Automatic recognition of physiological parameters in the human voice: heart rate and skin conductance

Björn Schuller, Felix Friedmann, Florian Eyben

Angaben zur Veröffentlichung / Publication details:

Schuller, Björn, Felix Friedmann, and Florian Eyben. 2013. "Automatic recognition of physiological parameters in the human voice: heart rate and skin conductance." In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 26-31 May 2013, Vancouver, BC, Canada*, edited by Rabab Ward, Li Deng, Michael Adams, and Vicky Zhao, 7219–23. Piscataway, NJ: IEEE. https://doi.org/10.1109/icassp.2013.6639064.

AUTOMATIC RECOGNITION OF PHYSIOLOGICAL PARAMETERS IN THE HUMAN VOICE: HEART RATE AND SKIN CONDUCTANCE

Björn Schuller[1,2], Felix Friedmann[2], Florian Eyben[2]

[1]Institute for Sensor Systems, University of Passau, Germany
[2]Machine Intelligence & Signal Processing Group, MMK, Technische Universität München, Germany
Bjoern.Schuller@uni-passau.de

ABSTRACT

We show that high pulse/low pulse, heart rate and skin conductance recognition can reach good accuracies using classification on a large group of 4k audio features extracted from sustained vowels and breathing periods. A database containing audio, heart rate and skin conductance recordings from 19 subjects is established for evaluation of audio-based bio-signal recognition. On this database in speaker-dependent testing, heart rate and skin conductance can be determined with a correlation coefficient of .861/.960 and mean absolute error of 8.1 BPM/88.2 µMhO for regression based on sustained vowels recorded from a room microphone. Using the same set-up, a high pulse/low pulse classification can reach an unweighted accuracy of 82.7%. The results are largely independent from microphone type and the two bio-signals can be determined from breathing periods as well. Performance does, however, degrade in speaker-independent setting.

Index Terms— Speech Analysis, Computational Paralinguistics, Heart Rate, Skin Conductance

1. INTRODUCTION

The traditional model of a person visiting a doctor to receive medical treatment is being revolutionized at this moment. The variety of affordable and portable medical devices allowing a person to actively contribute to diagnosis and treatment is permanently increasing. Particularly for persons whose motility is limited or who are living remotely this trend may significantly improve quality of life. There are devices for measuring blood pressure, heart rate, body core temperature, respiration rate and many other physiological parameters autonomously, but yet they are still rather expensive and inconvenient for an everyday use. Ideally, monitoring of vital signals should require a minimal effort by a user and cause minimal disturbance; a user should not have to spend time thinking about the use of a monitoring device or even notice monitoring particularly. Monitoring should be easy enough to perform it in emergency situations, for example when calling a hospital, and monitoring should also be carried out over longer periods of time so that it could spontaneously react on emergencies or collect vital data for creating a health profile. For these reasons, physiological data should be recorded by sensors that are best 'unnoticed' in terms of intrusiveness. Disturbance of daily life would be minimal, if not a separate device had to be carried, but if monitoring of physiological parameters could be performed by computers or mobile phones which provide computational power in reach most of the time already. These considerations draw attention on signal types which can be recorded by mobile phones and computers easily: video and audio. In addition to the advantages mentioned above, video- and audio-based recognition can also be performed on past recordings, e.g., movies, songs and other voice recordings. A major advantage of audio-based over video-based recognition is that a microphone does not have to be directed towards a user's face or skin. It can also be employed as a complimentary technology in any situation in which a video camera is not available or able to record, e.g., in the dark.

