On the Significance of Speech Pauses in Depressive Disorders: Results on Read and Spontaneous Narratives

Anna Esposito, Antonietta M. Esposito, Laurence Likforman-Sulem, Mauro N. Maldonato and Alessandro Vinciarelli

Abstract This paper investigates whether and how depressive disorders affect speech and in particular timing strategies for speech pauses (empty and filled pauses, as well as, phoneme lengthening). The investigation is made exploiting read and spontaneous narratives. The collected data are from 24 subjects, divided into two groups (depressed and control) asked to read a tale, as well as, spontaneously report on their daily activities. Ten different frequency and duration measures for pauses and clauses are proposed and have been collected using the PRAAT software on the speech recordings produced by the participants. A T-Student test for independent samples was applied on the collected frequency and duration measures in order to ascertain whether significant differences between healthy and depressed speech measures are observed. In the "spontaneous narrative" condition, depressed patients exhibited significant differences in: the average duration of their empty pauses, the average frequency, and the average duration of their clauses. In the read narratives, only the average pause's frequency of the clauses was significantly lower in the depressed subjects with respect to the healthy ones. The results suggest that depressive disorders affect speech quality and speech production through pause and clause

A. Esposito (✉)
Department of Psychology, Seconda Università di Napoli, and IIASS, Naples, Italy
e-mail: iiass.annaesp@tin.it

A.M. Esposito
Istituto Nazionale di Geofisica E Vulcanologia, Sezione di Napoli,
Osservatorio Vesuviano, Naples, Italy
e-mail: antonietta.esposito@ingv.it

L. Likforman-Sulem
Telecom ParisTech, Paris, France
e-mail: laurence.likforman@telecom-paristech.fr

M.N. Maldonato
Department of European and Mediterranean Cultures,
Università della Basilicata, Potenza, Italy
e-mail: m.maldonato@unibas.it

A. Vinciarelli
School of Computing Science, University of Glasgow, Glasgow, UK
e-mail: Alessandro.Vinciarelli@glasgow.ac.uk

A. Esposito et al. (eds.), *Recent Advances in Nonlinear Speech Processing*,
Smart Innovation, Systems and Technologies 48,
DOI 10.1007/978-3-319-28109-4_8

durations, as well as, clause quantities. In particular, the significant differences in clause quantities (observed both in the read and spontaneous narratives), suggest a strong general effect of depressive symptoms on cognitive and psychomotor functions. Depressive symptoms produce changes in the planned timing of pauses, even when reading, modifying the timing of pausing strategies.

Keywords Depressive disorders · Speech pauses · Filled and empty pauses · Vowel and consonant lengthening · Clauses · Pausing strategies · Phonation time · Silences

1 Introduction

Speech structure is characterized by pauses, the role and cognitive dimension of which were first underlined by Rochester [33], Butterworth [7], Chafe [10], Clark and Fox Tree [11] and more recently by Esposito et al. [17] and Benus et al. [4], attributing to them the function of signaling complex cognitive planning processes related to the effort to communicate new information content.

Coarsely, speech pauses have been described as a multi-determined phenomenon serving different communicative social functions, among these, synchronizing the verbal and nonverbal communication modes [18, 19], marking the boundaries of narrative units [27–30], predicting a person's status [32], as well as, identifying the caller/receiver's status in phone calls, the caller doing significantly more pauses than the receiver [37].

Pauses in speech are distinguished into three coarse categories as "empty", and "filled" pauses and "consonant/vowel lengthening" (this first categorization was due to Maclay and Osgood [23]). Empty pauses are defined as silent speech intervals, the length of which may vary depending on the task under consideration [15, 16]. Filled pauses are vocalizations such as *uhm*, *uh*, *ehm*, etc. Some authors tend to attribute different communicative functions to empty and filled pauses [9]. For example, it is supposed that empty pauses are mostly used to signal phrase boundaries while filled pauses serve to keep the speaker's turn [11]. The lengthening of vowels and/or consonants in word final or central positions is conceptualized as due to pre-articulatory planning processes [24]. Clauses are considered by definition "a sequence of words grouped together on a semantic or functional basis". Clauses are tied to pausing strategies because they are segmented through the three abovementioned different pause categories.

