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

Combining Acoustic-Prosodic, Lexical, and Phonotactic Features for Automatic Deception Detection

Sarah Ita Levitan[1], *Guozhen An*[2], *Min Ma*[2], *Rivka Levitan*[3], *Andrew Rosenberg*[4], *Julia Hirschberg*[2]

[1]Department of Computer Science, Columbia University, USA
[2]Department of Computer Science, CUNY Graduate Center, USA
[3]Department of Computer and Information Science, Brooklyn College (CUNY), USA
[4]Department of Computer Science, Queens College (CUNY), USA

sarahita@cs.columbia.edu, gan@gradcenter.cuny.edu, mma@gradcenter.cuny.edu,
rlevitan@brooklyn.cuny.edu, andrew@cs.qc.cuny.edu, julia@cs.columbia.edu

Abstract

Improving methods of automatic deception detection is an important goal of many researchers from a variety of disciplines, including psychology, computational linguistics, and criminology. We present a system to automatically identify deceptive utterances using acoustic-prosodic, lexical, syntactic, and phonotactic features. We train and test our system on the Interspeech 2016 ComParE challenge corpus, and find that our combined features result in performance well above the challenge baseline on the development data. We also perform feature ranking experiments to evaluate the usefulness of each of our feature sets. Finally, we conduct a cross-corpus evaluation by training on another deception corpus and testing on the ComParE corpus.

Index Terms: deception detection, computational paralinguistics

1. Introduction

Automatic deception detection is an important goal for many researchers, from psychologists and computational linguists to practitioners in law enforcement, military, and intelligence agencies. Despite many attempts to develop automated deception detection technologies, there have been few objective successes. Researchers have explored the use of several modalities for deception detection. Perhaps most typically, biometric indicators are measured by the polygraph. In addition, facial expressions [1], gestures and posture [2], brain imaging [3], and linguistic information have all been explored as indicators of deception. Many of these features (e.g. facial expressions, gestures) are expensive to automatically capture, and some (e.g. brain imaging) are too invasive to be practical for general use. Language cues have the advantage of being inexpensive and easy to collect. More importantly, prior research examining linguistic cues to deception has been promising. Such cues include speech-based and text-based features.

In this paper we present a language-based system for automatic deception detection. This work was performed as a submission to the Interspeech 2016 ComParE Deception Sub-Challenge [4]. We extract acoustic-prosodic, lexical, syntactic, duration, and phonotactic features from the Deceptive Speech Database (DSD) provided by the challenge, and we compare the performance of the feature sets using a variety of machine learning algorithms. To further evaluate our method, we perform a cross-corpus evaluation, training on the Columbia Deception Corpus (CDC) [5] and testing on the DSD corpus. Our results are above the challenge baseline, and our feature ranking and machine learning experiments provide insight into useful techniques for deception detection.

In Section 2 we review related work in language-based deception detection. In Section 3 we describe the multiple feature sets used in our experiments. Section 4 presents the results of our classification experiments, on both the development set and the blind test set, and we include an analysis of the contributions of the different features. In Section 5 we report on our cross-corpus experiments. We conclude in Section 6 with a discussion of the results and future directions.

2. Related Work

There have been a number of studies of linguistic cues to deceptive speech and text, mostly conducted by psychologists, and more recently by computer scientists. Early work by Ekman et al. [6] and Streeter et al. [7] found pitch increases in deceptive speech. An increased effect was observed when subjects were highly motivated to deceive. Linguistic Inquiry and Word Count (LIWC) categories were found to be useful in deception detection studies across five corpora, where subjects lied or told the truth about their opinions on controversial topics [8]. They achieved as high as 67% accuracy using LIWC categories. Bachenko et al. [9] analyzed linguistic indicators of deception in criminal narratives, interrogations, and legal testimony. They found that a mixture of automatic and manually assigned linguistic indicator tags, including hedges, verb tense, and negative expressions, were highly predictive of the truth value of statements. Other studies report similar findings: deceptive statements can be distinguished from truthful statements using language based cues [10, 11].

