Tone-of-voice : message in the medium

Tone-of-Voice: Message in the Medium

Mark CAMPANA

I. Introduction*

The title of this paper is a play on words, specifically those of Canadian philosopher Marshall McLuhan who famously said “the medium is the message.” Basically this means that the form of a medium (video, text, etc.) is embedded in every message that it transmits—the medium thus influences how the message is perceived. In our case, the medium is the human voice that colors the words and phrases that make up the spoken utterance. How it does so is of some interest: the medium, as we shall see, comes in more than just one form, hence we might have to speak of a plural ‘media’ instead of a single medium.

Tone-of-voice is an interesting concept. Everyone seems to know what it is (as in the often-repeated admonition “Don’t talk to me in that tone-of-voice!” spoken by a frustrated parent to their sarcastic son or daughter), but it is not a technical term in the speech sciences. Linguists will research tones or intonation, but rarely tone-of-voice. This may be because its ‘parts’ interact in subtle, little-understood ways. Humans are capable of producing sounds of different orders, such as tempo, melody, timbre and rhythm (all of which are familiar to students of music). In terms of meaning, tone-of-voice often signals the speaker’s attitude or social stance.

We thus assume that speakers ‘embed’ their vocal sounds in several channels, one of which (at least) is dedicated to tone, or pitch. Clearly though, there is more than just one tone in the repertoire of the speaker, just as there are many notes in music (the scale DO-RE-MI-SO-FA-LA-TI-DO names eight, but these are merely conventional posts along a continuum; there are many more). The title then should really be “Tones of voice: messages in the media.” This doesn’t sound very elegant though, and for good reason: titles and other types of spoken text are more appealing when their metrical structure—another feature of tone-of-voice—is ‘balanced’ (more on this

* This paper is the written version of a talk given on March 7, 2018 at Kobe City University of Foreign Studies. Many thanks to the participants of that event, and to the university itself which has nourished an invigorating intellectual and pedagogical environment from the beginning of my tenure to the end.

below). Sometimes we say or write things because we want them to *sound* good, not just because the words are interesting. This is not to say that words are unimportant; without the scaffolding they provide, vocal sounds could signal almost anything, and the listener would be clueless. For communication to be successful, there must some agreed-upon format whereby words are pronounced in a recognizable way. Beyond producing basic structure, however, speakers are free to influence how their words should be interpreted. They do this through their tone-of-voice.

II. Elements of tone-of-voice

Tone-of-voice is not a simple thing, but is composed of several types of sound. This is not a mystery. A good place to start the discussion is with the concept of prosody—a word invented by the Greeks, who first brought attention to vocalic features floating above the segments of their language. Prosody refers to those *extra* properties of consonants and vowels. Stress, for example, is *extra* energy placed upon a vowel, the center of a syllable. Tone is an *extra* pitch value (frequency) placed upon a vowel—traditionally described as high, medium or low, but allowing for contours between these levels too. Sometimes vowels or consonants are given *extra time* to be pronounced. Stress, tone and duration are called prosodic features.

Languages incorporate prosodic features into their phonological systems to establish lexical meaning. Stress has been incorporated in English to distinguish nouns from verbs: *cónstruct* (noun) vs. *constrúct* (verb). Chinese uses four different tones to distinguish the meaning of segmentally identical words: *mā* 'mother'; *mǎ* 'horse'; *má* 'hemp' and *mà* 'scold'. Japanese uses both vowel and consonant length (duration) for the same reason: *koto* 'thing', *kōto* 'coat', *kōtō* 'high class', *kotō* 'solitary island' and *haka* 'tombstone' vs. *hakka*, a kind of leaf. In its original sense (that is, according to the Greeks), prosody is a suprasegmental feature (lit: "above the segment"). It is a property of words.

