Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Heart Rate as a Proxy Measure to Assess Human Confidence in Spoken Speech

Harish Battula, Gauri Deshpande, Yagna Gudipalli, Sachin Patel

TCS Reseach and Innovation Pune, India

(battula.h, gauri1.d, yagna.gudipalli, sachin.patel)@tcs.com

Abstract

Human confidence reflects a positive self-perception and balanced autonomic nervous system response. In this paper, we present a three stage approach to detect human confidence level by computing the heart rate from speech signals. First stage involves extraction of breathing patterns from speech using a pre-trained model followed by stage 2, where the heart rate is extracted using Independent Component Analysis (ICA) on the breathing patterns. Finally in stage 3, an analysis of heart rate values indicating human confidence levels is done. To the best of our knowledge, this is the first time ever, with our experiments, empirically it is found that heart rate extracted from speech carries information related to the confidence of the candidate. We observe that, on an average, confident speakers have 10 beats per minute lower heart rate as compared to non-confident speakers.

Index Terms: health informatics, heart-rate, independent component analysis, speech-breathing, human-confidence

1. Introduction

The ability to determine the confidence of a person is of paramount importance in several soft-skill assessing use-cases, for example during an interview. Speech is a communication modality that is non-intrusive and easy to gather during a human-human or a human-machine conversation. Though not directly related, it is well known that there is a correlation between the physiological signal, namely heart rate (HR), and self-esteem, a psychological state of an individual [1]. In this paper, we propose deriving HR, from spoken speech signal to assess human confidence. While it is essential to utilize multiple modalities to gain a comprehensive understanding of an individual's psychological state, HR extracted from speech can serve as an additional valuable indicator of human confidence.

1.1. Human Confidence & HR:

The relationship between confidence and HR has been explored in literature. A thorough investigation carried out in [1], shows a consistent pattern that individuals with heightened self-esteem consistently exhibit lower HR levels across diverse time points. Another study extends this exploration to unveil the nuanced connections between self-esteem and HR during a social performance task [2]. Notably, speakers displaying elevated HR prior to a jury presentation are found to experience a simultaneous decline in confidence and a surge in self-reported anxiety [3]. Moreover, individuals characterized by high socially prescribed perfectionism showcase heightened HR responses when confronted with negative feedback, coinciding with diminished confidence levels [4].

1.2. HR Detection from Speech:

Numerous studies have highlighted the relationship between human speech and physiological parameters, HR. For instance, though the accuracy of the method was not discussed, estimation of HR through statistical analysis of speech signal was explored in [5]. A data mining approach was employed by [6], to establish a connection between HR and vocal frequency, using multiple speech recordings, albeit from only two speakers. Along similar lines, [7] demonstrated that fluctuations in vocal fundamental frequency (F0) could be attributed to the influence of heartbeats. Additionally, [8] investigated the estimation of HR and skin conductance using speech features and machine learning techniques. While the performance was moderate in terms of performance accuracy, they established the relevance of MFCC features for HR estimation from speech. Another approach, presented in [9], focused on detecting HR using speech-derived breathing (SDB) patterns for seven speakers.

1.3. Human Confidence Detection from Speech:

The studies [10, 11, 12] focused on detecting confidence expressions from speech signals. They found that parameters like F0, amplitude, speech rate, duration, and harmonic-to-noise ratio played a crucial role in classifying confidence levels, achieving a 62% accuracy in speaker-independent analysis. Another study [13] explored the influence of vocal speed, intonation, and pitch on the perception of confidence; they concluded that higher speech rates, falling intonation, and lower pitch indicated high speaker confidence. Confidence expressions in 195 children aged between 10-14 years was studied in [14] by analyzing a spoken paragraph. They used acoustic features like pause, pitch, and speech rate with a random forest regressor, achieving 65% accuracy for three-class classification and 82% for binary classification (combining high and medium classes as high). However, this analysis focuses on evaluating students' comfort with the language and may not directly assess a speaker's self-efficacy in responding to unknown scenarios.

