



Elicitation Design for Acoustic Depression Classification: An Investigation of Articulation Effort, Linguistic Complexity, and Word Affect

Brian Stasak^{1,2}, Julien Epps^{1,2}, and Roland Goecke³

¹University of New South Wales, Australia

²Data61-CSIRO, Sydney, Australia

³Human-Centred Technology, University of Canberra, Canberra, Australia

b.stasak@unsw.edu.au, j.epps@unsw.edu.au, roland.goecke@ieee.org

Abstract

Assessment of neurological and psychiatric disorders like depression are unusual from a speech processing perspective, in that speakers can be prompted or instructed in what they should say (e.g. as part of a clinical assessment). Despite prior speech-based depression studies that have used a variety of speech elicitation methods, there has been little evaluation of the best elicitation mode. One approach to understand this better is to analyze an existing database from the perspective of articulation effort, word affect, and linguistic complexity measures as proxies for depression sub-symptoms (e.g. psychomotor retardation, negative stimulus suppression, cognitive impairment). Here a novel measure for quantifying articulation effort is introduced, and when applied experimentally to the DAIC corpus shows promise for identifying speech data that are more discriminative of depression. Interestingly, experiment results demonstrate that by selecting speech with higher articulation effort, linguistic complexity, or word-based arousal/valence, improvements in acoustic speech-based feature depression classification performance can be achieved, serving as a guide for future elicitation design.

Index Terms: Affect, computational paralinguistics, depression classification, sentiment, speech production

1. Introduction

As the prevalence of depression disorders continues to rise, new automated techniques for assessing these illnesses will be in high demand. Patient interviewing skills differ amongst physicians, and medical literature concurs that ineffective interviewing techniques result in less complete knowledge of patients' backgrounds and their major concerns [1-5]. Research asserts the quantity and quality of information gathered is greatly dependent on physician's unique elicitation style, range of pertinent enquiries, and individual protocol guidelines [6]. For instance, studies have demonstrated that open-ended questions (i.e. "Tell me about your family") rather than closed questions (i.e. "Do you like your family") gather more pertinent information about patients' health status [7, 8].

Examples of speech elicitation in automated assessment and monitoring systems have included smartphone prompts or a virtual human interviewer. An interactive, objective virtual human interview assessment tool [9-13] has several practical advantages in the diagnosis of depression disorders. For example, a virtual human can provide a non-biased, uniform, automated prescreening assessment before a patient arrives to meet with his/her physician. This can help reduce medical

costs, prioritize at-risk patients, and facilitate expedited diagnosis of clinical depression - potentially resulting in a decrease in the number of annual suicides. There are several other dialogue advantages to the automated virtual human interviewer: effective open-ended enquires without typical clinical session time constraints; questions using consistent verbal and non-verbal behaviors, thus imparting less influence on a patient [14]; and attuning enquiry stylization to better associate with different patients' backgrounds [6, 10, 15, 16].

For speech-based depression analysis, there are five types of commonly accepted non-invasive elicitation methods: (1) articulatory-based speech (e.g. single phoneme, held-vowel, word-level) [13, 17]; (2) non-speech (e.g. diadochokinesis patterns, nonsense words) [13]; (3) automatic speech (e.g. alphabet, counting) [13, 18]; (4) read speech (e.g. read excerpt aloud) [19, 20]; and (5) spontaneous speech (e.g. conversational, unrehearsed) [13, 19, 21]. Studies have yet to indicate which elicitation method(s) produces the most reliable acoustic speech-based depression classification performance [9, 13, 19, 22]. However, it is recognized that spontaneous speech situations permit a wider range of natural affect, topic content, and organizational cognitive prerequisites than the more limited aforementioned verbal tasks [9, 23]. Further, literature has also discovered that non-speech, automatic speech, and read speech do not activate the identical areas of the brain as found during natural spontaneous speech [23].

