Development of a Model to Identify Empathy in the Vocals of Mental Health Helpline Counsellors.

Ruvini Sanjeewa, Ravi Iyer, Pragalathan Apputhurai, Nilmini Wickramasinghe, Denny Meyer

Submitted to: JMIR Formative Research
on: October 22, 2024

Table of Contents

Original Manuscript **5**
Supplementary Files **26**
Figures **27**
Figure 1 ..... **28**
Multimedia Appendixes **29**
Multimedia Appendix 1 ..... **30**
Multimedia Appendix 2 ..... **30**
Multimedia Appendix 3 ..... **30**
Multimedia Appendix 4 ..... **30**
Multimedia Appendix 5 ..... **30**
Multimedia Appendix 6 ..... **30**

Development of a Model to Identify Empathy in the Vocals of Mental Health Helpline Counsellors.

Ruvini Sanjeewa[1] BSc; Ravi Iyer[1] PhD; Pragalathan Apputhurai[1] PhD; Nilmini Wickramasinghe[2] PhD; Denny Meyer[1] PhD

[1]School of Health Sciences Swinburne University of Technology Hawthorn Melbourne AU
[2]School of Computing, Engineering & Mathematical Sciences La Trobe University Melbourne AU

Corresponding Author:
Ruvini Sanjeewa BSc
School of Health Sciences
Swinburne University of Technology
Hawthorn
PO Box 218
John Street
Melbourne
AU

Abstract

Background: The research study aimed to detect the vocal features immersed in empathic counsellor speech using samples of calls to a mental health (MH) helpline service.

Objective: The study aimed to produce an algorithm for the identification of empathy from these features, which could act as a training guide for counsellors and conversational agents who need to transmit empathy in their vocals.

Methods: Two annotators with a psychology background and English heritage provided empathy ratings for 57 calls involving female counsellors, as well as multiple short call segments within each of these calls. These ratings were found to be well correlated between the two raters in a sample of six common calls. Using vocal feature extraction from call segments and statistical variable selection methods, such as L1 penalised Least Absolute Shrinkage and Selection Operator (LASSO) and forward selection, a total of 14 significant vocal features were associated with empathic speech.

Results: Generalised additive mixed models (GAMM), binary logistics regression with splines and random forest models were employed to obtain an algorithm that differentiated between high and low empathy call segments. Slightly higher predictive accuracies of empathy were reported from the binary logistics regression model (AUC=0.617) than the GAMM (AUC=0.605) and the random forest model (AUC= 0.600).

Conclusions: This study suggests that the identification of empathy from vocal features alone is challenging and further research involving multi-modal models (e.g. models incorporating facial expression, words used and vocal features) are encouraged for detecting empathy in the future.
This study has several limitations including a relatively small sample of calls and only two empathy raters. Future research should focus on accommodating multiple raters with varied backgrounds, to explore these effects on perceptions of empathy. In addition, considering counsellor vocals from larger more heterogeneous populations, including mixed-gender samples, will allow an exploration of the factors influencing the level of empathy projected in counsellor voices more generally.

(JMIR Preprints 22/10/2024:67835)
DOI: https://doi.org/10.2196/preprints.67835

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Development of a Model to Identify Empathy in the Vocals of Mental Health Helpline Counsellors.

Ruvini Sanjeewa[1], BSc; Ravi Iyer[1], PhD; Pragalathan Apputhurai[1], PhD; Nilmini Wickramasinghe[2], PhD; Denny Meyer[1], PhD
[1] School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia.
[2] School of Computing, Engineering & Mathematical Sciences, La Trobe University, Australia.
Corresponding Author:
Ruvini Sanjeewa,
Swinburne University of Technology,
Hawthorn, 3122
Australia
Phone 0422587030
Email: rsanjeewa@swin.edu.au

Abstract (450 words)

The research study aimed to detect the vocal features immersed in empathic counsellor speech using samples of calls to a mental health (MH) helpline service. The study aimed to produce an algorithm for the identification of empathy from these features, which could act as a training guide for counsellors and conversational agents who need to transmit empathy in their vocals.

Two annotators with a psychology background and English heritage provided empathy ratings for 57 calls involving female counsellors, as well as multiple short call segments within each of these calls. These ratings were found to be well correlated between the two raters in a sample of six common calls. Using vocal feature extraction from call segments and statistical variable selection methods, such as L1 penalised Least Absolute Shrinkage and Selection Operator (LASSO) and forward selection, a total of 14 significant vocal features were associated with empathic speech. Generalised additive mixed models (GAMM), binary logistics regression with splines and random forest models were employed to obtain an algorithm that differentiated between high and low empathy call segments. Slightly higher predictive accuracies of empathy were reported from the binary logistics regression model (AUC=0.617) than the GAMM (AUC=0.605) and the random forest model (AUC= 0.600).

This study suggests that the identification of empathy from vocal features alone is challenging and further research involving multi-modal models (e.g. models incorporating facial expression, words used and vocal features) are encouraged for detecting empathy in the future.

This study has several limitations including a relatively small sample of calls and only two empathy raters. Future research should focus on accommodating multiple raters with varied backgrounds, to explore these effects on perceptions of empathy. In addition, considering counsellor vocals from larger more heterogeneous populations, including mixed-gender samples, will allow an exploration of the factors influencing the level of empathy projected in counsellor voices more generally.

Keywords: Vocal features, voice characteristics, empathy, mental healthcare, crisis helpline services

Introduction

Background

Empathy is defined as experiencing the emotions (emotional empathy) and cognitions (cognitive empathy) of others and responding to them appropriately[1]. Empathy is especially important for patient care, where it is essential that the lived experience of the patient is understood by responding healthcare professionals, while also conveying this understanding in conjunction with a desire to help the patient[2]. The effectiveness of physician empathy has been shown to improve patient satisfaction and commitment to recovery while reducing anxiety and distress levels, leading to better clinical results[3]. Furthermore, empathic behaviour by mental health (MH) care providers reduces their own risk of burnout[4].

Telephone helpline services offer an effective means of supporting those who need immediate MH care[5]. The demand for such services has increased dramatically since the outbreak of the COVID-19 pandemic[6], increasing the expectations of counselling staff to provide support for people with complex MH concerns[7]. As a basic counselling skill, empathy is key to a successful engagement with patients in the context of complex psychosocial needs.

Besides emotional and cognitive empathy in understanding the status of a patient, contextual awareness is equally important for therapeutic engagement[8]. This means that empathic responses need to be contextually appropriate by considering environmental cues, culture, demographic factors and the specific circumstances of the patient, to understand the broader context of their MH status[9, 10]. This allows counsellors to tailor responses, based on context, to engage in effective communication with distressed patients, thereby delivering better outcomes[3].

Verbal cues and tone of voice are crucial when communicating empathy[11]. For example, reduced speech rate and lower pitch are perceived as more empathic by patients when receiving bad news from healthcare providers in an oncology setting[12] and while actively listening to telephone callers, nurses have been found to express empathy through their choice of words, voice and intonation, projection of compassion and warmth, as well as “tuning in” to the caller’s story and identifying with the caller’s emotions[13].

Unfortunately, the global demand for MH support is not being met by the existing workforce[14]. This service gap is leading to growing interest in alternative interactive technological solutions. Technological innovations, such as the design of conversational agents, has demonstrated potential in facilitating effective and immediate patient care. However, to optimise upon end user acceptance, conversational agents need to display empathy[15, 16] . However, we are yet to identify the precise

vocal features most associated with an empathic human response, and it is not known if it is possible to categorise empathy levels using such vocal features. Thus, the aim of this paper is to:

i. identify the vocal features significantly associated with empathy in a large collection of telephone helpline counselling call recordings and to;
ii. evaluate the accuracy of a machine learning algorithm to correctly designate short segments of each recording to categories of low and high empathy.

