

## RESEARCH ARTICLE

# Talking Less during Social Interactions Predicts Enjoyment: A Mobile Sensing Pilot Study

Gillian M. Sandstrom<sup>1\*</sup>, Vincent Wen-Sheng Tseng<sup>2</sup>, Jean Costa<sup>2</sup>, Fabian Okeke<sup>2</sup>, Tanzeem Choudhury<sup>2</sup>, Elizabeth W. Dunn<sup>1</sup>

**1** Department of Psychology, University of British Columbia, Vancouver, B.C., Canada, **2** Department of Information Science, Cornell University, Ithaca, N.Y., United States of America

✉ Current address: Department of Psychology, University of Essex, Colchester, United Kingdom

\* [gsands@essex.ac.uk](mailto:gsands@essex.ac.uk)



CrossMark  
click for updates

## Abstract

Can we predict which conversations are enjoyable without hearing the words that are spoken? A total of 36 participants used a mobile app, My Social Ties, which collected data about 473 conversations that the participants engaged in as they went about their daily lives. We tested whether conversational properties (conversation length, rate of turn taking, proportion of speaking time) and acoustical properties (volume, pitch) could predict enjoyment of a conversation. Surprisingly, people enjoyed their conversations more when they spoke a smaller proportion of the time. This pilot study demonstrates how conversational properties of social interactions can predict psychologically meaningful outcomes, such as how much a person enjoys the conversation. It also illustrates how mobile phones can provide a window into everyday social experiences and well-being.

## OPEN ACCESS

**Citation:** Sandstrom GM, Tseng VW-S, Costa J, Okeke F, Choudhury T, Dunn EW (2016) Talking Less during Social Interactions Predicts Enjoyment: A Mobile Sensing Pilot Study. PLoS ONE 11(7): e0158834. doi:10.1371/journal.pone.0158834

**Editor:** Christopher M. Danforth, University of Vermont, UNITED STATES

**Received:** December 17, 2015

**Accepted:** June 22, 2016

**Published:** July 20, 2016

**Copyright:** © 2016 Sandstrom et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data files are available on the Open Science Framework: [osf.io/qwmx4/](https://osf.io/qwmx4/).

**Funding:** This work was funded by the Social Sciences and Humanities Research Council of Canada, both through a Doctoral Fellowship to GMS [752-2011-2192] and a Standard Research Grant to EWD [410-2011-0582]. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

## Introduction

People generally enjoy talking to one another. When asked how they are currently feeling, people report being happier during social activities/interactions than during non-social activities [1]. When people think back on their day, they remember being happier during times in which they were socializing than during times in which they were doing other activities [2,3]. Further, people report being happier on days in which they recall more social activities [4–7]. These effects extend not only to interactions with close others, but also to interactions with people who are more peripheral in our social networks: people report being happier on days when they interact with more close friends and family, but also on days when they interact with more acquaintances [8].

Although, in general, the more social interactions a person has, the happier they feel, this conclusion ignores the fact that interactions differ in quality: Not every interaction results in equally positive feelings. Conversations that involve receiving help or support, and conversations that involve arguing or confrontation are associated with increases in negative affect [6].

**Competing Interests:** The authors have declared that no competing interests exist.

In contrast, the more enjoyable a conversation is, the more positive affect a person feels after the conversation [4]. These findings suggest that the quality of a conversation is related to the emotional response to that conversation. In turn, these emotional responses have implications for well-being, especially for older adults [9,10]. Given the difficulty of tracking every conversation that a person has, and the burden of reporting on the emotional response to each conversation, past research has often relied on retrospective, aggregate reports. Could there be another way to assess conversation quality?

