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

Sincerity and Deception in Speech: Two Sides of the Same Coin? A Transfer- and Multi-Task Learning Perspective

Yue Zhang[1], *Felix Weninger*[2], *Zhao Ren*[3], *Björn Schuller*[1]

[1]Department of Computing, Imperial College London, London, UK
[2]Nuance Communications, Ulm, Germany
[3]Northwestern Polytechnical University, Xi'an, P. R. China
yue.zhang1@imperial.ac.uk

Abstract

In this work, we investigate the coherence between inferable deception and perceived sincerity in speech, as featured in the Deception and Sincerity tasks of the INTERSPEECH 2016 Computational Paralinguistics ChallengE (ComParE). We demonstrate an effective approach that combines the corpora of both Challenge tasks to achieve higher classification accuracy. We show that the naïve label mapping method based on the assumption that sincerity and deception are just 'two sides of the same coin', i. e., taking deceptive speech as equivalent to non-sincere speech and vice versa, does not yield satisfactory results. However, we can exploit the interplay and synergies between these characteristics. To achieve this, we combine our previously introduced approach for data aggregation by semi-supervised cross-task label completion with multi-task learning, and knowledge-based instance selection. In the result, our approach achieves significant error rate reductions compared to the official Challenge baseline.

Index Terms: Computational Paralinguistics, Transfer Learning, Cross-Task Labelling, Multi-Task Learning, Deception and Sincerity Identification

1. Introduction

Building multi-faceted descriptions of speech including speaker states, traits and speaking styles has immediate applications in content-based speech retrieval, as well as crime prevention and forensics. This can be highly relevant to filter important conversations from large amounts of speech recordings, where the issue of social signals detection particularly comes into play. The lion's share of today's studies considers speech characteristics in isolation, i. e., single or only few attributes are analysed at once. There is very little exploitation of the interplay and synergies between different characteristics, yet in reality, strong interdependency between bits of paralinguistic information exists.

The INTERSPEECH 2016 Computational Paralinguistics ChallengE (ComParE) [1] provides an interesting test bed for holistic speech analysis by featuring two classification tasks that are – intuitively – correlated but subtly different: the manifestation of (quasi) objective deception in speech (induced by a specifically designed scenario), as well as sincerity of speech as subjectively perceived and estimated by human listeners. From an engineering point of view, this motivates the usage of data aggregation to improve performance on either task. Since the training databases for each task are only labelled in one dimension, we rely on semi-autonomous annotation to complete the missing target labels. Furthermore, we use multi-task learning to exploit the hypothesised correlation between these tasks. Besides improving performance, our aim is to shed light on the question whether these tasks are so highly correlated that one can be directly mapped to each other – and hence, they should be considered as a single task, as is often done in emotion recognition, where data sources with different underlying emotion concepts can be united [2] –, or, whether there is a more complex interplay, justifying to treat them as separate tasks. For this we will investigate the accuracy of direct label mapping and transfer learning.

Related Work: In the context of computational paralinguistics, multi-task learning [3, 4] has been exploited by [5, 6, 7], and semi-supervised learning (SSL) [8, 9] has been investigated for emotion recognition in [10, 11] and for sentiment analysis in [12, 13]. In [14], we extended SSL to the Cross-Task Labelling (CTL) method, yielding fully automatic annotations of speaker trait databases with multiple labels. Related work in speech recognition covered the completion of missing features [15], however, these were not used as classification targets. As far as we know, the application of all these techniques to deception or sincerity classification is new.

Deception detection has been investigated in a few studies involving audio and acoustic feature analysis [16, 17, 18]. The prevailling view amongst philosophers is that a speech act is sincere if the speaker is in the state of mind that the speech act functions to express [19]. An alternative definition of sincerity in speech as Spontaneity and as a Communicative Virtue can be found in [20]. An interesting view is presented in [21], according to which sincerity is anchored in the speaker's and hearer's attitudes towards the propositions and thus to the illocutionary level. The nature of sincerity as a subjective speech phenomenon has been studied in [22, 23]. However, there has been in the literature little work on the automatic identification of deceptive and sincere speech from acoustic, prosodic, and lexical cues.