Methods

Data collection

Recordings of telephone helpline calls (*n*=57) were obtained from On The Line, Australia, a suicide helpline counselling service. The level of suicide risk of each caller had been previously assessed by counsellors using the Columbia Suicide Severity Rating Scale (C-SSRS) to differentiate between calls featuring high suicide risk (with C-SSRS ratings of 6-7) and calls with low risk of suicide (with C-C-SSRS ratings of 1-2) (Please refer to Iyer et al., 2022 for further details[17, 18]).

Annotation of Call Segments and Overall Call Empathy for Counsellors

Annotations of counsellor empathy were conducted by two independent researchers (IG, SD) using RStudio (Version 2024.04.2, Build 764)[19]. Both IG and SD had extensive experience in counselling, obtained during post-graduate psychology training programs. Segments of the counsellor voices were selected from each call using Audacity (Version 3.5.1, CMake Release Build) [20], ensuring that overlap between the caller and responding counsellor voices were minimised. Empathy displayed by the counsellor within each call segment, was rated using the Carkhuff and Truax Empathy (CTE) scale[21]. A weekly project team meeting, attended by a clinical psychologist (MN) was used to reconcile any disparities in ratings. A Qualtrics online questionnaire was used to also collect data on the overall level of empathy displayed by the responding counsellor during each call and to evaluate caller distress at the commencement and conclusion of each call. Figure 1 shows the flow of the voice analysis process during the study.

Measures of Counsellor Empathy and Caller Distress

The CTE scale, was used to rate the audio segments selected from each call on a 5-point Likert-style scale (1=’low empathy’, 5=’high empathy’)[21]. In addition, three measures were used to assess overall counsellor empathy for each call. These three scales were modified to suit counsellor-caller conversations through an iterative process, in which members of the research team provided independent feedback to achieve the final questionnaire. Examples were developed by the annotators

for each item included in these scales, ensuring clarity and consistency of ratings. The details for these scales are provided in the supplementary materials 1. The measurement scales included:

1. The Perceived Emotional Intelligence (PEI) Scale[22] identified variations in PEI in the counsellor's vocals. The PEI is a 20-item scale with each item scored with 1=’Never or almost never true’ and 7=’Almost or almost always true’.

2. The Active-Empathic Listening (AEL) Scale[23] was modified appropriately to produce 10 items measuring empathic listening using a 7-point Likert-style scale with 1=’Never or almost never true’ and 7=’Almost or almost always true’.
3. RS7: A single item 7-point Likert Scale[24] was used to rate overall empathy with 1=’Low empathy’ and 7=’High empathy’.

Finally, at the start and end of each call the annotators assessed the level of caller distress using the Distress thermometer[25], a visual analogue 11-point scale (0=‘no distress’ to 10=‘extreme distress’).

Data validation through inter-rater consistency check

Six calls (10%) were chosen at random to measure inter-rater reliability. The empathy ratings of the more experienced rater (IG) were used as the reference, against which the SD ratings were compared. Spearman's correlation[26] was calculated for each of the three scales used to rate overall counsellor empathy for each call. The Mann-Whitney U test was then used to check for the significance of differences between the ratings provided by the annotators.

Relationships between perceptions of empathy and call context

The associations between perceived counsellor empathy and call context were explored using a combination of empathy ratings, caller distress at the beginning and end of the call, and caller suicide risk. Caller distress and suicide risk were correlated with perceived counsellor empathy to evaluate the relationship between level of empathy and caller disposition.

Preprocessing stage

Audio file format conversion and vocal features extraction

The input call recordings were obtained as 8kHz sample rate, 8-bit depth .wav files. The encoding type of the files were transformed to PCM float format with 32-bit depth to ensure compatibility with RStudio for analysis. Vocal features (n=55) were extracted per 30ms speech frames (50% overlap; Blackman windows) within each annotated segment using Rstudio(Version 2.7.0,Soundgen package[27]).

Remove moderate ratings and binary coding empathy level

The vocal segments that scored a rating of three out of five on the CTE scale were removed (n=142, 18%) from further analysis because of their neutral empathic character. A binary response variable was then created for each of the *n*=643 remaining segments (190,345 30ms speech frames) with an empathy rating of 4-5 coded as high empathy (n=146 segments) and a rating of 1-2 coded as low empathy (n=497 segments).

