Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 7 (2015) 62-79

Seasonal Affective Disorder Speech Detection on the Base of Acoustic-Phonetic Speech Parameters

Gábor KISS, Klára VICSI

Department of Telecommunication and Media Informatics,
Budapest University of Technology and Economics, Hungary
e-mail: {kiss.gabor; vicsi}@tmit.bme.hu

Manuscript received February 11, 2016; revised April 11, 2016.

Abstract: The development of an online monitoring system is shown in order to track the physiological and cognitive condition of crew members of the Concordia Research Station in Antarctica, with specific regard to depression. Follow-up studies were carried out on recorded speech material in such a way that segmental and supra-segmental speech parameters were measured for individual researchers weekly, and the changes in these parameters were detected over time. Typical acoustic-phonetic parameters were selected on the base of statistical analyses. On the base of the selected parameters, a function was developed which indicates the likelihood of depression at the examined person.

Keywords: Acoustic-phonetic speech analysis, seasonal affective depression, cognitive status monitoring, statistical analysis

1. Introduction

Speech reflects the physiological and cognitive condition of humans: therefore, there are changes in the acoustic phonetic parameters of speech when the condition of humans changes. For example, in case of vocal disorders, the acoustical parameters of disordered speech are significantly different from normal speech [1], [2], [3].

Psychology says experience of failure can cause depression, which has emotional, cognitive, physical and motivational symptoms; thus these symptoms may also be observed in speech. Depression is a major public health challenge, its prevalence is high and it has a major impact on sufferers [4]. Effective treatment is available, but depression care is facing barriers at several levels, such as under-recognition, stigmatization, inadequate treatment and mistreatment.

Because speech reflects the physiological and cognitive condition of humans, doctors can diagnose depression not only from patients' speech content but also from their speech quality. They often characterize depressed speech as faded, slow, monotonous, lifeless and metallic. We can link these properties with acoustic characteristics such as fundamental frequency, amplitude modulation, formant structure, energy distribution, etc.

Numerous studies have identified acoustic features that can be linked to depression. In some of the studies the differences between the speech of healthy and depressed people is measured; others perform follow-up monitoring to gather features with high classification performance. Prosodic parameters like rhythm, intonation, accent and timing are sensitive to changes in mood states and emotions [5], [6], [7], [8]. A lot of the research that has been done over the last few years deals with the relationship between depression and different acoustic phonetic parameters [9], [10], [11], [12]. One early study identified fundamental frequency as one of the most important acoustic features of depressed speech [13]. Nowadays many parameters are investigated at different levels of speech production: fundamental frequency, variation of fundamental frequencies, formants, power spectral density [14], cepstrum [15] or MFC coefficients [16], speech rate [17], glottal features [18], amplitude modulation and other different prosodic parameters [19].

The data suggest that depressed patients take more time to express themselves. They speak with greater hesitation, but they do not vocalize more. In this way they are producing more cumulative and variable pauses. Consequently, voice acoustic measures are examined that reflect depression severity such as the percentage of pause time, vocalization or pause ratio, and speaking rates. Pitch variability and first and second formants correlate significantly with overall depression severity.

Our goal is to develop a monitoring system with specific regard to Seasonal Affective Disorder (SAD) type of a depression. The monitoring system's aim is to track the physiological and cognitive condition of crew members of the Concordia Research Station in Antarctica. For the segmentation we used an automatic language-independent program, developed earlier, to segment the records in phoneme level for measurement [20].

For this research we have developed four types of databases: Seasonal Affective Disorder Speech Database and Healthy Reference Speech database are used for the examination of the sensitivity of acoustic-phonetic parameters of speech regarding depression and Concordia Speech Database 2013 and Concordia Speech Database 2014 are used for the physiological and cognitive status monitoring of the crew members in the Concordia research station; and the Healthy Follow-Up Speech Database is used for the physiological and cognitive status monitoring of healthy Hungarian speakers.

The paper is structured as follows. The descriptions of the databases used are presented in Section 2, detailed descriptions of evaluation methods in Sections 3 and 4, followed by results and conclusions in Sections 5 and 6.

2. Databases

In this part we give short descriptions of the databases that we developed and used in our research.

A.Seasonal Affective Disorder Speech Database

The database contains 55 sufferers: 35 female and 20 male. A psychiatrist from the Neurology Department of Semmelweis University, Hungary assisted in the selection of the patients. The database consists of two parts: the first part is a collection of spontaneous speech obtained from the discussion between the patient and the doctor; in the second part patients read a standard phonetically balanced short folk tale (about 6 sentences altogether) called “The North Wind and the Sun”, frequently used in phoniatry practice for all European languages. The recordings were recorded with clip-on microphones (Audio-Technica ATR3350), an external USB sound card, at 44,100 Hz at a 16 kHz sampling rate, quantized at 16 bits.

In order to measure depression severity, Beck Depression Inventory (BDI) was used [21]. The BDI indices are within the range of 14 to 43. The mean age of subjects was 31,5 years, with a standard deviation of 12,3 years and a range of 18 to 63 years. For each patient we noted additional information about smoking habits, illnesses and prescribed medication.