In this paper, we present a three stage framework to estimate human confidence from spoken speech. We first derive the breathing patterns from the speech signal (Stage #1), followed by HR from the breathing signal (Stage #2) and mapping HR to confidence (Stage #3). This is the main contribution of the paper. We explain in details the three stage approach to derive confidence from speech signal and evaluate the performance on real data.

2. Proposed Approach

The proposed three stage approach is shown in Figure 1. In Stage #1, breathing patterns $\boldsymbol{b}$ are derived from the speech sig-

 10.21437/Interspeech.2025-1570

nal $s(t)$. A pre-trained model, `SBreathNet` [15] takes a speech signal $s(t)$, sampled at 16 kHz as input and generates a breathing pattern $\boldsymbol{b}$ at 50 Hz. In Stage #2 HR $\boldsymbol{h}$ is extracted at 50 Hz (f_{hr}) from the $\boldsymbol{b}$. The final Stage #3 adopts these $\boldsymbol{h}$ to estimate the human confidence level (binary class: c^+, c^-).

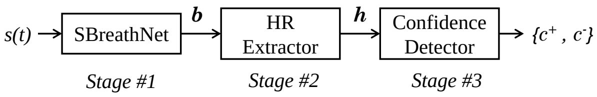

Figure 1: *Stage #1 detects breathing patterns $\boldsymbol{b}$ from speech signal $s(t)$,* HR *$\boldsymbol{h}$ is detected in Stage #2 from $\boldsymbol{b}$ and finally in Stage #3 $\boldsymbol{h}$ is analysed for human confidence*

2.1. Datasets

The three datasets used for conducting this study were collected by authors. The study analyses HR values, a physiological signal, to assess human confidence levels derived from $s(t)$. Hence, extracting $\boldsymbol{b}$, another physiological signal, from $s(t)$ appeared to align closely with our objectives. Considering that the labeled confidence dataset reflects the Indian demographic, we opted for a pre-trained model `SBreathNet`, developed on a dataset from the same demographic for consistency.

Stage #1 Dataset (D1): In Stage #1, the pre-trained model `SBreathNet` is trained using data from 100 college students (69 male; 31 female) aged between 18 to 23 years as described in [16]. Time-synchronized speech $s(t)$ and breathing signal $b(t)$ is captured using a condenser microphone and ADInstruments' respiratory belt transducer. The transducer is positioned 4 centimeters (cm) below the collarbone, and the microphone is placed approximately 4 cm from the mouth. Additionally, physiological parameters like pulse rate and blood pressure are also recorded, to ensure all speakers showed normal readings. speakers are seated and given 2 minutes to relax before they are asked to read a couple of phonetically balanced Harvard sentences [17]. Each speaker takes around 2-3 minutes to read them. Since the device supports single sampling rate for all the channels, both speech and breathing signals are sampled at 40 kHz; speech is down-sampled to 16 kHz and breathing to 50 Hz.

Stage #2 Dataset (D2): A speech dataset is collected from 41 speakers to investigate the relationship between speech signals and HR dynamics. The dataset is collected in two parts. First part uses a commonly available wearable device (Apple Watch) to collect HR data hence referred to as wearable device-based HR dataset (WD-HRD). The second part follows a clinical setup where HR is measured using highly accurate instruments, hence referred to as Clinical ECG-Based HR dataset (C-ECG-HRD). The goal is also to assess if wearable, while being commonly available and cost-effective, can serve as an alternative to expensive clinical setups, which are not always readily accessible. In WD-HRD, HR is continuously measured from 15 speakers while in C-ECG-HRD, ECG signals are acquired from 26 speakers using the ADInstruments ECG kit, consisting of four limb leads placed on the left and right wrists and ankles. HR is extracted for every 5 seconds (s) of ECG data using a peak detection algorithm [18]. In both setups, the speakers are physically fit individuals with no known cardiovascular conditions, and informed consent is obtained prior to data collection. Each speaker reads aloud a phonetically balanced text passage, as used in [16], while synchronized $s(t)$ and $h(t)$ are recorded. Speech is captured using a condenser microphone positioned approximately 4 cm from the speaker's mouth. and the recorded duration varies from 60 to 120 s per speaker resulting in 12 $(60/5)$ to 24 $(120/5)$ HR samples. The cohort includes 10 female and 30 male subjects aged between 25 and 40 years. This age range is chosen to minimize confounding factors related to age-associated physiological changes affecting cardiovascular and speech production mechanisms. All speech recordings are in `wav` format, sampled at 40 kHz, and down sampled to 16 kHz using `sox`.