2. RELATION TO PRIOR WORK

Given video- and audio-based bio-signal recognition's manifold advantages over competing technologies, there is already research and development in this area though, recently, the efforts in research have mainly focused on video-based recognition. Poh et al. [1] show that heart rate, breath rate and heart rate variability can be determined by a laptop's built-in video camera with great accuracy. Scully et al. [2] use the video of a subject's finger resting on a mobile phone's camera lens to extract heart rate, heart rate variability, breathing rate and blood oxygen saturation. There are several applications for mobile phones performing video-based heart rate and breath rate recognition on the market right now, based on a webcam as well as based on a mobile phone's camera.
For audio-based recognition, comparably few efforts have been made recently, but in 1989, Orlikoff and Baken [3] substantiated the connection between human voice and heartbeat. In the study, six male and six female participants had to produce sustained vowels while measured with an electroglottograph (EGG). By signal-averaging and autocorrelation, the study found that the heartbeat accounts for approximately 0.2% to 19% of absolute perturbation of the fundamental frequency (jitter) measured on pronunciations of sustained vowels. The influence of heart beats on jitter shown by Orlifkoff and Baken implies that heart rate information can be found in periodic changes of F_0. The present study evaluates heart rate (HR) recognition, skin conductance (SC) recognition and high pulse / low pulse (HP/LP) classification on features extracted from audio recordings. For this purpose an experiment is carried out during which subjects' HR and SC is recorded during breathing, pronunciation of sustained vowels and text reading. The recordings are connected with the subjects' personal data to create a database allowing further processing and evaluation. This study aims at providing a general evaluation of audio-signals for HR and SC recognition to determine whether audio-based recognition can be used as an alternative or supplement to current technologies. This has so far to our best

knowledge only been attempted for HR in vowels [4,5]. Furthermore, the study aims to examine in which settings audio-based recognition performs particularly bad or well in order to identify possible fields of application. Additionally possibilities for further improvements of the technology shall be identified. In a long-term perspective the outcomes of this study shall support the current development towards accurate, affordable, available and autonomous bio-signal monitoring, with the vision that usage of mobile medical diagnosis will ultimately become as common and natural as making a phone call. In the following sections, the experiment designed and carried out for HR and SC acquisition is described together with the methods employed to evaluate HR, HP/LP and SC recognition before results of the evaluation are shown and interpreted.

3. THE MUNICH BIOVOICE CORPUS

To collect the data for evaluation of audio-signal based recognition of physiological parameters, an experiment in which HR and SC are recorded simultaneously with vocal expressions was carried out with suited equipment: Wild Divine Inc.'s "iom" is a lightweight hardware device that records HR and SC data. It was initially designed for the "Journey to the Wild Devine" video game, but can also be used independently from the game for HR and SC recording. Data is collected from 3 sensors attached to a subject's fingers. The sensors are connected to a computer via USB. A Zoom Q3Hd camcorder equipped with an X-Y hd microphone was used to record audio ("room microphone") with a sampling rate of 92 kHz in PCM-wave format. In addition, a Logitech Clearchat Headset ("close-talk microphone") was used as representative of a typical headset available on the market connected to the laptop via USB. Overall, 19 subjects (4/15 female/male, 3 Chinese, 15 German, 1 Italian) gave their consent and participated in the experiment. All were free of temporary diseases, but the subjects include smokers and such with cardiac and neurological disorders. All subjects had to sign a letter of consent and fill out a questionnaire about their height, weight, nationality and health condition as well as the BFI-10 short personality test [6]. All subjects were recorded breathing, pronouncing the sustained vowel /a/ and reading a text with low pulse and with high pulse under constant, pre-defined conditions. For this purpose the subjects had to undergo a training period first before being recorded. The subjects raised their pulse by physically exercising (cf. Fig. 1, left).

Fig. 1: Left: a subject exercising, middle: a subject being recorded, right: a subject's hand grounded and connected to the "iom" sensors.

They were recorded sitting on a chair in front of a desk with the laptop used for recording and feedback (cf. Fig. 1, middle). The Zoom Q3Hd was placed on the desk in a distance of 50 centimeters from the subjects' lips, the Logitech Clearchat headset was head-worn. The Wild Divine iom's sensors were attached to the subject's left hand to measure BPM and SCL (cf. Fig. 1, right). The iom's heart rate sensor was connected to the middle finger and the two skin conductance sensors to the ring and forefinger of the subject. Iom data is collected with the Wild Divine Grapher. The subjects' left hands are grounded to minimize influence of noise by the computer's power supply on the headset recordings.

For each subject, a comfortable frequency F_c for the pronunciation of a sustained /a/ vowel is determined through a live frequency analysis. F_c is marked, and the subject has to train repeating the /a/ vowel in comfortable frequency (/a/$_c$) as precise as possible for several times. After the subject is able to intentionally produce /a/$_c$ within a tolerance of ± 7 Hz during 5 subsequent attempts, it is assumed that /a/$_c$ can be produced reliably during the recording. Following [3], the subject has to undergo the same training for /a/$_l$, an /a/ vowel with a frequency F_l 4 semi tone levels below F_c. In equal temperament, raising a frequency by one octave equals multiplying the frequency with 2, and raising it by one semitone equals multiplying it with $\sqrt[12]{2}$. Therefore, F_l is determined by $F_l = \sqrt[12]{2^{-4}} \cdot F_c$.