B. Healthy Reference Speech Database

For the Healthy Reference Speech Database72 healthy speakers (28 male and 44 female) were asked to read the same tale, “The North Wind and the Sun”. The recording conditions were the same as in the Seasonal Affective Disorder Speech Database, using clip-on microphones (Audio-Technical ATR3350) at a sampling rate of 44,100 Hz, quantized at 16 bits; the reference database recordings were also annotated and segmented on phoneme level, using the SAMPA phonetic alphabet. The BDI indices are within the range of 0 to 13. The mean age of subjects was 28,7 years, with a standard deviation of 10,4 years and a range of 18 to 52 years.

C. Concordia Speech Database 2013 and 2014

We are participating in an international ESA project, AO-11-Concordia, entitled Psychological Status Monitoring by Computerized Analysis of Language Phenomena (COALA). The records for this database were collected from crew members using their mother tongue. The records were made in a weekly period during their stay at the Concordia Station. Baseline recordings were made from the same crew members in normal circumstances in Europe before their departure for Antarctica. This way we had the opportunity to monitor the impact of hypoxia on speech and occasional occurrence of SAD symptoms.

The database consists of two parts: the first part is a collection of spontaneous speech obtained from the recorded diaries; in the second part the participants read the same phonetically balanced short folk tale (as described above) in their mother tongue.

The experiment was planned to be carried out in two seasons: 2013 and 2014. The crew members were native French, Italian or Greek a total of 20 people. In both seasons the doctor (always a member of the crew) was a native Greek speaking in English. The recordings were made with clip-on microphones (Audio-Technica ATR3350) at a sampling rate of 44,100 Hz, using 16 bits.

D. Healthy Follow-Up Speech Database

In this database we gathered the recordings of healthy people in everyday conditions. In our everyday lives there are many factors that can affect our speech (mood, fatigue, stress); thus the effect of these factors on speech is an important consideration. The creation of the database was necessary in order to have a good reference point for our studies to be able to compare the results, in this case when monitoring depression.

The database contains recordings of 10 participants: 5 female and 5 male. The male age distribution is between 21 and 29, and the female age distribution is between 26 and 57. The participants read the same phonetically balanced short folk tale as in the previous two databases, in Hungarian. The recordings were collected over three months at weekly intervals, gathering a total of 156 recordings from 9 people. For each person we noted additional information about smoking habits, illnesses and prescribed medication, and before the first recording we checked the participants’ BDI score. As expected, all participants had low BDI scores: their distribution was between 0 and 8.

The recordings were made with the same equipment as in the previous two databases, using clip-on microphones (Audio-Technica ATR3350), with external USB sound card, at 44,100 Hz at a 16 kHz sampling rate, quantized at 16 bits.

E. Segmentation and Labelling

All the databases' recordings need to be segmented into phoneme level in order to measure the acoustic-phonetic parameters. For this reason we used an automatic language-independent segmentation program, developed earlier in our laboratory. For each recording, manual correction was done.

3. Analysis of the acoustic-phonetic parameters

Our final goal is to monitor the speech of the crew members at the Concordia Station, and indicate if any of the crew members show any sign of depression which could lead to a serious level of depression. Speech, as an acoustic product, is very diverse. We can distinguish between inter-individual differences, that is, variety in speech among different people, and intra-individual differences, i.e., variety in the speech of one person. This variety can arise for many reasons, primarily the physical condition of the speaker (for example flu', alcohol condition, emotional state, stress, sleepiness, etc.), but it has some natural variety too, as speech is a non-deterministic process. This is the reason that our work includes the examination of the relationship between the acoustic-phonetic parameters and other biological and psychological data of the crew members measured at the station, such as long term medical survey data (LTMS) and oxygen saturation data.

In spite of the considerable variety in speech parameters, we would like to identify specific changes which indicate depression. This task is not obvious. We can eliminate the variety between speakers by monitoring only the changes of the acoustic parameters for each person over time.

In order to distinguish the changes in acoustic parameters caused by depression from the natural intra-individual variety we investigated the changes in acoustic parameters in speech of nine healthy people in normal circumstances. For this investigation the Healthy Follow-Up Speech Database was used.

A. Selection of acoustic-phonetic parameters

Having previously carried out a study to decide what kind of acoustic parameters can indicate depression [22], the variation in these acoustic parameters was examined over time: variance of intensity (VI), fundamental frequency (F0), variance of fundamental frequency (VF0), first formant (F1), variance of first formant (VF1), second formant (F2), variance of second formant (VF2), jitter (J), shimmer (S), articulation rate (AR), speech rate (SR), and total length of pauses (TLoP). We added one further acoustic parameter which showed significant difference between healthy speech and depressed speech: the rate of transient (Rot) [23].

B. Pre-processing and segmentation

The recorded speech was automatically segmented into phoneme units with an automatic segmentation program, developed in our laboratory [23].

The acoustic-phonetic parameters were examined in two groups according to their segmental and supra-segmental (prosodic) features.