There has also been progress in identifying cues to deception drawn from the speech signal. Hirschberg et al. [12] collected the Columbia-SRI-Colorado (CSC) corpus, the first cleanly recorded large-scale corpus of deceptive and non-deceptive speech. They automatically extracted acoustic-prosodic and lexical features and achieved about 70% accuracy. They found that subject-dependent features were especially useful in capturing individual differences in deceptive behavior. Building on this work, with a focus on individual differences, Levitan et al. [5] collected a large-scale corpus of cross-cultural deceptive and non deceptive speech. They also collected personality information for participants, and found

http://dx.doi.org/10.21437/Interspeech.2016-1519

that including gender, native language, and personality scores along with acoustic-prosodic features improved classification accuracy, supporting the notion that deceptive behavior varies across different groups of people.

A meta-analysis [13] identified cues to deception that were significant across many studies. Some of these cues were linguistic, including duration, vocal tension, F0, and negative emotion words. It is clear from the literature that acoustic-prosodic and lexical features are predictive of deceptive speech. In this work we extract features that have been found useful in previous work, introduce some new features, and systematically evaluate machine learning classifiers trained on acoustic-prosodic and lexical feature sets. We also use the cross-cultural deception corpus for a cross corpus evaluation of our work.

3. Features

3.1. ComParE Baseline Features

The COMPARE baseline feature set contains 6373 static features from the computation of various functionals over low-level descriptor (LLD) contours extracted from openSMILE [4, 14]. The LLD features include pitch (fundamental frequency), intensity (energy), spectral, cepstral (MFCC), duration, voice quality (jitter, shimmer, and harmonics-to-noise ratio), spectral harmonicity, and psychoacoustic spectral sharpness.

3.2. Acoustic-Prosodic Features

In accordance with previous research, we hypothesize that, when people are deceptive, their pitch, energy, and rhythm patterns may unconsciously change. We measure the pitch and energy variations by modeling their contours. We extract F0 measurements for each response in the corpus using Snack [15]. Post-processing on the F0 measurements corrects implausible F0 jumps and interpolates over unvoiced regions with smoothing using a Butterworth filter. Finally, all raw F0 measurements are normalized to z-scores by speaker. We then fit the normalized F0 measurements using 1- to 7-order polynomial models, thus developing 48 pitch contour features. Additional details can be found in [16]. A similar procedure is applied to log-energy measurements extracted by openSMILE [17].

We measure speaking rate by calculating the ratio of syllables to the utterance duration. Intra-syllable pause and syllable duration are used to design duration-related features. We detect the pseudosyllable regions based on the Villing envelope based approach [18] as implemented in AuToBI [19], and derive the following features for each utterance: 1) total number of syllables,N; 2) total duration of syllable regions, sum(syl) and total duration of the utterance, sum(utt); 3) silence ratio defined as the ratio of sum(syl) to sum(utt); 4) speaking rate on the syllable level, namely, $\frac{N}{\text{sum}(utt)}$; 5) average duration per syllable, namely, $\frac{\text{sum}(syl)}{N}$; 6) the maximum, minimum, range, standard deviation, mean and median values of the duration of the syllables within each utterance. All the duration values used in 1) to 5) are raw time duration, while the duration values used in 6) is normalized by sum(utt).

In addition to our acoustic-prosodic feature set, we extract lexical, syntactic, and phonotactic features. In order to extract these feature sets, we required a transcript of each utterance in the corpus. We used a web-based API available through wit.ai (`https://wit.ai`), to acquire ASR output for the audio samples in the corpus. We hand-corrected some of the transcripts, but the ASR seemed reasonably good for our purposes.

3.3. Linguistic Inquiry and Word Count Features

Previous work has found that deceivers tend to use different patterns of word usage when they are lying [8]. Inspired by [20], we used LIWC [21] to extract the lexical features from each utterance. LIWC is a text analysis program that computes word counts for 72 linguistic dimensions. LIWC dimensions have been used in many studies to predict outcomes including personality [20], deception [8], and health [22]. We extracted a total of 130 LIWC features based on 64 LIWC categories: 64 features based upon the ratio of words appearing in each LIWC categories over total word count; 64 features based on the ratio of words appearing in each LIWC category over the total words appearing in any LIWC category; the total number of words appearing in any LIWC category; and the total word count.

3.4. Dictionary of Affective Language Features

[12] found that Dictionary of Affect in Language (DAL) [23] scores are useful to distinguish between deceptive and truthful speech. The DAL is a lexical analysis tool that is used for investigating emotive content of speech. It lists approximately 4500 English words, along with a rating for Pleasantness (Evaluation) and for Activation (Arousal) associated with each word. We extract nineteen features derived from the DAL scores for each word in each subject's baseline interview transcript. From all words' pleasantness, activation and imagery scores, we calculated the mean, minimum, maximum, median, standard deviation, and variance. We also added the number of words in the transcript that appear in the DAL.