After the subjects finished their training, recording of both microphones was started. A short beep tone was produced simultaneously with starting recording with the Wild Divine iom. The subjects then had to pronounce /a/$_l$ and /a/$_c$ 4 times each and read a text with resting pulse. Native German speakers read out loud the text *"Der Nordwind und die Sonne"* – a standard text frequently used in phonetics – other subjects read the English version of the text *"The Northwind and the Sun"*. Next, the subjects had to repeat these three tasks, but each of them preceded by a physical exercise break – mostly running including staircases – to raise their pulse. During exercise breaks, the iom's sensors are removed from the subject's fingers and the subjects had to physically exercise until their pulse exceeded 90 BPM. After the subjects finished reading the text with high pulse, a second beep was generated and recording was stopped.

HR and SC data from the Wild Divine Grapher was stored in a table together with values of the fundamental frequency (F_0) over time for the audio recordings. F_0 is used for distinction between vowels pronounced in comfortable frequency and vowels pronounced in low frequency of a subject. The timestamps for HR and SC data show delay and protraction when compared to the audio recordings. This was corrected by a linear transformation that takes into account that 'beep' sounds in the audio recordings should be in sync with the beginning and end of recorded SC data as well as that 'clac' sounds in the audio recordings produced by removing the iom sensors from a subject's fingers should be in sync with sudden drops of SC data to zero. Sound chunks (vowel, breath or text) are selected and named according to their type. HR, SC and F_0 were looked up in the table for each sound chunk and are included in the chunk's name. Audio was then cut manually according to the beginning and end of according name tags. Overall, the final database – referred to as Munich BioVoice Corpus (or MBC for short) in the ongoing – consists of 1,420 BPM- and SCL-labeled audio recordings from 19 speakers. The instances are divided into 74 text periods, 644 breath periods and 630 sustained vowel expressions. They are further divided into low pulse and high pulse recordings and into headset and Q3Hd microphone recordings. Sustained vowels are labeled with F_0 data and divided into vowels sustained in comfortable or low frequency. Personal information, health state and the results of the psychological test of the subjects are included in the database which is available for scientific studies as per request including partitioning for reproduction of oncoming results.

Table 1: Results for the automatic classification of high/low pulse (HP/LP), and regression of heart rate and skin conductance by unweighted accuracy (UA), correlation coefficient (CC) and mean absolute error (MAE) on the Munich BioVoice Corpus. min/mean/max values: Heart Rate (BPM): 51.6/86.5/158.6, Skin Conductance (µMhO): 115.3/921.1/3,311.2.

Setting			HP/LP	Heart Rate		Skin Conductance	
			UA [%]	CC	MAE [BPM]	CC	MAE [µMhO]
Sustained Vowels	Close-talk Headset	Speaker independent (LOSO)	64.0	.343	17.5	.298	571.2
		Speaker only (SCV)	83.1	.809[f]	8.4[f]	.978[f]	84.4[f]
		All speakers (SCV)	79.6	.770	10.6	.891	265.3
	Room Microphone	Speaker independent (LOSO)	63.0	.366[f]	17.0[f]	.170[f]	626.9[f]
		Speaker only (SCV)	82.7	.861[f]	8.1[f]	.960[f]	88.2[f]
		All speakers (SCV)	76.0	.574	11.7	.633	311.2
Breathing Periods	Close-talk Headset	Speaker independent (LOSO)	70.2	.382	17.5	.131[f]	732.0[f]
		Speaker only (SCV)	84.1	.722[f]	10.7[f]	.908[fi]	153.7[fi]
		All speakers (SCV)	78.6	.629[f]	13.1[f]	.632[f]	469.7[f]
	Room Microphone	Speaker independent (LOSO)	62.8	.382[f]	17.3[f]	-.204[f]	881.3[f]
		Speaker only (SCV)	81.9	.718[f]	10.6[f]	.905[fi]	165.3[fi]
		All speakers (SCV)	72.9	.521	14.8	.483	570.8

[f] Predictions with an obvious error greater than 3000 µMhO or 200 BPM were excluded (cf. Section 6)
[i] Three subjects were excluded from analysis, due to sparse breath recordings