While prior studies have shown acoustic features can be used as an indicator for depression [9]; there has been little investigation into aspects of articulation effort as a depression indicator, despite its likely link with psychomotor retardation. One impediment could be the lack of quantified descriptors of articulation effort. In addition, for depression classification, the application of linguistic complexity based on transcripts, and its relation to the acoustic speech signal has received minimal attention, despite its possible link with cognitive impairment. However, more recently in [21], it has been strongly advocated that all speech-based depression analysis should integrate speech-to-text linguistic analysis due to its considerable depression classification performance.

It is believed that affect will be an important part of speech-based depression analysis and identification. For example, in [24], it was discovered that depressed patients neglect to suppress pessimism when relating to negative stimuli. This abnormal change in affect caused by a depressive state could be quantitatively measured by examining affective dimensions, such as arousal and valence, using a word-by-word rating scale. While a word-level affect approach is quite common in data mining sentiment analysis [25], its application is emerging quickly in other areas, such as medicine [26].

In this paper, we study the use of three different measures to help select speech segments for acoustic depression classification, with the following hypotheses: (1) speech representing higher overall articulatory effort or linguistic complexity will provide increased discrimination of depressed speakers; and (2) phrases containing lower valence/arousal word affect can provide increased discrimination of depressed speakers.

2. Experiment Data

A subset of the training and development from the Distress Analysis Interview Corpus (DAIC) [27] was used for all experiments. The DAIC ‘Wizard-of-Oz’ subset utilized an animated virtual interviewer and its data collection was explicitly created to assist the diagnosis of psychological distress disorders (e.g. anxiety, depression, post-traumatic stress). This experimental subset was chosen because it has a reasonable number of speakers (82 male/female), natural speech in a clinical room environment, a virtual human interviewer, a finite number of patient enquiries, phrase-level spoken transcriptions with time stamps, and a Patient Health Questionnaire (PHQ-8) score per participant.

For help in the diagnosis of depression disorders, physicians commonly use the eight-questioned PHQ-8 self-assessment [28]. The PHQ-8 score is between 0 and 24, where larger scores imply a greater depression severity. In research herein, two different classification speaker groups were analyzed based on a limited PHQ-8 score range: non-depressed (0-4) and depressed (15-24). According to the PHQ-8, the non-depressed group is considered to have “no current depression”, whereas the depressed group has “current moderately severe to severe depression”. Note that approximately 20% of speakers in this experimental database were clinically depressed. See [27] for more details about the DAIC specific speaker content and transcription conventions.

3. Methods

3.1. Articulation Effort Measures

It is generally agreed that American English has between 38 to 46 phonemes depending on the regional accent [29]. During conversational speech, the respiratory system, larynx, velum, jaw, tongue, and lip motor are essentially controlled in an unconscious manner just to generate just a single phrase. For more than a century, there have been studies describing which phonemes and/or phonemic syllables are more complex than others [30, 31]. Without delving into too much articulatory detail (cf. [32, 33]), the effort complexity of consonants depends on several factors, such as the formant transition from the preceding vowel, formant location, and voicing type.

The production of voiced consonants is generally more complex than voiceless consonants. Thus, due to their reduced physical motor planning effort, it would seem that voiceless sounds occur earlier in age and in greater numbers when compared with voiced sounds. Indeed studies across many different languages have indicated that voiceless plosives occur more frequently than voiced ones [34]; and further, that there is a highly significant positive correlation between a larger consonant phoneme inventory size and greater syllable complexity across languages throughout the world [35].

For decades, speech pathologists have recognized that several phonemes in the English language require more time

to master than others [30, 36, 37] (see Figure 1) and adult-like phonemic recognition/categorization mastery is achieved beyond early childhood, especially with consonants [38].