3.5. Fundamental Frequency Variation (FFV) Features

Previous work [7] found that there are some correlations between deception and fundamental frequency. In order to capture the frequency information, we extracted 42 features which come from fundamental frequency variation (FFV) spectrum with 7 components [24]. From each of the 7 spectrum components, we extract 6 features: mean, minimum, maximum, median, standard deviation, and variance. These features have been found to be helpful in characterizing dialogues [24] and also in acoustic modeling for speech recognition [25].

3.6. Phonotactic Variation Features

Phonotactic modeling has been shown effective for language recognition [26]. In this paper, we hypothesize that phonotactic modeling will also be useful for deception detection, since deceptive speech may result in pronunciation differences, $i.e.$ a deceptive speaker may tend to choose certain phonotactic variants or words more frequently than others. We first apply an English phoneme recognizer developed by Brno University of Technology (BUT) [27] to generate the phoneme hypotheses for each audio instance. After excluding all the non-phonetic symbols (*'oth'*, *'pau'*, *'sil'*,*'spk'*,*'int'*), we build a trigram language model with Witten-Bell smoothing [28] for each class over training set, using SRILM [29]. Both of the models ("phonLM") are used to assign log-probability for each phoneme sequence in the dev and test sets. Because the perplexity reflects the degree of "uncertainty" of language model, which is approximately to the confidence of the assignment, we collect the perplexity calculate as $10^{\frac{log-prob}{\#sent+\#words}}$ and $10^{\frac{log-prob}{\#words}}$ along with the log-probability to construct the feature set. We develop similar language models ("wordLM") and derived features on the word level as well, using the transcripts

produced by ASR.

3.7. Additional Lexico-Syntactic Features

We implement a number of lexical features described in [30] which were used for deception detection. We estimate complexity of an utterance by computing the number of syllables per speech segment and dividing by the number of words. We include binary features capturing whether the utterance contains a hedge word, feeling word, number, or date. We also encode whether the utterance is only "yes" or "no", has a direct denial such as "I did not do it", or contains a contraction. Finally, we develop a bag of words model using part of speech tags obtained using NLTK's built in POS tagger [31].

4. Deception Detection Experiments

4.1. Development Set Evaluation

After preparing all of the feature sets, we ran classification experiments using each feature set independently, to get an initial sense of the usefulness of each feature set. We use the SMO model provided as a baseline (set to the same parameters), training on the train set and and evaluating the performance on the dev set, using the UAR metric. In total we used seven feature sets described in section 3: DAL, LIWC, FFV, phonotactic (phonLM and wordLM), POS (bag of part-of-speech tags), lexical (complexity, binary indicator features), and duration (syllable-based). We compare the performance to two baselines – a majority class baseline (UAR=50%) and a less trivial baseline using the openSMILE features provided by the challenge (UAR=61.9%). As shown in Table 1, three of our feature sets outperform the openSMILE baseline: LIWC, DAL, and Phonotactic. The remaining four feature sets yield better results than the majority class baseline.

Table 1: *Deception classification results on dev set, using single feature sets and single+baseline feature sets*

Features	UAR	+baseline
DAL	63.1	61.5
LIWC	63.9	61.9
FFV	54.3	60.7
Phonotactic	**67.7**	61.9
POS	57.8	57.8
Lexical	59.3	61.3
Duration	56.9	**62.2**
Baseline (majority)	50	-
Baseline (openSMILE)	61.9	-

Next, we repeated the experiments, this time combining each of our feature sets with the baseline openSMILE feature sets, in order to evaluate their contribution over the baseline feature set. As shown in the second column of Table 1, only the duration features combined with the baseline features improved over using the baseline set alone. Although the DAL, LIWC, and phonotactic features perform well on their own, combining them with the baseline feature set decreases performance. Additionally, we note that the baseline feature set improves performance when combined with FFV and lexical features.

