The imprint of disposition in social interaction

Mark Campana
Dept of English and American Studies, Kobe City University, Japan
https://doi.org/10.36505/ExLing-2016/07/0008/000267

Abstract

This study considers how listeners perceive and interpret the disposition of others through non-linguistic vocal cues. Changes in F0 and pitch span (measured against a 'running' mean of the previous 15 seconds), constellations of sequential tones, and emergent speech rhythms index recognizable states of positive/negative valency, desire, knowledge and/or processing, which together constitute emotional display (these same states correlate with mental predicates in the composition of emotion words. Excerpts of natural conversation were converted to 'iterant speech', i.e. speech devoid of lexical content. Listeners were invited to identify speaker disposition, and their ability to do so was remarkably accurate. The results lend support to a theory of vocal affect based on sound-types, rather than sounds.

Keywords: disposition; emotional display; mental predicates; iterant speech

Introduction

This paper addresses the imprint of disposition in social interaction. Disposition is taken to mean something like a frame-of-mind which both governs the behavior of the person who has it, and is evaluated by those who witness it. One can *have* a certain disposition (where *certain* is replaced by an adjective) or *be of* a certain disposition (idem). The question we address here is how the listener comes to realize that a speaker has (or is of) a certain disposition based on tone-of-voice. In principle, the answer will be the same as how the listener grasps the shifting mental states of the speaker in the course of interaction, but there are some differences.

To illustrate the phenomena, a person can have e.g. a *sunny* disposition or a *surly* one, be of a *grumpy* or a *fearful* disposition. Other plausible/attested collocations are *thoughtful, cheerful, kind, easy-going*—with positive valency—or *angry, taciturn*—with negative. One can also be *predisposed* towards a proposition with positive/negative content—e.g. judging someone *harshly* or with *kindness*.

Consider next the concept of an 'imprint', which is different from an impression. An impression of e.g. a person's character can be formed after a single encounter. An imprint typically results from several encounters, i.e. it includes memories of previous ones. In it, impressions are weighed and integrated in a more substantive schema. Basically, it takes longer to make an imprint of disposition, but in theory it can be appraised after a single encounter.

Social interaction is something that everybody understands. The focus here is on talk-in-interaction, considering utterances that are highly affective. These are called stances, or spoken actions in which a speaker displays his/her thoughts or feelings about some object, and communicates them to the listener with inevitable social consequences. We will concentrate on the prosodic features of stance utterances, paying special attention to pitches and pitch combinations (in sequence or together) rhythms, tempos and timbres. Different types of vocal are suitable conveyors of mental states, which in turn constitute emotions and dispositions.

Theory

What is tone-of-voice? First, it is 'about' tones (or pitches) but the array of sounds at the disposal of the speaker has a temporal aspect, an organizational one, and then there is the issue of voice quality. At the same time, we understand tone-of-voice to be the audible analogue of emotion. The litany of speech sounds in social interaction is essentially infinite. The oral cavity alone is designed such that even minute flexions of a single muscle (or muscle-group) can produce a complex, distinctive sound that is potentially 'meaningful' for the assessment of the speaker's mental state. In terms of efficiency, it would make sense for such sounds to be organized into sound-types for the purpose of transmitting and understanding vocalized meaning. Categorization is a cognitive skill at which humans (and some other species) have proved to be adept. In this paper, we test a specific theory of sound-types that index mental 'sub-states' of positive/negative valency, desire, knowledge and processing, which together constitute an emotional display (Wierzbicka 1999). Inasmuch as changes in perceived disposition correlate with controlled modulation of sound-type parameters, the theory can be verified. What then is emotion? This is not a simple question either, but we may start by following Wierzbicka (1999) and others in assuming that most 'emotions' include a 'thinking' part, as well as a 'feeling' one. In her model of semantics (NSM), words like *disappointment, afraid, happiness,* etc. are cast as 'cognitive scenarios', short narratives made up of simple words and propositions. Among the set of 'mental predicates' which play a key role in every scenario are *want, know, feel* and *think*. Together with *good, bad* and and *not* (also from the metalanguage) we derive the following mental states, any combination of which can be heard in the expression of emotion itself (abbreviated as WXYZ):

(1) Mental states (adapted from Weirzbicka 1999)
W wanting/not wanting (takes an object)
X knowing/not knowing (takes an object)
Y feeling good/bad (about something)
Z thinking (no negative counterpart)

The next step is to match the types of sounds that make up tone-of-voice with WXYZ. This is only an approximation, whereby a given sound type is just a 'leading indicator' of a mental state, not necessarily the only one. Combinations of sounds (as well as the meaning of words) can also index a mental state. That said, we propose that voice qualities—broadly defined—are used to signal states of *wanting* or *not wanting*. Intensity of F0 (volume) counts as a voice quality, along with upper partials (timbres) and non-standard vocal gestures, such as 'clipped' endings, etc.