In the following, we detail the holistic speech analytic method applied in this study, as well as the investigated data sets, before presenting experimental results.

2. Methodology

Let us first introduce some required notation: Assuming a set of L different classification tasks to be performed, $\mathbf{x}_i^{(l)} \in \mathcal{X}$ denotes the i-th feature vector for classification task l, while $y_i^{(l)} \in \mathcal{Y}^{(l)}$ denotes its gold standard label and $\mathcal{Y}^{(l)}$ the set of possible labels for task l. $[\cdot;\cdot]$ denotes the concatenation of features.

http://dx.doi.org/10.21437/Interspeech.2016-1305

Table 1: *Overview of the presented learning schemes.*

	Training data	Cross-task training data	Test data
Transfer learning	–	$\mathbf{x}_j^{(m)}, h^{m\to l}(y_j^{(m)})$	
Data aggregation:			
– direct mapping	$\mathbf{x}_i^{(l)}, y_i^{(l)}$	$\mathbf{x}_j^{(m)}, h^{m\to l}(y_j^{(m)})$	$\mathbf{x}_k^{(l)}, y_k^{(l)}$
– semi-supervised learning	$\mathbf{x}_i^{(l)}, y_i^{(l)}$	$\mathbf{x}_j^{(m)}, \tilde{y}^{(l)}(\mathbf{x}_j^{(m)})$	
– CTL, single-task learning	$\mathbf{x}_i^{(l)}, y_i^{(l)}$	$\mathbf{x}_j^{(m)}, \hat{y}_j^{(l)}$	
– CTL, multi-task learning	$[\mathbf{x}_i^{(l)}; \hat{y}_i^{(m)}], y_i^{(l)}$	$[\mathbf{x}_j^{(m)}; y_j^{(m)}], \hat{y}_j^{(l)}$	$[\mathbf{x}_k^{(l)}; \hat{y}_k^{(m)}], y_k^{(l)}$

2.1. Transfer Learning

For transfer learning, we make use of the assumption that the Sincerity and Deception tasks are just 'two sides of the same coin', i.e., non-sincere speech is the same as deceptive speech and vice versa. We can formalise such expert rules as mapping functions $h^{m\to l} : \mathcal{Y}^{(m)} \to \mathcal{Y}^{(l)}$ that maps a label for task m to the corresponding label for task l, $l \neq m$. In the scope of this paper, $l, m \in \{\text{Deception}, \text{Sincerity}\}$, and the mappings are bijective. Using $h^{m\to l}$, we can derive a straightforward transfer learning scheme for the task l of interest, by re-labelling the training data from task m. Note that, this is similar in spirit to the strategies proposed for cross-corpus emotion recognition in [2], where a hand-crafted mapping between emotional speech labels in various categories and dimensions was defined and applied to a set of corpora prior to joining them to obtain best performance.

2.2. Data Aggregation: Baseline Strategies

Transfer learning is mainly useful when there is little or no labelled training data for the task of interest. However, in the scope of this paper, we can assume enough labelled training data for both domains, and the question is how to combine them for best performance. We can thus treat the problem similar to cross-corpus data aggregation, cf. [2, 24]. As a baseline strategy for the task l, we join the re-labelled training sets from the transfer learning experiment with the original training set of the task l. Furthermore, we can exploit simple semi-supervised learning, exploiting a classifier $\tilde{y}^{(l)} : \mathcal{X} \to \mathcal{Y}^{(l)}$ trained on the original training set of task l.