4. EXPERIMENTS

We use our openSMILE toolkit [7] to perform extraction of 4,368 acoustic features that we had defined as baseline features for the INTERSPEECH 2011 Speaker State Challenge [8]. The features consist of 4 energy-, 50 spectral- and 5 voice-related Low Level Descriptors (LLD) to which functionals are applied. To the energy related and spectral LLD and their first order deltas, base functionals are applied together with min, mean, max and the standard derivation of the segment length. To the voice related LLD and their first order deltas, the base functionals are applied together with quadratic mean, rise duration and fall duration of the signal in case of voicing probability greater than .7. The F_0 functionals are applied on the F_0 LLD and its first order derivate.
In detail, the LLD are:

- *4 energy related LLD*: Sum of auditory spectrum (loudness), Sum of RASTA-style filtered auditory spectrum, RMS Energy, Zero-Crossing Rate,
- *50 spectral LLD*: RASTA-style filt. auditory spectrum, bands 1–26 (0–8 kHz), MFCC 1–12, Spectral energy 25–650 Hz, 1 k–4 kHz, Spectral Roll Off Point 0.25, 0.50, 0.75, 0.90, Spectral Flux, Entropy, Variance, Skewness, Kurtosis, Slope
- *5 voice related LLD*: F_0, Probability of voicing, Jitter (local, delta), Shimmer (local)

Accordingly, functionals consist of:

- *33 base functionals*: quartiles 1–3, 3 inter-quartile ranges, 1% percentile (≈min), 99% percentile (≈max), percentile range 1 %–99%, arithmetic mean, standard deviation, skewness, kurtosis, mean of peak distances, standard deviation of peak distances, mean value of peaks, mean value of peaks – arithmetic mean, linear regression slope and quadratic error, quadratic regression a and b and quadratic error, contour centroid, duration signal is below 25% range, duration signal is above 90% range, duration signal is rising/falling, gain of linear prediction (LP), LP Coefficients 1–5
- *6 F_0 functionals*: percentage of non-zero frames, mean, max, min, standard deviation of segment length, input duration in seconds

For further good reproducibility of findings, we decided for the open-source Weka implementations [9] of support vector regression (SVR) for regression and support vector machines (SVM) for classification trained with the sequential minimal optimization algorithm using a linear kernel. Intra-speaker and speaker-independent classification performance measures are calculated from the distribution of the individual results for each speaker. For the nominal class HP/LP, unweighted accuracy (UA, i.e., recalls per classes added and divided by number of classes to cope with imbalance) is used as evaluation measure. Correlation coefficient (CC), and mean absolute error (MAE) are determined from the distribution of prediction results for numeric classes (HR and SC) following the standards set in the INTERSPEECH series of Computational Paralinguistics.

5. EXPERIMENTAL RESULTS

In the ongoing, we want to explore the influence of microphone setting, sound type and inclusion of a specific speaker's training data on the performance of recognition. The influence of microphone type and sound type on classification performance has

been investigated and is summarized in Table 1 for HR, SC and HP/LP. Recognition performance analysis was performed for vowels and breathing recordings via close-talk or room microphone with subject dependent testing (10-fold stratified cross-validation (SCV) with standard random seed in Weka) using either only the speaker or all data, and speaker independent testing (leave-one-speaker-out, LOSO). With a speaker's own training input, HR prediction shows a maximal CC of .861 with a MAE of 8.1 BPM for vowels via room microphone. Best CC reached for SC resembles .978 for vowels via close-talk microphone together with the lowest MAE of 84.4 µMhO. HP/LP classification shows a maximal UA of 84.1% for breathing via close-talk microphone. Speaker independent classification shows comparably low results with CCs smaller than .5 for HR and SC value prediction, but UAs above 60% for HP/LP classification, with a maximal UA of 70.2% observed for breathing via close-talk microphone.