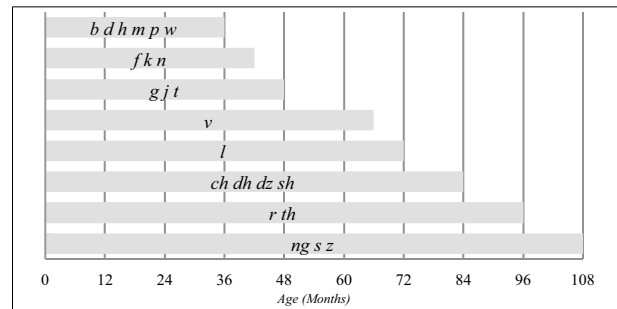


Figure 1: Recommended age at which 90% of children acquire articulation mastery of English consonants based on [37].

Numerous studies [9, 39, 40] have indicated unusual pause phonation times, reduced prosody, lax articulation and sluggish verbal motor coordination (e.g. psychomotor retardation) in individuals with depression disorders. Moreover, [41, 42] found significant changes in motor articulation transitions in depressed speakers, including poor intelligibility and imprecise consonant production. Due to the aforesaid articulatory abnormalities exhibited by depressed speakers, a phoneme-related articulation effort measure is proposed to help identify these speakers.

Our proposed novel indicator of articulation effort is based on the age of articulation mastery for males shown previously in Figure 1. For phoneme effort scoring, vowels/diphthongs were given a score of 0, whereas the eight different consonant groupings were given a score based on the age of mastery as indicated in Figure 1. A phonetic dictionary was utilized to convert all DAIC transcript text into English phoneme representations. This allowed proper scoring based on spoken phoneme rather than letters (i.e. ‘three’ as ‘TH-R-IY’) as further illustrated by Figure 2.

During experiments, each speaker’s articulation effort measures per phrase were averaged to create a non-partitioned feature set. In addition, the mean, standard deviation, 50th-percentile mean grouped, and variance were calculated per speaker transcript for consonants only, allowing for disparate utterance lengths to be equally compared. Partitioned experiments used the mean per-utterance measure to separate the data into four partitions of low, mid-low, mid-high, and high articulation effort values.

Text Transcript		"It is nice talking to you."															
Phoneme	IH	T	IH	Z	N	AY	S	T	AO	K	IH	NG	T	UW	J	UW	
Art. Effort Value	0	48	0	108	42	0	108	48	0	42	0	108	48	0	48	0	
	Mean 66.66								Std. Dev. 29.31								

Figure 2: Articulation effort measure values for a spoken phrase, phonetic conversion, and corresponding articulation effort values; blue indicates only consonants calculated.

3.2. Linguistic Complexity Measures

Articulatory effort produces an acoustic signal, which is intrinsically linked to quantitative linguistic representations. It is well established that aspects of high-level language (e.g. lexical choice, syntax, pragmatics) are affected by depression [43-45]. While computational text-processing techniques were originally developed for written language, they have been

successfully transitioned for application on spoken language [21, 45]. For instance, in [21], features derived from semantic analysis of spoken transcripts performed better than acoustic and video features for depression recognition.

For research herein, linguistic measures, such as lexical sophistication, syntax phrase indices, grammar structure, and genre-related language were extracted from each transcript by using Simple Natural Language Processing Tool (SiNLP) [46], Tool for the Automatic Analysis of Lexical Sophistication (TAALES) [47], and Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC) [48]. Note that this is the first research to apply these particular toolkits for depression classification, and all transcripts were processed in original text format.

For experiments, each speaker’s linguistic complexity measures per phrase were averaged to create a non-partitioned feature set. During the partitioned experiments, we used a sorted mean per-utterance SiNLP word letter length measure to separate the data into four partitions of low, mid-low, mid-high, and high linguistic complexity values. Afterwards, per partition, each speaker’s linguistic complexity values were averaged to obtain a partitioned feature set.

3.3. Word Affect Measures

While word affect ratings are a qualitative measure, they can be useful in broadly interpreting text-based content. From a speaker’s transcript, text-based sentiment analysis can provide insight into broad emotional context expressed. In [45], it was shown that word affect text processing information can be transformed into a feature to help recognize individuals with depression. Moreover, they found that word affect features, when complemented by additional types of features (i.e. linguistic) can increase depression recognition performance.