The proposed investigation aims at identifying speech pause duration and pause's frequency measures robust enough to show significant differences between depressed and healthy speech. Such measures will serve for the development of algorithms implemented in automatic diagnostic tools for the early detection and diagnosis of depressive disorders.

2 Methods

2.1 Subjects

The 12 depressed patients are Italian and were recruited with the help of psychiatrists at the Department of Mental Health and the General Hospital in Caserta, the Institute for Mental Health and the General Hospital in Santa Maria Capua Vetere, the Centre for Psychological Listening in Aversa (Italy) and in a private psychiatric office. They were already diagnosed as depressed and some of them were under treatment. In addition they were administered the Beck Depressive Inventory Second Edition (BDI II) proposed by Beck et al. [3] in the Italian language version assessed by Ghisi et al. [20]. Table 1 reports their code (because of anonymity constraints all the subjects received a subject identification code when enrolled in the experiment and were identified by it), gender, age, and BDI II scores. Before being enrolled in the experiment, informed consensus were gathered from all of them. Moreover, the subjects were informed by the experimenter that the aims and the expectations of the proposed research were to find voice and face (also facial expressions were recorded, even though, not for all the involved participants) features that may serve for detecting depressive states. No mention to speech pauses was indeed made.

The speech was first collected for the depressed patients and then the control group was matched to them for age and gender. The BDI II was also administered to the healthy group. Table 1 reports both the experimental and control group demographic variables. Participants were local inhabitants of the Campania Region (Italy), and therefore they matched for geographical/cultural areas and social rule sharing.

2.2 Procedure and Experimental Set-up

Participants were conducted to a quiet room and seated in front of a PC. They were provided with headphones and asked to provide their speech in two different experimental conditions. In the former, they were asked to spontaneously report on their weekly activities. In the latter, they were asked to read from the PC monitor the tale "The North Wind and the Sun" by Esopo. The tale is a standard phonetically balanced short folk tale (about six sentences all together), frequently used in phoniatric practices.

The recordings were made using a clip-on microphone (Audio-Technica ATR3350), with external USB sound card. Speech was sampled at 16 kHz and quantized at 16 bits. For each subject, the recording procedure did not last more than 15 min.

Table 1 Demographic variables describing the experimental (depressed) and control group paired by age

Experimental group							Control group			
CODE	GENDER	AGE	BDI II Scores	Diagnosis	Personality	TREATMENT	CODE	GENDER	AGE	BDI II Scores
3	M	60	33	Severe	Bipolar	yes	44	M	55	2
17	M	53	20	Moderate		no	6	M	49	5
18	M	29	25	Moderate		yes	42	M	33	2
4	F	63	27	Moderate		no	39	F	60	1
15	F	49	20	Moderate		yes	36	F	39	4
13	F	48	16	Mild		no	32	F	48	5
12	F	46	31	Severe		no	10	F	46	4
24	F	41	16	Mild		no	34	F	40	12
8	F	35	36	Severe		no	35	F	36	0
9	F	33	38	Severe		yes	33	F	36	9
14	F	27	19	Mild	Bipolar	no	40	F	27	8
19	F	27	25	Moderate		no	30	F	27	10

2.3 Measurements

The recordings were analyzed through the PRAAT (http://www.fon.hum.uva.nl/praat/) software, which allows to display at the same time both the spectrogram and waveform of each recording. The pauses (either filled, or empty or consonant/vowel lengthening) were identified manually through their spectral characteristics (from the spectrogram) and then boundaries were identified through the listening of these spectral chunks. The details of the criteria applied to identify the boundaries in the speech waveform are accurately described in [14].

The 10 measurements taken on each participant' (both depressed and healthy subjects) recordings (both read and spontaneous narratives) are described as follow:

1. Empty pause rate1 (EPR1): It was calculated as the ratio between the number of empty pauses (frequency) and the total duration of each spontaneous and read narrative (measured in seconds);
2. Empty pause rate2 (EPR2): It was calculated as the ratio between the total duration of empty pauses and the total duration of each spontaneous and read narrative;
3. Filled Pause Rate1 (FPR1): It was calculated as the ratio between the number of filled pauses (frequency) and the total duration of each spontaneous and read narrative;
4. Filled Pause Rate2 (FPR2): It was calculated as the ratio between the total duration of filled pauses and the total duration of each spontaneous and read narrative;
5. Lengthening Rate1 (LR1): It was calculated as the ratio between the number of consonant/vowel lengthening (frequency) and the total duration of each spontaneous and read narrative (in seconds);
6. Lengthening Rate2 (LR2): It was calculated as the ratio between the total duration of consonant/vowel lengthening and the total duration of each spontaneous and read narrative;
7. Clause Rate1 (CR1): It was calculated as the ratio between the number of clauses (frequency) and the total duration of each spontaneous and read narrative;
8. Clause Rate2 (CR2): It was calculated as the ratio between the total duration of clauses and the total duration of each spontaneous and read narrative;
9. Pause Rate1 (PR1): It was calculated as the ratio between the sum of EPR1, FPR1, and LR1 (that are frequencies) and the total duration of each spontaneous and read narrative;
10. Duration Pause Rate1 (DPR1): It was calculated as the ratio between the sum of empty, filled pauses, and consonant/vowel lengthening durations and the total duration of each spontaneous and read narrative.

For sake of clarity, Table 2 exemplifies the description of the above measurements.

Table 2 Description and acronyms assigned to the measurements under examination

Duration in s/duration in s	Acronym	Frequency/duration in s	Acronym
Empty pause durations/Speech total duration	*EPR2*	Empty pause frequency/Speech total duration	**EPR1**
Filled pause durations/Speech total duration	**FPR2**	Filled pause frequency/Speech total duration	**FPR1**
Lengthening durations/Speech total duration	**LR2**	Lengthening frequency/Speech total duration	**LR1**
Pause total durations/Speech total duration	*DPR1*	Pause total frequency/Speech total duration	**PR1**
Clause durations/Speech total duration	*CR2*	*Clause frequency/Speech total duration*	*CR1*

3 Results

A T-Student test for independent samples was applied on the collected measures in order to ascertain significant differences between healthy and depressed speech.

In the "spontaneous narratives" it was found that:

(a) DPR1, the total duration of speech pauses (empty, filled and vowel lengthening taken all together) is significantly longer ($\alpha < 0.05$) for depressed subjects with respect to healthy ones, under one ($T(22) = 2.438257$, $\rho = 0.011646$) and two tailed testing hypotheses ($T(22) = 2.438257$, $\rho = 0.023293$);

(b) EPR2, the total duration of empty pauses is significantly longer ($\alpha < 0.05$) for depressed subjects with respect to healthy ones under one ($T(22) = 2.282619$, $\rho = 0.016239$) and two tailed testing hypotheses ($T(22) = 2.282619$, $\rho = 0.032479$);

(c) CR2, the clause duration is significantly shorter ($\alpha < 0.05$), for depressed subjects with respect to healthy ones under one ($T(22) = 2.434987$, $\rho = 0.011729$) and two tailed testing hypotheses ($T(22) = 2.434987$, $\rho = 0.023458$);

(d) CR1, the clause frequency is significantly lower ($\alpha < 0.05$) for depressed subjects with respect to healthy ones under one ($T(22) = 4.028679$, $\rho = 0.000281$) and two tailed testing hypotheses ($T(22) = 4.028679$, $\rho = 0.000562$);

Figure 1 illustrates the differences between depressed and healthy subjects in CR1 frequency rate and CR2, DPR1, EPR2 duration rates. Differences were not significant for the remaining measurements.