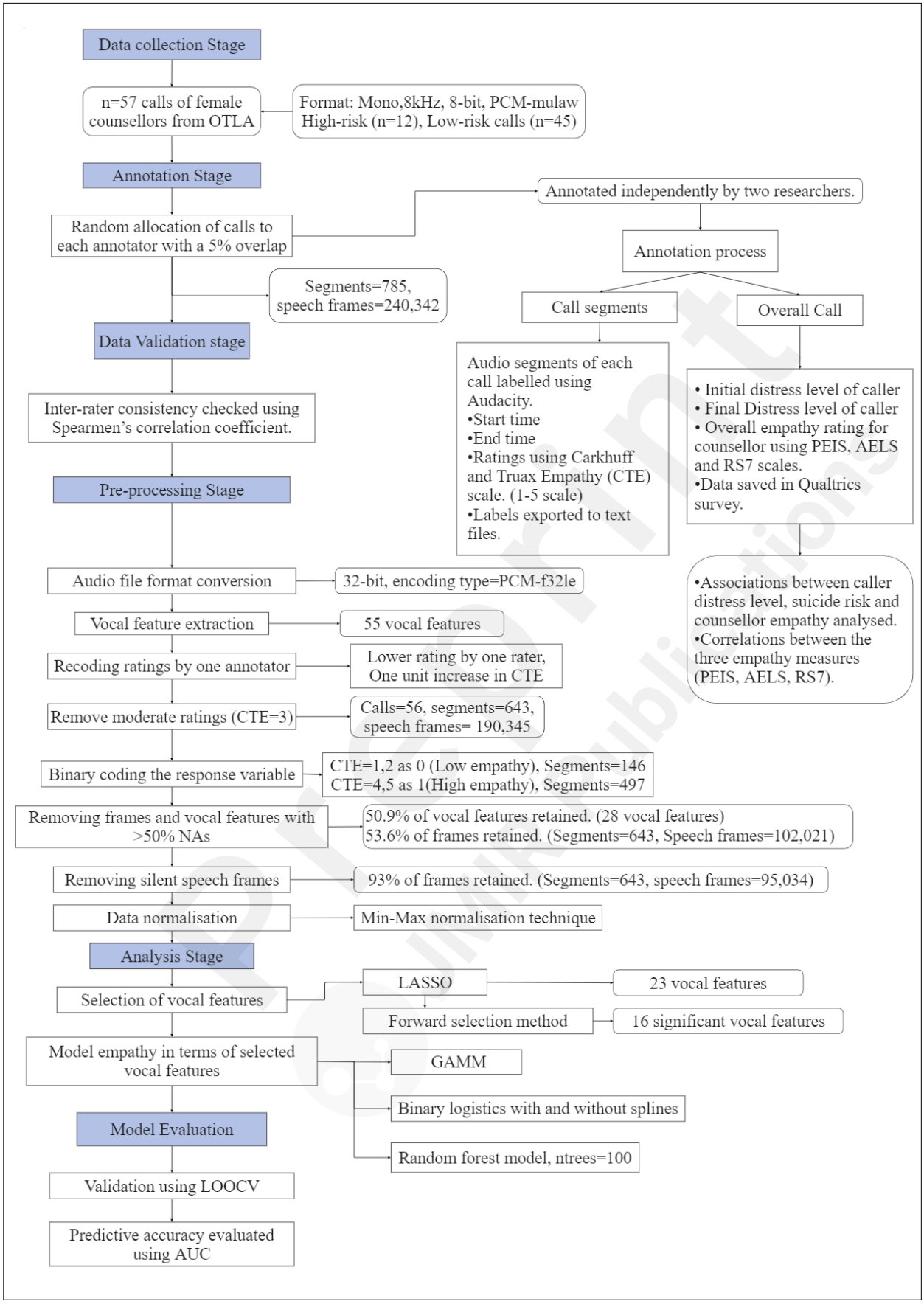


Figure 1: Voice analysis flow chart

Removing missing values in speech frames and vocal features

Vocal features and 30ms speech frames with more than 50% missing values were removed to maintain the quality of data and improve the overall accuracy of the results. The resulting data retained 50.9% (n=28) of the original vocal features and 53.6% (n=102,021) of the original speech frames.

Removing silent speech frames and normalisation

The silent speech frames were also removed from each of the 643 segments leaving 95,034 speech frames in the final analysis sample. Finally, the min/max normalising technique[28]was applied to reduce the influence of background noise.

Analysis Stage

Selection of vocal features

Variable selection was performed to identify vocal features that were strongly associated with empathy. L1 penalised Least Absolute Shrinkage and Selection Operator (LASSO) regression was used to select the most relevant variables by shrinking the coefficients of the least relevant variables to zero[29]. Tenfold cross-validation was used to optimise the tuning parameter, lambda. Further refinement in terms of the selected vocal features was then conducted using a forward stepwise regression model.

Models for identifying empathy level with selected vocal features

Three methods were used to classify low and high empathy segments based on this final set of selected vocal features. A Generalized Additive Mixed Model (GAMM) included vocal features as fixed effects, with each call treated as a random effect. Spline functions for the selected vocal features were used to account for non-linearity[30]. The GAM function[31] of package mgcv[32] in Rstudio was used for the analysis.

Random forest classification also allowed for non-linear relationships using step functions, while more efficiently processing large datasets[33]. This has been a prominent classification model used in studies involving vocal analysis[34-36]. The binary logistic regression model was the third model considered, again accounting for non-linearity using splines and was used in the present study as the baseline model[37].

Model Evaluation

Probabilistic predictions for high versus low empathy levels were obtained for each segment using Leave-One-Call-Out-Cross-validation (LOCOCV). Based on these probabilities, receiver operating characteristic curves were created and areas under the curves (AUCs)[38, 39]were used to compare

the reliability of these models. The Youden index[40] was employed to decide the optimal cut-point for classifying segments based on their estimated high versus low empathy probabilities.

Results

The following results were obtained from the 56 calls for female counsellors. This sample of calls included 12 calls at high risk of suicide (12/56, 21%) and 44 calls at low risk of suicide (44/56,79%).

Annotation of Call Segments.

Using the CTE scale, 146 segments (18.6%) showed low empathy, 142 (18.1%) showed medium empathy and 497 segments (63.3%) demonstrated high empathy.

Overall Call Empathy Ratings

Supplementary Materials 2 provides descriptive statistics for the overall empathy ratings for the 56 calls using the three scales. Excellent reliability is observed from both the raters, IG and SD for the PEI and AEL scales, with Cronbach alpha values above 0.9. The descriptive statistics and the Spearman correlation statistics between the annotators are shown in supplementary materials 3. A strong agreement between raters was observed with the PEI measure and the RS7 empathy measure approaching statistical significance.

Relationships between perceptions of empathy and call context

The relationship of counsellor empathy ratings with caller's distress at the start and end of the call and suicide risk are shown in Table 1. While the correlations between the initial distress and the three empathy measures were not significant, a moderate, statistically significant negative correlation between the final distress of the caller and the empathy of the counsellor was observed for both the PEI and RS7. The suicide risk of callers as measured using the C-SSRS had a statistically significant but weak positive correlation with counsellor empathy across the PEI and AEL measures. Strong statistically significant correlations among the three empathy measures were found, validating the empathy measurement process.

Table 1: Spearman’s Correlation Coefficients for Caller Distress and Suicide Risk with Counsellor Empathy

	Caller Context			Ratings for Counsellor Empathy		
Empathy Rating	Initial distress	Final distress	Suicide risk	PEIS	AELS	RS7
PEIS[a]	-0.011	-0.414**	0.310*	1		
AELS[b]	0.102	-0.233	0.308*	.822***	1	

RS7[c]	-0.024	-0.443***	0.055	.847***	.851***	1

[a]PEIS = The Perceived Emotional Intelligence Scale

[b]AELS = The Active-Empathic Listening Scale

[c] RS7 = Rating Scale 7 (7-item)

Note: *** p<.001 (two-tailed) ** p<0.01 (two-tailed), * p<0.05 (two-tailed).

Analysis Stage

The selection of vocal features using LASSO to predict high versus low empathy employed a cross-validation of the training dataset to reveal an optimum Log (Lambda) parameter = -8.542. The relationship between this parameter and the Binomial deviance is shown in supplementary materials 4. Using this lambda value, 23 vocal features were retained. These 23 vocal features were then passed on to the forward selection binary logistic regression model, which identified 16 significant vocal features as shown in Table 2.

Table 2: Results of the forward selection binary logistic regression model for vocal feature selection for predicting high versus low empathy in counsellor voices

Vocal features	β	z-value	p-value
Depth of amplitude	-1.468	-15.994	< 0.000
Frequency of amplitude (Hz)	0.068	2.524	0.012
Frequency of amplitude (Hz) via Modulation Spectrum (MS).	-0.462	-12.350	< 0.000
Purity of amplitude via MS.	-0.834	-9.596	< 0.000
Amplitude (dB)	-3.391	-51.688	< 0.000
Dominant frequency (Hz)	0.985	3.136	0.002
Entropy	3.470	20.843	< 0.000
Shannon entropy	-6.016	-25.270	< 0.000
Epoch	0.289	5.040	0.000
First formant frequency (Hz)	1.244	5.080	0.000
First formant width (Hz)	0.194	4.181	0.000
Second formant frequency (Hz)	-0.191	-2.808	0.005
Second formant width (Hz)	-0.063	-1.802	0.072
Third formant frequency (Hz)	-0.099	-1.521	0.128
Third formant width (Hz)	0.004	0.118	0.906
Spectral flux	0.110	1.435	0.151
Harmonics-to-noise ratio -HNR (dB)	-0.370	-2.869	0.004
Spectral novelty	0.035	0.715	0.475
Peak frequency (Hz)	-0.293	-1.827	0.06
25th percentile frequency (Hz)	-1.311	-4.720	0.000