But tone-of-voice uses the same features in a different way. After all, English doesn't *own* stress; every other language uses it for one purpose or another, usually when the speaker wants to emphasize something, as in sug*GOKU konde ta!* ("It was *very* crowded"). At the same time, tones make up intonation (melody+rhythm), and every language has that. And who can't make a word sound longer to express doubt? (*Ikeru kanaaa…*) The point here is that these same so-called prosodic features are present in everything we say.

So what else is there to tone-of-voice besides prosody? Well, consider this: everybody has a different mouth, a different tongue, a different set of lungs. Clearly, these factors result in a different sound. Speakers are aware of this, of course, and can even imitate the sound of others. They can speak in unique ways to different people. The bottom line is, we consciously control the muscles that produce vocalization. Somehow this has to figure into tone-of-voice as well. Linguists call this voice quality, much like timbre in music. Middle C doesn't sound the same on a piano as it does on a trombone.

There is still more to tone-of-voice than prosody and voice quality. Consider the temporal aspects of speech, the timing mechanisms. Tempo has to do with speech rates: slow, fast, or very fast. Rhythm marks intervals of so-called prominent syllables—i.e. those with stress, higher pitch, or longer duration (otherwise known as 'beats'). If the intervals in speech are regular (that is, the same or similar), we say it has a certain rhythm.

Finally, there is a mathematical element to tone-of-voice which goes by the name of meter. Basically, this means counting the number of beats: syllables, prominent syllables, or the time it takes to pronounce a group of syllables. We are always counting. If a speaker repeats the same number and arrangement of beats, listeners will take notice: *The bigger they are, the harder they fall*. They also pay attention when led to believe that an intonation phrase will be repeated but is not: *He said he'd be on time, and he was!* The first case can be described as 'resolution', the second as 'surprise'. Both terms are also familiar to students of music.

III. Organization of elements

Having identified the types of sounds inherent in tone-of-voice, we now attempt to show how they might match up with types of meaning. To the knowledge of this author, no such attempt has been made before, beyond common generalizations like "a rising tone at the end of sentence indicates uncertainty" (as in a question) and the converse. In what follows, I will outline a system which seems to hold some promise. It is based on a semantic theory of emotion words — a plausible approach since, for the most part, tone-of-voice is the audible analogue of emotion. In this system, the types of sounds discussed above align with general types of meaning. But what kind of meaning?

The semantic theory adopted here is the Natural Semantics Metalanguage

(NSM) of Wierzbicka (1999). Like many other translators, Wierzbicka realized that certain cultural concepts—in particular, emotion words like *disappointed, hopeful* or *satisfied*—do not mean the same in different languages. Even more seriously, scientific studies of emotion proper are often carelessly based on the researchers' own native language emotion words, effectively undermining any insight into human psychology. To address this problem, Wierzbicka proposed that all emotion words be defined in simple terms that *can* be translated accurately and which are found in every language.

The metalanguage is thus made up of universal terms. Every culture has a word for *want*, so this word would be included in the metalanguage. Not every language has a perfect translation for *disappointment*, however, so this emotion word would not be listed. Instead it would be 'defined' via a short narrative composed of simple universal terms, e.g. "I thought that something good was going to happen, and I felt pretty good about that. Now I know it *isn't* going to happen, and I feel bad." In Wierzbicka's system (and as seen here), emotion words are presented as if the perceiver were experiencing the emotion in a real-life situation. It is only a short cognitive step to assume that perceivers really *do* experience the states and processes laid out here.

Each line of a narrative (called a cognitive scenario) is constructed from words listed in the NSM. Certain key words are ubiquitous, the so-called mental predicates *see, hear, want, know, think,* and *feel*. Seeing and hearing do not pertain directly to emotion, since they are less involved in inner states than others; we omit them from consideration here. Note too that *feeling* usually means "feeling good or bad" both of which can be heard in the voice of a speaker. The negative counterpart of each mental predicate also plays important role in the system. In the following cognitive scenario the key terms are shown in boldface:

(1) Cognitive scenario: *fear* (adapted from Wierzbicka 1999)

a. "I **don't know** what will happen.
b. **Bad** things can happen, but I **don't want** them to.
c. I **want** to do something, but I **don't know** if I can."
d. When a person **thinks** this, they **feel** something **bad**.