2.3. Data Aggregation: Cross-Task Labelling

Cross-task labelling (CTL) [14] can be understood as a generalisation of semi-supervised learning to L-dimensional labels, where each dimension corresponds to a classification task. The algorithm is depicted in Figure 1. Starting from labelled training data from various 'isolated' tasks, we construct a 'holistic' database whose feature vectors comprise the union of instances from all the tasks, and whose labels are defined for all instances in all dimensions. In Figure 1, $\mathcal{L}^{(l)}$ denotes the labelled training set of a specific task l. $\mathcal{U}^{(l)}$ comprises all the data where one or more labels for the task l are still missing ($\perp$). In a double nested loop, a form of iterative semi-supervised learning is applied for each task. In the inner loop, refined classifiers $\tilde{y}^{(l)}$ are constructed, based on a training algorithm denoted by Train(). The Select() function returns the indices of instances from the set $\mathcal{U}^{(l)}$, where the classifier outputs have a high confidence. For simplicity, we assume that all instance indices are unique and every instance j belongs to exactly one 'original' task, which is denoted by l_j. For the selected instances, the labels obtained by the classifier $\tilde{y}^{(l)}$ are taken as gold standard and are added to the multi-dimensional labels $\hat{y}_j$ of the holistic database.

Algorithm: *Cross-Task Labelling*
Input: original data sets $\{(\mathbf{x}_i^{(l)}, y_i^{(l)})\}, 1 \leq l \leq L$
Output: cross-labelled data sets
$\{(\mathbf{x}_i^{(l)}, \hat{\mathbf{y}}_i = (\hat{y}_i^{(1)}, \ldots, \hat{y}_i^{(L)})^\top)\}, 1 \leq l \leq L$
Initialisation: **For all** i, l':

$$\hat{y}_i^{(l')} := \begin{cases} y_i^{(l)} & l = l' \\ \perp & \text{otherwise} \end{cases}$$

For $l = 1, \ldots, L$:
  Do:
    $\mathcal{U}^{(l)} := \{\mathbf{x}_i^{(l')} \mid 1 \leq l' \leq L, \hat{y}_i^{(l)} = \perp\}$
    $\mathcal{L}^{(l)} := \{(\mathbf{x}_i^{(l')}, \hat{y}_i^{(l)}) \mid 1 \leq l' \leq L, \hat{y}_i^{(l)} \neq \perp\}$
    $\tilde{y}^{(l)} := \text{Train}(\mathcal{L}^{(l)})$
    $\mathcal{J} := \text{Select}(\mathcal{U}^{(l)}, \tilde{y}^{(l)})$ // according to highest confidence
    For $j \in \mathcal{J}$: $\hat{y}_j^{(l)} := \tilde{y}^{(l)}(\mathbf{x}_j^{(l_j)})$
  While $\mathcal{U}^{(l)} \neq \emptyset$

Figure 1: Pseudocode description of the Cross-Task Labelling (CTL) algorithm.

2.4. Multi-Task Learning

While all the previous methods can be used with standard single-task learning, it seems conducive to combine data aggregation and CTL with multi-task learning via auxiliary features in a classifier chain, similar to [25]. The idea is to append labels for a task m that is expected to be related to task l as features to the vectors $\mathbf{x}^{(l)}$. In case that the feature vector belongs to the training or test set for task l, we have to rely on a hypothesised label $\hat{y}^{(m)}$ obtained by CTL[1]. However, if the feature vector belongs to the cross-task training set (from task m), we can use the gold standard for m ($y^{(m)}$) as attribute.

2.5. Knowledge-Based Instance Selection in CTL

The mapping $h^{m\to l}$, here from deception to sincerity labels and vice versa, can also be interpreted as an expert rule that can be exploited for instance selection in CTL. In doing so, we assume that instances that violate the constraint imposed by $h^{m\to l}$, i. e., $\hat{y}^{(l)}(\mathbf{x}_j^{(m)}) \neq h^{m\to l}(y_j^{(m)})$, should be dropped from the cross-task training set, as they are considered outliers.

Table 1 shows a summary of the learning schemes presented in the previous sub-sections. We depict the pairs of feature vectors and labels used from the training data of the task of interest l, as well as a task m that is used for cross-task data

[1]For determining the auxiliary feature on the test set, we cross-label the test set separately from the training set.