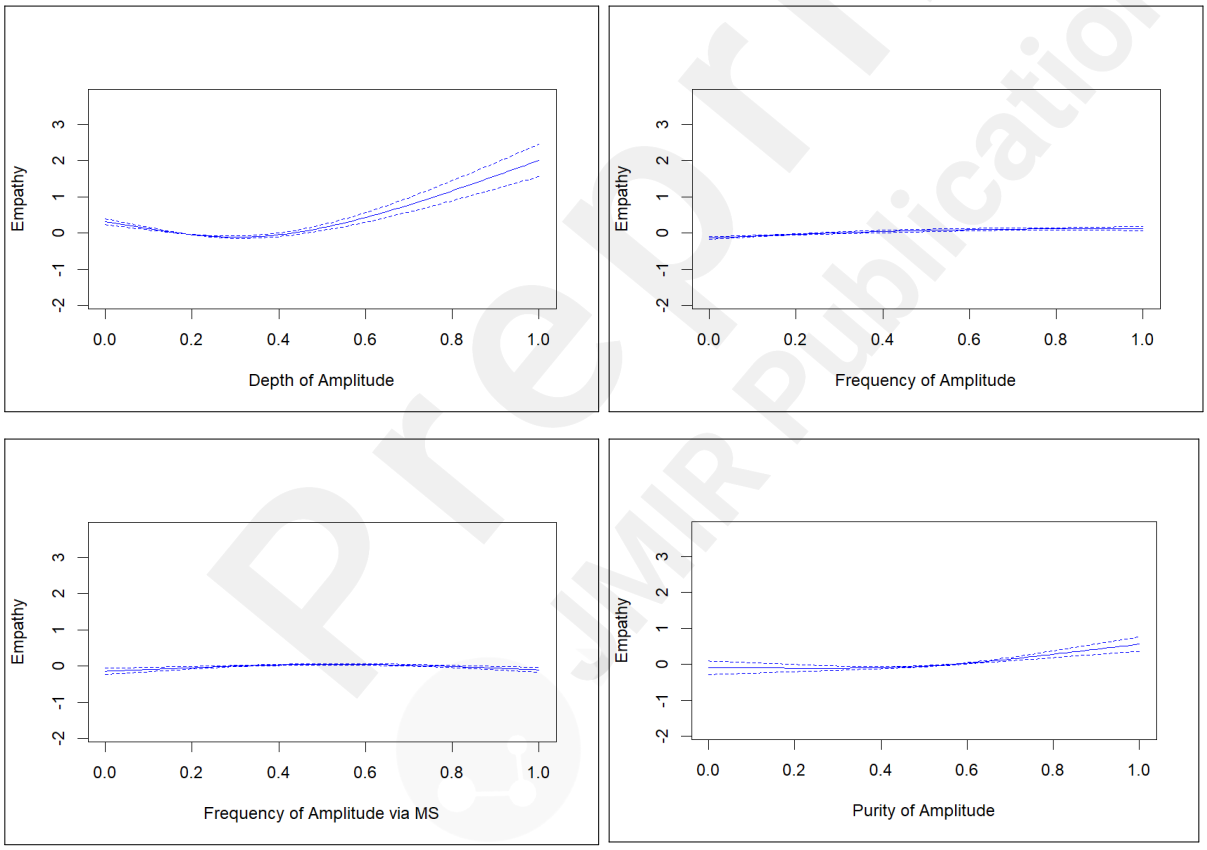
50th percentile frequency (Hz)	-0.262	-1.586	0.113
Spectral centroid (Hz)	5.782	15.717	< 0.000
Spectral slope (Hz)	-3.944	-17.881	< 0.000

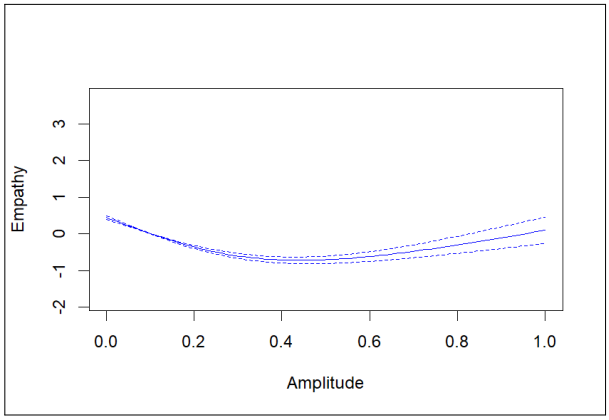
Hz = Hertz; dB = Decibels

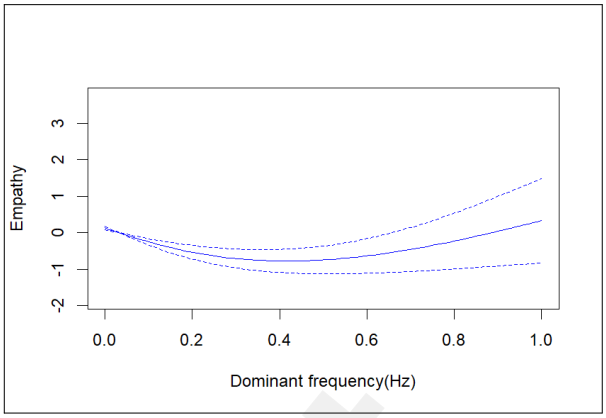
Feature extraction and Classification

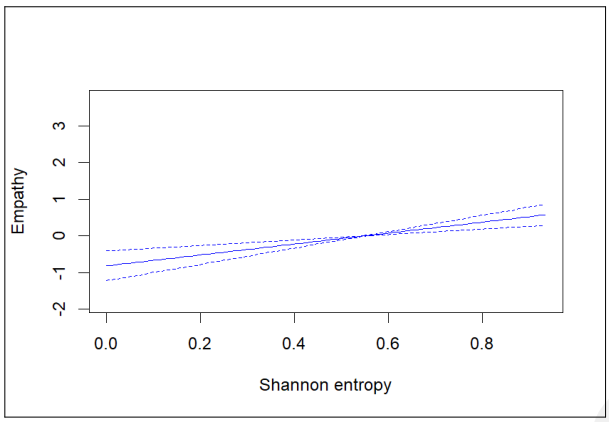
The results of the GAMM model are shown in Supplementary Material 5. Out of the 16 selected vocal features, the GAMM used only 14. Based on the effective degrees of freedom values, two of the vocal features (Shannon entropy and the 25th percentile frequency) show a linear relationship in the GAMM model. Figure 2 illustrates the non-linear nature of all other relationships. An AUC value of 0.605 was obtained for the GAMM model[41].

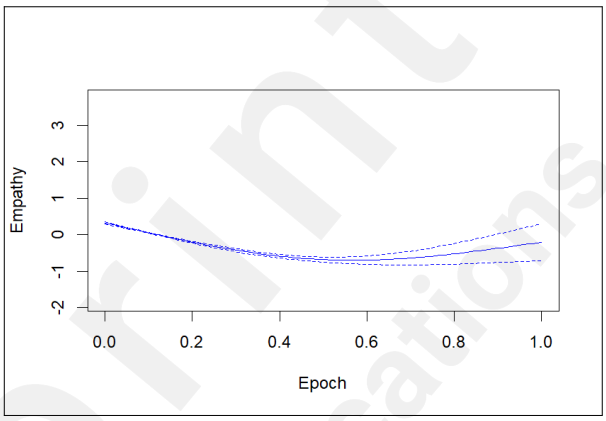
Figure 2: Association of empathy to each vocal feature

