The theme of this issue is “emotion in speech,” a topic in phonetic studies that is currently gaining popularity although still not rigorously defined. Professor Hiroya Fujisaki has suggested making a distinction between linguistic information (i.e., discrete categorical information which can be explicitly represented by the written language or uniquely inferred from context) from paralinguistic information (discrete and continuous information added by the speaker to modify or supplement the linguistic information) and non-linguistic information (information not generally controlled by the speaker, such as the speaker’s gender, age, etc.). This would also include “emotion,” both spontaneous and acted, even though with the later, the actor controls the expression of emotion.

The articles in this issue fall generally into the category of paralinguistic information or non-linguistic information. However, as more studies come forth about what is conveyed in the speech signal to the listener, more questions arise about how to define what we are looking at. The introduction that follows is intended to encourage thinking about how to fine tune our definitions in order to explore this large, difficult and very important area of “emotion” in speech.

The first article in this issue, by Nick Campbell and Donna Erickson, talks about non-verbal affective information and what people hear when they listen to various productions of the (Japanese) interjection “eh” as uttered by a single speaker in conversations, with both the same and different conversation partners. The interjection varies considerably in terms of its prosodic manifestations of pitch, duration, loudness, and voice quality, and listeners assign about 30 “affective-type” labels to this single expression. A principal component analysis shows that about 16 labels can be used to describe the speakers’ percepts; moreover, the percepts tend to group according to valences (positive or negative affect) and strengths (active or passive affect). In terms of a definition of paralinguistic information, it is not obvious what is “discrete and continuous.”

The study by Campbell and Erickson also asks the question how the listener’s language background affects perception of non-verbal affective information. Somewhat surprisingly, the results suggested a consensus among Japanese, Korean and American English listeners in assigning affective labels to “eh,” with, of course, some interesting exceptions. Is language background of the speaker/listener best defined as part of paralinguistic information or because it cannot be controlled, as part of nonlinguistic information?

The next article by Toshiyuki Sadanobu examines how a specific voice quality, i.e., pressed voice, is used to let the listener know that the speaker is experiencing suffering or admiring. Use of pressed voice makes the suffering or admiring more sincere, and in this sense it can be said to convey paralinguistic information; yet, Sadanobu also convincingly argues that pressed voice is an expression of an action itself which can only be made by the sufferer or admirer. Does this belong to the same category...
as “emotion,” since the speaker is undergoing an experience while speaking, and as such conveys non-linguistic information?

The third article is by Caroline Menezes and examines the articulatory changes, specifically jaw movement involved in producing syllables and phrasing of utterances by speakers expressing irritation. Using jaw movement data from X-Ray Microbeam recordings from 4 speakers of General American English, she derived syllable and boundary magnitudes as well as syllable time values for each utterance. She reported a change in rhythmic pattern for irritated vs. non-irritated utterances, but also noted considerable speaker-dependent effects. This study perhaps deals with non-linguistic information in that it examines the articulatory underpinnings of expressions of emotion.

The first three articles in this issue provide landmarks for developing methods for collecting acoustic and articulatory data with colloquial, non-acted speech utterances. However, acted speech also provides important information about phonetic characteristics of speech. The fourth article in this issue, by Miho Teshigawara, describes a method of auditory analysis (based on the models developed by John Laver and Klaus Scherer) to study vocally expressed emotions and stereotypes of heroes and villains in Japanese animated movies. She reports that the auditorily critical vocal component that differentiates heroes’ and villains’ voices is the epilaryngeal setting. The majority of heroes’ voices had absence of pharyngeal constriction and presence of breathy voice while the majority of villains’ voices had laryngeal sphinctering and pharyngeal expansion. This paper deals with non-linguistic information, since it describes the character of the speaker, supposedly not changeable by the speaker. The techniques used in this study can be used to analyze colloquial speech also. An advantage of the technique is that it is non-invasive, and as such, substantial amounts of speech data can be analyzed. But it is important that corollary physiological and acoustic studies be done to substantiate the auditory deductions performed by the trained listener.

The final paper is by Kyoko Sakuraba in which she examines vocally-expressed emotions (happy, sad, angry, and neutral) as produced on the utterance “pikachu” elicited from Japanese and American children looking at different pictures of the cartoon character “pikachu” portraying happy, sad, angry, and neutral faces. She reports an interesting interplay between linguistic and non-linguistic information in that the dynamic range of F0 was found to be similar in the productions of the Japanese and American children, but the length of syllables varied, reflecting language-specific characteristics.

Methods in collecting data, effects of language and cultural background of speakers and listeners, analyses from various standpoints of perception, audition, acoustics, and articulation of “emotion in speech” are a few of the components of a challenging and important goal for understanding what speakers do and what listeners hear during spoken interchanges. This special issue is but a tip of the iceberg in this burgeoning area of study about “emotion in speech,” the interplay between linguistic, paralinguistic, non-linguistic information, and how speakers and listeners interact using spoken signals. The articles in this issue challenge us to continue to define what we are looking at, and to continue to collect more and more, and yet more, spoken utterances for the purpose of phonetic (auditory, acoustic, articulatory and perceptual) analyses. Why? To communicate more effectively using spoken language.