Stage #3 Dataset (D3): A dataset of 51 individuals in the age group 22 to 30 years is collected with labels of confidence levels as described in [19]. The speech data is collected through video phone calls at a sampling rate of 16 kHz. Each interview session consists of five questions designed to elicit varying levels of confidence. speakers are unaware of the questions beforehand, ensuring spontaneous responses. Responses are labeled by the speaker as well as three additional researchers into two categories: confident (c^+) or non-confident (c^-). The annotators, two females and one male, aged between 30 and 40, work in behavioral studies and are Indian. Labels are determined by majority voting, with one label per response. Speech extracted from these audio-visual responses is used henceforth for the current analysis. Out of 255 responses (51 individuals $\times$ 5 questions), at least two researchers' labels match those of the candidates. This results in a majority vote of three out of four for all responses. There are 37 c^-responses (14% of total responses) and 218 c^+responses.

2.2. Methodology

Figure 2 explains the three stage approach to analyse human confidence levels computed from HR derived from speech signal $s(t)$.

2.2.1. Extracting b from $s(t)$

In Stage #1, `SBreathNet`, the pre-built model, extracts SDB patterns ($\boldsymbol{b}$) for every non-overlapping 5 s of $s(t)$. The $\boldsymbol{b}$ has a dimension of $(N, 250)$ where N represents the number of 5 s chunks in $s(t)$ and 250 is the embedding size output by `SBreathNet` model. We use `SBreathNet` to extract SDB patterns from datasets D2 and D3. Note that this yields $\boldsymbol{b}$ for dataset D2 along with the ground truth HR and $\boldsymbol{b}$ for dataset D3 with ground truth confidence labels (c^+, c^-).

2.2.2. Extracting h from b

As seen in Figure 2, the SDB patterns ($\{\boldsymbol{b}^i\}_{i=1}^N$) are passed through the HR extractor. A signal processing technique, described below allows for extraction of $\boldsymbol{h}$ from $\boldsymbol{b}$.

As a first step, we employ the well known Independent Component Analysis (ICA) to extract the HR and respiration signals from the SDB patterns, $\boldsymbol{b}^i$. Note that, $\boldsymbol{b}^i$ can be regarded as a linear combination of multiple underlying signals, such as the cardiac, chest motion, noise, and many more. We assume the $\boldsymbol{b}^i$ to be formed of three components: respiration ($\boldsymbol{s}_1$), HR ($\boldsymbol{s}_2$), and residual ($\boldsymbol{s}_3$). Hence, an observation for a speaker can be expressed as:

$$\boldsymbol{b}^i = \sum_{k=1}^{3} a_k^i \boldsymbol{s}_k^i \quad (1)$$

for $i = 1, \cdots N$. Here $\boldsymbol{b}^i$ represents the i^{th} 5 s segment of the SDB pattern, $\boldsymbol{b}$ for a given subject. Note that $\boldsymbol{s}_k^i$ for $k = 1, 2, 3$, are the original unknown statistically independent components of the i^{th} SDB pattern $\boldsymbol{b}^i$, and a_k^i are the weights corresponding