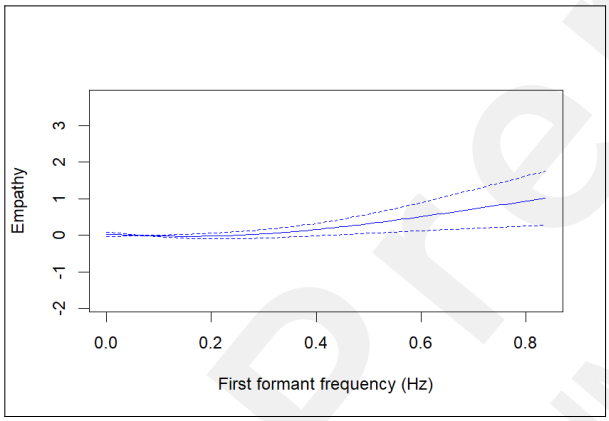


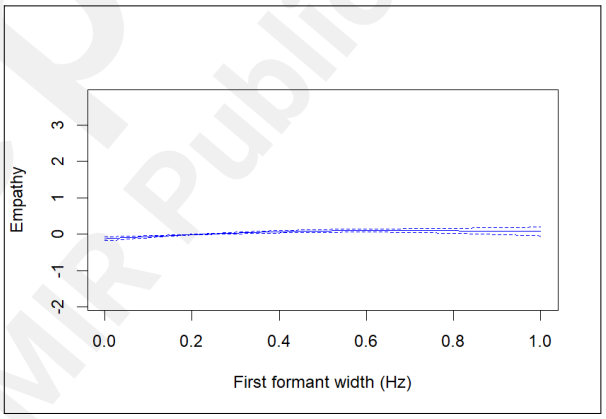


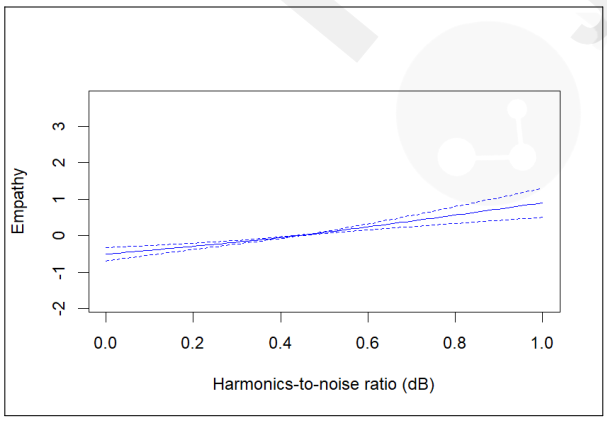


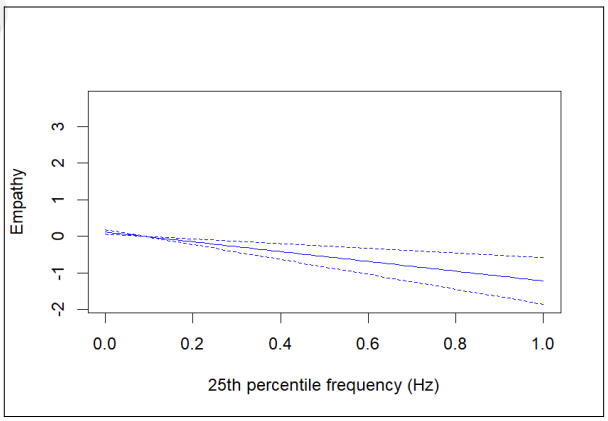


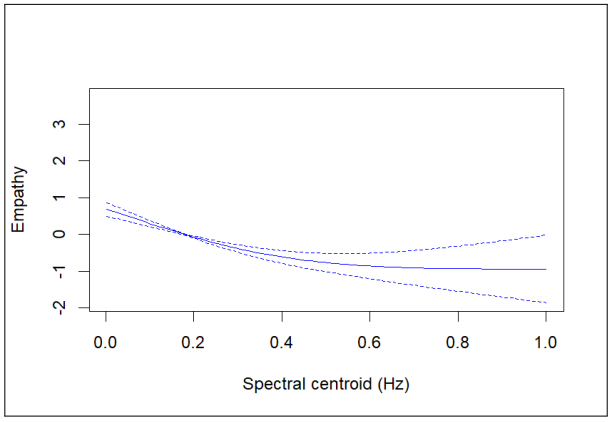


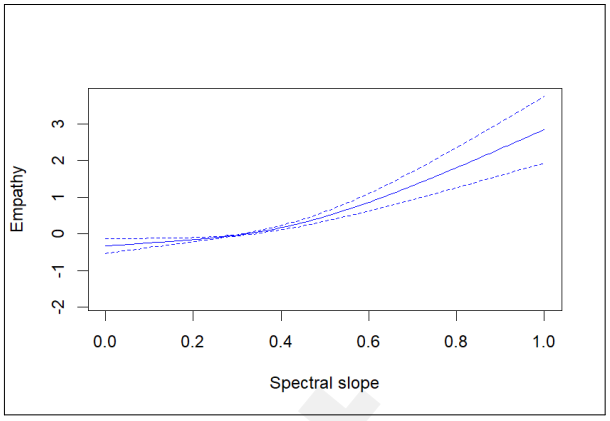












As illustrated in Figure 2 higher empathy was associated with higher values for First formant frequency, Dominant frequency, Shannon entropy, Spectral slope, and Harmonic-to-noise ratio. In contrast, higher empathy was associated with lower values for 25th percentile frequency, and spectral centroid. Finally, lower empathy was associated with intermediate values for depth of amplitude, amplitude, dominant frequency, epoch and first formant frequency of speech.

The GAMM was able to differentiate between low and high 30ms segments of speech to a classification accuracy of 75%. This was superior to both Random Forest and Binary Logistic Regression models (74% and 69% respectively).

Epoch (39%), amplitude (22%) and depth of amplitude (7%) were the top three vocal features contributing to empathic speech in the GAMM model. The vocal features that contribute the most towards the identification of empathic speech varied across the three methods used, as shown in Multimedia Appendix 5.

Discussion

The current study was undertaken to identify a range of vocal features that can predict the level of empathy exhibited in the recordings of female counsellors and to accurately classify short segments of each recording according to low or high empathy ratings. We were successful in identifying 14 unique vocal features that significantly distinguished between low and high empathy ratings in the GAMM model. Furthermore, we were able to successfully classify short segments of speech using these vocal features to an accuracy level of 75% using this model.

Empathy is an important component of human interactions and social connections that promotes general wellbeing[42]. It is an essential component of MH care support, required to enhance therapeutic alliance and rapport building. Empathy embodies the ability to understand and compassionately reflect the range of feelings and experiences communicated by others. Empathic communication relies upon verbal (words and vocal features) and non-verbal means such as body language and facial expression[43]. However, it is only verbal expressions of empathy through vocal

feature that are the subject of this study.

The detection of empathy is traditionally based upon subjective human perception captured using standardised empathy scales or questionnaires[44]. Our study is unique in identifying a range of vocal features that identify the level of empathy expressed during a counselling interaction. This has involved developing an algorithm that detects the human vocal features that are associated with empathic speech. This algorithm has the potential to enhance the training of counsellors in the use of empathic speech and offers valuable insights into effective human communication in the MH care domain.

Call context and counsellor empathy in a crisis helpline setting

The analysis revealed that there is a strong negative correlation between the final distress level of the caller and the counsellor's level of empathy, suggesting that empathic communication with a caller can lower their level of distress by the end of the call. Higher levels of empathy allow the counsellors to build rapport and trust with their patients, allowing for effective emotional support during a crisis. These findings align with the existing literature about the benefits of empathic interactions. The suicide risk of the caller was positively related to counsellor empathy. This is an indication that counsellors effectively recognise crisis situations, exhibiting higher levels of empathy when speaking to callers with high risk of suicide. It also confirms that the level of empathy displayed by counsellors is adapted to the situation of the caller. These relationships of counsellor empathy with caller distress and suicide risk confirm the importance of counsellor empathy in the context of crisis helpline services.