For research herein, the Sentiment Analysis and Cognition Engine (SEANCE) [49] was used to extract word affect-based features per individual transcript phrases. Several different affective word-rating references are included in SEANCE, such as the Affective Norms for English Words (ANEW), EmoLex, Laswell, and Hu Lui Polarity. Note that this is the first research to apply these particular toolkits for depression classification, and all transcripts were processed in original text format.

For experiments, each speaker’s word affect measures per phrase were averaged to create a non-partitioned feature set. During partitioned experiments, we used a sorted mean per-utterance SEANCE ANEW arousal/valence word affect measure to separate the data into four partitions of low, mid-low, mid-high, and high based on arousal/valence values. Afterwards, per partition, each speaker’s word affect values were averaged to obtain a partitioned feature set.

3.4. Acoustic System Configuration

As explained in Section 1, the focus of this experimental work involves acoustic speech-based data selection by leveraging text-based measures (e.g. articulation effort, linguistic complexity, word affect) using partitioned data analysis (see Figure 3). However, comparatively, non-partitioned experiment results were also included. For experiments herein, the open-source COVAREP speech toolkit [50] was used to extract 74 acoustic features including glottal flow, wavelet phase, MFCC, pitch, and creakiness. Each feature had its mean, standard deviation, kurtosis, and skewness calculated

by aggregating 10-ms frame-level features across individual utterances. COVAREP was chosen as an acoustic feature set since it is the baseline for the AVEC 2016 challenge [21], and other depression analysis research.

Similarly to [51, 52], depression classification was conducted using decision trees, which performed well in preliminary experiments. All experiments used the MATLAB simple decision tree classification toolkit using a few leaves with coarse decision-making and a maximum of 4 splits. Experiments utilized 2-class classification (non-depressed/depressed) with 5-fold cross validation using a 20/80 training/test split.

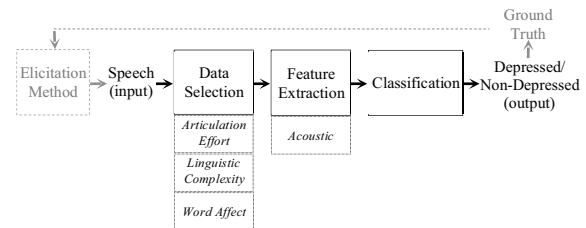


Figure 3: Experimental design for comparing data selection measures for acoustic depression classification. The dashed line suggests the implications of the experimental results for elicitation design.

4. Results and Discussion

It was first hypothesized that depressed speakers will use words/phrases with phonemes that take less effort and/or demonstrate a decrease in articulatory precision - thus, resulting in better discrimination of these speakers. Evidently, partitioned results shown in Figure 4 confirm that using acoustic features from phrases with greater articulation effort or linguistic complexity produce higher depression classification results than using all utterances. Concerning affect, contrary to our second hypothesis, it seems that words with higher arousal and/or more positive valence are more discriminative. This could be due to depressed speakers’ difficulty producing more energetic (aroused) or happier (more positive valence) speech than non-depressed speakers. In partitioned experiment results shown in Figure 4, absolute gains in depression classification were recorded when compared to the non-partitioned acoustic-based results shown in Figure 5; but, comparison is not straightforward because these accuracies are from different data subsets.

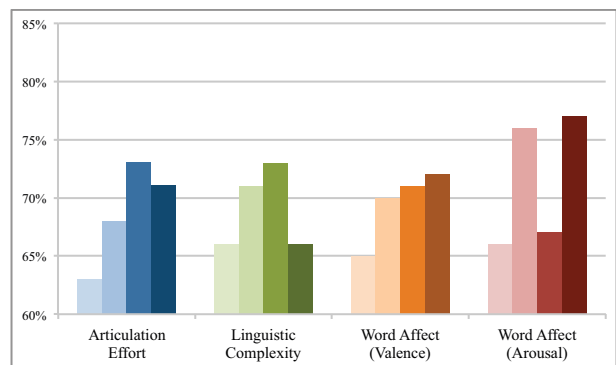


Figure 4: COVAREP acoustic features average depression classification accuracy results per measure and partition shown from left to right: low, mid-low, mid-high, and high.