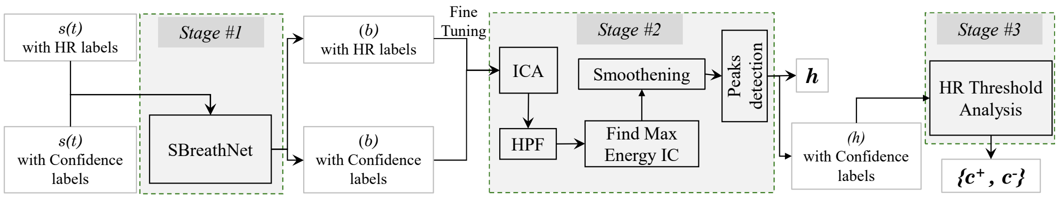

Figure 2: *Stage #1:* `SBreathNet` *extracts breathing patterns* ($\boldsymbol{b}$) *for D2 and D3 datasets. Stage #2: The* HR *labels used to fine-tune* HR *estimation. Stage #3: Predicting the confidence labels for the predicted* HR .

to each unknown component. The original SDB pattern ($\boldsymbol{b}$), $\boldsymbol{s}_1$, and $\boldsymbol{s}_2$ are as seen in Figure 3.

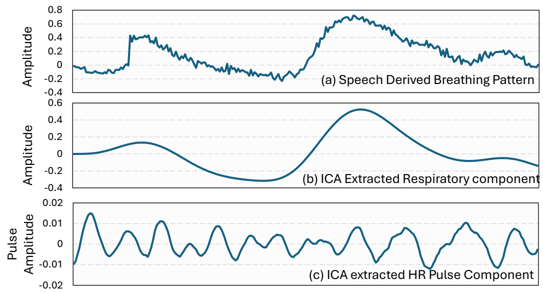

Figure 3: *(Top)* $\boldsymbol{b}$ *extracted from speech, (middle) respiratory signal* ($\boldsymbol{s}_1$)*. (bottom)* HR *pulse* ($\boldsymbol{s}_2$)*.*

We used FastICA [20] due to its computational efficiency and robustness, leveraging its nonlinear approach to separate the cardiac and respiratory components from the SDB pattern.

We now high-pass filter (HPF) the extracted IC's, namely, $\{\boldsymbol{s}^i\}_{k=1}^3$, as shown in Figure 2. We select the HPF's cutoff frequency at 2.2 Hz, this translates to 132 beats per minute (bpm) to eliminate frequencies associated with respiration and its key correlates. Notably, in many instances involving heartbeat harmonics, their power surpasses that of respiratory harmonics [21]. Therefore, we compare the power of ICs post-HPF and select the IC exhibiting the greatest power as the heart signal.

We now extract peaks in the HR signal ($\boldsymbol{s}_2^i$) for every $i = 1, \cdots N$ (produced for every 5 s) using four hyper-parameters, namely, height (δ), threshold (τ), distance (d), and smoothing window (s_{win}). The parameter d determines the minimum distance between two consecutive peaks, while the s_{win} improves peak detection accuracy by smoothing the HR signal. The parameter δ sets the minimum peak height, and the τ defines the minimum relative height for peak detection. The HR is computed as

$$\boldsymbol{h}^i = \frac{60 \times \alpha f_{hr}}{d_{\text{average}} \text{ in } \boldsymbol{s}_2^i} \tag{2}$$

for all $\{h^i\}_{i=1}^N$ where, α = 2 is the scaling factor. This method of extracting HR from breathing patterns is applied to the SDB patterns with ground truth HR values of D2 dataset and ground truth confidence values of D3 dataset. The δ, τ, d, s_{win} are fine-tuned using grid search method for each of the specified dataset. Both the WD-HRD and C-ECG-HRD are used for fine-tuning to determine the optimal hyper-parameter values. Hyper-parameters obtained with fine-tuning WD-HRD are tested on C-ECG-HRD and vice-versa.

2.2.3. *Analysing* HR *values for human confidence dataset*

The estimated HR $\boldsymbol{h}^i$ are analyzed with the c^+ or c^- label to understand the correlation between human confidence and HR. We adopt the strategy from [19] to analyze the speaker-independent five folds ($SI5F$) of the D3 dataset. A threshold value of HR (HR_{thr}) ranging between 80 to 110 is determined, effectively distinguishing between c^+ and c^- classes of the four training folds. The HR_{thr} value that yields the best combined performance across all the selected metrics - precision, recall, F1-score ($F1$), accuracy, and area under the curve (AUC) in discriminating the two classes of training partition is considered. This HR_{thr} is then used to classify c^+ and c^- samples in the test data of the fifth fold. This process is repeated five times, with each fold serving as the testing partition once.