Characteristics of empathic vocals

This study has identified several vocal features associated with empathy. Depth of amplitude in speech reflects varying levels of loudness, emphasising the expressiveness and dynamic nature of the human vocals apparent in empathic speech. A stable, more consistent emotional delivery during speech (purity of amplitude), also helps to convey empathy. Quieter vocals (Amplitude and lower tonal frequency, dominant frequency) are also associated with higher empathy.

Higher first formant frequencies are associated with “a” vowel sound, which correspond with high ratings of empathy. Additionally, a higher Harmonic-to-Noise Ratio (HNR) indicating greater clarity and more pleasant-sounding vocals is also associated with greater empathy. The spectral slope has a strong positive relationship to empathy, while the spectral centroid shows the opposite relationship. This indicates that a lower spectral centroid, with more low-frequency components makes a speaker sound more empathic.

Based on these findings, it is evident that empathy in vocals is provided by a combination of multiple human vocal features, and variations in each of the features exert a different impact on empathy. In particular, the way that a specific threshold of loudness in the vocals decides the delivery of perceived empathy in the context of effective counselling provides compelling evidence. This further suggests that the right balance of each of these vocal features is needed, where stability, energy and clarity play a pivotal role.

However, the study of empathy in vocals is a complex topic and has challenges. Especially, the subjective nature of empathy perception is an area that requires further study. This study relied on empathy ratings provided by two psychology trained female raters of English heritage. Different results may have been obtained if raters with different cultural, social and educational backgrounds had been included[45].

Vocal Feature Extraction and Empathy Classification

Three methods were employed to study the association between empathy and relevant vocal features. A GAMM, a random forest and a logistic regression approach with splines were fitted and compared using AUC values, using the LOCOCV method for evaluating these classifiers. The accuracy of empathy level classifications achieved was similar (75%, 74% and 69% respectively) when Youden's index was used to choose the probability cut point, as were their AUC values (0.605, 0.600 and 0.617 respectively). The GAMM and binary logistic regression with splines required significantly more computational time compared to the Random Forest, which utilised 100 trees. Despite the significant improvement in AUC value with the logistic regression with splines in comparison to the random forest model, the AUC results are not that dissimilar. However, these results do highlight a trade-off between model performance and computational efficiency for future research.

These relatively low AUC values can perhaps be partly attributed to the difficulties encountered in providing accurate ratings of empathy. The algorithms developed for detecting empathy from vocal features were reliant on the quality of the input data provided for the empathy ratings. A larger sample of raters and a larger sample of calls might have produced more reliable data, and this is recommended for future research in this area.

However, the multi-modal approaches commonly used for empathy recognition through the use of words, vocals, visual signs and psychological signals reflect the multi-faceted nature of empathy[46]. The importance of facial expressions for recognising empathy is particularly emphasised in this literature, where factors such as observation time and the type of emotion expressed significantly

influence the accuracy of identification[47]. These multi-modal approaches suggest that a higher accuracy in detecting empathy can be achieved when all these factors are collectively considered, rather than vocal approaches on their own.

Limitations

The inherent differences of empathy perception among the annotators of the study were a concern in this study. An additional analysis was conducted to explore this further, incorporating a third annotator without psychological expertise and from a different cultural background (Supplementary Materials 6). The findings from this analysis produced even lower model performance confirming that perceptions of empathy do vary between individuals and cultures. This analysis included both female and non-female counsellor vocals, which may also have contributed to this poorer performance. While this finding underscores the complexity of recognising empathy, it also highlights how cultural differences, personal experience and psychological knowledge of individuals contribute to subjective perceptions. Therefore, future research would benefit from accommodating these differences in empathy perception within the model by including multiple annotators with varied backgrounds.

Another limitation of the study was due to the very small number of available male counsellor recordings (n=13), the main analysis could only be conducted on female vocals, limiting the generalisation of our findings. Ideally it would have been possible to develop separate models for empathy in male and female counsellors and to test whether there were significant differences between these models. Larger samples of counsellor vocals would also have been preferable, providing more diversity in the data. A balanced representation of cultural background, empathy levels and individual characteristics should ideally be considered for the counsellor recordings in future research.

A further limitation is the source of the calls used for this study. The context of a suicide helpline service is very specific, and it may be that more algorithmic success would have been possible in a less stressed environment. In addition, for most of the calls the level of empathy was assessed by a single rater. As mentioned above, it would have been preferable if greater duplication of ratings could have been used to provide the dependent variable for the models that have been used to identify the level of empathy in counsellor vocals.

Implications of the study

The importance of empathy in reducing the distress of callers confirms the need for the incorporation of empathic communication skills in training programs for counsellors. Additionally, a statistically significant

positive correlation was found between suicide risk and counsellor empathy. This suggests that counsellors tend to be more empathic towards high-risk callers. This perhaps highlights the need for counsellors to adhere to a more caller-centred approach, ensuring that empathy is consistently exhibited for both low-risk and high-risk callers. Resources should perhaps be allocated equally for all callers rather than having a crisis intervention strategy that is tailored to prioritise callers needing emergency support.

To the best of the authors' knowledge, this is the first study that has identified the unique features of human vocals that are associated with the communication of empathy in a MH care setting. The results of this study have implications for the training of counsellors and psychologists working for MH-related telephone helpline services. Additionally, these findings can serve as a training resource for MH professionals more broadly, enhancing the quality of care provided. The engineering of empathic chatbots, especially within a triage capacity, is another significant area of research that would benefit from the findings of this study[48].

However, collecting vocal data of individuals for research purposes raises important ethical concerns. It is essential for this research to prioritise user consent and caller privacy. It is recommended that people with lived experience in telephone counselling and MH be asked to assist with the co-design and co-production of such research, to ensure that any resulting training programs or monitoring systems are acceptable to users and meet consumer needs.

Acknowledgements

The research was funded by Swinburne University of Technology. The data call recordings were provided by On The Line Australia who was the main research partner of this study.

Author’s Contributions

RS developed the first draft of the manuscript and conducted a voice analysis including the creation of the algorithm. RI contributed to developing the algorithm for detecting voice features that relate to empathy. IG, SD and RS conducted the call annotations where variation in counsellor empathy within the calls were identified. MJ, DM and RI contributed their supervision during the entire annotation process confirming the wellbeing of the researchers involved. DM, RI, PA and NW were involved in the manuscript revisions.

Conflicts of interest

None

Abbreviations

AUC: areas under the curves

C-SSRS: Columbia Suicide Severity Rating Scale

CTE: The Carkhuff and Truax empathy

MH: Mental Health

LASSO: Least Absolute Shrinkage and Selection Operator

LOCOCV: Leave one caller out cross validation

GAMM: Generalized Additive Mixed Model

PEI: Perceived Emotional Intelligence

AEL: Active-Empathic Listening

Multimedia Appendix 1: Online Questionnaires used for data collection

Multimedia Appendix 2: Descriptive Statistics of the analysis

Multimedia Appendix 3: Results of consistency between the raters.

Multimedia Appendix 4: The results of LASSO regression

Multimedia Appendix 5: Detailed results of the significance of the vocal features through GAMM, Random Forest and Binary Logistics Regression models.

Multimedia Appendix 6: Impact of expanded dataset and rater diversity on voice analysis results.