Short tunes or melodies—sequences of tones—are used to signal *knowing* or *not knowing*. *Aizuchi* (backchannels) are typical: even when the 'tune' appears to have a single tone, it is juxtaposed against that of previous speech. Consider what it sounds like to say "I don't know" in your language. Echoes of the same can be heard in longer stretches of speech as well.

Next, consider the mental states of *feeling good* or *feeling bad*. These correspond most closely to valency, as it is known in emotion research. Pitches and pitch combinations are primarily responsible for signaling these states. Cook (2003) develops the idea that valency follows from three-tone chordal structure, and there is no reason to dispute this. Emotional displays do unfold quickly, so it is likely that even tones in sequence are perceived as simultaneous, i.e. in the 'psychological now'.

Finally, we propose that rhythms and timing units in general (tempos, pauses, hesitations etc.) accurately reflect the mental activity of *thinking*. It is not enough to simply demonstrate that thinking is taking place; the style presentation and grouping of syllables is important too, influenced in part by the choice of words.

To summarize, the mental predicates that serve to characterize emotion words in Wierzbicka's semantic system correspond to real mental states that occur in the display of emotion. In theory, such states could be indicated by facial expression, body movements (including gesture), or simply words. Tone-of-voice is just another means of expression, where each mental state/activity is indicated by a sound type, shown below (wxyz):

(2) Mental states and leading sonic indicators (sound types)

W wanting/not wanting	w voice qualities
X knowing/not knowing	x short tunes/melodies
Y feeling good/bad	y pitches/pitch combinations
Z thinking	z timing units (rhythm, tempo)

Given that at least one display of emotion is necessary to appraise a speaker's disposition, it follows that the same elements listed here will contribute to it. In the following section, we outline how such events can be discerned in a controlled experiment.

Data, methods

In the course of daily interaction, listeners can appraise the disposition of a speaker based on tone-of-voice. Can naïve subjects reach similar conclusions in a clinical experiment? Possibly, but not necessarily: every action depends on individual experience, social consequences, and other factors. It isn't fruitful to devise an experiment along these lines. Nevertheless, listeners may be able to recognize repeated patterns in a speaker's voice on different occasions, and trained ones can identify and describe them. Gathering such data from a longitudinal study is optimal, but impractical. In the tasks reported on here, listeners were presented with stance utterances from speakers over a range of topics, and asked to appraise their disposition. In order to control for word meaning though, the stance utterances were converted to 'iterant' form, leaving only prosody.

In its core meaning, a stance is a physical event whereby the stance-taker assumes a bodily position that signals a clear intention to the audience. One can easily imagine how something like 'defiance' is acted out by assuming a defensive posture. In current sociolinguistics, the concept of stance has been extended to talk-in-interaction. Many researchers refer to the seminal work of DuBois (2007), who proposes that every stance has a subjective dimension (i.e. about the speaker), an objective one concerning the person or thing being evaluated, and an intersubjective dimension which pertaining to the social relationship between speaker and hearer. He refers to this as the "stance triangle". A stance utterance encapsulates the stance, and can be regarded as its core element. Stance utterances make good objects for study because a) they are usually short and succinct, and b) they tend to summarize a speaker's story or narrative. Typical stance utterances might be "I'm sorry, but that's *not* exactly what I had in mind", "There's a *reason* why we do this", or "I don't even know if that's *enough*" (emphasis added). Further examples are given below, with purported effects (punctuation omitted):

(3) Typical stance utterances (all negative valency)	TOPIC
a. The worst is yet to come	[global warming]
b. Hillary (Clinton) does not inspire confidence	[politics]
c. Frankly, I can't understand how people put up with this	[migration]
d. The Internet hasn't enriched my life in any significant way	[modern life]
e. Keeping up relations takes a lot of work	[social obligations]
f. Every day I eat the same thing	[food]

Judgements of disposition are based on tone-of-voice as well as words, however. In order to test for it, it is necessary to expunge all lexical content. Nooteboom (2000) suggests using 'iterant' speech, that is substituting nonsense syllables for words, thus preserving prosodic features. At present this can only be done by humans, and is most effective when the forms are

produced immediately after voicing. To illustrate, the same utterances in (3) are repeated below as iterant speech:

(4) Iterant speech
a. da**DA** da*Da* da***Da:***
b. ***Dadada*** da**DA** da**Da**da **Da**dada
c. **Da**da | da**Da** dada***Da*** da dada da*Da* dada
d. *daDadada* dada dada **DA**da dadada ***daDadada***
e. ***dadada*** da*Da*da da dadada **DA**
f. dada**Da** dada dada **DA**

'Prominent' syllables appear in in upper case letters, with two degrees of prominence (onset or onset+vowel). These are all stressed syllables in English which might be represented by some other prosodic feature in another language. Prominence, or sentential stress is itself a kind of voice quality, pointing to extremely rapid displays of *wanting* or *not wanting*—[W] in the syntax of mental states). Metrical structure—and some hint of rhythm—is preserved in the grouping of syllables ([Z]). Most of the prominent syllables in (4)—and some non-prominent ones—are show relative pitch levels: bold (non-italic) stands for highest, bold italic for lowest, and italic for mid. The intervals between the tones are significant, but cannot be depicted in this transcription system. Tones in sequence and in harmony are responsible for the communication of melody and valency—[X] and [Y] in the theory of emotion we are assuming).

One Japanese and one English speaker produced scripted, 'emotional' utterances in reference to several topics. For each topic, one utterance was characterized by positive valency, another by negative valency. These were then converted to iterant speech and presented to separate groups of Japanese and American subjects. In one test, subjects were asked to appraise the disposition of the speaker (same and different languages). Only speech forms of one valency ([±]) were presented; no choices were offered. In a control test, speech forms of both valences were 'mixed'.

Subjects were prompted with a lexical 'introduction' to each topic, before hearing converted (iterant) utterances. Samples included *He was real bastard, didn't give a fig about the people who elected him* ([–]) vs. *Actually, he didn't do anything that everyone else before him had done* [+] (in reference to Masuzoe, the former mayor of Tokyo); *It doesn't taste the same, and it kills off all the nutrition* [–] vs. *I use it all the time* [+] (RE food/microwave ovens); *It sucks. Worst thing to hit the planet* [–] vs. *It's raining now, but it should be better soon* [+] (weather), etc.

Discussion

The results of these tests were predictable. Subjects could easily determine valency based on their choice of terms to describe perceived disposition, e.g.

grumpy, cheerful, or Japanese *ganko* 'stubborn', *rakutenteki* 'optimistic', etc. The 'mixed' test of utterances with positive/negative valency produced no consensus as to what kind of person the speaker was. While it is unfortunate that more nuanced appraisals of disposition beyond valency could not be obtained, to do so would be difficult given limited exposure to the speakers' tone-of-voice, the varied experiences of the participants, and the different conceptualizations of emotion in the languages themselves.

Listeners gather their impressions though repeated verbal exchanges. Not only through words (lexis), they may rely on prosodic features to build an imprint. Experiments have shown that listeners can do this based on iterant speech where lexical/semantic meaning has been stripped away. We have proposed that disposition is indeed analyzable in the same terms as 'emotions' generally, where the latter are understood as composites of mental states WXYZ related to types of sound (wxyz): voice qualities, sequential tones, tones produced simultaneously, and timing units.

What distinguishes 'disposition' from rapid, continuous displays of mental states is time. Given the similarity of (theoretically quantifiable) frequent displays, the listener will store them economically in terms of a general impression or 'imprint' with regard to the speaker. To judge someone's disposition then, is to have such an imprint. Regardless of topic, a speaker with a certain attitude will voice similar prosodic outlays over time. This can be shown with a more precise examination of interval sizes and 'harmonic' effects that arise between and among prominent tones. Listeners can recognize previously-heard constellations of sounds, and base their appraisal of speaker disposition on them. Speakers may also gravitate towards topics that facilitate the expression of their attitudes. This implies they sometimes choose words based as much on how they sound as on the meaning of words themselves. It is certainly a topic worthy of future study.

References

Cook, N. 2003. *Tone of Voice and Mind.* John Benjamins Publishing Co., Amsterdam.

Crystal, D. 1975. *The English Tone of Voice.* Edward Arnold, London.

DuBois, J. 2007. The stance triangle. In *Stancetaking in Discourse.* R. Englebretson (ed.), John Benjamins Publishing Co., Amsterdam.

Laver, J. 1980. The Phonetic Description of Voice Quality. CUP.

Nooteboom, S. 2000. The prosody of speech: Melody and rhythm'. MS, Research Institute for Language and Speech, Utrecht.

Wichmann, A. 2000. The attitudinal effects of prosody and how they relate to emotion. *Proc. of ISCA Workshop on Speech and Emotion;* Cowie, R., E. Douglas-Cowie, & N. Schroder (eds.)

Wierzbicka, A. 1999. Emotion Across Languages and Cultures. CUP.