Looking at the terms *wanting* (or *not wanting*), *knowing* (*not knowing*), *feeling good* or *feeling bad* and *thinking,* we may ask if any particular type of sound is a good medium for their transmission. Oftentimes when someone wants something badly (or adamantly does *not* want that thing), they signal their commitment through voice

quality, that is, greater intensity (volume) or wider pitch span (deeply embodied). When a speaker wishes to show that they know something (or have no idea about it), they often use short tunes, or melodies. Think about it: How does one say "I don't know" in English or Japanese? The pitch changes can go up down, but there is always a pitch change: *Shiranakatta!* When a speaker feels good, s/he uses higher, 'brighter' pitches, whereas feeling bad invokes lower, 'darker' ones. When a speaker is thinking (or at least appearing to think), s/he will rely on rhythm, meter and/or tempo to demonstrate their understanding (note that there is no sound correlate of *not thinking*).

To summarize then, it appears that tone-of-voice is organized in a systematic way, based on the mental predicates of *wanting, knowing, feeling* and *thinking*. Let us next consider this proposal in more detail.

IV. Sound and meaning

How does tone-of-voice work in everyday conversation? In this section we look at each of the four sound-types and match them with the mental predicates *want, know, think,* and *feel*. The sound-types in question are voice quality (more accurately, qualities); short tunes or melodies; pitches (and in particular pitch combinations), and timing units: rhythm, meter, and tempo.

Voice quality

There are several aspects to consider regarding voice quality. In phonetics people talk about "whispery voice" or "creaky voice", or in Japanese, *kyōshuku*: the tone-of-voice a female employee might use to tell her older male boss that he screwed up. These are special cases, but voice qualities are ubiquitous and diverse. It takes muscles to produce them, hence they are best understood in terms of gesture. When people make a visual gesture, they move their fingers, hands and arms; when such movement is held in place, it's called a stance. When speakers make a vocal gesture, they activate the muscles that control their lips and tongue; when they set the muscles of the mouth and larynx to a non-default setting—a kind of stance—they produce a unique voice quality.

Voice begins with the vocal cords, which are also under the control of the speaker. When they vibrate, they produce the unique sound of each individual called the fundamental frequency, or F0. When speakers increase the power (intensity) of F0, it is perceived as loudness. Volume then, is one kind of voice quality. Other (higher) frequencies are produced in the throat and mouth which are constantly changing in

shape (often due to the position of the tongue). Increasing the intensity on these frequencies results in different vowel qualities ([i, e, u, e, o]) as well as others that the speaker chooses, for effect. The latter are somewhat akin to musical timbre. In short, voice quality means controlling the muscles that alter the shape of the vocal tract for the purpose of highlighting a certain sound.

Speakers can also activate multiple frequencies at the same time. When only one is activated they sound like a robot. When several frequencies are activated, they are perceived as having a 'rich' voice, like that of a seductress. Try saying wonderful, with a deep, rich sound on the first syllable. Does this not sound like *wanting* or *not wanting?*

Pitch variation (tones in sequence)

Let us now consider 'short tunes' or melodies. These are created as pitch levels rise or fall over a stretch of time, the length of which can either be long or short: a syllable (very short), a word or phrase ('short') or a sentence (long). The sole claim here is that the contrast between pitches is important: if there are no changes in pitch level, the sequence will have no meaning. When speakers end their utterance on a higher pitch preceded by a lower one, there is an air of uncertainty about it, as in a yes/no question. The opposite is true when a lower pitch is preceded by a higher one, as in a statement of fact. Clearly, these reflect mental states of *knowing* or *not knowing.* The story is not quite so simple though: in effect, *any* tune can reflect a knowledge state if the speaker wants it to. Consider the case of *wakaranái*, with a higher pitch on the final syllable. While the word itself denotes uncertainty (not knowing something), the prosody indicates the opposite, in effect "knowing about not knowing." There are times too when we say to ourselves "That person (he or she) sure seems to know what they're talking about." How is it that we get this feeling? It is likely that the speaker is repeating the same falling pitch pattern over several utterances. In this case, they are demonstrating their knowledge in a general sense, not necessarily about any single item.