The emotional quality of a conversation is not simply a function of what is said, but also *how* it is said. Research on the communication of emotion in music highlights the importance of prosody: linguistic features such as intonation, tone, stress and rhythm [11,12]. Auditory signals of pitch and loudness can be assessed through acoustical information (e.g., fundamental frequency and intensity [13]). Individual acoustical features (e.g., pitch, volume, and speech rate) have been linked to judgments of psychological constructs, such as power/dominance and competence. When men lower the pitch of their voice, others attribute higher social dominance [14–16]. Pitch also affects judgments of competence. In a forced choice design, male targets with lower pitched voices were judged to be significantly more competent (better leaders, more intelligent) than targets with higher pitched voices [16]. Volume is another acoustical feature associated with judgments of power/dominance; people associate trait dominance with loud voices [17]. Although these judgments could be merely a result of unfounded stereotypes, in fact people are capable of making relatively accurate judgments of others' personalities if they hear, but don't see the person [18].

The emotional quality of a social interaction may be related to not only acoustical features, but also structural features of the conversation. Computer scientists in the emerging area of social signal processing posit that computers can be empowered with the ability to sense and understand human social signals [19]. Experiments seeking automated ways to detect social signals have found that speaking time and interruptions are related to dominance [20–22], and that turn-taking patterns are related to social influence [23,24]. Turn-taking, as a measure of engagement, is also related to liking (e.g., in one study: feelings towards a speed-dating partner [25]).

Taken together, past research suggests that the acoustical and conversational properties of a social interaction might be related to psychological outcomes, such as emotional responses. Mobile phones provide an ideal means of capturing both kinds of properties because they are portable and are equipped with a wide array of unobtrusive sensors. The microphones built into mobile phones can pick up on in-person conversations even when the phone is not in use, providing an acoustical trace from which conversational and acoustical properties can be extracted.

Further, by using a phone app, this acoustical information can be collected unobtrusively, providing a window into real-world, everyday social experiences instead of into artificial social experiences created in a laboratory. All of this can be done while maintaining the privacy of both conversation partners; the auditory signal can be pre-processed on the phone so that only information like volume and pitch is sent to the researchers (i.e., formant information is not preserved, so that no raw acoustical data, such as voices/words, leave the phone [26]).

We used a mobile phone app, My Social Ties, to capture information about the social interactions people had as they went about their daily lives. In this pilot study, we explored whether the conversational and acoustical properties we extracted from a social interaction could predict the emotional response to the interaction. In essence, we wanted to know whether we could predict which conversations were enjoyable without hearing the words that were spoken.

## Methods

### Participants

We recruited 60 undergraduate students with Android phones, who participated in exchange for class credit or \$30. One student was removed from the study due to non-compliance and 4 students withdrew from the study because they were unhappy with the app (the audio files took up a lot of space on their phones and the app depleted their phone's battery). Due to technical difficulties related to downloading the enormous audio files from students' diverse phones (which included Samsung, HTC, LG, Dell, Motorola, and Sony), we had no acoustical information for 2 students. Due to file corruption, we could not process the acoustical information for an additional 9 students. These kinds of technical issues are not unexpected with non-commercial apps that are developed for small-scale use. Finally, given that hierarchical linear modelling has a practical minimum of three data points per person, we dropped 8 participants who had fewer than three conversations each, leaving us with a sample of 36 participants (21 females, 14 males, 1 did not report their sex;  $M_{\text{age}} = 20.6$ ,  $SD_{\text{age}} = 4.92$ ). These participants had a total of 473 conversations (range = 3 to 58;  $M = 13$ ,  $SD = 12$ ).

### Procedure

This research involving human participants was approved by the Behavioural Research Ethics Board of the University of British Columbia [H12-00469]. Participants came to the lab and provided written consent. Participants then filled out a survey with demographic information (including sex and age), and completed an abbreviated 21-item version of the Big Five Inventory measuring Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [27], plus three perceived intelligence items (for exact items see [28]).

