CHAPTER 15

Voice and Affect in Speech Communication

Geneviève Caelen-Haumont and Branka Zei Pollermann

Abstract

After clarifying the semiotic status of vocal indicators of affect, we trace the role of affective components of speech communication from ancient oratory up to the present. Specific attention is given to the often neglected contribution of linguists to the study of emotional components of speech (the contribution very much present already at the turn of the 19th–20th centuries). Recent approaches to the study of both cognitive and affective components of speech prosody are also described. MELISM, a tool for automatic segmentation and coding of prosodic features of speech, is presented at the end.

The Semiotic Status of Vocal Indicators of Affect: Vocal Expression or Vocal Communication of Affect?

All observable changes in human behavior (regardless of the type of sensorial modality) can be informative of the state of the individual. However, most of such changes are not produced in order to inform other members of the species of the person's psychological or physical state. From the perceived features, the observer infers the meaning without explicit learning of the code involved. Such features are informative without having been produced to this effect. By contrast, those features that *are* produced in order to transmit information have a communicative function and are conventionalized. Wharton (2003) illustrates this distinction by comparing shivering with smiling. Shivering is a spontaneous behavior whose function is to generate heat. To an observer, shivering provides information that the individual is feeling cold. However, the function of shivering is not to signal this information. Smiling, by contrast, could be considered as intentional signaling, whose function *is* to convey information (Ekman, 1999). As Wilson and Wharton (2005) write:

> For instance, a speaker's mental or physical state may affect the prosodic properties of her utterance, enabling a hearer with the appropriate experience or background knowledge to infer whether she is drunk or sober, sick or healthy, tired or alert, hesitant or assured. As with shivering, these prosodic properties carry information about the speaker's mental or physical state, but it is not their function to do so: they are natural signs, interpreted by inference rather than decoding. (p. 1561)

In everyday speech communication, the interpretation of speech signal relies heavily on the information provided by the shared knowledge that is part of the context (concrete circumstances and culture). In the interest of the economy of communication, the spontaneously produced information found in the speaker's vocal and facial behavior is part of the shared contextual knowledge that serves as interpretative framework —the context that provides the information not contained in the verbal message itself. This information is specifically geared at guiding the hearer during the inferential phase of auditory comprehension. In this process the speaker may also use nonverbal signals intentionally in order to facilitate access to meaning, to divert the hearer from a particular interpretation or even mislead him.

Unlike in typewritten communication, in oral communication the sender of the message produces his own tool for communication. By this fact the sounds produced *on-line* necessarily carry the "signature" of their producer. The information thus provided is related to the speaker's physical conditions at the moment of speech production, as well as to his or her cognitive-affective states such as attitudes, interpersonal stances, emotions, or personality traits. Such information can be produced spontaneously or intentionally and may vary according to social context and the speaker's communicative intentions. Some vocal features can thus be interpreted as the expression (symptom) of the speaker's inner state while others will be considered as intentional communicative signals or styles (Zei-Pollermann, 2002).

Voice and Affect in Ancient Oratory

Since antiquity, the study of emotional aspects of speech has been part of philosophical, religious, literary, and artistic works. It was only later that affective speech became a subject in its own right, investigated by psychologists, linguists, neurologists, physiologists, and other scientists. Going back to Aristotle, in Book 3 of *Rhetoric*, Aristotle points out the importance of studying the subject of verbal expression and delivery, which he felt was a powerfully effective means of conducting good oratory. Aristotle writes (Russel & Winterbottom, 1972, p. 135): "This study is about the proper use of voice (loud, soft, and moderate, to express individual emotions), the proper use of accents (acute, grave and circumflex), and the rhythms appropriate to different things." While for Aristotle the whole study of rhetoric is directed towards producing belief (that is, not knowledge), for Cicero (Russel & Winterbottom, p. 250), the purpose is triple: "For the best orator is the one, who by his oratory instructs, pleases, and moves the minds of his audience. To instruct is a debt to be paid, to give pleasure a gratuity to confer, to rouse emotion a sheer necessity." At that time, and for many years to come, emotional effects of voice on the listener played a major part in oratory. Emotional effects are produced for a purpose and the delivery is one of the means.

In this context the speaker's true emotions are not the focus of attention. Only the emotional impact created on the hearer is clearly pointed out. The orator's genuine feelings are mentioned only indirectly in the context of their usefulness in making the orator appear more credible and creating the impression of naturalness for "Naturalness is convincing, artificiality the reverse. . . . One should aim at the effect attained by Theodorus' voice in comparison with other actors'; his seems to belong to the character, theirs to be imposed on it" (Russel & Winterbottom, p. 137). The instrumental use of oratory to elicit emotions in others is well described by Cicero (p. 217) in the following paragraph of his text *The Brutus:*

> The orator's audience believes his words, thinks them true, assents, approves; his speech carries conviction, . . . the crowd rejoices, grieves, laughs, cries, likes, dislikes. . . . It is angered and soothed, it hopes and fears. These effects take place according to the way in which the minds of those present are worked on by words, thoughts and delivery.

