then, two methods for the measurements of prosodic and temporal cues are presented: (1) usual prosodic measures using speech rates and latency response times, (2) audio features based on a multi-resolution paradigm. In Section 5, the computation of audio features related to the lexical content is described. This method is based on keyword-spotters. In Section 6, experiments and results on the Test set are presented. The last section concludes the study.

2. Speech material

The Deceptive Speech Database (DSD) created at the University of Arizona [25] is used for the Deceptive sub-challenge. The DSD consists of the audio recordings of student participants randomly assigned to two roles: guilty (G) and non-guilty (NG). Preliminary condition before recording: guilty participants were asked to steal an exam key with a false

http://dx.doi.org/10.21437/Interspeech.2016-33

identity at the department's office, non-guilty participants were asked to retrieve a leaflet from the same office under their own identity. Structured interviews were conducted by an Embodied Conversational Agent (ECA) with each participant. The interviews consisted of a fixed set of short-answers and open-ended questions. Questions are unknown.

For each audio file of the Training (Train) and Development (Devel) sets, four kinds of metadata are provided: (1) the identification, gender and ethnicity of the speaker, (2) the label (Deceptive (D) vs Non-Deceptive (ND)). There are no metadata available on the Test set.

Table 1. *Statistics on the speech material.*

Corpora	Train	Devel	Test
# of speakers (female, male) (G, NG)	26 (11, 15) (14, 12)	23 (11, 12) (10, 13)	? ? ?
# of audio files (G, NG) (D, ND)	572 (308, 264) (182, 390)	487 (220, 267) (130, 357)	497 ? ?
Interaction dur. (s) Average: min-max	7.6: 0.4-220	6.3: 1.1-236	6.5: 0.1-220
Speech dur. (s) Average: min-max	5.1: 0.2-214	4.0: 0.1-227	3.8: 0.1-211
# of phonemes Average: min-max	32.2: 1-1480	24.5: 1-1330	23.9: 1-1268

Table 1 gives some characteristics and statistics of DSD database on the Train, Devel and Test sets. G and NG audio files are well-balanced but D and ND audio files are significantly unbalanced. G and NG audio files are uttered by G and NG participants respectively. The duration of the audio files is very variable (from 0.1 s to 236 s). The last two lines of the table have been obtained by an automatic speech transcription (cf. §4.1). The statistics have been computed on the time interval during which the participant speaks.

3. Baseline and Extended systems

Audio feature set [27] allows the representation of the audio files in terms of spectral, cepstral, prosodic and voice quality information. The Official Baseline System (OBS) [25] is as follows. The Standard Feature set (Std) was extracted from the complete audio file using the open source openSmile [27] and the configuration file ComParE. Standard feature sets (6,373 features) have been provided for the Train, Devel and Test sets. Support Vector Machines (SVM) classifier with linear Kernel and Sequential Minimal Optimization (SMO) [28] was used for the D/ND prediction. The SVM complexity parameter was chosen to 10^{-4}. To account for the imbalanced class distribution of the DSD database, the D class was up-sampled by a factor of 100%. The Baseline performances in terms of Unweighted Average Recall (UAR) are 61.9% on the Devel set and 68.3% on the Test set.

3.1. Filtering of the training set

The NG-participants always tell the truth while the G-participants lie and tell the truth. On the assumption that the audio files of the G-participants are more useful for searching a decision surface between the D and ND audio files, experiments were carried out on the effect of the filtering of the NG audio files.

Table 2 gives for different percentages of NG-filtering the accuracy of the Deception classifier (D-classifier) on the Devel set in UAR.

Table 2. *UAR on the Devel set of the D-classifier for different percentages of NG-filtering.*

Filtering percentage	0%	33%	66%	100%
# of audio files	572	484	418	308
UAR	61.9%	61.9%	63.8%	**66.3**%

An improvement (+4.4%) is obtained by removing all the NG-audio files compared to OBS. The corresponding training set is made of the 308 audio files of the Train set. This set is used as Train set for the next experiments and the corresponding D-classifier is called Extended Baseline System (EBS).