During the course of speech, displays of mental states can change from one moment to the next. This is especially true of knowledge states, which aren't really about what the speaker knows (e.g. that the capital of India is New Delhi), but what s/he *thinks* they know. As such, this component of tone-of-voice is often taken for confidence (or lack of it), technically a feature of personality.

Pitches and pitch combinations

Individual tones (or pitches) are associated with positive or negative emotions. Speakers usually convey a positive emotion (e.g. joy) when using higher pitches, a negative one (dissatisfaction) with lower ones. In tone-of-voice, this is the basis for *feeling good* or *feeling bad.* Still, higher pitches can also be heard with anger, lower ones with contentment. Moreover, we don't always communicate our feelings with single tones. Pitches are heard in combination, for instance, much like musical chords. In the course of one or two utterances, the impression of the speaker's inner state may be described as 'bright' (as with a major chord) or 'dark' (a minor one). In music, the tones that make up a chord are produced at the same time, yet in speech they are sequential. How then is it possible to produce a chord-like combination in speech? According to Levitin (2006), we often speak so fast that sequences of pitches (which would otherwise constitute a melody) are interpreted as being spoken simultaneously, as in harmony.

As every music student knows, it takes a minimum of three notes to form a chord. When this happens, the middle note is cast against a background defined by the distance between the other two (the longest interval). Longer utterances (more than just a couple of words) typically begin at F0 and climb to an intonation 'peak' before descending to a final note. If the final note is also F0, there aren't enough pitches to form a chord, and the sentence (if it is a sentence) is perceived as having little or no emotion. If the final note is something besides F0, the resulting chord receives a value—positive or negative—depending on the coincidence of the frequencies involved. The following illustrates how a sequence of pitches can 'sonify' a complex emotion:

(2) Harmony in a single utterance: 'frustrated'

L　H　　M

I cán't get this damn thing to work

[L]----------[M]----------[H]

F0　　7.2　　10

In (2), LMH stand for low, middle and high tones (pitches), relative to the speaker's voice (they do not depict lexical tones, as in Chinese). The largest interval between F0 and the high tone covers the first two words. The sentence ends on a mid tone M,

acoustically located somewhere between L and H—here arbitrarily pegged at 7.2. Because it doesn't end on F0 we know that some affect is present, but what kind depends on the precise coincidence (or lack of it) between the frequencies of M&L and M&H. Sometimes pitches do not coincide, and the effect is neither bright nor dark—just awful. This is known as dissonance, and it is attested with feelings of anger, fear and/or confusion. It is thus possible to feel bad, yet still sound 'consonant' as in sadness.

How big can a chord be? In theory many pitches could be involved, but there are probably limits to how quickly they can be processed. Chords could also conceivably be formed *across* utterances and speakers, but little research has been done on this. To summarize then, the mental states of *feeling good* and *feeling bad*, as well as their nuanced variants, are predominantly signaled by tones (pitches) in isolation and/or combination. They lie at the very heart of tone-of-voice.

Rhythm

Tone-of-voice may also signal that the speaker is *thinking*, setting out the words in exactly the way that he or she intends. It doesn't necessarily inform the listener what the speaker is thinking about (although this could be known), just that they are thinking.

Rhythms arise when a repetitive pattern of strong vs. weak syllables is established ('strong' is the same as prominent, as when a syllable is stressed, enriched, heightened or prolonged.) It takes some talent to create a speech rhythm: speakers must select the right combination of words, each with its own inherent structure. In the examples below, syllables are represented as x's: small ones are weak (e.g. unstressed), capitals are strong (stressed). Strength is also indicated by boldface, so there are in fact two levels of prominence. [:] indicates length and [|] represents a break between intonation phrases (IPs).