Table 2: *Class distribution among the original instances in the DSD (Deception task), and those from the SSC (Sincerity task) added by means of cross-task labelling (CTL).*

#	Train	Test	Σ	Added by CTL	Σ
D	311	121	432	297	729
ND	747	376	1 123	614	1 737
Σ	1 058	497	1 555	911	2 466

Table 3: *Class distribution among original instances in the SSC, and those from the DSD added by means of CTL.*

#	Train	Test	Σ	Added by CTL	Σ
S	326	151	477	661	1 138
NS	329	105	434	894	1 328
Σ	655	256	911	1 555	2 466

aggregation. In particular, when referring to experiments on the Deception test set, the index l represents Deception and the index m (cross-task) represents Sincerity. Conversely, in the context of the Sincerity task, the roles of l and m are swapped.

3. Corpora

3.1. Deception

In this work, we use the official DECEPTIVE SPEECH DATABASE (DSD) from the *Deception Sub-Challenge* of the ComParE 2016 Challenge [1], which is designed to investigate the manifestation of *inferable deception* in speech. The full set of recordings includes approximately 162 minutes of speech from 72 speakers, leading to a total of 1 555 instances (see Table 2). We use the union of the training and development sets of the Challenge as training set, and the official Challenge test set for evaluation. The recordings were obtained in an empirical study where participants were randomly assigned to one of two groups (thieves and innocents) in a role-playing scenario. Participants were asked to answer truthfully, or lie, in a structured interview, depending on which experimental condition they were in. By the experimental design, the gold standard for each utterance can be defined as a binary non-deceptive (ND) or deceptive (D) label. Note that, participants who failed to behave in accordance with the experimental condition were removed from the data set.

3.2. Sincerity

The SINCERITY SPEECH CORPUS (SSC) as used in the ComParE 2016 Challenge [1] covers the *perceived sincerity* in speech. It contains approximately 72 minutes of speech by 32 speakers and a total of 911 instances (see Table 3). Test subjects were asked to read six predefined sentences, each a form of apology, in various prosodic styles. Each instance was rated in terms of perceived sincerity using an ordinal rating scale by at least 13 annotators (up to a maximum of 19). The ratings were standardised to zero mean and unit standard deviation on a per annotator basis in order to eliminate individual biases. To determine the gold standard for the Challenge, the average sincerity rating was taken across all annotators. For the purpose of this study, the resulting real number was discretised to a binary label: sincere (S) or non-sincere (NS), based on whether the average normalised rating is positive or negative.

Table 4: *Unweighted average recalls (UARs) of the D/ND classes (Deception task) and S/NS classes (Sincerity task), using the learning schemes depicted in Table 1 with SVMs. Note that, the Sincerity baseline is not an official baseline, as it uses discrete instead of continuous labels.*

UAR [%]	Deception	Sincerity
Baseline	68.3	70.9
Cross-task methods		
Transfer learning	50.7	51.2
Data aggregation:		
– direct mapping	66.8	69.4
– semi-supervised learning	69.7	69.1
– CTL, single-task learning	71.8	71.0
– CTL, multi-task learning	**72.2**	**71.3**
– with instance selection	68.9	70.4

4. Experiments and Results

4.1. Acoustic Features

The COMPARE feature set is the same as has been used in the three previous editions of the INTERSPEECH ComParE challenges [26, 27, 28], and contains 6 375 static features resulting from the computation of various functionals over low-level descriptor (LLD) contours. The configuration file is IS13_ComParE.conf, which is included in the 2.1 public release of openSMILE [29]. A full description of the feature set can be found in [30].

4.2. Classifier Training

As classifier, we exclusively use Support Vector Machines (SVMs) in this paper, as they are known to be robust to overfitting in large feature spaces such as the above named feature set. We use the Sequential Minimal Optimisation (SMO) algorithm to train SVMs, as implemented in Weka [31]. In the first set of experiments, we keep the complexity parameter constant at $C = 10^{-4}$ for all learning schemes, which is optimal for the ComParE challenge baseline [1].

Feature standardisation is crucial in cross-task experiments [2, 32]. Here we follow a straightforward scheme that lends itself to utterance-based processing at test time (i.e., no batch processing required). The training set of task l is standardised to zero mean and unit variance. The test set of task l is standardised using the same scales and offsets. Then, the entire data set of task m is standardised separately to zero mean and unit variance prior to applying the labelling schemes (direct mapping, semi-supervised learning or CTL). Finally, the standardised training set and the cross-task data set are jointly used to train a classifier. No further standardisation is applied during training and testing.