Next, research assistants explained that the study involved installing an app, My Social Ties, on their Android phones, which they would use for 6 days. (NOTE: My Social Ties is not publicly available, but if you are interested in the possibility of using it for research purposes, please contact [tanzeem.choudhury@cornell.edu](mailto:tanzeem.choudhury@cornell.edu).) The research assistants explained that the app would store audio data collected during participants' conversations, but not any raw audio (i.e., their voice and conversation content could never be heard). The app was then installed on each participant's phone. Each participant read part of a story out loud to provide training data for the app, then sat in silence for one minute so that the app could detect the end of the conversation (i.e., the participant reading the story to the research assistant). Upon detecting the end of a conversation, the app triggered a survey, which asked participants to rate how they had felt during the conversation (1 = *very unhappy*, 7 = *very happy*;  $M = 4.31$ ,  $SD = 1.25$ ) (see [29] for the full list of questions on the momentary survey).

The audio files collected from participants' phones were parsed through a two-step process. First, we identified the voiced segments of the conversation and eliminated potential noises from the environment by using a method that has been validated on privacy-sensitive audio information [26]. Second, we performed speaker diarization, to identify the voiced segments where the participant was speaking, where other people were speaking, and where there was non-speaking noise and silence. Given that we did not retain raw audio data, we could not do speaker diarization manually. Instead, we used k-means clustering (with a random seed) [30] to break each individual conversation into segments based on volume (i.e., energy intensity). Previous studies have shown that k-means clustering is capable of achieving good results on conversations containing any number of speakers [31,32]. Since we were interested in the data of only one speaker (the person using the My Social Ties app), we were able to use a fixed set of clusters ( $k = 4$ ): 1) Extremely high (i.e., might not be heard by human ears)—noise of the

phone rubbing against clothing; 2) High—Voice of the person closest to the phone (i.e., the participant); 3) Low—Voices of other people; 4) Extremely low—Silence. Our first attempt at diarization resulted in an artificially high number of segments, as a result of being overly sensitive to the ups and downs and pauses within the volume fluctuations of a single speaker (e.g., misclassifying a participant's lowest volume conversation segments as belonging instead to their conversation partner; see [33]). Consequently, we ran a smoothing algorithm that assumed a minimum speaking time of 1.5 seconds. Finally, we removed the chunk of silence at the end of each conversation that was needed for the app to determine whether or not the conversation was terminated.

The acoustical and conversational properties of interest were extracted or computed based on the output of the diarization process (i.e., information about which voiced segments corresponded to the participant speaking, and which corresponded to someone else speaking). The volume and pitch were extracted from each voiced segment and we computed an average ( $\min_{vol} = 39$  db,  $\max_{vol} = 65$  db;  $M_{vol} = 57$  db,  $SD_{vol} = 5$  db;  $\min_{pitch} = 92$  Hz,  $\max_{pitch} = 253$  Hz;  $M_{pitch} = 182$  Hz,  $SD_{pitch} = 36$  Hz) and standard deviation ( $M_{vol} = 53$  db,  $SD_{vol} = 5$  db;  $M_{pitch} = 39$  Hz,  $SD_{pitch} = 23$  Hz) over all the voiced segments that corresponded to the participant speaking. The conversation length (min = 30 sec, max = 128 min;  $M = 8.6$  min,  $SD = 15.5$  min) indicates the length of the entire audio file (which corresponds to a detected conversation), except for the chunk of silence at the end. The percentage of time that the participant was speaking (min = 1%, max = 95%;  $M = 40\%$ ,  $SD = 22\%$ ) indicates the total amount of time that the participant spoke, divided by the conversation length. Finally, the rate of turn-taking (min = 1 per min, max = 18 per min;  $M = 7$  per min,  $SD = 4$  per min) indicates how many times people took turns speaking per minute (i.e., the number of voiced segments from the diarization process). For example, if the participant spoke, then someone else, and then the participant again, that would represent 3 turns.