4. Prosodic and temporal features

Previous studies [7], [11], and [20] have shown that many prosodic measures are cues of deception. We looked for an improvement of the D-classifier using prosodic and temporal features. Two methods have been used for the measurements of prosodic cues: the first one consists in usual prosodic measures such as duration of silent and filled pauses, the second one uses the multi-resolution paradigm on the computation of the audio features. These two methods rely on an automatic speech transcription.

4.1. Automatic speech transcription

The transcription of the corpus was obtained by an acoustic-phonetic decoding system using a <phone | pause> loop search. The ASR system was based on the version 0.8 of the Pocketsphinx recognizer library [29]. The acoustic models were the pre-trained generic US-English acoustic models provided by CMU [29].

Table 3. *List of the unit transcriptions of the ASR.*

Phonemes
ɑ æ ʌ ɔ j aɪ b tʃ d ð ɛə ɜɹ eɪ f g h ɪ i dʒ k l m n ŋ oʊ ɔɪ p r s ʃ t θ ʊ u v w y z ʒ
Silent and filler pauses
SIL BREATH NOISE COUGH SMACK UH UM

Table 3 gives the list of the acoustic models which were used (phonemes, silent and filler pauses). The list of the phonemes is given in the International Phonetic Alphabet.

Speech boundaries were computed from the automatic transcription. Related to the interactions with the ECA agent, three time intervals were distinguished (cf. Figure 1):

- Speech Onset latency (SO): the time interval between the beginning of the interaction and the first phone (excluding the filled pauses) of the verbal interaction.
- Verbal Interaction (VI): the time interval during which the participant answers to the question.
- Interaction Ending latency (IE): the time interval between the last phone (excluding the filled pauses) of the verbal interaction and the end of the interaction.

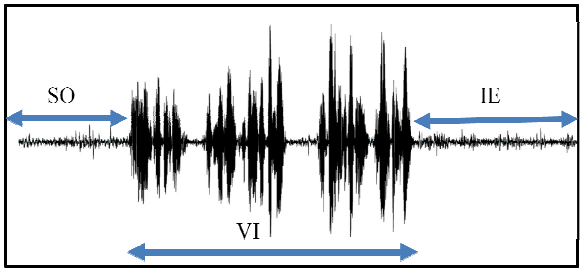

Figure 1: *SO, VI and IE segments of an audio file.*

4.2. Prosodic and temporal measures

In related work to the prosody assessment [20] and [30], various features have been investigated: speech rate measures, global interval proportions and pairwise variability index. We chose twelve features:

- the duration of the three time intervals previously defined (SO, VI and IE) and the duration of the audio file
- eight features computed from the transcription of the VI audio segment. These later features are the following: occurrence number of phones and pauses, duration of the phones and pauses, speaking and articulation rates, and pause and filled pause ratios.

The relevance of all these features is computed by the information gain [31] which was computed on the Train set with the following formula:

$$H(class) - H(class/feature) \quad (1)$$

where Shannon entropy H is estimated from a table of contingency and the class ={D, ND}. Features for which the information gain is greater than zero are considered to be relevant.

On the Train set, four features out of 12 were relevant and selected for the D-classifier in the following ranking order: (1) the duration of IE, (2) the duration of the audio file, (3) the articulation rate and (4) the occurrence number of phones.

The IE duration has been investigated. Figure 1 shows for the two classes N and ND the IE duration histogram computed for each point *t* of the time abscissa as the percentage of utterances having an IE duration in the interval [t, t + 0.3 s]; a spline curve was drawn from the values of the histogram.

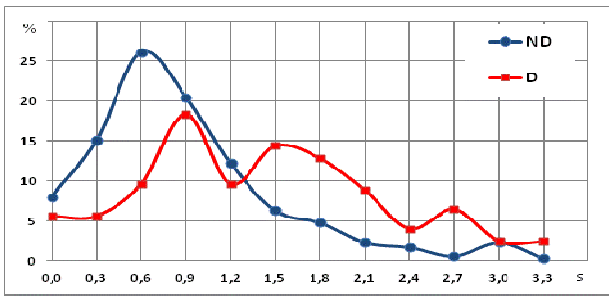

Figure 2: *IE duration histogram for the D and ND-classes of the audio files of the Devel set.*

We remark that the IE duration distributions are very different for the D and ND classes. The average of the IE duration is equal to 1.07 s for the ND-class and 1.68 s for the D-class. Moreover, the ND-distribution is unimodal (maximum around 0.9 s) while the D-distribution is multimodal: a first mode around 0.9 s, a second mode around 1.6 s and a third one around 2.7 s. This difference between the two distributions supports the hypothesis that deception brings cognitive load.