The segmental features were measured at the middle of the same vowel, in our case the vowel was "E". The following segmental features were measured as previously mentioned: F0 of 'E' vowels, F1 and F2 of 'E' vowels (F1, F2), VF1 and VF2 of 'E' vowels, jitter of the 'E' vowels, and shimmer of the 'E' vowels. For the measurement of formants, fundamental frequency and the spectral values, a Hamming window was used with 25 ms frame size; these features were always evaluated from the middle of each vowel 'E'.

The supra-segmental (prosodic) features were measured by the total length of each recording. The following features were measured: VI as volume dynamics of speech (range of intensity), VF0 as fundamental frequency dynamics of speech (range of fundamental frequency), ratio of total length of pauses and the total length of the recording, articulation and speech rate and rate of transient (Rot).

C. Variation of the selected acoustic-phonetic parameters in case of healthy speech

The selected acoustic parameters were evaluated for each record for each person. The difference from the mean value for each parameter was calculated. An example is presented in *Fig. 1*. Interpolation was applied between the measuring points.

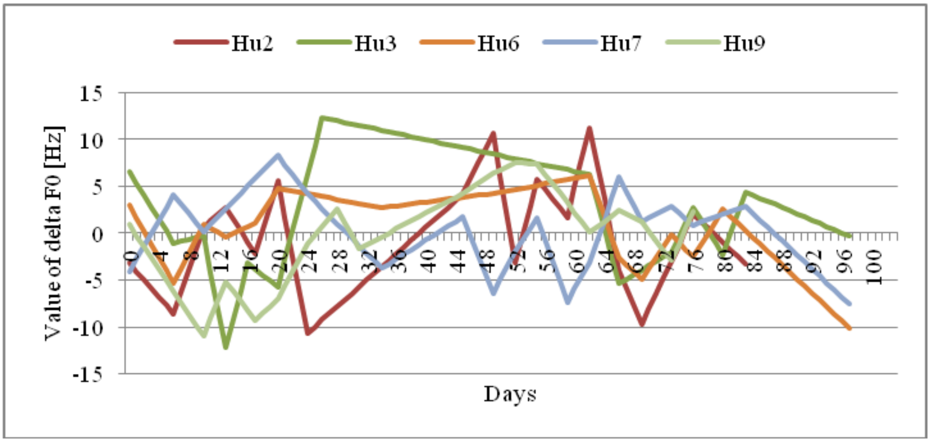

Figure 1: An example for the delta F0 changing in time for nine healthy males. Day 0 is the beginning of the examination

Maximum difference, from the mean value of the selected parameters, was calculated for each person for each parameter, and presented in *Table 1*.

Table 1: Absolute difference of the intra-individual variety in case of healthy speech

	Average		Maximum		Minimum	
	of the Absolute Difference of the Personal Variety					
	Female	Male	Female	Male	Female	Male
VI [dB]	0.58	0.55	1.39	1.27	0.18	0.15
F0 [Hz]	9	5	12	10	6	3
VF0 [Hz]	8	7	14	14	2.5	2
F1 [Hz]	23	24	34	36	11	10
VF1 [Hz]	11	10	21	19	5	5
F2 [Hz]	38	36	57	53	22	23
VF2 [Hz]	28	24	53	49	14	14
J [%]	0.6	0.5	1.1	1.0	0.2	0.2
S [%]	1.5	1.3	2.7	2.6	0.4	0.4
AR [phon. / s]	0.8	0.7	1.5	1.5	0.4	0.3
SR [unit / s]	0.9	0.9	1.4	1.5	0.4	0.4
TLoP [s]	2.6	2.8	3.5	4.1	1.5	1.6
RoT [%]	3.4	3.8	4.7	5.1	2.2	2.3

D. Statistical difference of the acoustical phonetic parameters between healthy speech and depressed speech

For our study we examined depressed speech, and selected a group of speech parameters which differ significantly in depressed speech when compared to normal speech [22]. We examined the weight of these parameters in the discrimination of healthy and depressed speech. Principal component analysis was carried out using Matlab on both the Seasonal Affective Disorder Speech Database and the Healthy Reference Speech Database.

Principal component analysis is an established tool for multivariate data analysis. Its goal is to predict which parameters address the greatest weight in the description of the information content of the data set.

During the individual examination of the components we did not find characteristic patterns. This indicates that we did not use features that are irrelevant in terms of depression; indeed, all the parameters are useful indicators of depression.

We investigated differences in the mean values of the selected parameters between the healthy and the depressed speech, especially in view of the intra-individual variation in the normal case. The results are presented in Table 2.