In order to evaluate which features were the most useful, we used Weka's attribute evaluator to rank all 6644 of our features using Information Gain[32]. The rank was determined by the information gain using four-fold cross validation over the training data. We then selected all features with an information gain above zero – there were 172 such features. The top ranked features were the phonotactic features, many of the auditory spectrum features from openSMILE, selected LIWC and DAL features, FFV features, and some of the additional lexical features. The POS tags were not included in the top ranked features, nor were the duration features. It was interesting to observe that 'hasDate' (a binary feature indicating whether the speech segment included a date) and the LIWC number category were both particularly useful features, and both were more frequent in non-deceptive speech. In the DSD corpus, subjects were asked to describe their activities, and some subjects provided great details including dates, times, and room numbers. Our observation that the truthful students used these details more supports the finding that liars provide fewer details than truth-tellers [13]. Another useful feature was 'isYes' – a binary feature that indicated whether the response was a one word 'yes'. We found that these affirmative statements were more frequent in truthful responses, supporting the finding that truth-tellers are generally more positive [13].

After obtaining the top-ranked features, we compared the performance of a variety of machine learning classifiers trained on these top features, using Weka. We do the same for the baseline feature set. Table 2 displays the top five classifiers along with the UAR for the top feature set and the baseline feature set. The five algorithms that yielded the best performance were SMO, Bagging, Dagging, BayesNet and NaiveBayes. All of the models were trained and tested using the default parameters. Two of the classifiers are ensemble-based learning algorithms (Bagging and Dagging), both of which are robust in noisy conditions. It is interesting that both these models perform well on our feature set, but poorly on the baseline features – perhaps because our feature set encompasses many types of features and therefore has more variance. Another two of the classifiers are Bayesian network models, which are known to perform well on text classification problems. For all five algorithms, our top feature set outperforms the baseline feature set.

Table 2: *Deception classification results (UAR) on dev set, using top 172 features vs. baseline feature set*

Algorithm	Top Features	Baseline Features
SMO	63.8	61.9
Bagging	63.5	55.4
Dagging	63.9	54.9
BayesNet	64.1	62.2
NaiveBayes	**64.7**	61.7
Majority voting	63.8	**62.9**

We also evaluate a classifier which takes as input predictions of the five best classifiers and outputs the majority vote. For our top feature set, this approach results in a UAR of 63.8%, which is the same as using the SMO classifier alone. On the other hand, the majority voting classifier yields a higher UAR (62.9%) than any of the single classifiers for the baseline feature set. It is not obvious why this is true only for this feature set, but it suggests combining the predictions of multiple independent classifiers can indeed improve performance.

4.2. Test Set Evaluation

After experimenting with several feature sets and machine learning algorithms using the dev data, we used our best models to get predictions on the test data. We submitted five predictions: (1) DAL+LIWC (2) Phonotactic (3) openSMILE (OS)+DAL+LIWC (4) OS+DAL+LIWC+Phonotactic (5) Majority voting (Ensemble) using top ranked features. Results are

shown in Table 3.

Table 3: *Deception classification results on test set*

Algorithm	Features	UAR test	UAR dev
SMO	DAL+LIWC	64.7	**66.3**
SMO	Phonotactic	64.2	**67.7**
SMO	OS+DAL+LIWC	**69.3**	61.5
SMO	OS+DAL+LIWC+Phon	**69.4**	61.9
Ensemble	top ranked	65.4	**63.8**
Baseline	openSMILE	68.3	61.9

Our results indicate that the phonotactic features and DAL+LIWC features on their own do not generalize well to the unseen test data. Although they performed best on the dev data, with 67.7% and 66.3% UAR, they performed worse on the test data. It is possible that the language used in the test set is different from the train and dev data, and therefore these lexical features do not work well on the test data. Our best performing model on the test set use an SMO classifier and a combination of the baseline openSMILE features with DAL+LIWC+Phonotactic. This model achieved a UAR of 69.4%, which was better than the results on the dev set (61.9%) and than the challenge baseline for the test set (68.3%). We were unable to perform error analysis on the test set (due to unknown class labels), but our results suggest that a combination of acoustic-prosodic features and lexical features are the most robust for deception detection.

We included our phonotactic features based on the hypothesis that deception would affect a speaker's phonotactic patterns; the performance of these features on the dev data supports our hypothesis to some extent. However, it is possible that these features are speaker-dependent, so the resulting language models might not accurately capture the deceptive phonotactic styles of speakers not present in training data. This might explain why these features, which performed best on the dev set, were not as effective on the test set, and did not make better predictions on the dev data when combined with the acoustic-prosodic and lexical features.