3. Observations and Results

3.1. Performance of `SBreathNet`

The performance of `SBreathNet`, is calculated using the Pearson's correlation coefficient (r-value) as metric. An average r-value of 0.61 is achieved across the 100 speakers with concordance correlation coefficient (CCC) loss functions. The breaths-per-minute (BPM) count for every speaker is calculated on the breathing predictions obtained and compared with that of the true breathing pattern. An average BPM error obtained is 2.50 across 100 speakers.

3.2. Heart Rate Detection

Table 1: *Hyper-parameters for peak detection algorithm*

Dataset	δ	τ	d	s_{win}
WD-HRD	mean	0	53	25
C-ECG-HRD	mean	0	45	25

Table 1 presents the fine-tuned hyper-parameters for both the WD-HRD and C-ECG-HRD datasets. The WD-HRD is fine-tuned with a larger d value for optimal performance. On both datasets, the δ is set to the mean of the HR signal over 5 s & the τ is fixed to zero. Additionally, setting s_{win} to 25 gives the optimal performance across both the datasets.

Table 2: *Tuning and Testing MAE (bpm) for* HR *Estimation*

Tuning/Testing	**Tuning MAE**	**Testing MAE**
WD-HRD/C-ECG-HRD	10.96	10.31
C-ECG-HRD/WD-HRD	7.16	16.95

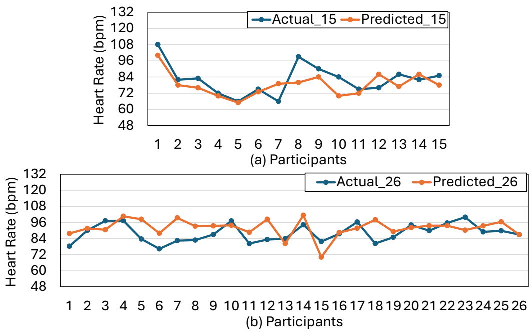

Figure 4: *Comparison of actual versus predicted average* HR *(a) WD-HRD tuning, (b) C-ECG-HRD testing.*

The performance of the HR detection algorithm using the hyper-parameters (see Table 1) is shown in Table 2 depicting the mean absolute error (MAE) results. As observed, the performance remains nearly identical when fine-tuned on the WD-HRD and tested on the C-ECG-HRD. This suggests a stronger generalization capability. Therefore, for further analysis, we have chosen the hyper-parameters obtained from fine-tuning on the WD-HRD. The actual versus the predicted HRs for the WD-HRD (Tuning) are shown in Figure 4a, and for the C-ECG-HRD (Testing) are in Figure 4b.

3.3. Heart-rate of Confident and Non-confident Speakers

Dataset D3 contains 217 c^+ and 37 c^- samples. Figure 5 shows the distribution of average HR for c^+ and c^- speakers. It can be observed that the HR is lower by at least 10 bpm across all data for c^+ speakers compared to c^- speakers. Certain c^+ speakers exhibit unusually high HR values, ranging from 120 bpm to 191 bpm, that are not explained by the available metadata such as age and gender. As mentioned earlier, HR_{thr} values are

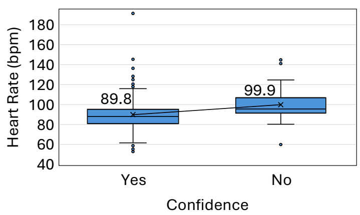

Figure 5: *Average* HR *of* c^+ *speakers is* 90 *bpm and* 100 *bpm for* c^- *speakers.*

identified such that optimal performance is obtained in discriminating between c^+ and c^- speakers across $SI5F$. Figure 6 displays the average values of five performance metrics: 1) precision, 2) recall, 3) F1-score, 4) accuracy, and 5) AUC across a range of HR_{thr} values from 80 to 110 bpm. A HR_{thr} of 91 bpm was found to strike a balance between precision, recall, and AUC, making it a reliable threshold for distinguishing c^+ and c^- speakers. Specifically, c^- samples are those with HR values above 91 bpm, while c^+ samples are those with HR values below 91 bpm.

