Conveying Empathy in Spoken Language

Anushka Kulkarni, Andrea Lopez, Barnard College
Mentors: Run Chen, Professor Julia Hirschberg, Columbia University
Lab Team: Aruj Jain, Divya Tadimeti, Haozhe Chen, Linda Pang, Tejasri Kurapati

Introduction

Background
Much research has been done to analyze empathy in text, but little research has been done to identify the features that make a voice *sound* empathetic (1, 2, 3, 4).

Our Research
Our working definition of *empathy* is understanding another's pain as if we were having it ourselves and taking action to mitigate the problems producing it (1).

Our research aims to answer the following questions: What characterizes empathy? How can we train a model to identify empathetic speech?

This research will help design more empathetic AI, including personal assistants.

Goals & Hypotheses

The purpose of this project is to identify acoustic-prosodic and lexical features that distinguish empathetic speech from neutral speech, which will guide us in developing models for empathy detection and generation tasks.

Our hypothesis is that empathetic speech is different from neutral speech in key acoustic-prosodic or lexical dimensions.

Methods

We collected a dataset of empathetic and neutral speech segments and compared acoustic-prosodic and lexical features.

Data Collection:
- Found 289 empathetic YouTube videos and identified 718 empathetic segments with corresponding neutral segments
- Labeled each segment as having empathetic voice, empathetic text, or both
- Cleaned transcripts
- Annotated key features: speaker gender, emotion dealt with, stage of empathy

Acoustic-Prosodic Analysis:
- Used Praat (5) and Parselmouth (6) to extract features from audio segments
- Features: intensity, pitch, voice quality, speaking rate, etc.

Lexical Analysis:
- Used LIWC dictionary to find the frequencies of word categories in segment transcripts (7).
- Categories: linguistic dimensions, psychological processes, personal concerns, spoken categories, etc.

We began our analysis by comparing our collected empathetic segments to neutral segments from the MSP-Podcast corpus, which is a dataset of annotated podcast speech segments (8). This analysis gave us a baseline understanding of our data. Then, we collected additional audio segments so that we could pair many empathetic segments with neutral segments from the same speaker. With this data, we could conduct more authentic feature comparison using a paired t-test to find significant acoustic-prosodic & lexical features.

Limitations & Challenges:
- Inter-rater reliability due to subjective opinions about empathy
- Scarcity of equal numbers of neutral segments from speakers with empathetic segments has made it more challenging to compare a speaker's empathetic speech to their neutral speech

Results & Discussion

Acoustic-Prosodic Features

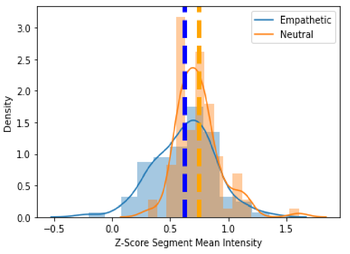
Significant Acoustic-Prosodic Features for Female Empathetic Speakers

Feature	t-score	p-value
Intensity mean	-3.818	0.000249
Intensity max	-2.685	0.00867
Jitter	4.117	8.626e-05
Shimmer	2.112	0.038

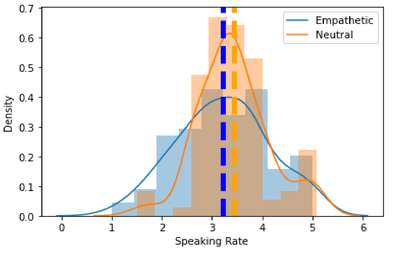
Significant Acoustic-Prosodic Features for Male Empathetic Speakers

Feature	t-score	p-value
Intensity mean	-3.910	0.00017
Intensity max	-2.182	0.031
Jitter	3.525	0.0006
Shimmer	2.613	0.010
Pitch mean	-3.797	0.00025
Pitch max	-3.194	0.0019
Speaking rate	-2.465	0.0154

This table shows the speech features that are significantly different across empathetic vs. neutral. A positive t-score means empathetic segments had a higher value on average.

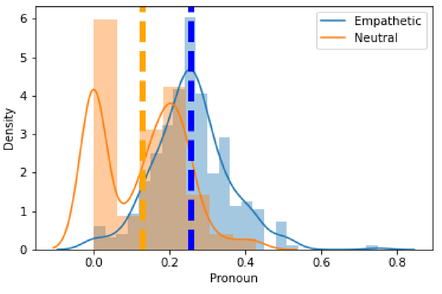


On average, for empathetic speech by *female* speakers, mean *intensity* was *lower* when compared to that of neutral speech by the same speaker.



On average, for empathetic speech by *male* speakers, *speaking rate* was *lower* when compared to that of neutral speech by the same speaker.

Lexical Features



On average, *pronouns* were used *more frequently* in empathetic speech than in neutral speech by the same speaker.

significant categories	t score	p value
you	15.31046802792000	4.70956132475776E-40
pronoun	15.23534152996220	9.16288311222893E-40
function	14.386386751819600	1.62838489050099E-36
affect	12.907723384358300	5.90581686437892E-31
cogproc	12.327772131384000	8.01920354305014E-29

This table shows the categories that are most significantly different across empathetic vs. neutral. A positive t-score means empathetic segments had a higher frequency of the category on average.

Conclusion

Preliminary Findings:
In our acoustic-prosodic analysis, we found the following categories were significant in empathetic speech compared to neutral speech: less intensity, more jitter, more shimmer and, for male speakers, lower pitch, slower speaking rate. This indicates that empathetic speech is less intense, or less loud, than neutral speech and that speech with slower speaking rate sounds more empathetic only with male speakers.

In our lexical analysis we found the following categories were used more in empathetic speech compared to neutral speech with a significant difference: you, pronoun, function. This indicates that empathetic speech uses more pronouns, especially second-person pronouns, than neutral speech.

Next Steps:
- Collect more data, including neutral and empathetic segments in Mandarin
- Look for most common n-grams in empathetic vs. neutral speech
- Consider the effect of other variables on empathetic speech
- Train a model to identify empathetic speech based on the patterns we find

References & Acknowledgements

1. Julia Hirschberg, "Conveying Empathy in Spoken Language," Amazon Gift.
2. Genta Indra Winata, Onno Kampman, Yang Yan, Anik Dey, Pascale Fung "Nora the Empathetic Psychologist," Interspeech 2017, Stockholm, 2017.
3. Zhaojian Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, Pascale Fung, "CAiRE: An End-to-End Empathetic Chatbot," 34th AAAI Conference on Artificial Intelligence (AAAI-20), 2020.
4. Mary Czerwinski, Javier Hernandex & Daniel McDuff, "Building an AI That Feels," IEEE Spectrum, 58 (50: 32-38, 2021.
5. Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.38, retrieved 2 January 2021 from http://www.praat.org/
6. Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. Journal of Phonetics, 71, 1-15. https://doi.org/10.1016/j.wocn.2018.07.001
7. James W. Pennebaker, Ryan L. Boyd, Kayla N Jordan, Kate Blackburn, "The Development and Psychometric Properties of LIWC2015," Technical Report, September 2015, DOI 10.15781/T29G6Z
8. Reza Lotfian and Carlos Busso, "Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings," IEEE Transactions on Affective Computing, 2019 (10.1109/TAFFC.2017.2736999).

It is a pleasure to acknowledge support for this research from Craig Newmark Philanthropies and Barnard College.

Department of Computer Science