Although partitions shown in Figure 4 were defined by retaining near-equal numbers of phrases per partition, there was variation in the total phrase duration per partition. Based on a more detailed investigation, this did not seem to have any bearing on low, mid-low, and mid-high results. However, among measures shown in Figure 4, the high linguistic complexity partition had the smallest duration, which might explain its decrease in classification accuracy. Additionally, in Figure 4, the decrease in the mid-high word affect arousal partition could be related to the narrower range of its partition.

The DAIC transcripts are limited because they do not contain exact phonetic representations. Albeit a speaker could pronounce a word with a deleted, added, or substituted phoneme/s, but the transcript does not indicate these speech attributes. For example, if a speaker said ‘*runnin*’, the DIAC transcripts would only render ‘*running*’. Ergo, a major advantage of referencing text-based measure values onto the *acoustic* features is that if abnormal articulatory production and/or affective states manifest in depressed speakers as complexity increases, these characteristics will be recorded acoustically. Currently, commercial automatic speech recognition may have sufficient performance to justify a text-based analysis application in combination with acoustic speech-based features for depression classification. Using the DAIC corpus, we have demonstrated that text-transcripts need only be an approximation of what was said to generate effective text-based measures as features and/or to use as data selection references for acoustic features.

Non-partitioned experiments comparing acoustic-based features versus text-based features are shown in Figure 5. Similarly to [21], the linguistic complexity features yielded superior classification accuracy than the acoustic-based features. The articulation effort feature performed surprisingly well despite being only 4-dimensional. While the SEANCE word affect features gave moderately low classification accuracy, this seemed to be because eight different affect rating references were included in this feature set (i.e. ANEW, EmoLex, GALC) and overfitting occurred. By contrast and also shown in Figure 5, if only single affect word-rating reference, such as the General Index were applied as features, depression classification performance could achieve 82% accuracy. It is important to note that accuracies based on linguistic features in Figure 5 may be optimistic relative to an automatic transcript.

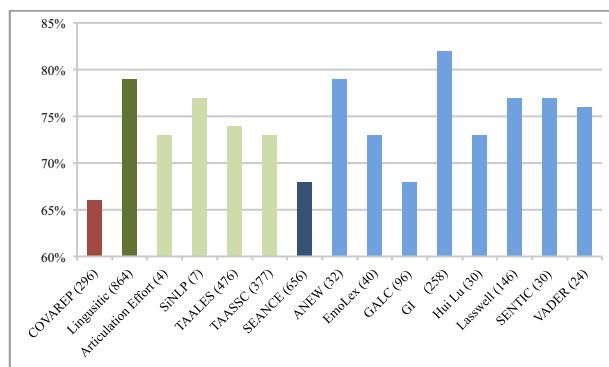


Figure 5: Average depression classification accuracy results for acoustic-based (red), linguistic-based (green), and affect-based (blue) features; the dark shade indicates results for all similar feature sets combined, while the number of total features are indicated in parenthesis.

5. Conclusions

The application of acoustic speech-based analysis in conjunction with text-based processing shows considerable promise for guiding clinical elicitation protocol design and interview speech analysis. For depression classification, we proposed analysis of acoustic speech-based feature segments wherein higher articulation effort, linguistic complexity, and word affect occurs. In particular, the newly derived articulation effort measure experimented on produced enhanced gains in depression classification performance along with other text-based measures.