Due to the skewed class distribution in the DSD corpus, the training instances of the D class were duplicated so as to reach a more balanced distribution on the training set.

4.3. Cross-Task Labelling Results

Tables 2 and 3 show the number of instances in the original Deception and Sincerity databases, as well as those that were added by the CTL algorithm to each of the classes. If we compare the numbers in the column 'Added by CTL' of Table 2 with the ones we would obtain by direct mapping (from NS to D, as well as from S to ND), it becomes apparent that there is a 'shift' of NS instances (434) away from the D (297) and into the ND (614)

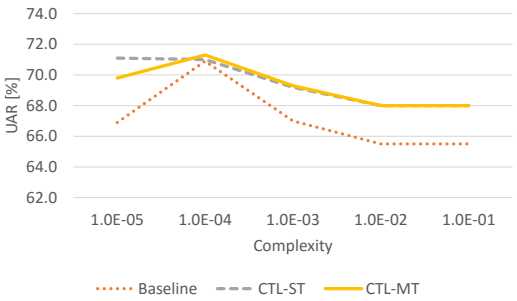

Figure 2: UAR on the Sincerity task by SVM complexity: baseline (official ComParE 2016 training sets), CTL-ST (cross-labelled training set, single-task learning), CTL-MT (cross-labelled training set, multi-task learning).

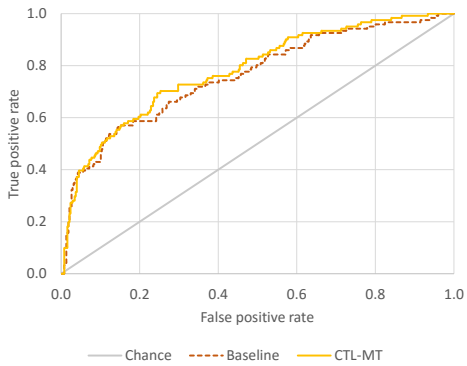

Figure 3: ROC of deception detection using the baseline and CTL-MT methods.

class. It is notable that, a similar trend (with swapped roles of sincerity and deception) is also to be observed in the Sincerity task (Table 3). We can thus conclude that the CTL algorithm performs labelling against the initial assumption of a one-to-one mapping between the classes of the Sincerity and Deception tasks. The classification results below will elucidate which of these 'competing' labellings is preferable in terms of accuracy.

4.4. Classification Performance

Table 4 shows the results in terms of an official ComParE 2016 challenge metric, unweighted average recall (UAR), where the average is taken over the D/ND and S/NS classes, respectively. First, it can be seen that transfer learning results hover around chance level. It is therefore a fitting result that adding the original training set to the transfer learning one, as done by the 'direct mapping' strategy, merely restores the baseline performance on both tasks.

For the Deception task, semi-supervised learning slightly outperforms the baseline, and CTL yields another gain on top. This can be attributed to the iterative self-training present in the CTL method. The best performance is obtained by combining CTL with multi-task learning, reaching 72.2 % UAR on the official test set of the Deception Sub-Challenge[2]. The improvement from the baseline (68.3 % UAR) to the CTL-MT method corresponds to a 12 % relative error rate reduction and is significant at $\alpha = 0.1$ according to a one sided z-test. Finally, the results with knowledge-based instance selection fall considerably behind the CTL-MT UAR. This indicates the importance of instances that are 'inconsistent' as regards their Sincerity and Deception labels.

On the Sincerity task, the baseline is considerably harder to outperform by our data aggregation methods. One difference in the nature of the SSC and DSD corpora is the speech style (read vs spontaneous), and in future work we might use the prosodic style annotation of the SSC corpus to further investigate into how this may affect performance. Besides, we investigated whether this result could be due to a sub-optimal tuning of the SVM complexity parameter. Figure 2 shows the UAR obtained with various complexity parameters set in SVM training. It can be seen that, while the optimal choice of C for the CTL-ST/-MT methods leads to slightly better performance than the optimal choice of C for the baseline, for other values of C the difference is more substantial. In particular, the superior performance of CTL-ST/-MT at $C = 10^{-5}$ provides evidence of added value of the cross-labelled instances when the optimisation algorithm is forced to select fewer support vectors.