6. DISCUSSION

The meaning of the results of this study is bound to the methods and tools employed to reach these. The influence of potentially erroneous reference recordings by the reference hardware, the validity of HP/LP classification and the meaning of the erroneous regressions are thus discussed now. A first aspect for discussion of the validity of this study's results is that HR and SC labels are inaccurate to a certain extent because of occasionally occurring temporary signal loss of the Wild Divine iom. Signal loss has been observed particularly for HR recordings, but also for SC recordings. As sound chunks were labeled with the mean HR and SC recorded by the iom during the duration of those chunks, the absence of measuring points led to reduced accuracies for the HR and SC values assigned to these sound chunks. Concerning the validity of HP/LP classification, it has to be considered that the employed threshold to decide between high pulse and low pulse was the mean of the HR labels of all recorded sustained vowels of a subject. The threshold is therefore based on a user-specific HR value, which means that the user-independent HP/LP classifications carried out in this study are in fact user-dependent classifications if considered as HR classifications. User independence of HP/LP classifications is valid on the other hand if the classification is not interpreted as a HR classification but as a classification for *physical excitement*. The underlying problem of this classification is that the resting pulse varies from person to person, making it impossible to find one fixed threshold that separates high from low pulses which is equally valid for all humans. For this reason, an individual threshold was chosen for each speaker for HP/LP classification. For comparison one examination was carried out employing a fixed threshold of 90 BPM for all subjects, and reached an UA of 78.6% for recognition on vowels via close-talk microphone. The prediction errors observed for speaker-only and speaker-independent recognition can be partially avoided by implementation of a predictor which classifies predictions as erroneous when they are not within a range of expected values (e.g., 20-250 BPM). Like this, a person performing audio-based recognition of bio-signals could be informed about the erroneous recognition and be queried to repeat the recognition. The erroneous predictions may be caused by a lack of training data. The examinations show a weak dependence of recognition accuracy on microphone type and a distinct, but not strong dependence on sound type, with vowel-based recognition of heart rate showing particularly good accuracy. CC's of HR and SC recognition are about .1 lower for breath periods than for sustained vowels, which might be explained by considering that the voice-related features (jitter, shimmer, F_0, probability of voicing) are not providing useful information but only add noise when extracted from breath periods. HP/LP classification shows little influence of sound type and even achieves better results for recognition on breath periods than for recognition on sound periods. This may be explained if HP/LP classification is interpreted as recognition of a state of physical excitement, since the influence of physical excitement on breath rate and depth can be recognized by humans.

7. CONCLUSION AND OUTLOOK

It was shown by this study that it can be determined by audio recordings of breath and sustained vowels whether a person's pulse is high or low. This was reached with good accuracy and largely independent from setting using large space acoustic feature extraction and support vector classification. Further, it was demonstrated that even heart rate value and skin conductance value can be recognized by according feature extraction and regression on breathing periods and sustained vowels. The performance observed was good, but far away from competing with the accuracy of medical equipment available today and also subject of partially severe outliers. Furthermore, good results for HR and SC values always required a speaker's training input, which limits spontaneous recognition as well as analysis of audio recordings without physiological reference data in possible applications. To improve future performance, we aim to increase recognition accuracy by selection of features, and to compare the importance of different feature groups. Recognition accuracy could be further increased by enlargement of the data set and development of a predictor able to recognize predictions out of an expected range of values. The fact that also breath periods allowed recognition with acceptable accuracy suggests that the technology can be used passively, for example by a mobile phone continuously recording a subject without its specific "contribution", i.e., talking. In this scenario, it would require less effort to provide sufficient training data, and a predictor could be implemented that considers the successive nature of predictions to recognize and eliminate outlier predictions within a more narrow range than the one possible with general upper and lower limits. Further, we had not used the read text chunks from the Munich BioVoice Corpus, yet. This section can be used in future studies to investigate performance on continuous speech, which may require an additional segmentation unit such as a phoneme or broader sound class recognizer, such as the one in [10]. Then, results as reached by fusion of physiological signals and speech, e.g., for stress [11] or emotion recognition [12,13] or related paralinguistic tasks [14] such as public speaking anxiety [15] could be reached using only speech, but deriving physiological data as "higher level features" in addition to typical speech features. Also, one could attempt to predict the HR and SC induced in a listener, such as babies listening to their mother's voice [16], any persons listening to stutterers [17] or in dependence of the text one reads [18], e.g., poems. Further, it is believed that stuttering can be expected according to HR [19]. If HR can be derived from speech, an Automatic Speech Recogniser can use this information to "expect" increased or decreased stuttering from speech. Altogether the study finds that audio-based recognition has the potential to be used in a software application on mobile phones and computers for remote monitoring of heart rate and skin conductance. If the technology can be further improved, it could be used for passive non-contact monitoring requiring a minimum of attendance by its user and improving live quality for many people.

