Communication study

Using music[al] knowledge to represent expressions of emotions

Stewart C. Alexander[a,b,c,*], David Kirkland Garner[d], Matthew Somoroff[d],
David J. Gramling[e], Sally A. Norton[f,g], Robert Gramling[f,g]

[a] Department of Consumer Science, Purdue University, West Lafayette, IN,USA
[b] Department of Medicine, Duke University Medical Center, Durham, NC, USA
[c] Health Services Research and Development Service, Durham VA Medical Center, Durham, NC, USA
[d] Department of Music, Duke University, Durham, NC, USA
[e] Department of German Studies, University of Arizona, USA
[f] School of Nursing, University of Rochester, USA
[g] Department of Family Medicine, University of Rochester, USA

ARTICLE INFO

ABSTRACT

Objective: Being able to identify expressions of emotion is crucial to effective clinical communication research. However, traditional linguistic coding systems often cannot represent emotions that are expressed nonlexically or phonologically (i.e., not through words themselves but through vocal pitch, speed/rhythm/tempo, and volume).

Methods: Using audio recording of a palliative care consultation in the natural hospital setting, two experienced music scholars employed Western musical notation, as well as the graphic realization of a digital audio program (Piano roll visualization), to visually represent the sonic features of conversation where a patient has an emotional "choke" moment.

Results: Western musical notation showed the ways that changes in pitch and rate correspond to the patient's emotion: rising sharply in intensity before slowly fading away. Piano roll visualization is a helpful supplement.

Conclusions: Using musical notation to illustrate palliative care conversations in the hospital setting can render visible for analysis several aspects of emotional expression that researchers otherwise experience as intuitive or subjective. Various forms and formats of musical notation techniques and sonic visualization technologies should be considered as fruitful and complementary alternatives to traditional coding tools in clinical communications research.

Practice implications: Musical notation offers opportunity for both researchers and learners to "see" how communication evolves in clinical encounters, particularly where the lexical and phonological features of interpersonal communication are concordant and discordant with one another.

© 2015 Published by Elsevier Ireland Ltd.

1. Introduction

Over the past two decades, we have contributed a number of studies examining emotions in medical encounters [1–10]. Over the course of this work, we have gradually moved away from a generic coding system that recognizes the literal, denotative expression of emotion through words [4] toward a more nuanced analytical approach that can track a subtle and diverse spectrum of types and intensities of emotion in conversation [2,11]. However, when we examine more emotion-intensive genres of talk such as palliative care consultations, we encountered a significant methodological problem: although the level of distress among patients is often quite high, and therefore audible in recordings of patients' voices, patients often express emotion phonologically – that is, not via their particular word choices, but rather in how they speak them. Such phonological expressions of fear, anxiety, happiness, calm etc., are often not captured by traditional coding systems, because they are registered in speech by way of volume, pitch, and phrasal shape. In addition, ambient noise in the natural clinical setting can preclude use of traditional laboratory acoustic analyses. Because of these challenges, indirect expressions of emotion (often referred to as cues) [12] that otherwise manifest audibly in human conversation – arguably the lion's share of emotional expression between people – tend to be neglected in large-scale studies requiring firm inter-rater reliability. We

therefore pose the question: are there other reliable ways to study the emotional expression in the natural setting of medical care that does not rely on words alone?

1.1. What are phonological features of language and why are they crucial components of emotional expression?

Scholarship from medicine, the humanities, and the social sciences has shown that communication entails much more than the mere words people speak to one another.

Yet classic research methods have tended to divide verbal communication into content (the lexical content of a message, the words spoken) and prosody (those features of speech that lie outside the denotative domain of words, i.e., inflection, emphasis and speech rate) [13–15]. The importance of prosody in human verbal expressions of emotion is axiomatic to this study.

When we represent the phonological elements of emotion, we are referring to specific key domains of prosody: pitch, speed/rhythm, and volume. Pitch refers to the frequency at which a given sound vibrates, usually measured in Hertz (Hz), or vibrations per second. Pitch allows us to distinguish between statements and questions, sarcasm and sincerity, and helps us perceive points of emphasis within utterances. Most importantly, changes in pitch can also indicate shifts in emotion. As pitch refers to the frequency of a given sound's vibrations, volume refers to size or amplitude of those vibrations. Changes in volume conventionally evince emotional shifts within the speaker or among multiple speakers. As a patient becomes depressed, her voice may soften as she starts to withdraw from the conversation, whereas when she becomes more upset during a conversation her volume may increase. Finally, speed/rhythm describes sounds in time and reflects speed, steadiness, and pauses in speech. For instance, speech increasing in speed/rhythm with agitation and decreases with psychomotor retardation from depression.