[2]Our results run out of competition since participants can perform only five evaluations on the test set and part of the authors are organisers.

4.5. Deception Detection Performance

Motivated from possible applications of deception detection in crime prevention and forensics, and the fact that deceptive speech is less prevalent in the present data set than non-deceptive speech, let us now investigate the performance of deception detection using the ComParE baseline as well as the CTL-MT method. The resulting receiver operating characteristic (ROC) curve can be seen in Figure 3. In line with the results above, CTL-MT yields superior or equivalent true positive rates at a given false positive rate w.r.t. the baseline, resulting in areas under the ROC curve (AUC) of 0.781 (CTL-MT) vs 0.760 (baseline).

5. Conclusions

We have presented a study on the fully automatic classification of deception and sincerity in speech, where the methodology was motivated by the assumption that these tasks are correlated. The hypothesis that de facto deception and perceived sincerity are 'two sides of the same coin' was falsified in its generality by the inferior performance of simple relabelling based on this hypothesis. In contrast, the performance analysis of the CTL method demonstrates that an iteratively self-trained classifier is able to select 'appropriate' instances from another task that complement the available intra-corpus training data, and that these labels need not match the intuitively expected ones. The performance of CTL reported in this study is all the more notable since CTL was previously not able to improve performance over standard single-task classification [14], which indicates that there is a significant amount of interplay and synergies between deception and sincerity. In future work, we will deepen the performance analysis as regards the features that differ between the 'well-behaved' instances (for which the label mapping assumption is true) and the others. Furthermore, we will extend our work to machine learning methods which are better capable of exploiting the multi-dimensional target labels in the holistic database obtained by CTL, such as deep neural networks.

6. Acknowledgements

The research leading to these results has received funding from the European Union's Framework Programme for Research and Innovation HORIZON 2020 under the Grant No. 645378 (ARIA-VALUSPA).

7. References

[1] B. Schuller, S. Steidl, A. Batliner, J. Hirschberg, J. K. Burgoon, A. Baird, A. Elkins, Y. Zhang, E. Coutinho, and K. Evanini, "The INTERSPEECH 2016 Computational Paralinguistics Challenge: Deception & Sincerity," in *Proc. of INTERSPEECH*. San Francsico, CA: ISCA, September 2016, 5 pages.

[2] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: Variances and strategies," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 119–131, July-December 2010.

[3] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014.

[4] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, pp. 1–13, 2007.

[5] B. Schuller, M. Wöllmer, F. Eyben, G. Rigoll, and D. Arsić, "Semantic speech tagging: Towards combined analysis of speaker traits," in *Proc. of the Audio Engineering Society Conference (AES)*. Ilmenau, Germany: Audio Engineering Society, July 2011, pp. 89–97.

[6] F. Eyben, M. Wöllmer, and B. Schuller, "A multi-task approach to continuous five-dimensional affect sensing in natural speech," *ACM Transactions on Interactive Intelligent Systems, Special Issue on Affective Interaction in Natural Environments*, vol. 2, no. 1, March 2012, 29 pages.

[7] B. Schuller, Y. Zhang, F. Eyben, and F. Weninger, "Intelligent systems' Holistic Evolving Analysis of Real-life Universal speaker characteristics," in *Proc. of the 5th International Workshop on Emotion Social Signals, Sentiment & Linked Open Data (ES^{3}LOD 2014), satellite of LREC*. Reykjavik, Iceland: ELRA, May 2014, pp. 14–20.

[8] X. Zhu, "Semi-supervised learning literature survey," Department of Computer Sciences, University of Wisconsin at Madison, Madison, WI, Tech. Rep. TR 1530, 2006.

[9] O. Chapelle, B. Schölkopf, A. Zien *et al.*, *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006.