An additional 65 conversations, not included in the descriptives, were discarded because of possible corruption or inaccurate diarization: the computed speaking time was less than 0 ( $N = 9$ ), there was no time when the participant was not speaking ( $N = 14$ ), the rate of turn-taking was abnormally high (more than 3 SD's above the mean;  $N = 6$ ), the average volume was abnormally high (more than 3 SD's above the mean;  $N = 8$ ), or the average pitch was higher than the maximum of the typical adult range (i.e., an average greater than 255 Hz;  $N = 28$ ). As mentioned earlier, these kinds of technical issues are not unexpected with non-commercial apps that are developed for small-scale use.

## Results

The conversational properties were marginally related to one another: conversation length was significantly correlated with percentage of time spent speaking,  $r(471) = -.29$ ,  $p < .001$ , but not with rate of turn-taking,  $r(471) = -.02$ ,  $p = .66$ . Percentage of time spent speaking was not significantly correlated with rate of turn-taking,  $r(471) = -.07$ ,  $p = .11$ . As to the acoustical properties, average volume was significantly correlated with average pitch,  $r(471) = -.37$ ,  $p < .001$ , and variability in volume was significantly correlated with variability in pitch,  $r(471) = .19$ ,  $p < .001$ .

Given the extremely large correlation between average volume and variability in volume,  $r(471) = .90$ ,  $p < .001$ , and the consequent likelihood of multicollinearity, it was important to use either average or variability in the subsequent analyses, but not both. Given that past research has focussed on variability [34], we used variability in volume and pitch as predictors in our analyses.

We capitalized on the fact that each person had multiple conversations by running within-person analyses using hierarchical linear modelling (HLM) via the lme4 package in R [35],

with conversation as the Level 1 variable, and person as the Level 2 variable. We predicted the emotional response to a conversation from conversation length, percentage of time spent speaking, rate of turn-taking, and variability in volume and pitch (all z-scored and entered simultaneously). Given that we lacked specific predictions, all analyses should be considered exploratory.

Conversations during which a person spent a smaller percentage of time speaking were enjoyed more,  $\beta = -.19$ ,  $t(30) = -3.98$ ,  $p < .001$ . In contrast, neither the conversation length,  $\beta = -.002$ ,  $t(30) = -0.04$ ,  $p = .98$ , nor the rate of turn-taking,  $\beta = -.001$ ,  $t(30) = -0.03$ ,  $p = .98$ , predicted how one felt after a conversation. Neither variations in volume  $\beta = -.05$ ,  $t(30) = -0.88$ ,  $p = .39$ , nor variations in pitch,  $\beta = .05$ ,  $t(30) = 1.09$ ,  $p = .28$ , predicted enjoyment.

Although our study is under-powered to test individual difference variables, and although individual differences were not the focus of our study, we examined the extent to which individual differences could predict conversation enjoyment. When all of the big-five personality traits, age, and gender were added to the model, none of these individual differences significantly predicted feelings,  $\beta$ 's  $< .10$ ,  $p$ 's  $> .39$ .

We also ran exploratory analyses to test for relationships between the acoustical and conversational properties (averaged across participants) and the individual difference variables (personality, age, gender). Neither average volume nor variability in volume was significantly correlated with any individual difference variables,  $r$ 's  $< .26$ ,  $p$ 's  $> .14$ , and did not differ by gender. As expected, average pitch was higher for women than for men,  $t(33) = 3.18$ ,  $p = .003$ . Additionally, older participants spoke with a somewhat lower average pitch than younger participants,  $r(33) = -.30$ ,  $p = .08$ . Pitch was not significantly correlated with any other individual difference variables,  $r$ 's  $< .21$ ,  $p$ 's  $> .24$ . Variability in pitch was not significantly correlated with any individual difference variables,  $r$ 's  $< .17$ ,  $p$ 's  $> .33$ , and did not differ by gender. Conversation length was not significantly correlated with age or personality,  $r$ 's  $< .23$ ,  $p$ 's  $> .19$ , but men had somewhat longer conversations than women  $t(33) = 1.99$ ,  $p = .06$ . The percentage of time spent speaking was marginally higher for older people,  $r(33) = .29$ ,  $p = .09$ , and, surprisingly, marginally lower for extraverted people,  $r(33) = -.30$ ,  $p = .08$ , but was not significantly correlated with any other individual difference variables,  $r$ 's  $< .28$ ,  $p$ 's  $> .11$ , and did not differ by gender. Finally, rate of turn-taking was not significantly correlated with any individual difference variables,  $r$ 's  $< .22$ ,  $p$ 's  $> .21$ , and did not differ by gender.