Table 2: The differences in the mean values of the changes of the acoustic phonetic parameters between healthy and depressed speech, compared to the normal intra-individual variation in case of healthy speech

	Mean difference between healthy and depressed speech		The average variation of the parameter	
	Female	Male	Female	Male
VI [dB]	-0.55	-0.1	+/- 0.58	+/-0.55
F0 [Hz]	-21	-6.6	+/- 9	+/-5
VF0 [Hz]	-4	-2.5	+/- 8	+/-7
F1 [Hz]	-18	-28.2	+/- 23	+/-24
VF1 [Hz]	+2.8	+6.3	+/- 11	+/-10
F2 [Hz]	+35	+42.1	+/- 38	+/-36
VF2 [Hz]	+11	+21.2	+/- 28	+/-24
J [%]	+0.3	+0.3	+/- 0.6	+/-0.5
S [%]	+1.5	+1.5	+/- 1.5	+/-1.3
AR [phon. / s]	-0.3	-1.7	+/- 0.8	+/-0.7
SR [unit / s]	-0.13	-1.6	+/- 0.9	+/-0.9
TLoP [s]	+1.1	+3.7	+/- 2.6	+/-2.8
RoT [%]	-1.5	-5.25	+/- 3.4	+/-3.8

As can be seen from the *Table 2*, the differences in the mean values of the acoustic phonetic parameters between the healthy and the depressed speech are almost as big as the variation of these parameters in normal, healthy speech. This can cause many problems in processing and separation. However, it does not mean that we are not able to distinguish between healthy speech and depressed speech, rather that the task is hard, because only the *likelihood* of depression can be predicted on the basis of the parameters.

4. Depression status monitoring of the crew members at Concordia

We built an online monitoring system that follows the acoustic-phonetic parameters of a subject's speech over time and compares the collected data with the mean value. The system is designed to create an alert if one of the subjects suffers depression. An automatic (semi-automatic) online alerting system was built which records speech at a given time rate and creates alerts to the appearance of depression. The flow chart of the online alerting system is shown in *Fig. 2*.

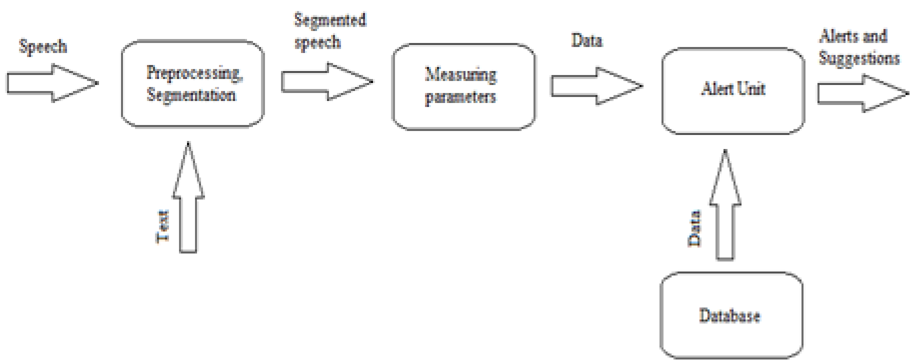

Figure 2: The online alerting system flowchart

The system has three main parts: Preprocessing and Segmentation unit, the unit that measures the acoustic parameters (Measuring Parameters), and the Alert unit. Preprocessing and Segmentation and Measuring the parameters were described in paragraph 3.2. For the alert unit we developed a method which can predict the likelihood of depression based on the acoustic phonetic parameters; this is described in paragraph 4.1.

A. Alert unit

Each acoustic-phonetic parameter is given as a function of time. The main question is what quantity of the difference in the actual measured speech parameters indicates depressed speech. The question is whether this difference indicates the presence of depression, and, if it does, to what extent: how likely is it that this value indicates depression?

We have developed a "Depression Probability Function" (DP) which takes the actual measured differences (compares the actual values with the mean value) of any acoustic-phonetic parameter, and suggests how likely it is that the value indicates depression at the given time. The operation of the function for a given acoustic-phonetic parameter *p* is described below.

The values of the measured *p* parameter are normalized to the mean of measured values from the corresponding healthy database separately for each gender. That is, the average value of parameter *p* (calculated from the healthy database) is subtracted from every other value of *p*. The result of this is that the mean of the parameter *p* is zero in the normalized healthy database and different (but close to zero) in the normalized depressed database. For each *p* four reference distributions are generated from the Healthy Reference Speech Database and Seasonal Affective Disorder Speech Database: healthy male and female distributions ($\mathrm{HD}_\mathrm{p}^{\mathrm{male}}$, $\mathrm{HD}_\mathrm{p}^{\mathrm{female}}$), and depressed male and female distributions ($\mathrm{DD}_\mathrm{p}^{\mathrm{male}}$, $\mathrm{DD}_\mathrm{p}^{\mathrm{female}}$).

The normalized distributions are expressed as a percentage as follows, where *p* is the measured acoustic parameter,*s* is gender (male, female)and $\mathrm{HD}_\mathrm{p}^{\mathrm{s}}(x)$ is the frequency of a given value (*x*) of the measured acoustic parameter:

$$\sum_{x=-\infty}^{+\infty} HD_p^s(x) = \sum_{x=-\infty}^{+\infty} DD_p^s(x) = 100\% \qquad (1)$$

Let the measured mean value of parameter *p* on the speech sample be *x*. Two variables, "Normal Chance" (NC) and "Depressed Chance" (DC), are calculated, which indicate the likelihood that the measured value *x* is derived from the HD and from the DD respectively. These variables are calculated differently according to the sign of the difference between the mean values in HD and DD. Let us assume that the difference of mean values for the depressed distribution is less than 0 (mean of HD – mean of DD). In this case DC and NC are calculated as follows:

$$NC_p(x) = \sum_{y=x}^{-\infty} HD_p^s(y) \qquad (2)$$

$$DC_p(x) = \sum_{y=x}^{+\infty} DD_p^s(y)\,, \tag{3}$$

where $NC_p(x)$ and $DC_p(x)$ is the Normal Chance and Depressed Chance for measured value *x* of acoustic parameter *p*. $HD_p^s(y)$ and $DD_p^s(y)$ are the probability values associated with *y* from the healthy and depressed distributions for gender *s* and parameter *p*. For a given *x* only the corresponding gender is used.

