Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

© 2022 European Federation for Medical Informatics (EFMI) and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI220498

Out-of-Hospital Cardiac Arrest Detection by Machine Learning Based on the Phonetic Characteristics of the Caller's Voice

Sonia RAFI[a]; Cedric GANGLOFF[a]; Etienne PAULHET[b]; Ollivier GRIMAULT[cd]; Louis SOULAT[b]; Guillaume BOUZILLÉ[a]; Marc CUGGIA [a]
[a] *Univ Rennes, CHU Rennes, INSERM, LTSI-UMR 1099, F-35000 Rennes, France*
[b] *Department of Emergency Medicine, CHU Pontchaillou, Rennes, France*
[c] *Search And Rescue team, Department of Emergency Medicine, CHU de la Cavale Blanche, Brest, France*
[d] *EA4324-ORPHY, Université de Bretagne Occidentale, UFR Sciences et Techniques, Brest, France*

Abstract. Introduction. Out-of-hospital cardiac arrest (OHCA) is a major public health issue. The prognosis is closely related to the time from collapse to return of spontaneous circulation. Resuscitation efforts are frequently initiated at the request of emergency call center professionals who are specifically trained to identify critical conditions over the phone. However, 25% of OHCAs are not recognized during the first call. Therefore, it would be interesting to develop automated computer systems to recognize OHCA on the phone. The aim of this study was to build and evaluate machine learning models for OHCA recognition based on the phonetic characteristics of the caller's voice. **Methods.** All patients for whom a call was done to the emergency call center of Rennes, France, between 01/01/2017 and 01/01/2019 were eligible. The predicted variable was OHCA presence. Predicting variables were collected by computer-automatized phonetic analysis of the call. They were based on the following voice parameters: fundamental frequency, formants, intensity, jitter, shimmer, harmonic to noise ratio, number of voice breaks, and number of periods. Three models were generated using binary logistic regression, random forest, and neural network. The area under the curve (AUC) was the primary outcome used to evaluate each model performance. **Results.** 820 patients were included in the study. The best model to predict OHCA was random forest (AUC=74.9, 95% CI=67.4-82.4). **Conclusion.** Machine learning models based on the acoustic characteristics of the caller's voice can recognize OHCA. The integration of the acoustic parameters identified in this study will help to design decision-making support systems to improve OHCA detection over the phone.

Keywords. cardiac arrest, resuscitation, dispatcher, call center, acoustic, phonetic, voice analysis, artificial intelligence, machine learning.

1. Introduction

Out-of-hospital cardiac arrest (OHCA) is a major public health concern [1]. The prognosis is closely related to the time from collapse to return of spontaneous circulation [2]. The resuscitation efforts are frequently initiated at the request of emergency call center professionals who are specifically trained to identify critical conditions over the phone. However, 25% of OHCAs are not recognized during the first call, most often because

emergency call centers are overwhelmed [3]. In this context, it would be interesting to develop automated computer systems to recognize OHCA based on the bystanders' speech on the phone.

Speech analysis can be decomposed in two fields: linguistic and phonetic. Linguistic analysis investigates the meaning of words and their relationships, while phonetic analysis focuses on the voice acoustic characteristics. Acoustic analysis is based on the following principle: the acoustic signal is generated in the glottis and passes through the vocal tract where it is modulated by the pharyngeal, buccal, labial and nasal cavities acting as filters [4]. The different frequency bandwidths emitted at the end of the vocal tract are called "formants". The human voice is composed of one fundamental frequency (F0) and four formants (F1 to F4) that correspond to each of the four cavities. Formant frequencies vary over time in function of the spatial conformation changes of the cavities driven by the phonatory muscles. Other characteristics, such as intensity variations and amount of noise contained in the acoustic signal, also are taken into account in the phonetic analysis. Software tools for fast and automated phonetic analysis have been recently developed [5].

The aim of this study was to build and evaluate machine learning models that can recognize OHCA based on the phonetic characteristics of the caller's voice.

2. Methods

The study protocol was approved by the Medical Ethics Committee of Rennes academic hospital (approval number 19.116, issued on December 4, 2019).

2.1. Software

Acoustic features of the recorded calls were extracted with WC-MDX Workstation version 11.6.0.0, UHERS Corporation, 2005. Phonetic analyses were performed with PRAAT v6.1.03, 2019, Institute of Phonetic Sciences, Amsterdam University. All statistical analyses and model building were performed with "R-studio", version 1.3.1093, RStudio PBC, 2009-2021. The following R packages were used: "Dplyr", version 1.0.0, for data manipulation; "MICE", version 3.9.0, for missing data implementation; and "Caret", version 6.0-90, for model building.

2.2. Setting and study population

Data were collected retrospectively from patients for whom a call was done to the emergency call center of Rennes academic hospital, France, between 01/01/2017 and 01/01/2019.