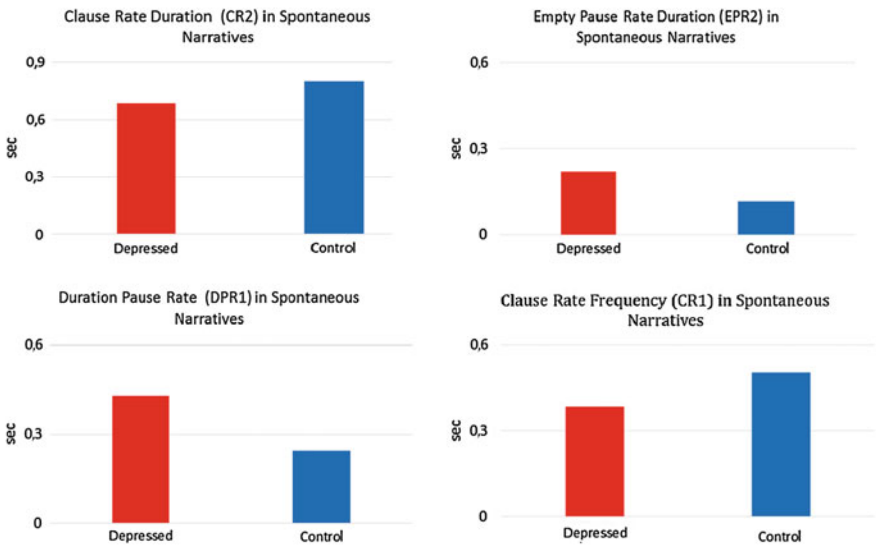

Fig. 1 Differences between depressed and healthy subjects in CR1, CR2, DPR1, EPR2 rates

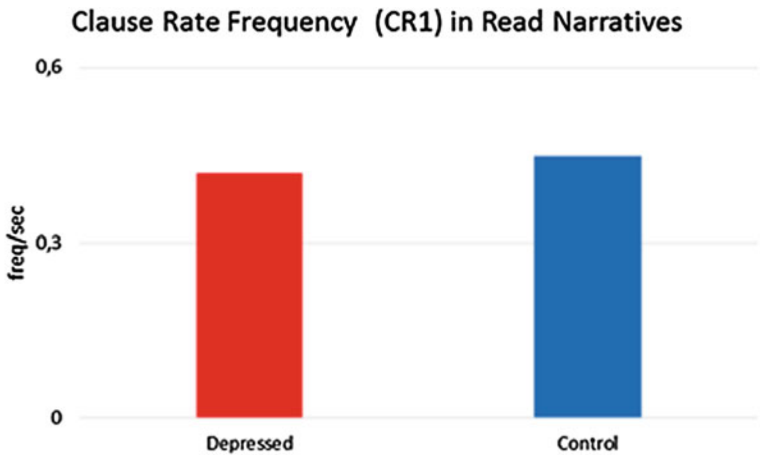

Fig. 2 Differences between depressed and healthy subjects in CR1 frequency rates

In the “read narratives” only the frequency of clauses (CR1) was significantly lower ($\alpha < 0.05$) for the depressed subjects with respect to healthy ones, under one tailed ($T(22) = 1.834149$, $\rho = 0.040778$) testing hypothesis. Figure 2 illustrates the CR1 differences between the two groups.

4 Discussions

This research investigates on how depressive disorders affect speech pausing strategies either in spontaneous or read narratives. In literature, the identified acoustic differences between depressed and healthy speech are attributed to changes in F0 frequency values and F0 related measures, as well as, formants, power spectral density, Mel Frequency Cepstral (MFCC) coefficients, speech rate and glottal parameters such as jitter and shimmer [2, 5, 13, 21, 22, 26, 27, 38]. In particular, it has been proven that depressed speech exhibits a "slow" auditory dimension [12, 25] and it is perceived as sluggish. According to our results, the perceived "sluggishness" can be primarily attributed to lengthened empty pauses (no significant differences were found for filled pause and consonant/vowel lengthening), and shortened phonation time (clauses are shorter in duration and less numerous). Similar results were found in "automatic" speech tasks (such as counting, reciting the alphabet, and/or reading (see [1, 6, 27, 35]) but not in "extemporaneous free speech samples" (see [27], p. 9). This different outcome can be attributed to the fact that in [27] the speech was collected through telephone and the noise on the communication line may have distorted the signal making difficult to automatically identify the pause boundaries. In addition, the production of spontaneous speech is more cognitively demanding with respect to automatic speech tasks, requiring preparation, word selection and motor articulatory control. Since speech pauses signal complex cognitive planning processes related to the effort to communicate new information content [10, 18], the lengthened silences and shorter phonation time can be considered an index of cognitive and psychomotor retardations related to the degree of severity of depressive states [8, 25, 31, 34, 36]. Longer silences in depressed speech suggest planning communicative efforts to maintain a regular conversation and the need for more cognitive elaboration time [25, 31, 36]. The extra time requirements may produce inadequacies in a typical interactional exchange, both for the listener that does not know how to react to these long silences and the speaker who feels the failure of the exchange. The requirements for extra time to express themselves reduce the phonation time, reducing the amount of information content to be communicated. The consequence is a social impairment that isolates these subjects and impairs the quality of their social life.