There is substantial evidence that phonological communication is a crucial component in the human interpretation of emotion [16–25] However, "considering its centrality to the care process, nonverbal behavior has received surprisingly little attention in the medical communication literature" [26]. Two recent studies indeed examined phonological features of the physician's voice in clinical interactions. Haskard et al. recorded 199 physician-patient conversations (51 primary care physicians) and 269 nurse–patient conversations (81 registered nurses) preceding an office visit with a primary care physician [27]. The team electronically filtered out all verbal content from the recordings, allowing naive raters the ability to assess affect from physicians' speech prosody. Their work illustrates how the emotional experiences of patients and physicians "reflect each others' emotional experience of satisfaction in their tone of voice" (p. 16) [27]. McHenry et al. examined 12 oncologists' speech in 17 clinical encounters to explore whether audible phonological features were associated with delivery of bad news [28]. The acoustic analysis included 30 s of physicians' speech (not patients), electronically obscured verbal content and excluded questions because the usual inflection could distort their interpretation of pitch. McHenry et al. argues that the "simultaneous assessment of verbal context and multi-parameter prosodic analysis of speech is necessary for a more thorough understanding of the expression and perception of empathy" [28]. This paper proposes such a novel approach that can integrate linguistic and phonological features of full-length conversations to better understand emotion expression in the natural medical setting.

1.2. How could researchers better represent emotional changes in medical encounters?

One possible solution is to use the same vocal analysis software that speech pathologists have used (e.g., PRAAT, University of Amsterdam) [29]. Yet using these instruments often results in a problem of scale. Built for in-depth acoustic analysis of words or short phrases, the software yields reams of data for every millisecond of audio input (i.e., providing pitch, duration, waveform, intensity, and spectrogram data for each millisecond of talk). This output is fruitful for speech pathologists studying subtle prosodic nuance in words or phrases. The problem for healthcare researchers is that programs like PRAAT do not have the ability to process and represent large amounts of audio data in a manageable way. Analyzing conversation-length data requires an ability to sift through data and draw out significant patterns. The PRAAT software often also requires a recording free of background noise in order to work properly; otherwise the software is not an effective tool for the analysis of conversation. Suppression of background noise in healthcare settings is impossible, given the various ambient sounds of ventilators, vital-sign measurement tools, and other human clinical interventions.

For the purposes of studying clinician-patient interaction, research has yet to develop elegant, standardized, and relatively accurate method that pulls out only a select and analyzable set of parameters: pitch, rate, and volume. We furthermore need a system that yields streamlined results in order to see large-scale phonological shifts in conversation. Finally, we need to represent the relative duration of words over lengthy strings of talk instead of specifying the exact number of milliseconds it takes a patient to say one word. The goal is thus a mid-range specificity: seeing the condition of the forest through various movements of its trees in the wind. There is indeed a convenient, flexible, and widely-used system that fulfills these criteria: music notation. For this study, we used two ways to represent emotion: Western musical notation and piano roll. Western music notation is a form of visual representation that musicians use to show how to play a piece of music (often referred to as Sheet Music). The second way is a piano roll, which is a music system that is used to operate a musical piece automatically.

The idea that conversations have poetic structures and musical qualities is not new [30–34]. However, we introduce the use of musical notation because it allows investigators to both visually notice and analyze phonological episodes of emotional expression that are not as readily available with existing methods. For example, this musical notation approach allows the analyst to oversee changes and shifts in pitch, volume, and rhythm/duration that are not nearly as apparent or objectively represented in common text-based transcriptions (e.g., Jefferson [35])."

2. Methods

This paper demonstrates how two musical notation techniques can be used to analyze prosodic features of a patient speech, using for example an excerpt of a longer audio-recorded discussion between a palliative care patient and a physician.

2.1. Example

This example from a palliative care consultation between a patient and a clinician serves to demonstrate the benefits of a musical notation approach to transcribing conversational data. This segment of the conversation was chosen because the patient does not utilize any specific words describing their emotion; however, they show audible signs of distress. Below is the transcript from the recording using the method developed by Jefferson [35] to render audio recordings of conversational speech:

P: No__ (0.4) ↓ I (0.7) and I don't even know if I_: (0.8) > ↓ I thou: ght about calling him and having him_< (0.7)< not bother to stop by because (.) I (1.0) few weeks ↓ago, ↑coup- >couple. weeks ago, I made an appoint:ment_: with him (0.8) uh:: for the eigh:teenth.