## Discussion

We used a mobile phone app to unobtrusively gather acoustical information about conversations that people had in their everyday lives. People's enjoyment of a social interaction can be predicted from conversational properties of that interaction. People enjoyed their conversations more when they spoke a smaller proportion of the time than usual. These effects were not moderated by personality, age, or gender.

Although our findings are based on only 36 people, those people had 473 conversations. Thus, the use of a more powerful within-person design bolsters the conclusions despite the small sample size. Given that the data for this study were collected via a mobile phone app, and given that mobile phone apps can be easily distributed via online app stores, future studies have the potential to collect large amounts of data from geographically distributed people who download and install the app on their own.

Future studies are needed to establish the generalizability of these findings. Indeed, several factors could moderate the relationship between acoustical/conversational properties and enjoyment. In a past study by Yuan and colleagues [36], speaking rate was found to vary by gender, age, and conversation partner: females, older people, and conversations with strangers

tend to exhibit slower speaking rates. Although we didn't ask participants specifically about conversations with strangers, we did ask whether each conversation was with a strong tie (e.g., close friends and family), a weak tie (e.g., acquaintances), or someone else. When we looked solely at the conversations with strong ties ( $N = 210$ ) and weak ties ( $N = 197$ ), we found no difference in how much participants enjoyed their conversations, and there were no differences in any of the acoustical (variability in volume, variability in pitch) or conversational features (conversation length, rate of turn-taking, or percentage of time spent speaking) depending on the conversation partner.

Culture is another possible moderator of the relationship between acoustical/conversational properties and enjoyment. On the predictor side of the equation, there is some evidence that women in various cultures exhibit differences in average pitch [15]. This might suggest that future replication efforts should focus on a single culture, and that the effect should be tested in several cultures that are known to vary in average pitch. On the outcome side of the equation, cultural differences in the extent to which people rely on internal speech result in differences in performance on reasoning tasks. It is not implausible that these differences might also have affective consequences, manifesting in differences in enjoyment.

At face value, the finding that talking a smaller proportion of the time resulted in more enjoyable conversations seems at odds with the fact that people who are depressed tend to talk less; the Center for Epidemiologic Studies' Depression Scale (CES-D) includes an item "I talked less than usual [during the past week]" [37]. However, we suspect that this item refers to the number of social interactions a person engages in, rather than the amount of talking during each social interaction. This interpretation is consistent with the finding that people are happier on days when they have more social interactions, whether with close friends and family or with acquaintances [8]. Future studies should further examine the distinction between these two constructs (i.e., amount of talking within a conversation vs. number of conversations). Also, it remains to be seen whether there is a minimum amount of talking within a conversation that yields benefits; in our experience, a conversation where you can never get a word in edgewise is not too enjoyable.

The current work has implications for Pentland's [34] theory of social signals. Our finding, that we can predict the emotional response to a social interaction from its conversational properties, is consistent with Pentland's idea that acoustical properties of interactions act as social signals. Pentland describes how to measure four types of social signals: activity level (proportion of time a person is speaking), engagement (the extent to which one person's turn-taking is influenced by the other's), stress/emphasis (variation in pitch and volume), and mirroring (mimicking the other's short utterances). The features that we examined map on quite closely to the features that he proposed: our proportion of speaking time maps onto his activity feature, our rate of turn-taking is similar to his engagement feature, and our variation in volume and pitch are analogous to his stress/emphasis feature. However, we could not analyze a feature that is similar to his mirroring feature, since participants' conversation partners generally did not have our mobile app, My Social Ties, installed on their phones.