Cicero is aware that passions induced in the audience by oratory modify people's judgment so that eventually, affect dominates over truth and justice.[1] A very similar view was developed by Quitilian who adds that vocal aspects of delivery can reveal the unadmitted passions of the speaker.

It appears that Cicero's oratory excellence includes induction of two kinds of

[1] *De Oratore*, II. XLII. 178: Haec properans ut et apud doctos et semidoctus ipse percurro, ut aliquando ad illa maiora veniamus: nihil est enim in dicendo, Catule, maius, quam ut faveat oratori is, qui audiet, utique ipse sic moveatur, ut impetu quodam animi et perturbatione magis quam iudicio aut consilio regatur: plura enim multo homines iudicant odio aut amore aut cupiditate aut iracundia aut dolore aut laetitia aut spe aut timore aut errore aut aliqua permotione mentis quam veritate aut praescripto aut iuris norma aliqua aut iudici formula aut legibus.

emotions: esthetic or disinterested and utilitarian (Scherer, 2005). The role of utilitarian emotions (such as fear or anger) is mainly to steer the receiver's behavior presumably through empathy or identification with the speaker.

The role of disinterested emotions would be to provide esthetic experience or disinterested pleasure (Kant, 2001) resulting either from intrinsic qualities of sensorial experience (visual or auditory) or from the pleasure of intellectual discovery of the truth. For Cicero the delivery skills do not only include the voice, they also include the orators' facial expressions and body movements, which "signify in what sense everything they say is to be understood" (Russel & Winterbottom, p. 242). This statement from Cicero clarifies the role of nonverbal aspects of communication, which is to guide the interpretation of the words by providing the psychological context (in terms of knowledge and affect). It is nevertheless important to keep in mind that in ancient oratory the main focus was on use of language and of the argument, while the delivery played a role of providing the right emotional context for winning an argument. Voice and feelings continued to be inherently related to each other in the centuries to come. St. Augustine considered the human voice as capable of stimulating the soul and producing physical pleasure as the voice is considered to contain the whole range of the feelings of one's soul: "*tutta la scala dei sentimenti della nostra anima*." (St Augustino, 397/1969). Following St. Augustine, the medieval scholastic tradition considered the speaker's voice as emanating from the soul and the rhetoric figures as superfluous because the preacher's strong emotion (*grandis affectus*) provided all the force to his discourse.

It is interesting to note that in medical writings on vocal disturbances, the voice was also regarded as a reflection of thought activities in general and thought disturbances in particular (Paparela, 1556).

Rhetoric's reputation suffered during the Age of Enlightenment, when it was condemned as meaningless bombast or unwelcome ornamentation—the view in agreement with the Enlightenment Movement that advocated rationality as a means to establish an authoritative system of ethics, esthetics, and knowledge.

In conclusion to this very sketchy presentation of the link between voice and emotion in ancient oratory, we must point out that voice was considered to have an emotion eliciting effect only in conjunction with the words and other nonverbal expressions.

The Study of Affect in Everyday Speech Communication: The Contribution of Linguists

Notwithstanding Darwin's (1872) impetus on research of emotional expression, the study of vocal indicators of affective states remains mainly in the domain of linguistics and literature. For example, Bourdon (1892) surfaces as the first one to engage in empirical study and theoretical modeling of the effects of emotions on speech communication. His work describes in quite some detail the effects of emotions and emotional tendencies on vocal intensity, pitch height, timing of speech, pauses, intonation contours, and accentuation. He defined the term *ten-*

dencies as a person's stable patterns of emotional reactions including behaviors; it is close to the concept of temperament.

At the beginning of the 20th century, as it is well known, Ferdinand de Saussure's teaching brings a decisive contribution to the constitution of linguistics as an independent science. The linguists' debate focalizes on the relation between the subjective, ever-changing aspects of speech vs. the stable and conventionalized aspects of language. Charles Bally (1905) studied the affective components of speech on both semantic and prosodic levels, and laid the foundations for the stylistics of ordinary spoken language. In the wake of Bally's stylistics, Troubetzkoy (1939)—an active member of the Prague linguistic circle—proposes a term *phonostylistics* for a science whose aim would be to study two domains of speech: (a) expression, as a symptom of the speaker's state and (b) appeal, as a set of conventionalized vocal features fulfilling specific social functions. His proposal is based on Buhler's (1934) *organon* model of the semiotics of speech communication. Troubetzkoy concedes that it is not always easy to separate the spontaneously expressed features related to the speaker's psychophysiological state from those used intentionally for purposes of communication—like for example the use of an accent characteristic of a social class.