6. Future Work

Based on our experimental results, more research on novel articulation effort features is warranted, such as exploring the 17 distinctive articulatory properties found in [31] as additional articulation effort measures. Also, a more extensive investigation of sorted measures and/or partitions could potentially generate better depression classification results – for research herein, only 4 basic measures (i.e. mean per utterance articulation effort, mean word letter length, arousal, valence) were examined of the 1,676 total possible. Furthermore, exploring various feature fusion methods of acoustic, articulation effort, linguistic complexity, and word affect measures is of high interest.

7. Acknowledgements

The work of Brian Stasak, Julien Epps, and Roland Goecke was partly funded by ARC Discovery Project DP130101094. Thanks to fellow UNSW PhD colleague Zhaocheng Huang.

8. References

- [1] Enelow, A.J., & Swisher, S.N., *Interviewing & Patient Care*, New York: Oxford University Press, 1979.
- [2] Ong, L., De Haes, J., Hoos, A., & Lammes, F., ‘Doctor-patient communication: a review of the literature’, *Social Science & Medicine*, 40(7), pp. 903-918, 1995.
- [3] Ani, C., Bazargan, M., Hindman, D., Bell, D., Farooq, M., Akhanjee, L., Yemofio, F., Baker, R., & Rodriguez, M., ‘Depression symptomatology and diagnosis: discordance between patients and physicians in primary care settings’, *BMC Family Practice* 2008, Vol. 9(1), Jan. 2008.
- [4] McCabe, R., Healey, P., Priebe, S., Lavelle, M., Dodwell, D., Laugharne, R., Snell, A., & Bremner, S., ‘Shared understanding in psychiatrist-patient communication: association with treatment adherence in schizophrenia’, *Pat. Ed. & Coun.*, 2013a.
- [5] McCabe, R., Khanom, H. Bailey, P. & Priebe, S., ‘Shared decision-making in ongoing outpatient psychiatric treatment’, *Pat. Ed. and Coun.*, 92(3), pp. 326-328, 2013b.
- [6] Howes, C., Purver, M., & McCabe, ‘Linguistic indicators of severity and progress online text-based therapy for depression’, *Workshop on Comp. Ling. and Clinical Psych.: From Ling. Signal to Clinic. Reality*, Baltimore, MD – USA, pp. 7-16, 2014.
- [7] Roter, D.L., Hall, & Hall, J.A., ‘Physicians’ interviewing styles and medical information obtained from patients’, *Journal of General Internal Medicine*, Vol. 2(Sept/Oct), pp.325-329, 1987.
- [8] Memon, A., & Bull, R., ‘The cognitive interview: its origins, empirical support, evaluation, and practical implications’, *J. of Community & Applied Social Psych.*, Vol. 1, pp. 291-307, 1991.
- [9] Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T., ‘A review of depression and suicide risk assessment using speech analysis’, *Speech Communication*, Vol. 71, pp. 10-49, 2015.
- [10] DeVault, D., Georgila, K., Artstein, R., Morbini, F., Traum, D., Sherer, S., Rizzo, A., & Morency, L., ‘Verbal indicators of