Indeed, one limitation of this study is that we did not report any acoustical features related to the conversation partner. Intuitively this is important, as each conversation partner will influence the other, and both parties seem likely to influence the emotional response to a conversation. Future studies could examine not only the volume and pitch of the conversation partner, but also measure synchronicity of volume and pitch.

Another limitation of this study is that the results are dependent on knowing when a participant is speaking during each conversation, and when they are not speaking (i.e., dependent on accurate speaker segmentation and diarization). We were unable to do these steps manually, due to the privacy-sensitive app that we used. Instead, we used automated methods that have



been validated against manual methods [26, 31], but no automated method is 100% accurate. Although inaccuracies are inevitable, we have no reason to believe that these inaccuracies would result in a spurious relationship between the enjoyment of a conversation and the percentage of speaking time (but not other conversational or acoustical properties).

The unobtrusive, privacy-maintaining method used in the current study shows vast potential as a tool for psychological study. With more than 2.5 billion people around the world already carrying around smartphones as they go about their daily lives [38], there is a huge opportunity to harness mobile apps for the psychological study of everyday behavior. Physicians and clinicians could use a mobile app to monitor patients who have difficulty communicating (e.g., people with social anxiety disorder, or people with Parkinson's disease; [39]), and use the data to potentially feed into treatment plans. Psychologists could use a mobile app to understand the ways in which people interact differently with outgroup members, to test whether an intervention (e.g., the "fast friends" procedure [40]) changes the way people interact with others, or to test myriad other questions.

This pilot study demonstrates how the conversational and acoustical properties of social interactions can predict psychologically meaningful outcomes, such as how much a person enjoys the conversation. In other words, even without hearing the content of a conversation, we can predict the emotional response to it. The current work also illustrates the potential of mobile sensing to provide a window into everyday social experiences and well-being.

## Acknowledgments

The authors would like to thank Victoria Lau for project management, Georgia Bradley, Navio Kwok, Jaden Lu, Sophia Ng, and Anna Podolsky for data collection, Mashfiqui Rabbi and Karen Eddy for computer programming, Tauhidur Rahman for data processing.

## Author Contributions

Conceived and designed the experiments: GMS TC EWD. Performed the experiments: GMS. Analyzed the data: GMS. Contributed reagents/materials/analysis tools: VT JC FO. Wrote the paper: GMS VT JC FO TC EWD.

## References

1. Pavot W, Diener E, Fujita F. Extraversion and happiness. *Pers Individ Differ*. 1990; 11:1299–1306. doi: [10.1016/0191-8869\(90\)90157-M](https://doi.org/10.1016/0191-8869(90)90157-M)
2. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: The day reconstruction method. *Science*. 2004; 306:1776–1780. doi: [10.1126/science.1103572](https://doi.org/10.1126/science.1103572) PMID: [15576620](https://pubmed.ncbi.nlm.nih.gov/15576620/)
3. Krueger AB, Kahneman D, Schkade D, Schwarz N, Stone AA. National time accounting: The currency of life. In: Krueger AB, editor. *Measuring the subjective well-being of nations: National accounts of time use and well-being*. Chicago, IL: University of Chicago Press; 2009. pp. 9–86.
4. Berry DS, Hansen JS. Positive affect, negative affect, and social interaction. *J Pers Soc Psychol*. 1996; 71:796–809. doi: [10.1037/0022-3514.71.4.796](https://doi.org/10.1037/0022-3514.71.4.796)
5. Clark LA, Watson D. Mood and the mundane: Relations between daily life events and self-reported mood. *J Pers Soc Psychol*. 1988; 54:296–308. doi: [10.1037/0022-3514.54.2.296](https://doi.org/10.1037/0022-3514.54.2.296) PMID: [3346815](https://pubmed.ncbi.nlm.nih.gov/3346815/)
6. Vittengl JR, Holt CS. A time-series diary study of mood and social interaction. *Motiv Emotion*. 1998; 22:255–275. doi: [10.1023/A:1022388123550](https://doi.org/10.1023/A:1022388123550)
7. Watson D, Clark LA, McIntyre CW, Hamaker S. Affect, personality, and social activity. *J Pers Soc Psychol*. 1992; 63:1011–1025. doi: [10.1037/0022-3514.63.6.1011](https://doi.org/10.1037/0022-3514.63.6.1011) PMID: [1460554](https://pubmed.ncbi.nlm.nih.gov/1460554/)
8. Sandstrom GM, Dunn EW. Social interactions and well-being: The surprising power of weak ties. *Pers Soc Psychol B*. 2014; 40(7):910–922. doi: [10.1177/0146167214529799](https://doi.org/10.1177/0146167214529799)