The study of phonostylistics continued in the works of Spitzer (1961) as well as Dámaso and Bousoño (1951), who classified speech styles according to the types of affect expressed.

Among the first empirical studies were those of Seashore (1927), Fairbanks and Pronovost (1939), as well as Fairbanks and Hoaglin (1941). The subsequent works of Williams and Stevens (1972), Bolinger (1945; 1946; 1965; 1972), and Crystal (1975) were fundamental. Their aim was to establish a repertory of acoustic features related to various emotional states. Fónagy (1971) makes a considerable contribution to the study of affect in speech. He suggests a model of *double coding* where speech communication involves two successive coding procedures: (a) linguistic coding by which an idea is transformed into a speech signal composed of a sequence of phonemes (e.g., "It's very kind of you") and (b) paralinguistic coding, which adds emotional components to the phonemic sequence (e.g., the feeling of hate or contempt). Fónagy's rich work on the prosody of emotions also includes experimental comparative studies of emotional prosody and music (Fónagy & Magdics, 1963).

Within the framework of phonostylistics and by analogy to the concept of a phoneme, Pierre Leon (1970; 1971; 1976) coined the term *phonostyleme,* defined as a bundle of distinctive features relevant for the vocal differentiation of emotions. These features were mean F_0, F_0 range, intonation contour, tempo, and vocal intensity. None of the features was considered to be able to characterize an emotion in an independent manner. Faure (1970; 1973), Fónagy (1973), I. Fónagy, Bérard, and J. Fónagy (1982), Fónagy and Sap (1997), and Fónagy (1982a, 1982b) contributed substantially to research on the acoustic patterns of emotions.

A frequent problem in social sciences is lack of consensus on the labeling of key concepts related to the object of the study. Drawing on the works of Crystal and Quirk (1964) as well as that of Laver (1968), Roach (2000) proposed a rather

comprehensive prosodic and paralinguistic labeling, comprising 36 labels. The prosodic coding presents four characteristics related to F_0 height and range, four for intensity, and eight for tempo. In addition, the paralinguistic code includes nine labels for voice quality, five for speech fluency, and three for expressions such as laughing, crying, and tremulous voice.

In the past 20 years many studies have tried to determine the acoustic patterns of emotions (Zei and Archinard, 2001). Prosodic features listed by the various authors overlap considerably. Table 15-1, elaborated by Murray and Arnott (1993, Table 1, pp. 1106), describes the main prosodic features in the English language for five basic emotions. These features seem to be applicable to many other languages.

Other numerous works were undertaken within the broader framework of linguistic or semantic studies. Notable research was done for Spanish and American English by Bolinger (1945; 1946; 1965; 1972), for British English by Crystal (1975), and for French by Faure (1970; 1973), Fónagy and Bérard (1973), I. Fónagy, Bérard, and J. Fónagy (1982), as well as Fónagy and Sap (1997).

These studies were characterized by a search of acoustic profiles based on the quantification of F_0, intensity, timing, and pace parameters related to the speakers' emotional states and attitudes.

For example Fónagy and Bérard (1973) and Fónagy et al. (1982) studied melodic

Table 15–1. Summary of human vocal emotion effects. The effects described are those most commonly associated with the emotions indicated, and are relative to neutral speech.

	Anger	Happiness	Sadness	Fear	Disgust
Speech rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Voice quality	breathy, chest tone	breathy, blaring	resonant	irregular voicing	grumbled, chest tone
Pitch changes	abrupt, on stressed syllables	smooth, upward inflections	downward inflections	normal	wide, downward terminal inflections
Articulation	tense	normal	slurring	precise	normal

Note. From "Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion," by L. R. Murray and J. L. Arnott, 1993, *Journal of the Acoustical Society of America, 93*(2), pp. 1097-1108. Copyright 1993 by American Institute of Physics. Reprinted with permission from L. R. Murray and J. L.Arnott.

clichés of Parisian French related to different speaking styles. Fónagy (1982a; 1982b) specialized in studying various speaking styles such as storytelling, reading the daily news, and expressing attitudes and emotions.