[10] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[11] B. Schuller, G. Rigoll, and M. Lang, "Hidden markov model-based speech emotion recognition," in *Proc. of of ICME*, vol. I. Baltimore, MD: IEEE, 2003, pp. 401–404.

[12] K. Bousmalis, M. Mehu, and M. Pantic, "Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools," *Image and Vision Computing*, vol. 31, no. 2, pp. 203–221, 2013.

[13] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, no. 2, pp. 15–21, 2013.

[14] Y. Zhang, Y. Zhou, J. Shen, and B. Schuller, "Semi-autonomous data enrichment based on cross-task labelling of missing targets for holistic speech analysis," in *Proc. of ICASSP*. Shanghai, P. R. China: IEEE, March 2016, 5 pages.

[15] J. F. Gemmeke, B. Cranen, and U. Remes, "Sparse imputation for large vocabulary noise robust ASR," *Computer Speech & Language*, vol. 25, no. 2, pp. 462–479, 2011.

[16] 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.

[17] M. Graciarena, E. Shriberg, A. Stolcke, F. Enos, J. Hirschberg, and S. Kajarekar, "Combining prosodic lexical and cepstral systems for deceptive speech detection," in *Proc. of ICASSP*, vol. 1. Toulouse, France: IEEE, May 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 *Proc. of Interspeech*. Lisbon, Portugal: ISCA, September 2005, pp. 805–809.

[19] M. Green, "How do speech acts express psychological states?" *John Searles Philosophy of Language: Force, Meaning and Thought*, pp. 267–84, 2007.

[20] J. Eriksson, "Straight talk: conceptions of sincerity in speech," *Philosophical studies*, vol. 153, no. 2, pp. 213–234, 2011.

[21] A. Fetzer, "Sincerity and credibility in political interviews," *Politics as Text and Talk: Analytic approaches to political discourse*, p. 173, 2002.

[22] S. Steidl, M. Levit, A. Batliner, E. Nöth, and H. Niemann, ""of all things the measure is man": Automatic classification of emotions and inter-labeler consistency," in *Proc. of ICASSP*, Philadelphia, PA, 2005, pp. 317–320.

[23] B. Schuller and A. Batliner, *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Wiley, November 2013.

[24] Z. Zhang, F. Weninger, M. Wöllmer, and B. Schuller, "Unsupervised learning in cross-corpus acoustic emotion recognition," in *Proc. of Automatic Speech Recognition and Understanding Workshop (ASRU)*. Big Island, HY: IEEE, December 2011, pp. 523–528.

[25] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Machine learning*, vol. 85, no. 3, pp. 333–359, 2011.

[26] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. 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 *Proc. of INTERSPEECH*. Lyon, France: ISCA, August 2013, pp. 148–152.

[27] B. Schuller, S. Steidl, A. Batliner, J. Epps, F. Eyben, F. Ringeval, E. Marchi, and Y. Zhang, "The INTERSPEECH 2014 Computational Paralinguistics Challenge: Cognitive & physical load," in *Proc. of INTERSPEECH*. Singapore, Singapore: ISCA, September 2014, pp. 427–431.

[28] B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hönig, J. R. Orozco-Arroyave, E. Nöth, Y. Zhang, and F. Weninger, "The INTERSPEECH 2015 Computational Paralinguistics Challenge: Degree of Nativeness, Parkinson's & Eating Condition," in *Proc. of INTERSPEECH*. Dresden, Germany: ISCA, September 2015, pp. 478–482.

[29] F. Eyben, F. Weninger, F. Groß, and B. Schuller, "Recent developments in openSMILE, the Munich open-source multimedia feature extractor," in *Proc. of ACM MM*. ACM, 2013, pp. 835–838.

[30] F. Weninger, F. Eyben, B. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: What speech, music and sound have in common," *Frontiers in Emotion Science*, vol. 4, no. 292, pp. 1–12, May 2013.

[31] 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.

[32] A. Hassan, R. Damper, and M. Niranjan, "On acoustic emotion recognition: Compensating for covariate shift," *IEEE Transactions on Audio, Speech, Language Processing*, vol. 21, no. 7, pp. 1458–1468, 2013.