2.3. Variables and groups

The predicted variable was OHCA presence. Patients included in the study were divided in two groups: i) OHCA group if they had OHCA, and ii) NO-OHCA group, if they did not have a diagnosis of OHCA.

Predicting variables were collected by automated phonetic analysis. They were based on the following parameters: fundamental frequency, formants, intensity, jitter, shimmer,

harmonic to noise ratio, number of voice breaks, and number of periods. A full description of all these variables is available in the PRAAT software documentation [5]. Briefly, the spectrogram of the human voice includes five high intensity bands. One corresponds to the fundamental frequency and the other four to formants (figure 1). The *fundamental frequency* (F0) is determined by the tension of the vocal cords and the subglottic pressure, and varies with the stress level. *Formant frequencies* (F1 to F4) are determined by the resonance system volume. *Jitter* measures the short-term variations in the fundamental frequency in seconds, and *shimmer* reflects short-term disturbances in signal intensity. The *noise to harmonic ratio* is defined by the ratio of the non-harmonic and harmonic intensities contained in the acoustic signal. The *number of voice breaks* is the number of distances between consecutive pulses that are longer than a defined duration divided by the pitch floor. The *number of periods* is calculated by counting the number of time intervals between glottal pulses.

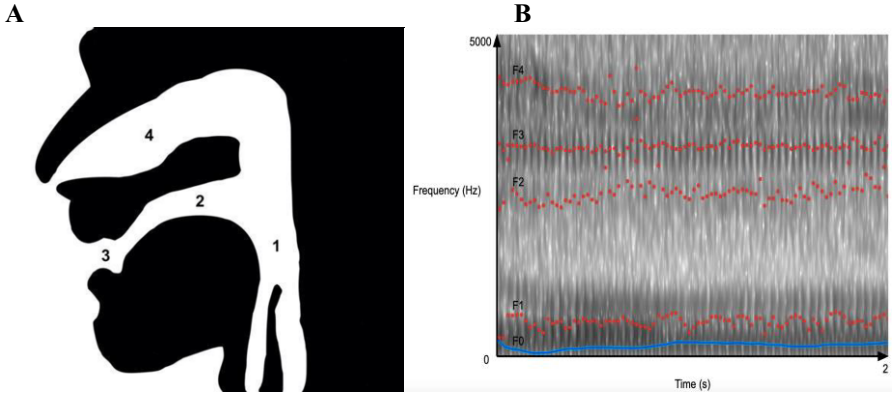

Figure 1: **Bases of the automated phonetic analysis. A: Voice signal.** The acoustic signal is generated in the glottis and modulated by four resonance systems: the pharyngeal (1), buccal (2), labial (3), and nasal (4) cavities. Each cavity acts as a filter, and the bandwidths emitted at the end of the vocal tract are called "formants". **B: Voice spectrogram.** The human voice is composed of one fundamental frequency (F0, blue line) and four formants (F1 to F4, red dots). Formants vary over time according to the spatial conformation changes of the four cavities driven by the phonatory muscles. A spectrogram is the graphic representation of these five bandwidths in function of time.

2.4. Statistics

The Student's *t*-test was used to compare means between groups. A p-value <0.05 was considered significant. To avoid multicollinearity, correlation coefficients were calculated for each pair of variables, and when the coefficient was higher than 0.8, one variable was excluded. *Data splitting*: Data were randomly divided in two parts: training dataset and test dataset. The training dataset corresponded to 80% of the complete dataset and was used to build the models. *Model training*: Three models were constructed: binary logistic regression, random forest (500 trees), and neural network (3 layers). *Performance measurement*: The predictions made by the three models were compared to the "OHCA" variable in the test dataset and receiver operating characteristic curves (ROC) were constructed accordingly. The area under the curve (AUC) was the primary outcome used to evaluate each model performance.

3. Results

3.1. Selected patients

820 patients were included in the study, 410 in each group.

3.2. Predicting variables

Table 1 shows the comparison of the mean values of each selected predicting variable between groups.

Table 1. Comparison of the selected predicting variables between groups. Values in brackets represent the 95% confidence intervals. *t-test significance was set at p <0.05 level. OHCA= Out of Hospital Cardiac Arrest, med=median, min=minimum, max=maximum, sd=standard deviation, n=number, NHR= noise to harmonic ratio.