To confirm these results, two audio feature sets have been assessed: the first one (T1) is made up of 6,385 features (12 prosodic features + 6,373 Std features), the second one (T2) of 6,377 features (4 relevant prosodic features + 6,373 Std features). Table 4 gives for the T1 and T2 audio feature sets the accuracy of the D-classifier on the Devel set in UAR.

Table 4. *UAR on the Devel set of the D-classifier using prosodic features*

Systems	EBS	T1	T2
UAR	66.3%	66.2%	**66.7**%

An improvement (+0.4%) is obtained by the prosodic features compared to EBS.

4.3. Multi-resolution-based audio features

To take into account the influence of the audio file segmentation on the computation of the audio feature set, we have investigated four segmentations: [SO∪VI∪IE], [SO∪VI], [VI], and [VI∪IE]. The Standard Feature set (Std) was computed from the first segmentation. The S1, S2 and S3 Multi-resolution-based (M-based) audio feature sets were computed respectively from the other three segmentations. All audio feature sets were extracted using the configuration file ComParE [25]. Table 5 gives for the different M-based audio feature sets the accuracy of the D-classifiers on the Devel set.

Table 5. *UAR on the Devel set of the D-classifiers for the M-based audio feature sets.*

Feat. set	Std	S1	S2	S3
Audio file	[SO∪VI∪IE]	[SO∪VI]	[VI]	[VI∪IE]
UAR	**66.3**%	64.7%	61.2%	65.9%

It is noticeable that S2-based D-classifier gave the worst accuracy. These results suggest that non-verbal information may be helpful in detecting deception.

Three combinations of the M-based audio feature sets have been assessed to determine whether or not synergies exist between M-based audio feature sets: (1) C1, the optimal combination of two sets in terms of accuracy on the Devel set, (2) C2, the optimal combination of three sets, and (3) C3, the combination of the four sets. Table 6 gives for the different audio feature sets the accuracy of the D-classifier on the Devel set in UAR.

Table 6. UAR *on the Devel set* of the *D-classifier using combination of M-based audio feature sets*

Feature set	Std	C1	C2	C3
Combination	{Std}	{Std,S3}	{Std,S1,S3}	{Std,S1,S2,S3}
# of features	6,373	12,746	19,119	25,492
UAR	66.3%	67.6%	68.4%	**68.5**%

An improvement (+2.2%) is obtained with the combination C3 of M-based audio feature sets compared to EBS.

5. Answer type detection and features

Several studies have been conducted on linguistic indicators of deception [32], [11], and [26]. In the meta-analysis [26], the most relevant cues out 79 were the word measures (word quantity, occurrences of exclusive and emotional word, distribution of personal pronouns by first, second and third person). As a result of such research, many automatic detectors of deceptive text have been developed such as Linguistic Inquiry and Word Count (LIWC) [33] and Agent 99 analyzer [34]. For example, the LIWC detector classifies the words in psychology-relevant dimensions using specialized lexicon. Applying a text-based approach for detecting deceptive speech is difficult because of the poor performances of the ASR systems. In [24], the accuracy of deception detection has been down 13% (73% to 60%) when using automatic compared to manual transcriptions of speech.

Aiming at using the lexical content for an automatic detection of deceptive speech, we have chosen to detect the answer type [35] corresponding to an audio file. This may enable the D-classifier to search a better decision surface between audio features corresponding to audio files with similar lexical content.

5.1. Answer types and bag-of-words

As part of Text Retrieval Conference (TREC) evaluation tasks, several classifications of answer types have been proposed [35]. We choose the Answer Type Hierarchy provided by the Ephyra question answering System [36]. A set of ten answer types {affirmation, date, duration, movement, negation, offense denial, opinion, self report, probability, time} was chosen. Each answer type was modeled by a bag-of-words.