If the difference of mean values for the depressed distribution is greater than 0 (mean of HD – mean of DD), then the -∞ and + ∞ are switched in the sums. The final measure of the likelihood of depression (Depression Probability, DP) for parameter *p* is calculated using NC and DC.

$$DP_p(x) = DC_p(x) - NC_p(x),\ \text{if}\ DP_p(x) > 0,\ \text{else}\ DP_p(x) = 0 \tag{4}$$

where $DP_p(x)$ is the likelihood of depression for parameter p for measured value *x*.

This calculation is done for all acoustic-phonetic parameters.

The Mean Depressed Probability Function (MDP) for one person for a given time is calculated in the following way:

$$MDP(x_i) = \frac{\sum_{i=0}^{n} DP_{p_i}(x_i)}{n} \tag{5}$$

where *n* is the count of the selected parameters.

This mean function has the advantage that it reduces inaccuracies caused by the natural variation of the parameters, assuming that the variables have independent diversity. Of course it may be that the simple mean is not the best, because some acoustic phonetic parameters can be more relevant than others. But, as we described in paragraph 3.3, all of these parameters seem equally important on the basis of principal component analysis. (At a later stage, when more data is available, weighted averages will improve the quality of the alert unit.)

In the case of monitoring, the MDP is measured during monitoring time at a given time rate.

5. Results

For the final examination AR, SR, TLoP parameters were excluded, as only the "tale" parts were used in the examination. However, because the subjects always read the same text, it naturally became faster and faster. We excluded VI and VF0 too, because their variances were too large, and showed no relevant results at all. Thus for the final examinations the following parameters were used: F0, F1, VF1, F2, VF2, Jitter, Shimmer, and RoT.

The MDP was calculated using Seasonal Affective Disorder Speech Database, and Healthy Speech Database, The evaluation was done on the basis of the final selected parameters for each person for the following databases: Healthy Follow-Up Speech Database, Concordia Speech Database 2013, and Concordia Speech Database 2014. While MDP was calculated on Hungarian databases, it is absolutely correct to use the Hungarian Healthy Follow-Up Speech Database for the evaluation. However our results on Hungarian language was compared with results in different European languages [24], [25], [26]. We found that the tendencies were the same for all the selected parameters, for each European languages used in Concordia Speech Databases. This is natural, since the changes of the studied parameters reflects in general human biological process and does not depend on the language. Thus probable we can use the results of the Hungarian databases to indicate depression in the Concordia Speech Database 2013 and Concordia Speech Database 2014.

A. Determination of an alert threshold for the detection of depression by the Mean Depressed Probability Function

For the determination of an alert threshold for the detection of depression the Healthy Follow-Up Speech Database was used. For each person in the Healthy Follow-Up Speech Database the MDP was measured for the finally selected parameters as a function of time.

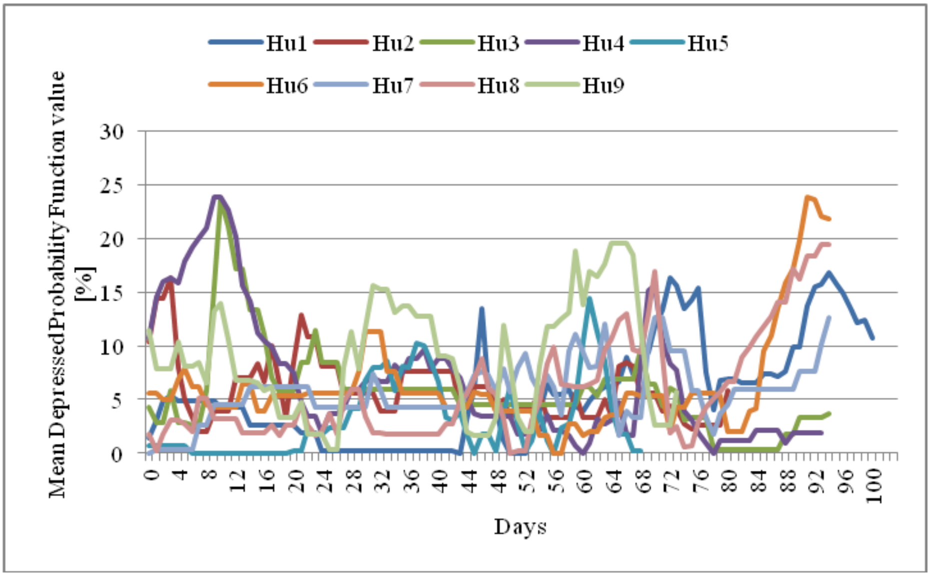

Figure 3: The Mean Depressed Probability Function values on the Healthy Follow-Up Speech Database

As can be seen in *Fig. 3.*, the maximum value of MDP was below 25%. Thus it can be stated that the value of the MDP is equal to or less than 25% in the case of healthy speech. Due to the small number of people, in order to make the method more robust the alert level has been set higher, at 35%.