9. Rook KS. Stressful aspects of older adults' social relationships: Current theory and research. In: Parris Stephens MA, Crowther JH, Hobfoll SE, Tennenbaum DL, editors. *Stress and coping in later-life families*. New York: Hemisphere; 1990. pp. 173–192.
10. Fiori KL, Windsor TD, Pearson EL, Crisp DA. Can positive social exchanges buffer the detrimental effects of negative social exchanges? Age and gender differences. *Gerontology*, 2013; 59(1):40–52. doi: [10.1159/000339747](https://doi.org/10.1159/000339747) PMID: [22814218](https://pubmed.ncbi.nlm.nih.gov/22814218/)
11. Frick RW. Communicating emotion: The role of prosodic features. *Psychol Bull.* 1985; 97(3):412. doi: [10.1037/0033-2909.97.3.412](https://doi.org/10.1037/0033-2909.97.3.412)
12. Thompson WF, Schellenberg EG, Husain G. Decoding speech prosody: Do music lessons help? *Emotion*. 2004; 4(1):46–64. doi: [10.1037/1528-3542.4.1.46](https://doi.org/10.1037/1528-3542.4.1.46) PMID: [15053726](https://pubmed.ncbi.nlm.nih.gov/15053726/)
13. Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol Bull.* 2003; 129(5):770–814. doi: [10.1037/0033-2909.129.5.770](https://doi.org/10.1037/0033-2909.129.5.770) PMID: [12956543](https://pubmed.ncbi.nlm.nih.gov/12956543/)
14. Cheng JT, Tracy JL, Ho S, Henrich J. Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *J Exp Psychol Gen.* 2016; 145(5):536–547. doi: [10.1037/xge0000166](https://doi.org/10.1037/xge0000166) PMID: [27019023](https://pubmed.ncbi.nlm.nih.gov/27019023/)
15. Van Bezooijen R. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Lang Speech.* 1995; 38(3):253–65.
16. Tigue CC, Borak DJ, O'Connor JJ, Schandl C, Feinberg DR. Voice pitch influences voting behavior. *Evol Hum Behav.* 2012; 33(3):210–216.
17. Carney DR, Hall JA, LeBeau LS. Beliefs about the nonverbal expression of social power. *J Nonverbal Behav.* 2005; 29(2):105–23.
18. Mehl MR, Gosling SD, Pennebaker JW. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *J Pers Soc Psychol.* 2006; 90(5):862–877. PMID: [16737378](https://pubmed.ncbi.nlm.nih.gov/16737378/)
19. Pantic M, Cowie R, D'Errico F, Heylen D, Mehu M, Pelachaud C, et al. Social signal processing: the research agenda. In: Moeslund TB, Hilton A, Krüger V, Sigal L, editors. *Visual analysis of humans*. Springer, 2011. pp. 511–538.
20. Mast MS. Dominance as expressed and inferred through speaking time. *Hum Commun Res.* 2002; 28(3):420–450. doi: [10.1111/j.1468-2958.2002.tb00814.x](https://doi.org/10.1111/j.1468-2958.2002.tb00814.x)
21. Littlepage GE, Schmidt GW, Whisler EW, Frost AG. An input-process-output analysis of influence and performance in problem-solving groups. *J Pers Soc Psychol.* 1995; 69(5):877. doi: [10.1037/0022-3514.69.5.877](https://doi.org/10.1037/0022-3514.69.5.877)
22. Wyatt D, Choudhury T, Kautz H. Capturing spontaneous conversation and social dynamics: A privacy-sensitive data collection effort. *Proc IEEE Int Conf Acoust Speech Signal Process*; 2007; Vol. 4, pp. IV–213.
23. Choudhury T, Basu S. Modeling conversational dynamics as a mixed-memory markov process. *Adv Neural Inf Process Syst*; 2004.
24. Choudhury T, Pentland A. Characterizing social interactions using the sociometer. *Proceedings of NAACOS*; 2004.
25. Madan A, Caneel R, Pentland A. Voices of attraction. In: *Proceedings of the International Conference on Augmented Cognition*; 2005; Las Vegas.
26. Wyatt D, Choudhury T, Bilmes JA. Conversation detection and speaker segmentation in privacy-sensitive situated speech data. *INTERSPEECH.* 2007;586–589.
27. John OP, Srivastava S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin LA, John OP, editors. *Handbook of personality: Theory and research* ( Vol. 2). New York: Guilford;1999. pp. 102–138.
28. Human LJ, Biesanz JC. Through the looking glass clearly: Accuracy and assumed similarity in well-adjusted individuals' first impressions. *J Pers Soc Psychol.* 2011; 100(2):349–364. doi: [10.1037/a0021850](https://doi.org/10.1037/a0021850) PMID: [21299315](https://pubmed.ncbi.nlm.nih.gov/21299315/)
29. Participants were also asked: to indicate whether they had been talking to a strong tie (i.e., a close friend or family member), a weak tie (i.e., an acquaintance), or neither; how many people they had been talking to; what was the nature of the conversation (Just for business/information; Both business and fun; Just for fun); and how surprised they were to have the conversation (1 = *not at all surprised*, 7 = *very surprised*).
30. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability.* 1967; 1(14):281–297.
31. Shum S, Dehak N, Chuangsuwanich E, Reynolds DA, Glass JR. Exploiting Intra-Conversation Variability for Speaker Diarization. *INTERSPEECH.* 2011; 11:945–948.