For each answer type, the bag-of-words was made of the words and expressions extracted from the transcription of the Train set. Semantic expansion [37] has been used in order to increase the coverage of the bag-of-words. Table 7 gives, for each answer type, examples of words and expressions extracted from the Train set.

Table 7. *Set of answer types with some examples.*

Answer type	Examples
affirmation	[yes] [absolutely] [sure]
date	[december nineteen ninety three]
duration	[twenty three years] [my all life]
movement	[I went upstair] [I step into]
negation	[no] [never] [not at all]
offense denial	[I did'nt do it] [I am not guilty] [I don't steal] [I am honest] [I did nothing wrong]
opinion	[they should be punished] [prosecuted]
probability	[maybe] [depends] [perhaps] [probably]
self report	[I don't feel nervous] [I'm not nervous]
time	[ten o'clock] [one twenty five][one PM]

5.2. Detection of answer types

Two methods of speech recognition were used to detect the answer type of an audio file: isolated word recognition and keyword spotting. The first method has been used for too short words or expressions (less than three syllables) such as “yes”, “no”, “never” or “maybe”. Keyword spotting has been used for the longer words (or expressions) such as “I am not guilty”. In the two methods, phone loop was used for rejection.

The two ASR systems were based on the 5prealpha CMU recognizer library. The acoustic models were the same as used in §4.1. The phonetic transcription of the words and expressions results from the CMU Pronouncing Dictionary [29]. For each word (or expression), a likelihood is computed from the lesser threshold allowing to detect the word in the audio file. For each answer type, one audio feature was computed from the likelihoods of the words and expressions of the corresponding bag-of-words. On the Train set, seven features out of 10 were considered relevant in terms of information gain (cf. §4.2) and selected for the D-classifier: date, duration, offense denial, self report, time, opinion and affirmation.

6. Experiments and Test results

To assess our approach, five audio feature sets have been defined as a combination or selection of audio feature sets. Let AT be the 7 relevant Answer Type features and PT be the 4 relevant Prosodic and Temporal features. The four audio feature sets corresponding to a combination are: D1 (AT+Std), D2 (AT+PT+Std), D3 (AT+PT+C1) and D4 (AT+PT+C3). The last set D5 was obtained by feature selection from D4 [38]. Table 8 gives for the five audio feature sets the accuracy of the D-classifier on the Devel set in UAR.

Table 8. *UAR on the Devel set of D-classifiers using additional features or feature selection.*

Feat. set	D1	D2	D3	D4	D5
# of feat.	6,380	6,384	12,757	25,503	11,510
UAR	68.6%	68.7%	69.0%	69.7%	**74.9%**

On the Devel set, a significant improvement (+13.0%) was obtained with the D5 set compared to OBS.

On the Test set, four submissions are described. The first one, using D2 and filtering of the Training set, gave a low UAR of 64.2% compared to the 68.3% (OBS). The second one, using D5 features and no filtering, gave an UAR of 67.9%. The third one, using D1 features and no filtering, gave an UAR of 70.6%. The last one, using D3 and no filtering, is our best result on the Test set with an UAR of 72.1% corresponding to an improvement of 3.8% compared to OBS.

7. Conclusion

In this paper, linguistic cues related to the prosody and lexical content have been investigated for deception detection. Audio features have been computed using the results of three ASR systems (acoustic-phonetic decoding, isolated word recognition and keyword spotting). A method to estimate audio features related to the likelihood of answer type have been introduced. These new features have been shown relevant to detect deceptive speech, in particular the interaction ending latency, the multi-resolution-based audio features and the answer type features.

Future works should include features related to the questions and the dialog states during the interview between the agent (human or machine) and the participant. The intra-speaker and inter-speaker variability of the deceptive speech should be also studied.

8. References

[1] J. Masip, E. Garrido, and C. Herrero, “Defining deception,” Anales de Psicología/Annals of Psychology,” vol. 20, no. 1, pp. pp. 147–172, 2004.

[2] A. Vrij, R. Fisher, S. Mann, and S. Leal, “A cognitive load approach to lie detection”, Journal of Investigative Psychology and Offender Profiling, vol. 5, no. 1-2, pp. 39–43, 2008.