Variable	OHCA group (n=410)	NO-OHCA group (n=410)	p (t-test)
Pitch mean (Hz)	244 (238 - 250)	197 (192 - 202)	< 0.001*
Pitch sd (Hz)	45.3 (43.5 - 47)	47 (44.9 - 49.1)	0.216
Pitch min (Hz)	136 (131 - 140)	109 (106 - 113)	< 0.001*
Pitch max (Hz)	417 (409 - 425)	395 (386 - 404)	< 0.001*
N of voice breaks	12.2 (11.9 - 12.5)	12.5 (12.2 - 12.9)	0.181
Jitter local absolute	$8.5e^{-5}$ ($8.2e^{-5}$ - $8.9e^{-5}$)	$10.7e^{-5}$ ($10.2e^{-5}$ - $11.2e^{-5}$)	< 0.001*
Jitter RAP (%)	0.82 (0.80 - 0.84)	0.84 (0.81 - 0.87)	0.399
Shimmer local (dB)	1.01 (1.00 - 1.03)	1.03 (1.01 - 1.04)	0.335
Shimmer APQ11 (%)	10.0 (9.7 - 10.2)	9.9 (9.6 - 10.3)	0.930
Mean NHR	0.153 (0.147 - 0.158)	0.154 (0.147 - 0.161)	0.741
Formant med H1 (Hz)	518 (512 - 525)	493 (487 - 498)	< 0.001*
Formant med H2 (Hz)	1482 (1469 - 1495)	1454 (1441 - 1468)	< 0.001*
Formant med H3 (Hz)	2288 (2279 - 2298)	2281 (2271 - 2290)	0.247
Formant med H4 (Hz)	3063 (3057 - 3069)	3074 (3068 - 3081)	0.014*
Intensity med (dB)	68.3 (67.6 – 69.0)	65.2 (64.5 - 65.9)	< 0.001*
Intensity sd (dB)	14.0 (12.9 - 15.2)	11.1 (10.8 - 11.4)	< 0.001*
Intensity min (dB)	27.6 (26.6 - 28.6)	27.7 (27.0 - 28.4)	0.887
Intensity max (dB)	79.6 (79.1 - 80.1)	78.2 (77.7 - 78.8)	< 0.001*

3.3. Model performance analysis

Table 2 shows the performance of the three models.

Table 2. ROC-AUC value for OHCA prediction by each model. Models used the acoustic features of the caller's voice to predict OHCA. Values in brackets represent the 95% confidence intervals.

Model	ROC-AUC
Binary logistic regression	71.4 (63.5-79.4)
Random forest	74.9 (67.4-82.4)
Neural network	64.5 (57.3-71.8)

4. Discussion

This study describes machine learning models that use the acoustic features of the bystander's voice to predict OHCA. These acoustic features were previously identified as stress markers. In 2010, Frampton and al. observed significant variations in pitch and

fundamental frequency in employees facing time restrictions to complete an order [6]. Similarly, Mendoza et al. [7] reported an increase in fundamental frequency, jitter and shimmer perturbations in individuals subject to work-related stressful conditions. To our knowledge, our study is the first to show similar results in the field of emergency healthcare. We observed statistically significant differences in the bystanders' voice acoustic parameters in the presence of OHCA (table 1). We also demonstrated that it is possible to create decision-making support models to recognize OHCA based on these parameters (table 2). These results provides perspectives for short-term applications of machine learning models that integrate semantic and acoustic parameters. Indeed, in 2019, Blomberg et al. developed a model based only on semantic elements that could reduce OHCA detection time [8]. The integration of the acoustic parameters described in the present study should increase the performance of such models.

5. Conclusion

This study demonstrates that machine learning models can recognize OCA based on the acoustic characteristics of the caller's voice. The integration of the acoustic parameters identified in this study could help to increase the performance of decision-making support systems that already integrate semantic parameters in order to improve OHCA detection over the phone.

References

[1] Berdowski J, Berg RA, Tijssen JGP, Koster RW. Global incidences of out-of-hospital cardiac arrest and survival rates: Systematic review of 67 prospective studies. Resuscitation 81, 1479–1487 (2010).
[2] Rossetti AO, Oddo M, Logroscino G, Kaplan PW. Prognostication after cardiac arrest and hypothermia: A prospective study. Ann. Neurol. 67, 301–307 (2010).
[3] Viereck S, Møller TP, Rothman JP, Folke F, Lippert FK. Recognition of out-of-hospital cardiac arrest during emergency calls — a systematic review of observational studies. Scand. J. Trauma Resusc. Emerg. Med. 25, 9 (2017).
[4] Titze IR. Some Consensus has been Reached on the Labeling of Harmonics, Formants, and Resonances. J. Voice 30, 129 (2016).
[5] Praat: doing Phonetics by Computer. https://www.fon.hum.uva.nl/praat/.
[6] Frampton M, Sripada S, Bion RAH, Peters S. Detection of time-pressure induced stress in speech via acoustic indicators.
[7] Mendoza E. Carballo G. Acoustic analysis of induced vocal stress by means of cognitive workload tasks. J. Voice Off. J. Voice Found. 12, 263–273 (1998).
[8] Blomberg SN, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. Resuscitation 138, 322–329 (2019).