B. Examination of the Mean Depressed Probability Function on the Concordia Crew Members Speech Database 2013 and 2014

The MDP values were calculated for Concordia Speech Database 2013 and Concordia Speech Database 2014. Day 0 was always 1st January of the year in question. The beginning of the winter was around day 150. Unfortunately, due to technical problems with Concordia Speech Database 2013, we only have data from day 60 (beginning of March). In both cases, the subjects arrived at the station around day 30 (beginning of December). The first values were the values of the reference records before arrival at Concordia Station, under normal circumstances in Europe.

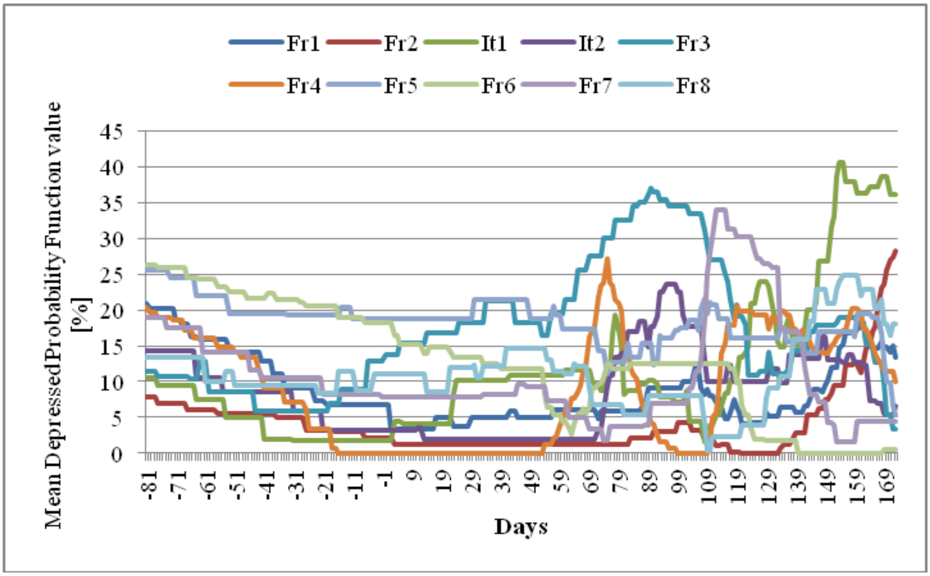

Figure 4: The Mean Depressed Probability Function values on the Concordia Speech Database 2013

As can be seen in *Fig. 4.*, there were two people where MDP value was above 35% (It1 and Fr3). It is noteworthy that several subjects' initial MDP level is around 20-25%, which is below alert level, but the acoustic parameters show large variations. The cause of this may be that the value of MDP not only indicates depression but also other cognitive states, such as stress.

Fr3's maximum MDP value was 36%, reached around April. This could be a sign some kind of depression or bad mood. It is probably not a sign of SAD, because it was at its maximum in April, became normal around May, and then remained there for the rest of the examination (till the middle of the winter): SAD is typical of the winter.

It1's maximum MDP value was 40%, reached around the beginning of June, which coincides with the beginning of winter. The depression probability value stayed around 35-40% during the rest of the examination (till the middle of the winter).

So it might be that there was only one person who suffered from SAD amongst the Concordia Crew members in 2013. Presumably the severity of the depression for this person was not high.

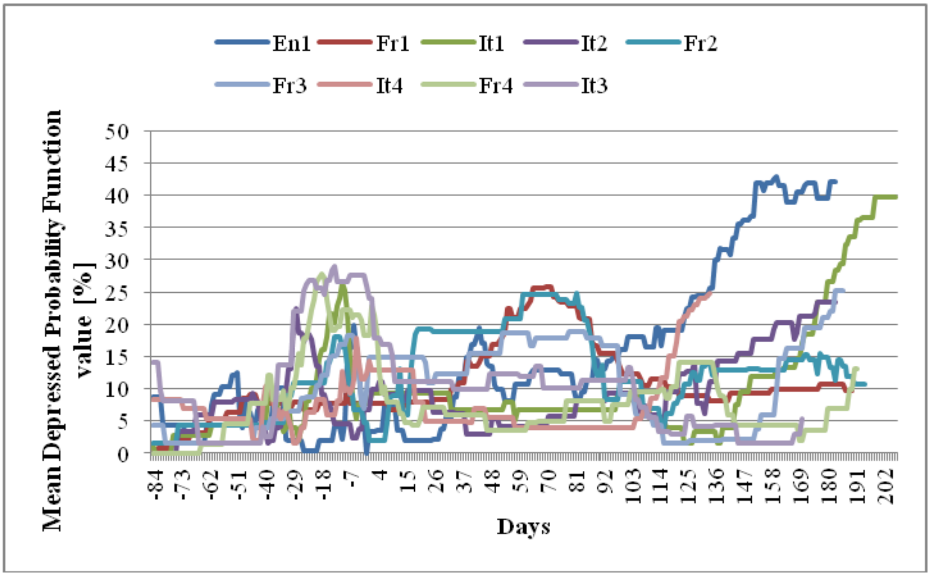

Figure 5:The Mean Depressed Probability Function values on the Concordia Speech Database 2014