Bolinger studied prosodic systems in various languages as of 1945. His position about an inherently paralinguistic emotive role of intonation is well illustrated in the following statement: " . . . while intonation may well play a grammatical role, it does so by virtue of its emotive power; intonation is supportive where grammar is concerned, but it is not definitional: there is no intonation that is the property of any one grammatical category" (Bolinger, 1986, p. 13). Similar findings were reported by Faure (1970; 1973) and Caelen-Haumont (1978/1981). Bolinger (1986) showed that accent or intonational prominence simultaneously signals the speaker's interest that a given word has for him (*interest accent*) and his emotional involvement such as earnestness (*power accent*). The latter type of accents obeys the principle of climax that explains why a final accent tends to be a strong one and why we so often encounter right-shifted accents, for example, "At Putnam dodge in Burlingàme" (Bolinger, 1986, p. 14). Conversely, if the speaker wants to tone down his utterance to reduce its power, he may move the word accent leftward: "It's bigger than Ténnessee" rather than "It's bigger than Tennessée." Bolinger also convincingly demonstrated his view that "melody expresses other forms of arousal, especially those of sustained feeling, and its opposite, rest" (p. 15). For him, it is much more important to apprehend the fundamentally emotive and metaphorical character of intonation denoting meanings, such as pacification, prompting, reprimanding, irony, deep concern, than to "pursue it down the byways of syntax and morphology" (p. 20).

The above-mentioned meanings of intonation are conceptually close to Scherer's set of design features of affective states (Scherer, 2004) such as moods, interpersonal stances, or personality traits. They also correspond to what Caelen-Haumont and Bel (2000) termed *ordinary emotion* states related to beliefs, values, and subjective feelings.

Empirical findings related to prosodic cues of emotions and attitudes are found in the work of Léon (1970), who presented evidence that high register and wide F_0 range were associated with joy, self-consciousness, and lightness, while a low register with a narrow F_0 range were associated with sadness, confidence, and graveness. The contour type was not considered to be discriminating by itself, whereas vocal intensity was proportional to the strength of the expressed feeling.

Much of the research done in this field was characterized by association of F_0 parameters to different phrase modalities. O'Connor and Arnold (1973), for instance, found that "wh-questions" realized with a *high drop* were perceived as *vivid, businesslike, not unfriendly, lively,* or *interested*. For Halliday (1994), wh-questions with a rising tone characterize a continuity and are tentative, whereas yes-no questions with a falling tone are peremptory. He also found that declarative sentences with a high rise contour could be perceived as "challenging, aggressive, defensive, or indignant."

Scherer and colleagues' "configuration" model (Scherer, Ladd, & Silverman, 1984) also explored the relationship between the intonation contour and the type of sentence. The authors found that the reversal of standard English patterns

(i.e., final pitch lowering for wh-questions, and final pitch rising for yes-no questions) produced negative connotations. As Wichmann (2002) pointed out, deviations from conventional tonal realizations may generate negative attitudes. In general, one notices that deviations from vocal stereotypes can often be perceived as vocal markers of affect. Wichmann also studied the influence of social context on affective prosody. She found that in private conversations among friends, the sentences ending with *please* have a final pitch rise, while final pitch fall is used in public situations demonstrating a display of social power or status. Wichmann concludes that the meaning thus depends to some extent on the power relationship between the speaker and hearer. These results are congruent with Ohala's *frequency code* (1983), which relates dominance to low-pitched and submissiveness to high-pitched voices.

Drawing on Ohala's frequency code, Piot (2001) investigated the prosodic expression of two cognitive parameters in French: *ignorance* and *desire to know*. He investigated whether the degree of pitch rise was related to the degree of *ignorance* and/or a *desire to know.* Using speech synthesis (by varying F_0 parameters, intensity, speech rate), he showed that (a) F_0 peak value was positively correlated with the degree of ignorance and the desire *to know or to inform*, (b) increased rate of delivery was related to the desire to know, but not to the desire to inform, (c) sentence final level of intensity was positively correlated with both cognitive parameters only in assertions, but not in questions. The experiment also revealed the importance of vocal range and the final pitch contour in the perception of attitudes, as well as the role of the tonal height in signaling novelty.

In the wake of Scherer's models of covariance and configuration, Ladd and colleagues (Ladd, Silverman, Tolkmitt, Bergman, & Scherer, 1985) undertook three experiments where subjects were to assess the emotion carried by statements in which the pitch range, the sentence contour type, and the voice quality were systematically changed. The results showed that pitch range and voice quality had the strongest influence on the subject's inference of the speaker's arousal and on the inference of cognition-related attitudes at the same time. The contour type alone was not significantly related to the classification of attitudes, but rather the classification of emotions. An important conclusion was drawn from these experiments, namely, that the three prosodic cues, pitch range, contour type, and voice quality, all function independently of each other.

Kehrein (2002) studied the relation between vocal parameters and emotional dimensions of valence, arousal, dominance, and unexpectedness. The latter dimension is known to be related to an increase in F_0 maximum (Caelen-Haumont, 1991, to be published). Based on intersubjective attribution of categorical emotional and attitudinal meanings, Kehrein measured the acoustic features of utterances whose speakers were judged as *excited/agitated, uncertain, eager or angry, calm/relaxed, content, delighted, uncertain/perplexed, apologetic, resigned, frustrated, and disappointed.* His findings confirm a *compositional approach to emotional meaning* because individual acoustic parameters contribute to the constitution of a variety of perceived emotions and attitudes.