- psychological distress in interactive dialogue with a virtual human', in *Proc. of the SIGDIAL Conf.*, pp. 193-202, 2013.
- [11] DeVault, D., Artstein, R., Benn, G., et al., L., 'SimSensei kiosk: a virtual human interviewer for healthcare decision support', *Proc. of the 2014 Intern. Conf. on Auton. Agents and Multi-agent Sys.*, Paris – France, pp. 1961-1968, 2014.
- [12] DeVault, D., Mell, J., & Gratch, J., 'Toward natural turn-taking in a virtual human negotiation agent', *Assoc. for the Advancement of Artificial Intelligence*, 2015.
- [13] Mundt, J., Snyder, P., Cannizzaro, M., Chappie, K., and Geralts, D., 'Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology', *J. Neuroling.*, Vol. 20(1), pp. 50-64, 2007.
- [14] Scherer, S., Hammal, Z., Yang, Y., Morency, L., & Cohn, J., 'Dyadic Behavior Analysis in Depression Severity Assessment Interviews', in *Proc. of the ACM Intern. Conf. on Multimodal Interaction*, 2014.
- [15] Yu, Z., Scherer, S., DeVault, D., Gratch, J., Stratou, G., Morency, L., & Cassell, J., 'Multimodal prediction of psychological disorder: learning nonverbal commonality in adjacency pairs', in *Proc. of the SemDial 2013 Workshop*, pp. 193-202, 2013.
- [16] Yang, Y., Fairbairn, C., & Cohn, J., 'Detecting depression severity from vocal prosody', *IEEE Trans. on Affective Computing*, pp. 1-8, 2013.
- [17] Kiss, G., & Vicsi, K., 'Physiological and cognitive status monitoring on the base of acoustic-phonetic speech parameters', L. Besacier et al. (eds.), *SLSP*, pp. 120-131, 2014.
- [18] Szabadi, E., Bradshaw, & C., Besson, J., 'Elongation of pause-time in speech: a simple, objective measure of motor retardation in depression', *Br. J. Psych.*, 1976.
- [19] Alghowinem, S., Goecke, R., Wagner, M., Epps, J., Breakspear, M., & Parker, G., 'Detecting depression: a comparison between spontaneous and read speech', *ICASSP*, pp. 7547-7551, 2013.
- [20] Moore, E., Clements, M., Peifer, J., & Weisser, L., 'Comparing objective feature statistics of speech for classifying clinical depression', in *26th Annual Intern. Conf. of the IEEE Eng. in Med. and Bio. Society, IEMBS '04*, pp. 17-20, 2004.
- [21] Williamson, J.R., Godoy, E., Cha, M., Schwarzenruber, A., Khorrami, P., Gwon, Y., Kung, H.T., Dagli, C., & Quatieri, T.F., 'Detecting depression using vocal, facial, and semantic communication cues', *AVEC '16*, Amsterdam, pp. 11-18, 2016.
- [22] Alpert, M., Pouget, E., & Silva, R., 'Reflections of depression in acoustic measures of the patient's speech', *J. of Affective Disorders*, 2001.
- [23] Bunton, K., 'Speech versus nonspeech: different task, different neural organization', *Semin Speech Lang.*, Vol. 29(4), pp. 267-275, Nov. 2008.
- [24] Goeleven, E., De Raedt, R., Baert, S., & Koster, E., 'Deficient inhibition of emotion information in depression', *J. of Affective Disorders*, Vol. 93, Issue 1-3, pp. 149-157, 2006.
- [25] Pang, & Lee, L., 'Opinion mining and sentiment analysis', *Found. and Trends in Inform. Retrieval*, 2(1-2), pp. 1-135, 2008.
- [26] Denecke, K., & Deng, Y., 'Sentiment analysis in medical settings: new opportunities and challenges', *Artificial Intelligence in Medicine*, Vol. 64, Issue 1, pp. 17-27, May 2015.
- [27] Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., Wood, R., Boberg, J., DeVault, D., Marsella, S., Traum, D., Rizzo, S., & Morency, L., "The distress analysis interview corpus of human and computer interviews", *LREC*, 3123-3128, 2014.
- [28] Kroenke, K., Strine, T., Spitzer, R., Williams, J., Berry, J., & Mokdad, A., "The PHQ-8 as a measure of current depression in general population", *J. of Affective Disorders*, Vol. 114, 163-173, 2009.
- [29] Ladeford, P. & Disner, S., *Vowels and Consonants* (3rd Ed.), John Wiley & Sons, West Sussex - UK, 2012.