Robust speech features for detecting depressive states might help in the implementation of automatic tools to support doctors in the early detection and diagnosis of depression. Speech silences and phonation time seem robust enough to be automatically exploited, given the significant differences measured in different speech tasks and contexts. An early detection of depression can help patients reduce the length of the depressive period and allow a faster recovery.

The strength of the reported results is that they are in agreement with data obtained through different experimental set-ups, supporting the robustness of speech pauses as feature based algorithms. However, there is the need for more data and more thorough analyses to exploit them as depression diagnosis support.

Acknowledgments This work contributes to the research developed in the European Space Agency (ESA) project: Psychological Status Monitoring by Computerised Analysis of Language phenomena (COALA, AO-11-Concordia).

References

1. Alpert, M., Pouget, E.R., Silva, R.R.: Reflection of depression in acoustic measures of the patient's speech. J. Affect. Disord. **66**, 59–69 (2001)
2. Alghowinem, S., et al.: Detecting depression: a comparison between spontaneous and read speech. In: IEEE 3 International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 7547–7551 (2013)
3. Beck, A.T., Steer, R.A., Brown, G.K.: Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio, TX (1996)
4. Benus, S., Enos, F., Hirschberg, J., Shriberg, E.: Pauses in Deceptive Speech, Prosody 2006, Dresden. http://www1.cs.columbia.edu/nlp/papers/2006/benus_al_06.pdf (2010)
5. Boonla, T., Yingthawornsuk, T.: Assessment of vocal correlates of clinical depression in female subjects with probabilistic mixture modeling of speech cepstrum. In: IEEE 11th International Conference on Control, Automation and Systems (ICCAS 2011), 26–29 Oct 2011
6. Bouhuys, A.L., et al.: Speech timing measures of severity, psychomotor retardation, and agitation in endogenously depressed patients. J. Common Disord. **17**, 277–288 (1984)
7. Butterworth, B.L.: Evidence for pauses in speech. In: Butterworth, B.L. (ed.) Language Production: Speech and Talk, vol. 1, pp. 155–176. London Academic Press (1980)
8. Caligiuri, M.P., Ellwanger, J.: Motor and cognitive aspects of motor retardation in depression. J. Affect. Disord. **57**, 83–93 (2000)
9. Campine, E., Véronis, J.: Pauses and hesitations in french spontaneous speech. In: Disfuency in Spontaneous Speech. Workshop in International Speech Communication Association, pp. 43–46 (2005)
10. Chafe, W.: Cognitive constraint on information flow. In: Tomlin, R. (ed.) Coherence and Grounding in Discourse John Benjamins, pp. 20–51 (1987)
11. Clark, H.H., Fox Tree, J.E.: Using uh and um in spontaneous speaking. Cognition **84**, 73–111 (2002). Goldman-Eisler, F.: Sequential temporal patterns and cognitive processes in speech. Lang. Speech **10**(2), 122 (1967)
12. Cohn, J.F., et al.: Detecting Depression from Facial Actions and Vocal Prosody. In: 3rd International Conference on Affective Computing and Intelligent Interaction (ACII 2009) (2009)
13. Daniel, J., et al.: Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Trans. Biomed. Eng. **47**(7), 829–837 (2000)
14. Esposito, A.: On vowel height and consonantal voicing effects: data from Italian. Phonetica **59**(4), 197–231 (2002)
15. Esposito, A.: Pausing strategies in children. In: Proceedings of NOLISP05, International Conference in Nonlinear Speech Processing, pp. 42–48. Cargraphics, Barcelona, Spain, 19–22 April 2005
16. Esposito, A.: Children's organization of discourse structure through pausing means. In: Faundez-Zanuy, M., et al. (ed.), Nonlinear Analyses and Algorithms for Speech Processing, Lecture Notes in Computer Science (LCNS), vol. 3817, pp. 108–115, Springer-Verlag, Berlin Heidelberg (2006)
17. Esposito, A., Stejskal, V., Smékal, Z., Bourbakis, N.: The significance of empty speech pauses: cognitive and algorithmic issues. In: Mele, F., et al. (eds.) Brain Vision Artificial Intelligence, LCNS, vol. 4729, pp. 542–554. Springer, Berlin Heidelberg (2007)
18. Esposito, A., Marinaro, M.: What pauses can tell us about speech and gesture partnership. In: Esposito, A., Bratanic, M., Keller, E., Marinaro, M. (Eds.), Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue, NATO Publishing Series, Sub-Series E: Human and Societal Dynamics, vol. 18, pp. 45–57. IOS press, The Netherlands (2007)