As the attribution of emotional meaning is dependent upon the verbal elements and the context of the interaction, Kehrein concludes that assessment of vocal expression of emotions can only be done in context.

We support the above conclusion in that the vocal aspects of emotions should be considered within the whole interactive context including the language, the concrete material context, and the speaker's nonverbal displays. As communication is a multichannel process, the contribution of each channel is a function of simultaneous presence of all the others.

Melism as Expression of the Speaker's Subjective Emotional Space

Prosody appears to be a particularly suitable means of expressing the subjective dimension of speech communication. Subjectively tinted prosody is a precious source of contextual information that helps disambiguate the interpretation of meaning.

As pointed out by Caelen-Haumont and Bel (2000), in the prosody of French language, the syntactic and pragmatic functions of intonation usually acquire normative strength leaving little space for the speaker's expression of his or her individuality. Nevertheless, the speaker's prosodic *decision latitude* can be exercised locally on the level of word prosody, mainly in lexical but also in grammatical words. Caelen-Haumont termed such prosodic marking as *melisms.* The notion of melism has been borrowed from the domain of singing and refers to pitch excursions (melodic movements) spread over the duration of the word, such that the number of notes perceived is higher than the number of syllables in the word. Melisms can coincide with syntactic and/or syntagmatic structure of the intonation, just as they can diverge significantly from the canonical intonation contour. The divergence is either local—on a single word or a syntagmatic unit—or it can spread over several units. Melisms are considered to be a prosodic reflection of the subject's basic affective states and attitudes such as doubts, beliefs, or value-invested thoughts as well as emotions. Such basic underlying affective states are conceptually close to Russel's *core affect* (Russel, 2003), where various shades of emotions are conceptualized as departures from a neutral point on two bipolar (valence-arousal) axes of his dimensional model of emotions. When projected into the prosodic space, Russel's arousal dimension appears to have an acoustic counterpart in the so-called *effort code* (Chen, Gussenhoven, & Rietveld, 2002) whereby the effort expended on speech production is proportional to the span of pitch excursions (de Jong 1995), the steepness of their rising or falling slopes (Caelen-Haumont, 1991, to be published; Caelen-Haumont & Bel, 2000), and the magnitude of intensity peaks. The effort code thus appears to reflect the speaker's personal involvement, which in turn is prosodically manifested in melisms realized on the lexical items relevant to the speaker's interaction with the hearer. Melisms can temporarily disturb the cohesion between intonation and the linguistic structure, but the underlying syntactic organization can allow the speaker to take this liberty and mark the utterance by his personal touch.

MELISM—An Automatic Method of Segmentation and Melodic Coding of Speech

To allow accurate descriptions of melodic salience, an automatic analysis tool, MELISM, was developed by Caelen-Haumont and Auran (2004; 2005). MELISM procedure is based on Praat software (Boersma & Weenink, 1996) and allows automatic detection of melisms, their segmentation into *tonal syllables*, and their positioning on a nine-level scale based on a stylized F_0 curve generated by MOMEL[2] (Hirst & Espesser, 1993). The tonal syllables are mono- or bitonal sequences obtained at the points of change of melodic slopes. MELISM requires (a) a preliminary segmentation of the signal into linguistic units considered as relevant (e.g., single words or prosodic words), and (b) a stylization of the F_0 contour by determining target points with MOMEL algorithm.

Table 15-2 illustrates the nine-level scale based on Delattre's four-level model (Delattre, 1966) and obtained by dividing the space between each of the four levels into three segments. The nine-level scale is expressed by the following symbols: a = Acute; s = Supra; h = High; s = Elevated; c = Central; b = Bottom; I = Infra; g = Grave. The more acute ones (A, S or H) are involved in the definition of melisms.

To illustrate the MELISM procedure, two spontaneous speech samples are presented here, one expressing the emotion of joy and the other two attitudes.

Table 15–2. Matrix of tonal sequences describing the melodic configurations of words

	Melisms								
Tone	Acute a	Supra s	High h	Elevated e	Middle m	Central c	Bottom b	Infra i	Grave g
a	*aa*	*as*	*ah*	*ae*	*am*	*ac*	*ab*	*ai*	*ag*
s	*sa*	*ss*	*sh*	*se*	*sm*	*sc*	*sb*	*si*	*sg*
h	*ha*	*hs*	hh	he	hm	hc	hb	hi	hg
e	*ea*	*es*	eh	ee	em	ec	eb	ei	eg
m	*ma*	*ms*	mh	me	mm	mc	mb	mi	mg
c	*ca*	*cs*	*ch*	ce	cm	cc	cb	ci	cg
b	*ba*	*bs*	*bh*	be	bm	bc	bb	bi	bg
i	*ia*	*is*	*ih*	ie	im	ic	ib	ii	ig
g	*ga*	*gs*	*gh*	ge	gm	gc	gb	gi	gg

[2]MOMEL (Hirst & Espesser, 1993) allows stylization of fundamental frequency contours as a combination of their macromelodic and a micromelodic components. This is assumed to correspond to the global pitch contour of the utterance, which is continuous and independent of the nature of the constituent phonemes. It corresponds approximately to what we produce if we hum an utterance instead of speaking it.