Table 3 exhibits the performance across $SI5F$ considering 91 bpm as the HR_{thr} value. These results highlight the reliability of HR based confidence classification when compared to the ground truth annotations provided by professionals, underscoring the potential of HR analysis as a robust feature for

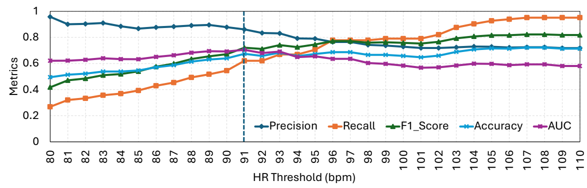

Figure 6: HR_{thr} *- versus metrics averaged over* $SI5F$.

Table 3: $SI5F$ *cross validation on D3 dataset with* HR *threshold -* HR_{thr} *as 91 bpm.*

Fold	Precision	Recall	F1	Accuracy	AUC
1	1.00	0.5	0.66	0.62	0.75
2	0.86	0.65	0.74	0.71	0.73
3	0.88	0.60	0.71	0.67	0.70
4	0.75	0.56	0.64	0.60	0.61
5	0.85	0.80	0.82	0.79	0.78
Avg	0.87	0.62	0.72	0.68	0.72

non-invasive confidence assessment.

3.4. Comparative Analysis

As demonstrated by [13] and [14], pause, speech rate, and pitch are effective in assessing human confidence. We calculated the average pause percentage (PP_{avg}), average pitch frequency (f_{0avg}), and pitch standard deviation (f_{0std}) for c^+ and c^- speaker groups, finding differences of 23%, 108 Hz, and 33 Hz, respectively, with c^- speakers exhibiting higher values. We also compared the $SI5F$ cross-validation performance of HR with SDB patterns from [19]. While SDB patterns performed better individually, their combination with HR led to improved recall, $F1$ score, and accuracy, as shown in Table 4. Although threshold-based analysis for pitch, pause, and pitch standard deviation was performed, it did not outperform HR-based analysis, which was considered optimal in the combined approach.

Table 4: *Performance of classifying human confidence levels exhibited using* HR *(HR Conf), SDB Patterns (SDB Conf), and the combination of both* HR *and SDB Patterns.*

Method	Precision	Recall	F1	Accuracy	AUC
HR Conf	0.87	0.62	0.72	0.68	0.72
SDB Conf	0.88	0.85	0.86	0.81	0.76
Combined	0.83	0.92	0.88	0.82	0.74

4. Conclusion and Future Work

In this paper, we propose a three-stage framework for detecting human confidence from speech, with a focus on HR as a key parameter. Our findings demonstrate that HR effectively reflects confidence levels, aligning with previous psychological studies [1], [2], [3], [4], and [22], which report lower HR in individuals with high self-esteem. By extracting HR from speech using SDB patterns, we provide empirical evidence supporting the link between physiological and psychological indicators of confidence. In future work, we aim to expand this framework by integrating HR into a broader set of behavioral indicators, thereby enhancing our understanding of human confidence and improving model interpretability.

5. Acknowledgments

The authors would like to express their sincere gratitude to Dr. Sunil Kumar Kopparapu for his invaluable feedback and guidance during the review of the manuscript and assistance in formatting the paper, which significantly improved the quality and presentation of this work.

6. References

[1] K. O'Donnell, L. Brydon, C. E. Wright, and A. Steptoe, "Self-esteem levels and cardiovascular and inflammatory responses to acute stress," *Brain, Behavior, and Immunity*, vol. 22, no. 8, pp. 1241–1247, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0889159108003036

[2] N. A. Rector and D. Roger, "The stress buffering effects of self-esteem," *Personality and Individual Differences*, vol. 23, no. 5, pp. 799–808, 1997.