As can be seen in *Fig. 5.*, there were two people whose MDP Function values were above 35% (It1 and En1). Here, and in contrast to the 2013 results, the starting MDP values were low for all the subjects, below 15%. It can be observed that on arrival at Concordia station several subjects' MDP value was relatively high: presumably this was caused by the change of environment.

En1's maximum depression probability value was 43%, reached around the end of May. This coincides with the beginning of winter. The depression probability value stayed around 39-44% during the rest of the examination (till the middle of the winter).

It1's maximum depression probability value was 40%, reached around the beginning of August, which coincides with the end of the winter, but the value started to rise from the middle of the winter (darkest day of the year). The depression probability value stayed around 40% during the rest of the examination (till the beginning of the spring).

So in 2014 it might be there were two people who suffered from SAD amongst the Concordia Crew. Presumably the severity of the depression for these subjects was not high, either.

6. Conclusions, future tasks

In this study we reviewed an online monitoring system to monitor the psychological condition of the Concordia Station’s crew members and to give alerts if any vocal disorders occur that could be a sign of cognitive dysfunction (especially SAD). We analysed segmental and supra-segmental acoustic-phonetic parameters from continuously read speech in order to show significant differences between the speech of depressed people and that of healthy people. In the speech of depressed people, we found that segmental parameters, such as fundamental frequency, F1, F2 formants frequencies, jitter, and shimmer, and supra-segmental parameters such as speech rate, length of pauses, intensity and fundamental frequency dynamics, show significant changes.

We examined the normal variety of the selected acoustic phonetic parameters and we found relatively large values (as big as the mean differences of the given parameter between healthy and depressed groups).

Our database is under continual expansion because the number of depressed people examined is underrepresented compared to the incidence of depression in the population. We are expanding our database with further people speaking in different languages as their mother tongue, in order to perform a full analysis and select a complete set of acoustic features that will enable more precise conclusions to be drawn. Our aim is to find a clear correlation between the severity of depression and the change of acoustic-phonetic parameters.

For further analysis free speech will also be used, in addition to the read folk tale, from the same people. The free speech recordings are stored in both databases, Seasonal Affective Disorder Speech Database and in Concordia Speech Database 2013 and Concordia Speech Database 2014.

As the result of this study, we suggest a method that can calculate depression probability as a function of time. We have found that, according to our method, a depression probability value below 25% should be considered as normal. We have monitored two groups, and in each group we have found people suffering from SAD. It would be very useful to verify our method results with other results, but sadly we are still waiting for LTMS data.

Acknowledgements

The authors would like to thank the COALA project: Psychological Status Monitoring by Computerised Analysis of Language phenomena (COALA) (AO-11-Concordia).

References

[1] Goberman, A. M. (2005), "Correlation between Acoustic Speech Characteristics and Non-speech Motor Performance in Parkinson's Disease", *Med Sci Monit*; vol. 11(3): pp. 109–116, 2005.

[2] Goberman, A. M., McMillan, J., "Relative Speech Timing in Parkinson's Disease", *Commun Science Disord*, vol.32, pp. 22–29, 2005.

[3] Metter, E. J., Hanson, W. R., "Clinical and Acoustical Variability in Hypokinetic Dysarthria", *Commun Disord*,vol. 19(5), pp. 347–366, 1986.

[4] Coppens, E., Van Audenhove, C., Scheerder, G., Arensman, E., Coffey, C., Costa, S., Koburger, N., Gottlebe, K., Gusmão, R., O'Connor, R., Postuvan, V., Sarchiapone, M., Sisask, M., Székely, A., van der Feltz-Cornelis, C., Hegerl, U., "Public attitudes toward depression and help-seeking in four European countries baseline survey prior to the OSPI-Europe intervention", *Journal of Affective Disorders*, vol 150(2), pp. 320–329, 2013.

[5] Ivry, R. B., Justus, T. C., Middleton, C., "The cerebellum, timing, and language: Implications for the study of dyslexia" In *M. Wolf (ed.), Dyslexia Fluency and the Brain. Timonium, MD: York Press*, pp. 198–211, 2001.

[6] Esposito, A., Bourbakis, N., "The role of Timing in Speech Perception and Speech Production Processes and its Effects on Language Impaired Individuals." *IEEE Computer Society, Sixth Symposium on BioInformatics and BioEngineering (BIBE'06)*, pp. 348–356, 2006.