Figure 15-1 presents the analysis of a spontaneous speech sample evoking a joyful childhood memory: *"Cela m'a marquée, petite (rires) je me rappelle le grenier (rires) / It struck me when I was little (laugh), I remember the attic (laugh)."* One notices that the F_0 contour evolves in the upper part of the speaker's register relative to her mean F_0 (170 Hz), with tonal peaks reaching high values: a, s, and h.

Figure 15-2 displays the analysis of an utterance expressing controlled irritation and irony: *"Non, c'est des matières sur l'étude de l'agronomie (pause), comme c'est intéressant . . . / No, they're courses in agronomy study (pause), how interesting . . . "* The highlighted part of the utterance expresses *"controlled irritation."* The latter is related to the fact that she was obliged to study agronomy despite her wish to pursue economic studies. One observes a sequence of F_0 rises—each marking a lexical word. Levels extend from *b* to *h*, mostly between *b* and *c*. The segment after the pause expresses *irony*. It is characterized by narrower F_0 range: between *b* and *m*, with an initial plateau at level *b*, which gives it a rather flat contour that is in contrast with the semantic content of the utterance *"how interesting."*

We believe that prosody is characterized by the interaction of two contradictory forces: one related to the grammaticalization of prosodic contours congruent with linguistic structures, and the other related to local deviations expressing the speaker's personal affective state. This *freedom* is linked to the potency dimension of affectivity (see Zei Pollermann & Izdebski, Chapter 3). Indeed, it is in the act of speaking that the subject can assert his or her personal *numerical identity* as opposed to a *collective identity* related to his or her belonging to a group of language users.

That everyday speech carries information about the speaker's cognitive-affective state is well known and has been adequately documented in the works of linguists and psychologists for well over a century. Some of the earliest influential readings in this field were those of Steel, 1775; Bourdon, 1892; Bréal, 1897; Bally, 1905; Marty, 1908; Sapir, 1927; Seashore, 1927; Buhler, 1934. Most of these authors—just as ancient Greek and Roman orators—pointed out the joint action of nonlinguistic and linguistic aspects of speech in speech communication. They had a global view of a coordinated action of verbal and nonverbal channels of communication, including facial and body movements. Such parallel coding is inherent to the speech act, and can serve to regulate interpersonal relationships on the level of the interaction partners' emotions, their mutual status and role, and the felt success of their communicative efforts.

In this chapter we have tried to highlight the historic continuity of the role assigned to vocal affect in speech communication. We could thus say that the study of vocally communicated affect started within a global framework of interaction between verbal and nonverbal aspects of communication. In ancient rhetoric, the semantic components of speech were interpreted in relation to the emotional messages carried by the human voice. With the introduction of experimentally more rigorous methods often referred to as the *standard content paradigms* (Davitz, 1964), much of research on the communication of affect was stripped off the semantic component of speech, in spite of evidence that