32. Shum S. Unsupervised methods for speaker diarization. Dissertation, Massachusetts Institute of Technology. 2011. Available: <http://dspace.mit.edu/handle/1721.1/66478>
33. Our first attempt at diarization resulted in a large number of segments voiced by parties other than the user (i.e., phone holder) that were shorter than 1 sec. in duration. We reasoned that these were more likely to be pauses while the user was speaking, rather than being other parties speaking for such a short length of time. We reviewed the literature to find an appropriate threshold for the minimum duration that was short enough to detect differences between segments, but long enough so that it wouldn't detect pauses as other people speaking. Given the wide range of past results, we decided to do our own benchmarking. We examined approximately 10 conversations, and determined that 1.5 sec was an appropriate duration; pauses longer than 1.5 sec might be awkward silence.
34. Pentland A. Social dynamics: Signals and behavior. In: International Conference on Developmental Learning (Vol. 5); 2004.
35. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–10. 2014.
36. Yuan J, Liberman M, Cieri C. Towards an integrated understanding of speaking rate in conversation. INTERSPEECH, 2006.
37. Radloff LS. The CES-D scale a self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1(3):385–401.
38. <http://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/>.
39. Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng. 2009; 56(4):1015–1022. doi: [10.1109/TBME.2008.2005954](https://doi.org/10.1109/TBME.2008.2005954) PMID: [21399744](https://pubmed.ncbi.nlm.nih.gov/21399744/)
40. Aron A, Melinat E, Aron EN, Vallone RD, Bator RJ. The experimental generation of interpersonal closeness: A procedure and some preliminary findings. Pers Soc Psychol B. 1997; 23(4):363–377. doi: [10.1177/0146167297234003](https://doi.org/10.1177/0146167297234003)