References

1. Smith A. Cognitive empathy and emotional empathy in human behavior and evolution. The Psychological Record. 2006;56(1):3-21. doi:10.1007/BF03395534
2. Hojat M. Empathy in health professions education and patient care. 2016. ISBN: 9783319801896
3. Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general practice: a systematic review. British journal of general practice. 2013;63(606):e76-e84. doi:10.3399/bjgp13X660814
4. Sturzu L, Lala A, Bisch M, Guitter M, Dobre D, Schwan R. Empathy and burnout–a cross-sectional study among mental healthcare providers in France. Journal of medicine and life. 2019;12(1):21. doi:10.25122/jml-2018-0050
5. Peppou LE, Economou M, Skali T, Papageorgiou C. From economic crisis to the COVID-19 pandemic crisis: evidence from a mental health helpline in Greece. European archives of psychiatry and clinical neuroscience. 2021;271:407-9. doi:10.1007/s00406-020-01165-4
6. Betancourt JA, Rosenberg MA, Zevallos A, Brown JR, Mileski M, editors. The impact of COVID-19 on telemedicine utilization across multiple service lines in the United States. Healthcare; 2020: MDPI. doi: 10.3390/healthcare8040380
7. Pavlova A, Scarth B, Witt K, Hetrick S, Fortune S. COVID-19 related innovation in Aotearoa/New Zealand mental health helplines and telehealth providers–mapping solutions and discussing sustainability from the perspective of service providers. Frontiers in Psychiatry.

2022;13:973261. doi:10.3389/fpsyt.2022.973261
8. Gibbons SB. Understanding empathy as a complex construct: A review of the literature. Clinical Social Work Journal. 2011;39:243-52. doi: 10.1007/s10615-010-0305-2
9. Angus L, Kagan F. Empathic relational bonds and personal agency in psychotherapy: Implications for psychotherapy supervision, practice, and research. Psychotherapy: Theory, Research, Practice, Training. 2007;44(4):371. doi:10.1037/0033-3204.44.4.371
10. Pugh MA, Vetere A. Lost in translation: An interpretative phenomenological analysis of mental health professionals' experiences of empathy in clinical work with an interpreter. Psychology and Psychotherapy: Theory, Research and Practice. 2009;82(3):305-21. doi:10.1348/147608308X397059.
11. Tan L, Le MK, Yu CC, Liaw SY, Tierney T, Ho YY, et al. Defining clinical empathy: a grounded theory approach from the perspective of healthcare workers and patients in a multicultural setting. BMJ open. 2021;11(9):e045224. doi:10.1136/bmjopen-2020-045224
12. McHenry M, Parker PA, Baile WF, Lenzi R. Voice analysis during bad news discussion in oncology: reduced pitch, decreased speaking rate, and nonverbal communication of empathy. Supportive Care in Cancer. 2012;20:1073-8. doi:10.1007/s00520-011-1187-8
13. Gustafsson SR, Wahlberg AC. The telephone nursing dialogue process: an integrative review. BMC Nurs. 2023;22(1):345. doi:10.1186/s12912-023-01509-0
14. Organization WH. Mental health atlas 2020: review of the Eastern Mediterranean Region. 2022. PMID: 9789240036703
15. Callejas Z, Griol D. Conversational agents for mental health and wellbeing. Dialog systems: a perspective from language, logic and computation. 2021:219-44. doi:10.1007/978-3-030-61438-6_11
16. Pelau C, Dabija D-C, Ene I. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. Computers in Human Behavior. 2021;122:106855. doi:10.1016/j.chb.2021.106855
17. Interian A, Chesin M, Kline A, Miller R, St. Hill L, Latorre M, et al. Use of the Columbia-Suicide Severity Rating Scale (C-SSRS) to classify suicidal behaviors. Archives of suicide research. 2018;22(2):278-94. doi: 10.1080/13811118.2017.1334610
18. Iyer R, Nedeljkovic M, Meyer D. Using voice biomarkers to classify suicide risk in adult telehealth callers: retrospective observational study. JMIR mental health. 2022;9(8):e39807. doi:10.2196/39807
19. RStudio. RStudio Builds. 2024; 2024.04.2:[https://chatgpt.com/c/95dcde54-c1c1-4206-b9be-9ea4f625e6f0.]
20. Audacity. Audacity Support. 2024; 3.5.1:[https://support.audacityteam.org/additional-resources/changelog/older-versions/audacity-3.5/audacity-3.5.1.]
21. Heck EJ, Davis CS. Differential expression of empathy in a counseling analogue. Journal of Counseling Psychology. 1973;20(2):101. doi:10.1037/h0034171
22. Mayer JD, Salovey P, Caruso DR, Sitarenios G. Measuring emotional intelligence with the MSCEIT V2. 0. Emotion. 2003;3(1):97. doi:10.1037/1528-3542.3.1.97
23. Bodie GD. The Active-Empathic Listening Scale (AELS): Conceptualization and evidence of validity within the interpersonal domain. Communication Quarterly. 2011;59(3):277-95. doi:10.1080/01463373.2011.583495
24. Joshi A, Kale S, Chandel S, Pal DK. Likert scale: Explored and explained. British journal of applied science & technology. 2015;7(4):396-403. doi:10.9734/BJAST/2015/14975
25. Ownby KK. Use of the distress thermometer in clinical practice. Journal of the advanced practitioner in oncology. 2019;10(2):175. PMID: 31538028
26. Bikker AP, Fitzpatrick B, Murphy D, Mercer SW. Measuring empathic, person-centred communication in primary care nurses: validity and reliability of the Consultation and Relational Empathy (CARE) Measure. BMC Family Practice. 2015;16:1-9. doi:10.1186/s12875-015-0374-y