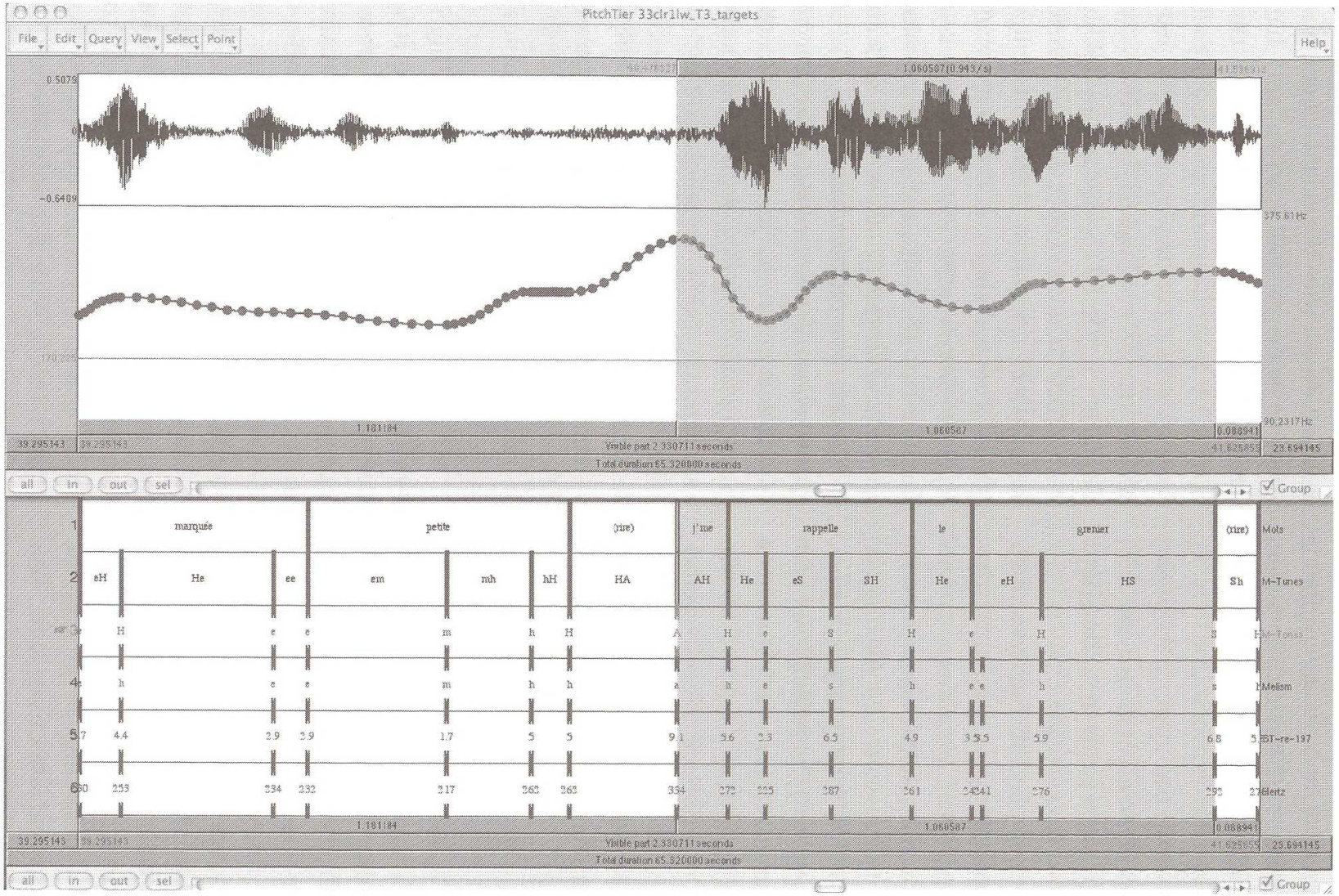

Figure 15–1. From top to bottom: Wave form of the sound; pitch contour generated with MOMEL (Hirst & Espesser, 1993); pitch contour with mean F_0 marked by a straight horizontal line; Tier1: text of utterance; Tier 2: melodic tunes of linguistic units; Tier 3: melodic tones—tonal targets of F_0 in semitones; Tier 4: alphabetic coding of melisms; Tier 5: F_0 in semitones relative to the subject's mean F_0 of 170 Hz; Tier 6: F_0 in Hertz.