19. Esposito, A., Esposito, A.M.: On speech and gesture synchrony. In: Esposito, A. et al. (Eds), Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issue. LNCS 6800, pp. 252–272. ISBN 978-3-642-25774-2, Springer-Verlag, Berlin Heidelberg (2011)
20. Ghisi, M., Flebus, G.B., Montano, A., Sanavio, E., Sica, C.: Beck Depression Inventory-II. Manuale italiano. Organizzazioni Speciali, Firenze (2006)
21. Helfer, B.S., et al.: Classification of depression state based on articulatory precision. In: 14th Annual Conference of the International Speech Communication Association, pp. 1–5 (2013)
22. Hewlett Sanchez, M., et al. Using prosodic and spectral parameters in detecting depression in elderly males. In: INTERSPEECH 2011, pp. 3001–3004. Florence, Italy, 27–31 Aug 2011
23. Maclay, H., Osgood, C.E.: Hesitation phenomena in spontaneous english speech. Word **15**, 19–44 (1959)
24. MacNeilage, P.: The motor control of the serial ordering of the speech. Psychol. Ref. **77**(3), 182 (1970)
25. Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., Faravelli, L.: Cognitive impairment in major depression. Eur. J. Pharmacol. **626**, 83–86 (2010)
26. Moore, E., Clements, M., Peifer, J., Weisser, L.: Investigating the role of glottal parameters in classifying clinical depression. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE 25th, vol. 3, pp. 2849–2852 (2003)
27. Mundt, J.C., Snyder, P.J., Cannizzaro, M.S., Chappie, K., Geralts, D.S.: Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. J. Neurolinguistics **20**, 50–64 (2007)
28. Oliveira, M.: Prosodic Features in Spontaneous Narratives, Ph.D. Thesis, Simon Fraser University (2000)
29. Oliveira, M.: Pausing strategies as means of information processing narratives. In: Proceeding of the International Conference on Speech Prosody, Ain-en-Provence, pp. 539–542 (2002)
30. O'Shaughnessy, D.: Timing patterns in fluent and disfluent spontaneous speech. In: Proceedings of ICASSP Conference, pp. 600–603. Detroit (1995)
31. Porter, R.J., Gallagher, P., Thompson, J.M., Young, A.H.: Neurocognitive impairment in drug free patients with major depressive disorder. Br. J. Psychiatry **182**, 214–220 (2003)
32. Richmond, V., Payne, McCroskey J., Payne, S.: Nonverbal Behavior in Interpersonal Relations. Prentice Hall, Upper Saddle River (1991)
33. Rochester, S.R.: The significance of pauses in spontaneous speech. J. Psycholinguist. Res. **2**(1), 51 (1973)
34. Sabbe, B., et al.: Retardation in depression: assessment by means of simple motor tasks. J. Affect. Disord. **55**, 39–44 (1999)
35. Szabadi, E., Bradshaw, C.M., Besson, A.O.: Elongation of pause-time in speech: a simple, objective measure of motor retardation in depression. Br. J. psychiatry **129**, 7–592 (1976)
36. Tsourtos, G., Thompson, J.C., Stough, C.: Evidence of an early information processing speed deficit in unipolar major depression. Psychol. Med. **32**, 259–265 (2002)
37. Vinciarelli, A., Chatziioannou, P., Esposito, A.: When the words are not everything: the use of laughter, fillers, back-channel, silence, and overlapping speech in phone calls. Frontiers in ICT, vol. 4, article 4, pp. 1–11 (2015). doi:10.3389/fict.2015.00004
38. Yingthawornsuk, T., et al.: Characterizing sub-band spectral entropy based acoustics as assessment of vocal correlate of depression. In: IEEE 10th International Conference on Control Automation and Systems (ICCAS 2010), p. 1183, 27–30 Oct 2010