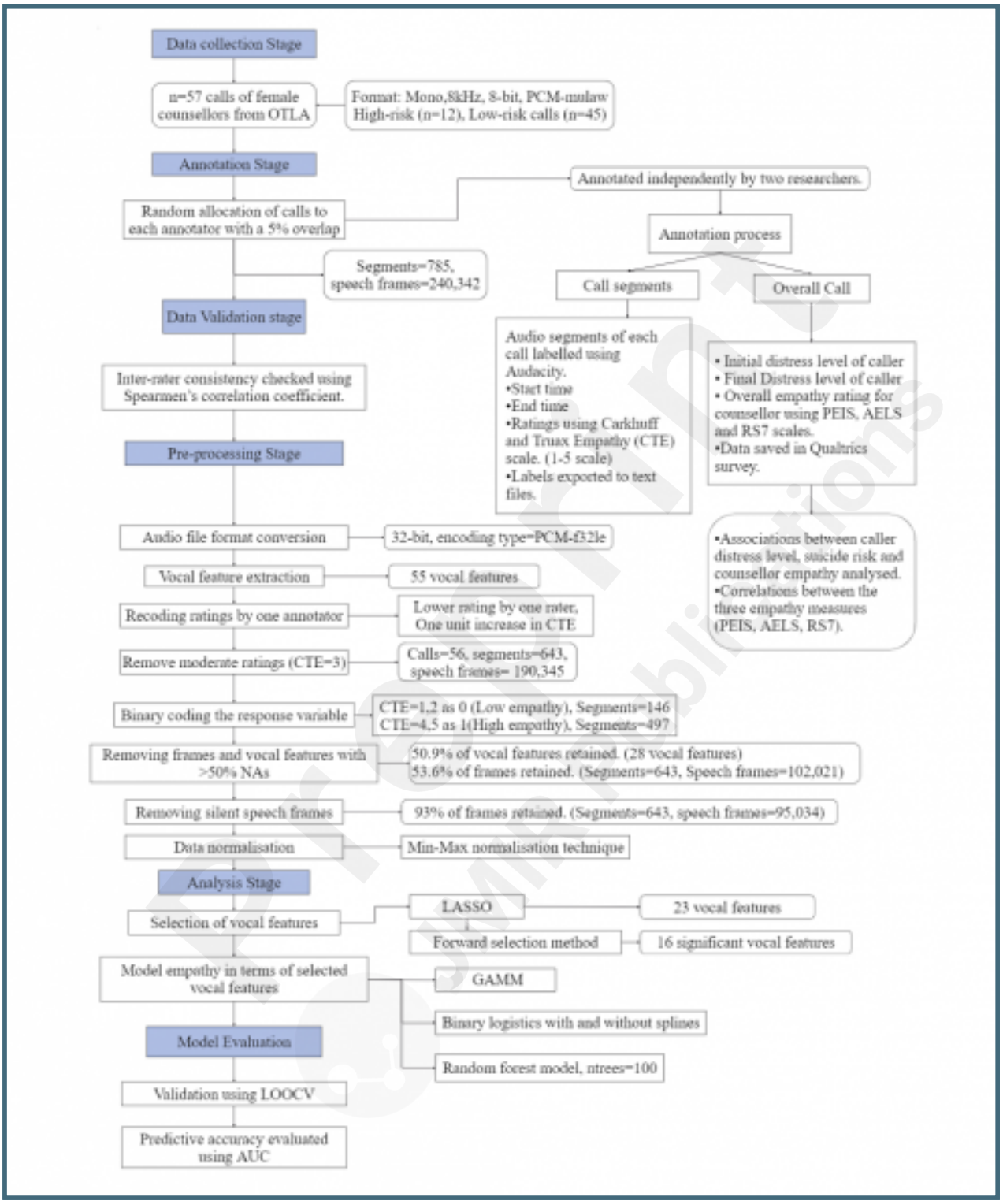
27. Anikin A. Soundgen: An open-source tool for synthesizing nonverbal vocalizations. Behavior research methods. 2019;51:778-92. doi: 10.3758/s13428-018-1095-7
28. Casale S, Russo A, Scebba G, Serrano S, editors. Speech emotion classification using machine learning algorithms. 2008 IEEE international conference on semantic computing; 2008: IEEE. doi:10.1109/ICSC.2008.43.
29. Rasmussen MA, Bro R. A tutorial on the Lasso approach to sparse modelling. Chemometrics and Intelligent Laboratory Systems. 2012;119:21-31. doi:10.1016/j.chemolab.2012.10.003
30. Sóskuthy M. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. arXiv preprint arXiv:170305339. 2017. doi:10.48550/arXiv.1703.05339
31. Wood SN, Pya N, Säfken B. Smoothing parameter and model selection for general smooth models. Journal of the American Statistical Association. 2016;111(516):1548-63. doi:10.1080/01621459.2016.1180986
32. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2011;73(1):3-36. doi:10.1111/j.1467-9868.2010.00749.x
33. Noroozi F, Sapinski T, Kaminska D, Anbarjafari G. Vocal-based emotion recognition using random forests and decision tree. International Journal of Speech Technology. 2017 Jun;20(2):239-46. doi:10.1007/s10772-017-9396-2
34. Guo CY, Chen F, Chang YJ, Yan JT. Applying Random Forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production. Biomedical Signal Processing and Control. 2022 Aug;77. doi:10.1016/j.bspc.2022.103811
35. Noroozi F, Sapiński T, Kamińska D, Anbarjafari G. Vocal-based emotion recognition using random forests and decision tree. International Journal of Speech Technology. 2017;20(2):239-46. doi:10.1007/s10772-017-9396-2
36. Raahul A, Sapthagiri R, Pankaj K, Vijayarajan V, editors. Voice based gender classification using machine learning. IOP Conference Series: Materials Science and Engineering; 2017: IOP Publishing. doi:10.1088/1757-899X/263/4/042083
37. Jacob A. Modelling speech emotion recognition using logistic regression and decision trees. International Journal of Speech Technology. 2017;20(4):897-905. doi:10.1007/s10772-017-9457-6
38. Tafiadis D, Kosma EI, Chronopoulos SK, Papadopoulos A, Drosos K, Siafaka V, et al. Voice handicap index and interpretation of the cutoff points using receiver operating characteristic curve as screening for young adult female smokers. Journal of Voice. 2018;32(1):64-9. doi:10.1016/j.jvoice.2017.03.009
39. Rice ME, Harris GT. What does it mean when age is related to recidivism among sex offenders? Law and Human Behavior. 2014;38(2):151. doi:10.1037/lhb0000052
40. Uloza V, Latoszek BBv, Ulozaite-Staniene N, Petrauskas T, Maryn Y. A comparison of Dysphonia Severity Index and Acoustic Voice Quality Index measures in differentiating normal and dysphonic voices. European Archives of Oto-Rhino-Laryngology. 2018 2018/04/01;275(4):949-58. doi:10.1007/s00405-018-4903-x
41. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surgery. 2016;159(6):1638-45. doi:10.1016/j.surg.2015.12.029
42. Gerdes KE, Segal E. Importance of empathy for social work practice: Integrating new science. Social work. 2011;56(2):141-8. doi:10.1093/sw/56.2.141
43. Regenbogen C, Schneider DA, Finkelmeyer A, Kohn N, Derntl B, Kellermann T, et al. The differential contribution of facial expressions, prosody, and speech content to empathy. Cognition & emotion. 2012;26(6):995-1014. doi:10.1080/02699931.2011.631296
44. Concannon S, Tomalin M. Measuring perceived empathy in dialogue systems. AI & SOCIETY. 2023:1-15. doi:10.1007/s00146-023-01715-z
45. Zhao Q, Neumann DL, Cao Y, Baron-Cohen S, Yan C, Chan RC, et al. Culture–sex

interaction and the self-report empathy in Australians and Mainland Chinese. Frontiers in psychology. 2019;10:378073. doi:10.3389/fpsyg.2019.00396
46. Hasan MR, Hossain MZ, Ghosh S, Soon S, Gedeon T. Empathy detection using machine learning on text, audiovisual, audio or physiological signals. arXiv preprint arXiv:231100721. 2023. doi:10.48550/arXiv.2311.00721
47. Besel LD, Yuille JC. Individual differences in empathy: The role of facial expression recognition. Personality and individual differences. 2010;49(2):107-12. doi:10.1016/j.paid.2010.03.013
48. Sanjeewa R, Iyer R, Apputhurai P, Wickramasinghe N, Meyer D. Empathic Conversational Agent Platform Designs and Their Evaluation in the Context of Mental Health: Systematic Review. JMIR Mental Health. 2024;11:e58974. doi:10.2196/58974

Supplementary Files

Figures

Voice analysis flow chart.



Multimedia Appendixes

Online Questionnaires used for data collection.
URL: http://asset.jmir.pub/assets/820eed388c39281e58ecb6b9578e36b8.docx

Descriptive Statistics of the analysis.
URL: http://asset.jmir.pub/assets/11b5dfcc85104931daff1e59af596728.docx

Results of consistency between the raters.
URL: http://asset.jmir.pub/assets/fd92db38d1a3ea5a31e9285b42707d67.docx

The results of LASSO regression.
URL: http://asset.jmir.pub/assets/6a82b716c33bc67c577e21d15587b20b.docx

Detailed results of the significance of the vocal features through GAMM, Random Forest and Binary Logistics Regression models.
URL: http://asset.jmir.pub/assets/f7c30320297249f5be608cca06f58d2e.docx

Impact of expanded dataset and rater diversity on voice analysis results.
URL: http://asset.jmir.pub/assets/5d51c79ef211ea258b19385be6bc6687.docx