[3] J. Masip and C. Herrero, “New approaches in deception detection I. Background and theoretical framework,” Papeles del Psicólogo, vol. 36, no. 2, pp. 83–95, 2015.

[4] M. Zuckerman, R. Koestner, and R. Driver, “Beliefs about cues associated with deception,” Journal of Non-Verbal Behavior, vol. 6, no. 1, pp. 105–114, 1981.

[5] I. Blandón-Gitlin, E. Fenn, J. Masip, and A. Yoo, “Cognitive-load approaches to detect deception: Searching for cognitive mechanisms,” Trends in Cognitive Sciences, vol. 18, pp. 441–444, 2014.

[6] R. Dilmon, “Between thinking and speaking – Linguistic tools for detecting a fabrication,” Journal of Pragmatics, vol. 41, no. 6, pp. 1152–1170, 2009.

[7] S. L. Sporer and B. Schwandt, “Paraverbal indicators of deception: A meta-analytic synthesis,” Applied Cognitive Psychology, vol. 20, no. 4, pp. 421–446, 2006.

[8] F. E. Inbau, “Scientific Evidence in Criminal Cases. II. Methods of Detecting Deception. Journal of Criminal Law and Criminology (1931-1951)”, vol. 24, no.6, pp. 1140–1158, 1934.

[9] J. J. Palmatier and L. Rovner, “Credibility assessment: Preliminary Process Theory, the polygraph process, and construct validity,” International Journal of Psychophysiology, vol; 95, no.1, pp. 3–13, 2015.

[10] H. Wang, W. Chang, and C. Zhang, “Functional brain network and multichannel analysis for the P300-based brain computer interface system of lying detection”, Expert Systems with Applications, 53, pp. 117–128, 2016.

[11] B. M. De Paulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, “Cues to deception,” Psychological bulletin, vol. 129, no. 1, pp. 74–118, 2003.

[12] J. K Burgoon, R. Schuetzler, and D. W. Wilson, “Kinesic patterning in deceptive and truthful interactions,” Journal of Nonverbal Behavior, vol. 39, no. 1, pp. 1–24, 2015.

[13] A. Vrij, P. A. Granhag, and S. Porter, “Pitfalls and opportunities in nonverbal and verbal lie detection,” Psychological Science in the Public Interest, vol. 11, no. 3, pp. 89–121, 2010.

[14] F. Horvath, J. McCloughan, D. Weatherman, and S. Slowik, “The Accuracy of Auditors' and Layered Voice Analysis (LVA) Operators' Judgments of Truth and Deception During Police Questioning,” Journal of forensic sciences, vol. 58, no. 2, pp. 385–392, 2013.

[15] A. Vrij, S. Leal, S. Mann, Z. Vernham, and F. Brankaert, “Translating theory into practice: Evaluating a cognitive lie detection training workshop,” Journal of Applied Research in Memory and Cognition, vol. 4, no. 2, pp. 110–120, 2015.

[16] J. F. Kubis, "Comparison of Voice Analysis and Polygraph as Lie Detection Procedures," U.S. Army Land Warfare Laboratory, LWL-CR-03B70, August 1973.

[17] M. Gamer, H. G. Rill, G. Vossel, and H. W. Gödert, “Psychophysiological and vocal measures in the detection of guilty knowledge,” International Journal of Psychophysiology, vol. 60, no.1, pp. 76–87, 2006.

[18] J. B. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis, B. Pellom, E. Shriberg, and A. Stolcke, “Distinguishing deceptive from non-deceptive speech,” in Proceedings of INTERSPEECH, Lisbon, Portugal, pp. 1833–1836, 2005.

[19] P. Ekman, M. O’Sullivan, W. V. Friesen, and K. R. Scherer, “Invited article: Face, voice, and body in detecting deceit,” Journal of nonverbal behavior, vol. 15, no. 2, pp. 125–135, 1991.

[20] S. Benus, F. Enos, J. B. Hirschberg, and E. Shriberg, “Pauses in deceptive speech,” Proceedings ISCA 3rd International Conference on Speech Prosody, 2006.