[3] J. L. Abel and K. T. Larkin, "Anticipation of performance among musicians: Physiological arousal, confidence, and state-anxiety," *Psychology of music*, vol. 18, no. 2, pp. 171–182, 1990.

[4] A. Besser, G. L. Flett, P. L. Hewitt, and J. Guez, "Perfectionism, and cognitions, affect, self-esteem, and physiological reactions in a performance situation," *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, vol. 26, pp. 206–228, 2008.

[5] J. Kaur and R. Kaur, "Extraction of heart rate parameters using speech analysis," 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:8139427

[6] M. Sakai, "Modeling the relationship between heart rate and features of vocal frequency," *International Journal of Computer Applications*, vol. 120, no. 6, pp. 32–37, June 2015. [Online]. Available: https://ijcaonline.org/archives/volume120/number6/21233-3986/

[7] R. F. Orlikoff and R. J. Baken, "The effect of the heartbeat on vocal fundamental frequency perturbation," *Journal of Speech, Language, and Hearing Research*, vol. 32, no. 3, pp. 576–582, 1989. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/jshr.3203.576

[8] B. Schuller, F. Friedmann, and F. Eyben, "The munich biovoice corpus: Effects of physical exercising, heart rate, and skin conductance on human speech production," in *International Conference on Language Resources and Evaluation*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:1500634

[9] H. Battula, G. Deshpande, S. Patel, and B. W. Schuller, "Heart rate from read-speech influenced by physical exercise," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 376–380.

[10] X. Jiang and M. D. Pell, "Encoding and decoding confidence information in speech," in *Proceedings of the 7th international conference in speech prosody (social and linguistic speech prosody)*, vol. 5762579, 2014.

[11] "The sound of confidence and doubt," *Speech Commun.*, vol. 88, no. C, p. 106–126, Apr. 2017. [Online]. Available: https://doi.org/10.1016/j.specom.2017.01.011

[12] X. Jiang and M. D. Pell, "Predicting confidence and doubt in accented speakers: Human perception and machine learning experiments," *Speech Prosody 2018*, 2018.

[13] J. J. Guyer, L. R. Fabrigar, and T. I. Vaughan-Johnston, "Speech rate, intonation, and pitch: Investigating the bias and cue effects of vocal confidence on persuasion," *Personality and Social Psychology Bulletin*, vol. 45, no. 3, pp. 389–405, 2019.

[14] K. Sabu and P. Rao, "Automatic prediction of confidence level from children's oral reading recordings." in *Proceedings of the 21st Annual Conference of the International Speech Communication Association*. Shanghai, China: INTERSPEECH Communication Association (ISCA), 2020, pp. 3141–3145.

[15] G. Deshpande, B. W. Schuller, P. Deshpande, A. R. Joshi, S. Oza, and S. Patel, "Analysing breathing patterns in reading and spontaneous speech," in *International Conference on Speech and Computer*. Springer, 2023, pp. 3–17.

[16] G. Deshpande, B. W. Schuller, P. Deshpande, and A. R. Joshi, "Automatic breathing pattern analysis from reading-speech signals," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2023, pp. 1–4.

[17] "Ieee recommended practice for speech quality measurements," *IEEE No 297-1969*, pp. 1–24, 1969.

[18] S. Community, "scipy.signal.findPeaks," 2008, [Accessed: Feb. 7, 2025]. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

[19] G. Deshpande, Y. Gudipalli, S. Patel, and B. W. Schuller, "Applying speech derived breathing patterns to automatically classify human confidence," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 1335–1339.

[20] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[21] Y. Rong and D. W. Bliss, "Remote sensing for vital information based on spectral-domain harmonic signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3454–3465, 2019.

[22] M. Allen, D. Frank, D. Samuel Schwarzkopf, F. Fardo, J. S. Winston, T. U. Hauser, and G. Rees, "Unexpected arousal modulates the influence of sensory noise on confidence. elife, 5 (october2016)," 2016.