(3) Rhythmic speech

a. pecha-pecha	[**X** x **X** x]
b. Thís is whát I sáw him dó	[**X** x **X** x **X** x **X**]
c. You cán't always gét what you wánt	[x **X** xx **X** xx **X**]
d. internátional státe of affáirs	[xx **X**xx **X**xx **X**]

Command of rhythm is a powerful tool that enhances the social position of the speaker.

In Japanese, *giongo* and *gitaigo* are inherently repetitive forms that draw the immediate attention of the listener (a). In (d), the rhythm is pronounced in a way that transcends word boundaries; more importantly, stressed syllables are evenly spaced over time. Because rhythms are known to the speaker milliseconds before they are pronounced, s/he can purposefully slow the tempo down to demonstrate the processing of thought. Doing so also provides extra time to add further nuance via tone or timbre.

Meter

Although rhythms are built on patterns of prominence, the number and organization of beats in an intonation phrase are also important. This is known as meter. Beats can be syllables (strong or weak), prominent syllables (as in rhythm) or morae (units of time). Speakers may signal *thinking* through a 'balanced' metrical structure where utterances consisting of two intonation phrases typically take the shape of a hat (cf. Bolinger 1986). In the first IP, pitch rises to a peak (the top of the hat) and in the second it steps back down. There is usually a brief pause between the IPs as well. If the number of beats in both IPs is the same, the structure is considered to be balanced. Here are some examples:

(4) Metrical sound models

a. You have to hear it in order to understand it

[x^1 x^2 x^3 $\mathbf{x}x^4$ | $xxxx^1$ $x{:}x^{23}$ $\mathbf{X}x^4$]

b. If it were up to me, I'd tell him to take a break

[xxx^1 $\mathbf{x}^2$ $x^3\mathbf{X}^4$ | $\mathbf{x}^1$ xx^2 xx^3 $x\mathbf{X}^4$]

c. I couldn't have put it any other way

[x^1 $\mathbf{x}x^2$ x^3 xx^4 | xx^1 xx^2 $\mathbf{x:}^{34}$]

d. If it isn't one thing it's another

[$xxxx^{12}$ $\mathbf{x}^3$ x^4 | x^1 $x^2\mathbf{x}^3x^4$]

In (a), the beats are arranged by morae, or timing units. IP-1 has four of them, the last of which spans two words spoken quickly. The morae in IP-2 are longer than those in IP-1, but equal in relation to each other.[1] The first mora spans four syllables and three words (*in order to*); the last three, four syllables and two words (*understand it*). The

[1] Strictly speaking, there is no universal mora length. In the literature, morae are considered as prosodic features, i.e. relevant to the phonology of a language, but nothing beyond that. It might then be more appropriate to describe timed units in IPs as 'tempo clusters'.

nasal coda of the syllable *un-* prolongs it slightly though, such that *under-* takes up two morae (*-stand* and the object clitic take up just one). Other pronunciations are possible of course, but this one is recognizable as a sign of *thinking*. The reader is invited to work out the meter of the other balanced examples.

Whenever possible, speakers tailor their delivery to maintain balance. While such structures may seem difficult to produce, they are in fact quite common, and listeners recognize them immediately. In balancing their speech, speakers demonstrate that they are socially 'engaged', or thinking.

V. Summary and conclusion

In this paper we have outlined a theory of tone-of-voice, first identifying its components, then matching them with components of meaning in the theory of NSM. The elements of tone-of-voice are not mysterious, but have never been organized in just this way. Although a large part of tone-of-voice is made of prosodic features and intonation, timbres, meter, and "tempo clusters" play a part as well. The elements that contribute to emotion words in NSM have never been considered with respect to sound, and herein lies the major innovation of our work. The following table sums up the links between the elements of sound and meaning:

(5) Mental/states activities and their audible exponents

Emotion meaning components	Tone-of-voice sound components
1. wanting/not wanting	A. voice qualities (F0, timbres)
2. knowing/not knowing	B. pitch movement: short tunes (melodies)
3. feeling good/feeling bad	C. pitch stance; pitch combinations
4. thinking	D. rhythm/meter/tempo

Much experimental work is needed to confirm or refute the links proposed above, which are mainly theoretical in nature; the overall picture is no doubt more complex than what is presented here. Useful data might be gleaned from a detailed description of the sound components in an online setting, e.g. with a five to ten second history of the same features for comparison. Subjects in a clinical setting might also report their feelings at the time of speech, elicited during playback. This will be the focus of future work.

References

Bolinger. Dwight (1986). *Intonation and Its Parts*. Stanford University Press.

Cook, Norman (2003). *Tone of Voice and Mind*. John Benjamins Publishing Co., Amsterdam.

Crystal, David (1975). *The English Tone of Voice.* Edward Arnold, London.

Laver, John (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.

Levitin, Dan (2007). *This is Your Brain on Music*. Plume (Penguin Group). New York.

Szczepek-Reed, C. (2006). *Prosodic Orientation in English Conversation*. Palgrave MacMillan, New York.

Tempest, Norton R. (1930). *The Rhythm of English Prose*. Cambridge University Press.

Wells, J.C. (2006). *English Intonation*. Cambridge University Press.

Wharton, Tim. (2009). *Pragmatics and Non-verbal Communication*. Cambridge University Press.

Wierzbicka, Anna (1999). *Emotion across Languages and Cultures*. Cambridge University Press.

Academic activities

During my years at Kobe Gaidai, I have read, slept, studied, taught, professed, conversed and learned. I have also engaged in many other things. In order to address my so-called academic achievements though, it is necessary to say something about the main events that shaped me. Here we go then.

I was born in the USA on November 25, 1952 (Showa 27), the son of a young professor at the university of Illinois. Later my family (mother and sister included) moved to Maine, way up in northern New England. Parts of Maine are so close to Canada that they're considered Indian territory. My father taught botany at the University of Maine in Orono, and at one point my mother started working at the library.

I attended public schools there and graduated from Orono High School in 1971. In the early years, there weren't many foreign languages to study — either French or Latin. French is spoken across the border in Quebec and everyone hated it (including me). Latin was (and still is) a dead language. In later years German was added to the curriculum, and I bonded to it right away. In 1970 the German club (to which I had been elected *Kanzler*) went on a school trip to Europe. The travel seed had been sown.

I went to the University of New Hampshire as an undergraduate. At first I majored in history and minored in German. It didn't take long to reverse these preferences. In 1972-3 I attended the University of Salzburg in Austria as an exchange student. I graduated from UNH with a BA in German Language and Literature in 1975.

I was accepted at the Monterrey Institute of Foreign Studies and Middlebury College graduate schools to study German, but joined the US Peace Corps instead. They sent me to South Korea to teach English at a middle school in the countryside and organize local workshops for English teachers. I lived there for two years, 1975-77. Such was my introduction to "the Orient" (westerners use this word to mean any country in Asia; it is an anachronism).

After the Peace Corps I spent one year doing nothing, then came to Japan to teach English (I had spent some time visiting friends in Tokyo while working in Korea). This lasted for about a year and led to my decision to study linguistics. Thus began my studies at McGill University in Montreal, Canada.

At the Department of Linguistics I studied almost everything, but ended up writing an MA thesis in morphology: "A unified analysis of the English suffix *-ing*." The secretary who typed the final manuscript became my wife (a native speaker of French), and we had one son together. I continued in the Ph.D program and wrote my dissertation in 1992, "A movement theory of ergativity."

During the McGill years (1979-92) I was a teaching assistant, a research assistant, and eventually a course lecturer for introductory linguistics and syntax. After graduating, our small family moved to Hawaii where my wife continued to work, our son went to school, and his father searched for academic positions. Eventually he found one as a Visiting Professor at Kobe University. That lasted for two years, after which I came to Kobe Gaidai.