8. REFERENCES

[1] M.Z. Poh, D.J. McDuff, and R.W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2012.

[2] C.G. Scully, J. Lee, J. Meyer, A.M. Gorbach, D. Granquist-Fraser, Y. Mendelson, and K.H. Chon, "Physiological parameter monitoring from optical recordings with a mobile phone," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 303–306, 2012.

[3] R.-F. Orlikoff, and R.J. Baken, "The Effect of the Heartbeat on Vocal Fundamental Frequency Perturbation," *Journal of Speech and Hearing Research*, vol. 32, no. 3, pp. 576–582, 1989.

[4] D. Skopin, and S. Baglikov, "Heartbeat feature extraction from vowel speech signal using 2D spectrum representation," in *Proc. 4th International Conference on Information Technology (ICIT)*, 6 pages, Amman, Jordan, June 2009.

[5] A. Mesleh, D. Skopin, S. Baglikov, and A. Quteishat, "Heart rate extraction from vowel speech signals," *Journal of Computer Science and Technology*, vol. 27, no. 6, pp. 1243-1251, November 2012.

[6] B. Rammstedt, and O.P. John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[7] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor," in *Proc. ACM Multimedia*, Florence, Italy, 2010, pp. 1459–1462.

[8] B. Schuller, A. Batliner, S. Steidl, F. Schiel, and J. Krajewski, "The INTERSPEECH 2011 Speaker State Challenge," in *Proc. Interspeech*, Florence, Italy, pp. 3201–3204, ISCA, 2011.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.

[10] F. Weninger, B. Schuller, M. Wöllmer, and G. Rigoll, "Localization of Non-Linguistic Events in Spontaneous Speech by Non-Negative Matrix Factorization and Long Short-Term Memory," in *Proc. 36th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, IEEE, Prague, Czech Republic, pp. 5840–5843, 2011.

[11] H.S. Hayre, and J.C. Holland, "Cross-correlation of voice and heartrate as stress measures," *Applied Acoustics*, vol. 13, no. 1, pp. 57–62, January–February 1980.

[12] J. Kim, E. André, M. Rehm, T. Vogt, and J. Wagner, "Integrating Information from Speech and Physiological Signals to Achieve Emotional Sensitivity," in *Proc. Interspeech*, pp. 809-812, ISCA, Lisbon, 2005.

[13] C.A. Frantzidis, C.D. Lithari, A.B. Vivas, C.L. Papadelis, C. Pappas, P.D. Bamidis, "Towards emotion aware computing: A study of arousal modulation with multichannel event-related potentials, delta oscillatory activity and skin conductivity responses," in *Proc. 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, IEEE, Athens, Greece, pp. 1-6, 2008.

[14] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Muller, and S. Narayanan, "Paralinguistics in Speech and Language – State-ofthe-Art and the Challenge," *Computer Speech and Language*, Special Issue on Paralinguistics in Naturalistic Speech and Language, vol. 27, pp. 4–39, January 2013.

[15] R.J. Croft, C.J. Gonsalvez, J. Gander, L. Lechem, R.J. Barry, "Differential relations between heart rate and skin conductance, and public speaking anxiety," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 35, no. 3, pp. 259-271, September 2004.

[16] B.S. Kisilevsky, S.M. Hains, K. Lee, X. Xie, H. Huang, H.H. Ye, K. Zhang, Z. Wang, "Effects of experience on fetal voice recognition," *Psychological Science*, vol. 14, no. 3, pp. 220-224, SAGE Publishers, May 2003.

[17] J. Zhang, J. Kalinowski, T. Saltuklaroglu, D. Hudock, "Stuttered and fluent speakers' heart rate and skin conductance in response to fluent and stuttered speech," *International Journal Language Communication Disorders*, vol. 45, no. 6, pp. 670-680, November-December 2010.

[18] H. Bettermann, D. von Bonin, M. Frühwirth, D. Cysarz, M. Moser, "Effects of speech therapy with poetry on heart rate rhythmicity and cardiorespiratory coordination," *International Journal of Cardiololgy*, vol. 84, no. 1, pp. 77-88, 2002.

[19] J.M. Baumgartner, and G.J. Brutten, "Expectancy and Heart Rate as Predictors of the Speech Performance of Stutterers," *Journal of Speech and Hearing Research*, vol. 26, pp. 383-388, September 1983.