[21] M. Graciarena, E. Shriberg, A. Stolcke, F. Enos, J. Hirschberg, and S. Kajarekar, “Combining prosodic lexical and cepstral systems for deceptive speech detection,” in Proceedings of ICASSP, vol. 1. Toulouse, France: IEEE, 2006.

[22] Y. Zhou, H. Zhao, X. Pan, and L. Shang, “Deception detecting from speech signal using relevance vector machine and non-linear dynamics features,” Neurocomputing, 2015, vol. 151, pp. 1042–1052, 2015.

[23] Enos, F., Shriberg, E., Graciarena, M., Hirschberg, J. B., & Stolcke, A. Detecting deception using critical segments, Proceedings INTERSPEECH 2007, ISCA, Antwerp, Belgium, pp. 2281–2284, 2007.

[24] R. Mihalcea, V. Pérez-Rosas, and M. Burzo, “Automatic detection of deceit in verbal communication,” in Proceedings of the 15th ACM on International conference on multimodal interaction, pp. 131–134, 2013.

[25] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanin, “The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception, Sincerity & Native Language", Proceedings INTERSPEECH 2016, ISCA, San Francisco, USA, 2016.

[26] V. Hauch, I. Blandón-Gitlin, J. Masip, and S. L. Sporer, “Are computers effective lie detectors? A meta-analysis of linguistic cues to deception”, Personality and Social Psychology Review, vol. 19, no. 4, pp. 307–342, 2015.

[27] F. Eyben, F. Weninger, F. Groß, and B. Schuller, “Recent developments in openSMILE, the Munich open-source multimedia feature extractor,” in Proceedings of ACM MM, Barcelona, Spain, pp. 835–838, 2013.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: An update,” ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10–18, 2009.

[29] A. Chan, E. Gouva, R. Singh, M. Ravishankar, R. Rosenfeld, Y. Sun, D. Huggins-Daines, M. Seltzer, “The Hieroglyphs: Building Speech Applications Using CMU Sphinx and Related Resources, ”www.cs.cmu.edu/~archan/share/sphinxDoc.pdf, 2007.

[30] F. Hönig, A. Batliner, K. Weilhammer, and E. Nöth, “Automatic assessment of non-native prosody for english as l2,” in Proc. Speech Prosody, Chicago, 4 pages, 2010.

[31] T. W. Rauber, A. S. Steiger-Garcao, “Feature selection of categorical attributes based on contingency table analysis,” paper presented at the Portuguese Conference on Pattern Recognition, Porto, Portugal, 8 pages, 1993.

[32] M. L. Knapp, R. P. Hart, and H. S. Dennis, “An exploration of deception as a communication construct,” Human communication research, vol. 1, no. 1, pp. 15–29, 1974..

[33] J. W. Pennebaker, M. E. Francis, and R. J. Booth, “Linguistic inquiry and word count: LIWC 2001,” Mahway: Lawrence Erlbaum Associates, pp. 1–11, 2001.

[34] J. Cao, J. M. Crews, M. Lin, J. Burgoon, and J. F. Nunamaker, “Designing Agent99 trainer: A learner-centered, web-based training system for deception detection,” in Intelligence and Security Informatics, Springer Berlin Heidelberg, pp. 358–365, 2003.

[35] M. M. Soubbotin and S. M. Soubbotin, “Patterns of potential answer expressions as clues to the right answers,” in Proceedings of TREC, 10 pages, 2001..

[36] N. Schlaefer, J. Ko, J. Betteridge, M. A. Pathak, E. Nyberg, and G. Sautter, “Semantic Extensions of the Ephyra QA System for TREC 2007,” in Proceedings of TREC, vol. 1, no. 2, 10 pages, 2007.

[37] E. M. Voorhees, “Using WordNet for text retrieval,” in: C. Fellbaum (Ed.), WordNet, an Electronic Lexical Database, MIT Press, pp. 285–303, 1998.

[38] M.-J. Caraty and C. Montacié, “Detecting Speech Interruptions for Automatic Conflict Detection,” in Conflict and Multimodal Communication, Springer International Publishing, pp. 377–401, 2015.