5. Cross-Corpus Evaluation

In order to assess whether these techniques are generalizable across domains for deception detection, we experiment with training on another deception corpus and testing on the ComParE corpus. We use the Columbia Deception Corpus (CDC) [33] for training. This corpus was collected using a fake resume paradigm, where subjects are asked 24 biographical questions and are told to lie for a random half of them. The dialogues in CDC are markedly different from the ComParE corpus: subject turns are longer, are asked more open-ended questions, and the vocabulary is less constrained. The CDC is also much larger: it includes data from 344 subjects, constituting about 122 hours of subject speech. Additionally, the recording conditions and sample rate of the audio files are not consistent. To minimize the differences between the two corpora, we randomly select a subset of 500 turns from the CDC that fulfill the following criteria: (1) has $<=$ 30 words and (2) contains "yes" or "no". This gives us a subset with similar length and style of turns. We extract the following feature sets from this subset of CDC: openSMILE acoustic-prosodic, LIWC, and DAL. The results, measured by UAR, are shown in Table 4. We compare the results to a simple baseline of 50%, obtained by predicting the majority class (ND). As expected, lexical features do not generalize across the two corpora. Although the LIWC and DAL features were particularly useful when training and testing on the ComParE corpus, they do not yield results above the baseline when training on the CDC. This is probably due to the difference in domain between the corpora. Another challenge is that we are using noisy ASR output as transcription for the ComParE corpus, while the CDC has clean transcription obtained via Amazon Mechanical Turk and hand-corrected by human transcribers. This difference is likely a factor in the poor performance of lexical features across corpora. On the other hand, the acoustic feature set performs surprisingly well. When using the same set of openSMILE features for training and testing on the ComParE data, we obtain a UAR of 60.7%, only slightly better than our cross-corpus performance of 59.8% UAR. It appears that the acoustic-prosodic features do generalize to a different domain. This is promising for deception detection applications outside of a laboratory experiment.

Table 4: *Cross-corpus results: train on CDC, test on ComParE dev*

Features	UAR
DAL	47.4
LIWC	47.6
DAL+LIWC	49.7
Acoustic	**59.8**
Baseline	50

6. Conclusion and Future Work

We have presented classification experiments for automatic deception detection using a combination of acoustic-prosodic, lexical, and phonotactic features. We experimented with feature combinations and ranking, and a variety of machine learning algorithms, and found that we can achieve results above the challenge baseline with our system. We extract lexical features by running ASR on the sound files, and performing minor hand correction of the ASR output. It is impressive that this simple approach results in lexical features that are ultimately useful in our classifer. This suggests that ASR output can be used for a flexible system to get real-time lexical features without waiting for a quality transcript of a speech sample.

We also present results of a cross-corpus evaluation, where we train a classifier using the CDC, and evaluate it on the ComParE corpus. Our results are quite promising; we find that we can obtain almost the same results as training and testing on the ComParE corpus – 59.8% compared to 60.7% UAR. This suggests that the features and models used are general enough to be applied to different data, which is important for any model that will be deployed in a real-world deception situation.

One area for improvement is to obtain quality transcription instead of relying on noisy ASR output. It would be interesting to see if lexical features extracted from quality transcription are more useful than the lexical features used in this work. Although this is less practical than using ASR, it can provide an upper bound on the performance. Another possible extension of this work is to model individual differences between subjects, perhaps by grouping similar speakers into clusters as a pre-processing step.

7. Acknowledgements

This work was partially funded by AFOSR FA9550-11-1-0120 and by NSF DGE-11-44155.

8. References

[1] P. Ekman, "Lie catching and microexpressions," *The philosophy of deception*, pp. 118–133, 2009.

[2] T. O. Meservy, M. L. Jensen, J. Kruse, J. K. Burgoon, J. F. Nunamaker Jr, D. P. Twitchell, G. Tsechpenakis, and D. N. Metaxas, "Deception detection through automatic, unobtrusive analysis of nonverbal behavior," *Intelligent Systems, IEEE*, vol. 20, no. 5, pp. 36–43, 2005.

[3] D. D. Langleben, J. W. Loughead, W. B. Bilker, K. Ruparel, A. R. Childress, S. I. Busch, and R. C. Gur, "Telling truth from lie in individual subjects with fast event-related fmri," *Human brain mapping*, vol. 26, no. 4, pp. 262–272, 2005.

[4] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and E. Keelan, "The interspeech 2016 computational paralinguistics challenge: Deception, sincerity & native language." in *INTERSPEECH*. ISCA, 2016, pp. I–I.