[7] Vicsi, K., Sztahó, D., "Problems of the Automatic Emotion Recognitions in Spontaneous Speech; An Example for the Recognition in a Dispatcher Centre", In: *Esposito, A., Martone, R., Müller, V., Scarpetta, G., Toward Autonomous, Adaptive, and Context-Aware Multimodal Interface, Theoretical and Practical Issues. (6456) Heidelberg: Springer*, pp. 331–339, 2011.

[8] Tóth, Sz. L, Sztahó, D, Vicsi, K., "Speech Emotion Perception by Human and Machine", In: *Proceeding of COST Action 2102 International Conference: Revised Papers in Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction. Patras, Springer*, pp. 213–224, 2007.

[9] France, D. J, Shiavi, R. G, Silverman, S., Silverman, M., Wilkes, D. M., "Acoustical properties of speech asindicators of depression and suicidal risk", *IEEE Transactions of Biomedical Engineering*, pp. 829–837, 2000.

[10] Cannizzaro, M., Harel, B., Reilly, N, Chappell, P, Snyder, P. J., "Voice acoustical measurement of the severity of major depression", *Brain and Cognition*, vol. 56, pp. 30–35, 2004.

[11] Cannizzaro, M, Reilly, N, Mundt, J. C, Snyder, P. J., "Remote capture of human voice acoustical data by telephone: A methods study", *Clinical Linguistics & Phonetics*, vol. 19, pp. 649–658, 2005.

[12] Garcia-toro, M., Talavera, J. A, Saiz-Ruiz, J., Gonzalez, A., "Prosody impairment in depression measured through acoustic analysis", *The Journal of Nervous and Mental Disease*, vol. 188, pp. 824–829, 2000.

[13] Askenfelt, A., Sjölin, A., "Voice analysis in depressed patients: Rate of change of fundamental frequency related to mental state", *Dept. for Speech, Music and Hearing, Quarterly Progress and Status Report*, vol. 21(2), pp. 71–84, 1980.

[14] Daniel J. et al, "Acoustical Properties of Speech as Indicators of Depression and Suicidal Risk", *IEEE Transactions on Biomedical Engineering*, vol. 47(7), pp. 829–837, 2000.

[15] Thaweesak Y. et al, “Characterizing Sub-Band Spectral Entropy Based Acoustics as Assessment of Vocal Correlate of Depression”, *International Conference on Control, Automation and Systems*, pp. 1179–1183, Oct. 2010.
[16] Terapong B. et al, “Assessment of Vocal Correlates of Clinical Depression in Female Subjects with Probabilistic Mixture Modelling of Speech Cepstrum”, *2011 11th International Conference on Control, Automation and Systems*, pp. 387–391, Oct. 2011.
[17] James C. et al, “Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology”, *J Neurolinguistics*, vol. 20(1), pp. 50–64, Jan. 2007.
[18] Elliot M. et al, “Investigating the Role of Glottal Features in Classifying Clinical Depression”, *Proceedings of the 25th Annual International Conference of the IEEE*, pp. 2849–2852, 2003.
[19] Michelle Hewlett Sanchez, et al, “Using Prosodic and Spectral Features in Detecting Depression in Elderly Males”, in *Proc. INTERSPEECH 2011*, pp. 3001–3004, 2011.
[20] Kiss, G., Sztahó, D., Vicsi, K., “Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features.” In: *4th IEEE International Conference on Cognitive Infocommunications – CogInfoCom*, pp. 579–582, 2013.
[21] Abela, J. R. Z., D’Allesandro, D. U., “Beck’s cognitive theory of depression: The diathesis-stress and causal mediation components.” *British Journal of Clinical Psychology*, vol. 41, pp. 111–128, 2002.
[22] Kiss, G., Vicsi K., “Physiological and cognitive status monitoring on the base of acoustic-phonetic speech parameters”, In: *L Besacier, A-H Dediu, C Martín-Vide (ed.) Lecture Notes in Computer Science: Statistical Language and Speech Processing*, pp.120-131, 2014.
[23] Kiss, G., Sztahó D., Vicsi, K., Golemis, A., “Connection between body condition and speech parameters - especially in the case of hypoxia”, In: *5th IEEE International Conference on Cognitive Infocommunications – CogInfoCom*, pp. 333–336, 2014.
[24] Cummins, N., Sethu, V., Epps, J., Krajewski, J., “Probabilistic Acoustic Volume Analysis for Speech Affected by Depression”, In: *15th Annual Conference of the International Speech Communication Association - InterSpeech*, pp. 1238–1242, 2014.
[25] Bozkurt, E., Toledo-Ronen, O., Sorin, A., Hoory, R., “Exploring Modulation Spectrum Features for Speech-Based Depression Level Classification”, In *15th Annual Conference of the International Speech Communication Association - InterSpeech*, pp. 1243–1247, 2014.
[26] Hönig, F., Batliner, A., Nöth, E., Schnieder.S., Krajewski, J., “*Automatic Modelling of Depressed Speech: Relevant Features and Relevance of Gender*”, In:*15th Annual Conference of the International Speech Communication Association – InterSpeech*, pp. 1248–1252, 2014.