- [30] Wellman, B., Case, I., Mengert, I., & Bradbury, D., 'Speech sounds of young children', University of Iowa Study, Child Welfare, Vol. 5(2), pp. 1-82, 1931.
- [31] Chomsky, N., & Halle, M., *The Sound Pattern of English*, Harper & Row, New York, NY - USA, 1968.
- [32] Hardcastle, W.J., *Physiology of Speech Production*, London: Academic Press, 1976.
- [33] Ladefoged, P., & Maddieson, I., *The Sounds of the World's Languages*, Oxford: Blackwell, 1996.
- [34] Zipf, G., *The Psycho-Biology of Language*, Houghton Mifflin, Boston, MA - USA, 1935.
- [35] Maddieson, I., 'Correlating phonological complexity: data and validation', *Linguistic Typology*, Vol. 10, pp. 108-125, 2006.
- [36] Sander, E.K., 'When are speech sounds learned?', *J. of Speech and Hear. Disord.*, Vol. 37, pp. 55-63, 1972.
- [37] Smit, A., Hand, L., Freilinger, J., Bernthal, J., & Bird, Ann, 'The Iowa articulation norms project and its Nebraska replication', *J. of Speech and Hear. Disord.*, Vol. 55, pp. 779-798, Nov. 1990.
- [38] Hazan, V., & Barrett, S., "The development of phonemic categorization in children aged 6-12", *J. of Phonetics*, Vol. 28(4), pp. 377-396, Oct. 2000.
- [39] Bennabi, D., Vandel, P., Papaxanthis, C., Pozzo, T., & Haffen, E., 'Psychomotor retardation in depression: a systematic review of diagnostic, pathophysiological, and therapeutic implications', *BioMed Research International*, Vol. 2013, 2013.
- [40] Scherer, K., & Zei, B., 'Vocal indicators of affective disorders', *Psychother Psychosom*, Vol. 49, pp. 179-186, 1988.
- [41] Flint, A., Black, S., Campbell-Taylor, I., Gailey, G., & Levinton, C., 'Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression', *J. Psychiat. Res.*, Vol. 27, No. 3, pp. 309-319, 1993.
- [42] Darby, J., Simmons, N., Berger, P., 'Speech and voice parameters of depression: a pilot study', *J. Commun. Disord.*, Vol. 17, pp. 75-85, 1984.
- [43] Andreasen, N., & Pfohl, B., 'Linguistic analysis of speech in affective disorders', *Arch. Gen. Psych.*, Vol. 33, pp. 1361-1367, Nov. 1976.
- [44] Rude, S, Gortner, E., & Pennebaker, J., 'Language use of depressed and depressed-vulnerable college students', *Cognition and Emotion*, pp. 1121-1133, 2004.
- [45] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E., 'Predicting depression via social media', Proc. of the 7th Intern. AAAI Conf. on Weblogs and Social Media, 2013.
- [46] Crossley, S.A., Allen, L.K., Kyle, K., & McNamara, D.S., 'Analyzing discourse processing using a simple natural language processing tool (SiNLP)', *Discourse Processes*, 51(5-6), pp. 511-534, 2014.
- [47] Kyle, K., & Crossley, S., 'Automatically assessing lexical sophistication: indices, tools, findings, and application', *TESOL Quarterly*, 49(4), pp. 757-786, 2015.
- [48] Kyle, K., 'Measuring syntactic development in L2 writing: fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication', (Doctoral Dissertation), 2016.
- [49] Crossley, S.A., Kyle, K., and McNamara, D.S., 'Sentiment analysis and social cognition engine (SEANCE): an automatic tool for sentiment, social cognition, and social order analysis', *Behavior Research Methods*, in press.
- [50] Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, S., 'COVAREP – a collaborative voice analysis repository for speech technologies', in *Acoustics, Speech and Signal Processing (ICASSP)*, pp. 960-964, 2014.
- [51] Mitra, V., Shriberg, E., McLaren, M., Kathol, A., Richey, C., Vergyri, D., & Graciarena, M., 'The SRI AVEC-2014 evaluation system', *AVEC '14*, Orlando, FL – USA, pp. 93-100, Nov. 2014.
- [52] Nadeem, M., Horn, M., Coppersmith, G., & Sen, S., 'Identifying Depression on Twitter', preprint *arXiv: 1607.07384*, July 2016.