[5] S. I. Levitan, M. Levine, J. Hirschberg, N. Cestero, G. An, and A. Rosenberg, "Individual differences in deception and deception detection," 2015.

[6] 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.

[7] L. A. Streeter, R. M. Krauss, V. Geller, C. Olson, and W. Apple, "Pitch changes during attempted deception." *Journal of personality and social psychology*, vol. 35, no. 5, p. 345, 1977.

[8] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, "Lying words: Predicting deception from linguistic styles," *Personality and social psychology bulletin*, vol. 29, no. 5, pp. 665–675, 2003.

[9] J. Bachenko, E. Fitzpatrick, and M. Schonwetter, "Verification and implementation of language-based deception indicators in civil and criminal narratives," in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2008, pp. 41–48.

[10] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell, "Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications," *Group decision and negotiation*, vol. 13, no. 1, pp. 81–106, 2004.

[11] R. Mihalcea and C. Strapparava, "The lie detector: Explorations in the automatic recognition of deceptive language," in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, 2009, pp. 309–312.

[12] J. B. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis *et al.*, "Distinguishing deceptive from non-deceptive speech," 2005.

[13] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception." *Psychological bulletin*, vol. 129, no. 1, p. 74, 2003.

[14] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*. ACM, 2013, pp. 835–838.

[15] K. Sjlander, "The snack sound extension for tcl/tk," 1997-99. [Online]. Available: http://www.speech.kth.se/SNACK/

[16] M. Ma, K. Evanini, A. Loukina, X. Wang, and K. Zechner, "Using f0 contours to assess nativeness in a sentence repeat task," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[17] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[18] R. Villing, J. Timoney, and T. Ward, "Automatic blind syllable segmentation for continuous speech," 2004.

[19] A. Rosenberg, "Autobi-a tool for automatic tobi annotation." in *INTERSPEECH*, 2010, pp. 146–149.

[20] J. W. Pennebaker and L. A. King, "Linguistic styles: language use as an individual difference." *Journal of personality and social psychology*, vol. 77, no. 6, p. 1296, 1999.

[21] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, p. 2001, 2001.

[22] J. W. Pennebaker, T. J. Mayne, and M. E. Francis, "Linguistic predictors of adaptive bereavement." *Journal of personality and social psychology*, vol. 72, no. 4, p. 863, 1997.

[23] C. Whissell, M. Fournier, R. Pelland, D. Weir, and K. Makarec, "A dictionary of affect in language: Iv. reliability, validity, and applications," *Perceptual and Motor Skills*, vol. 62, no. 3, pp. 875–888, 1986.

[24] K. Laskowski, M. Heldner, and J. Edlund, "The fundamental frequency variation spectrum," *Proceedings of FONETIK 2008*, pp. 29–32, 2008.

[25] X. Cui, B. Kingsbury, J. Cui, B. Ramabhadran, A. Rosenberg, M. S. Rasooli, O. Rambow, N. Habash, and V. Goel, "Improving deep neural network acoustic modeling for audio corpus indexing under the iarpa babel program," in *Interspeech*, 2014.

[26] M. A. Zissman *et al.*, "Comparison of four approaches to automatic language identification of telephone speech," *IEEE Transactions on Speech and Audio Processing*, vol. 4, no. 1, p. 31, 1996.

[27] P. Schwarz, P. Matejka, L. Burget, and O. Glembek, "Phoneme recognizer based on long temporal context," *Speech Processing Group, Faculty of Information Technology, Brno University of Technology.[Online]. Available: http://speech. fit. vutbr. cz/en/software*, 2006.

[28] I. H. Witten and T. C. Bell, "The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression," *Information Theory, IEEE Transactions on*, vol. 37, no. 4, pp. 1085–1094, 1991.

[29] A. Stolcke *et al.*, "Srilm-an extensible language modeling toolkit." in *INTERSPEECH*, vol. 2002, 2002, p. 2002.

[30] F. Enos, "Detecting deception in speech," Ph.D. dissertation, Citeseer, 2009.

[31] E. Loper and S. Bird, "Nltk: The natural language toolkit," in *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1*. Association for Computational Linguistics, 2002, pp. 63–70.

[32] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Information processing & management*, vol. 42, no. 1, pp. 155–165, 2006.

[33] S. I. Levitan, G. An, M. Wang, G. Mendels, J. Hirschberg, M. Levine, and A. Rosenberg, "Cross-cultural production and detection of deception from speech," in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*. ACM, 2015, pp. 1–8.