When I entered graduate school I first became interested in phonetics but quickly abandoned it for phonology because I liked the challenge of solving phonology problems. The phonologist at McGill, Glynn Piggott had done research on Amerindian languages, but when he became interested in morphology, I did too. The Linguistics Dept. at McGill is very theoretically-oriented, a satellite of M.I.T. and the generative enterprise. All of my syntax teachers had been Chomsky's students, so I was pulled into that maelstrom. One of them — Lisa Travis — worked on Austronesian languages and I found them interesting too. Eventually Mark Baker became my graduate supervisor, and he encouraged me to continue work on Chamorro (spoken on Guam),

Palauan and Tagalog. Amerind never disappeared completely though; my first published article was ‘The conjunct order in Algonquian’. Amazingly, it’s still the one most people want to read.

For the next 10 years (about 1995 to 2005) I wrote articles and gave talks on the syntax and morphology of Austronesian and Amerindian languages. Most of this research was carried out in the Principles and Parameters version of generative theory, but eventually that changed to the Minimalist Program. I was not ‘on board’ with minimalism, which meant re-working much of the material that had been analyzed before. At the same time, my fieldwork on Algonquian languages (some are spoken in Maine) led to some interesting things beyond the syntax, i.e. in the discourse. Little things, like discourse particles. Clearly these were governed by the context, and this is how I first became interested in pragmatics.

There are many ways to consider discourse particles — they seem to have multiple and ever-changing functions. I noticed right away that in real speech (never grounds for study in generative models) some of those tiny ‘words’ were little more than sounds, and that is how I wanted to study them. When a speaker begins his or her utterance with “Well”, the sound of the word alone can signal how they expect their message to be interpreted. Thus began my interest in tone-of-voice, which brings us up to date…

Farewell and acknowledgements

Overall, it has been a pleasant ride down the Gaidai road, not always straight but never very bumpy. This is to say that the university offered a comfortable environment in which to work, staffed by notable faculty members, hardworking staff, and students of every stripe and color. To all the members of this community I offer heartfelt thanks and wish you well.

There are a few names that stand out, but any omission should not be taken seriously. Wada-sensei (now retired) first brought me to Gaidai with some timely help from David Farrah and a never-ending welcome from Montserrat Sanz. Tagawa-san became my 保証人 while others made me laugh: Yamaguchi-san and Tatsuki-san come to mind. Kudos for putting up with me to Mitsunaga-san, Honda-san, Ueda-san, Qin-san and Edgar Franz, my 8th floor neighbors. Most members of the Ei-bei gakka are just plain cheerful: Saeko, Midori, Hitomi, Atsuko, Noriko, and Kyoko; Chuta, Jun, Kensei and Kaz; Henri, Adam and Matt. Former members of the department cannot be forgotten: Iyeiri-san, Okada-san, Oshima-san, Tsujimoto-san, and even Sugayama-san

who had always treated me kindly. Within the university, let's not forget Tamai-san, Okamoto-san, Fukushima-san, Mishima-san and Laurie! Nasu-san kept me sharp in matters of zen and syntax; Zamma-san was a best friend and colleague — I dedicate this paper to him.

Abstract

This paper is about tone-of-voice, a familiar aspect of spoken language, but a very complex array of different sound types. Its primary function is to signal the changing inner states of the speaker pertaining to knowledge and emotion. The latter are themselves complex, however. Borrowing from Natural Semantics Metalanguage, it is proposed that the meaning 'parts' of emotions match up with specific sound-types. 'Wanting', for instance, is a meaning component best conveyed by voice quality—much akin to musical timbre. 'Knowing' (or not knowing) is communicated via pitch movement, whereas 'feeling good' and 'feeling bad' are usually signaled by single pitches or pitch combinations. The leading indicators of 'thinking' are rhythm, tempo and metrical structure.

Keywords : tone-of-voice, emotion words, voice quality, rhythm, meter