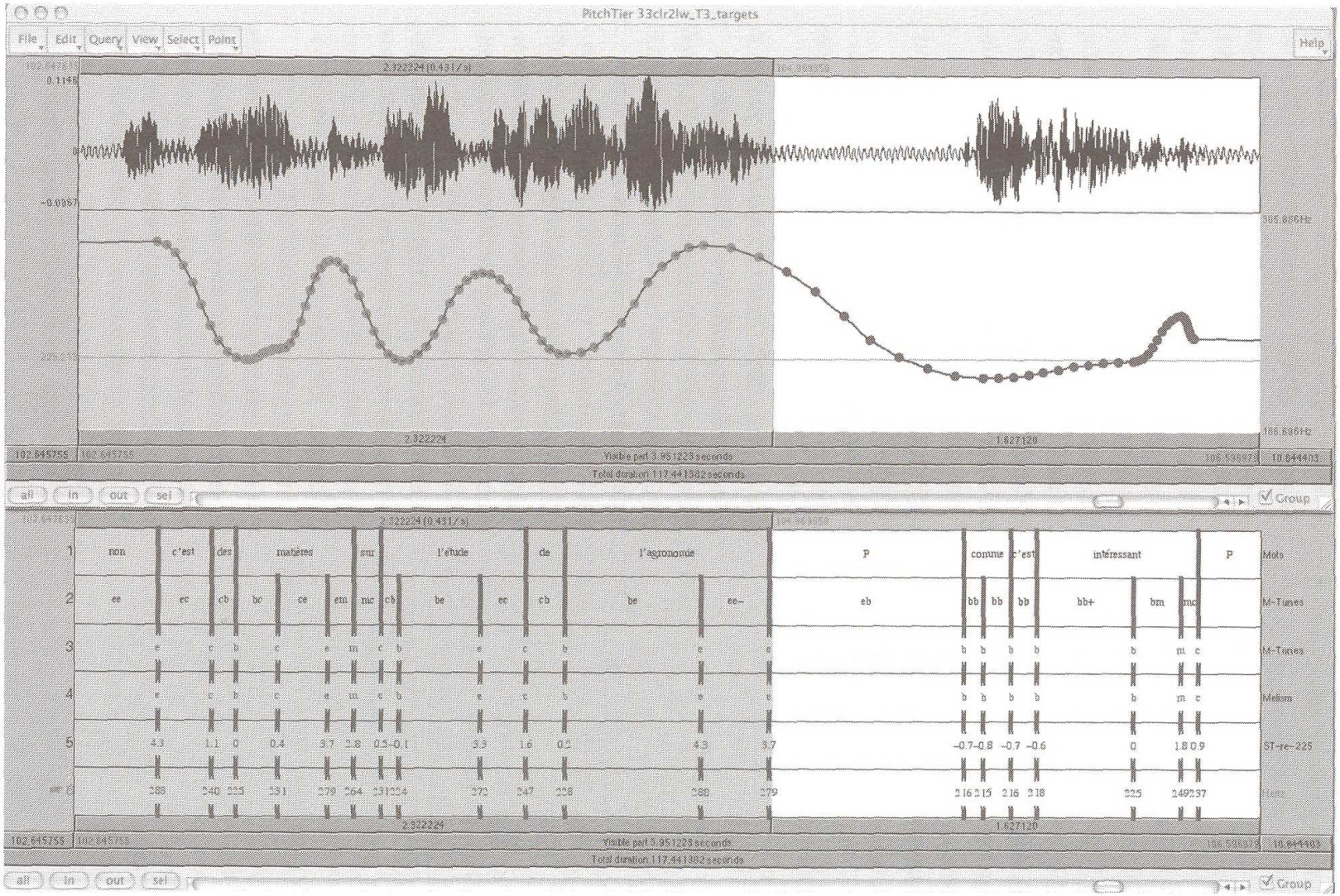

Figure 15–2. An excerpt from the spontaneous speech sample expressed with restrained irritation followed by irony: "Non, c'est des matières sur l'étude de l'agronomie (pause), comme c'est intéressant . . . No, they're courses on agronomy (pause) how interesting . . . " From top to bottom: Wave form of the sound; pitch contour with mean F_0 marked by a straight horizontal line; Tier1: text of utterance; Tier 2: melodic tunes of linguistic units; Tier 3: melodic tones—tonal targets of F_0 in semitones; Tier 4: alphabetic coding of melisms; Tier 5: F_0 in semitones relative to the subject's mean F_0 of 225 Hz; Tier 6: F_0 in Hertz.

linguistic and nonverbal cues can contribute in an additive fashion to listeners' judgments of affect (Ladd et al., 1985; Scherer & Oshinsky, 1977). More recent research on attitudes (Wichmann, 2002) as well as the work of Kehrein (2002) illustrate the necessity of integrating verbal elements and social context of interaction into the research paradigms. Caelen's study of melisms is an example of how the attribution of affective meaning depends on both the context and the semantic content of the utterance. In the concluding paragraph of their chapter on vocal expression of emotion, Scherer and colleagues wrote:

> Apart from studying vocal communication process as a whole, it may also be time to drop the assumption of separate linguistic and nonlinguistic channels, together with the hermeutic separation of the respective research traditions. As we have shown, there is much evidence that a large part of emotion signaling in voice and speech is dually coded, in both linguistic and nonlinguistic features. Thus a rapprochement between researchers interested in expression and those interested in language, is highly desirable, as is a more intensive interaction between researchers studying vocal and facial expression, two research areas that have had little contact so far, even though they have a common origin in the underlying emotion, and are often interpreted as a Gestalt by a perceiver (Scherer, Johnstone, & Klasmeyer, 2003, pp. 451–452).

References

Bally, C. (1905). *Précis de stylistique: Esquisse d'une méthode fondée sur l'étude du français moderne.* Geneva, Switzerland: A. Eggimann.

Boersma, P., & Weenink, D. J. M. (1996). *Praat, a system for doing phonetics by computer.* Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam.

Bolinger, D. (1945). Spanish intonation, Review of Tomás Navarro, Manual de entonación española. *American Speech, 20,* 128–130.

Bolinger, D. (1946). The intonation of quoted questions. *Quarterly Journal of Speech, 32,* 197–202.

Bolinger, D. (1965). *Forms of English: Accent, morpheme, order.* Cambridge, MA: Harvard University Press.

Bolinger, D. (1972). *Intonation.* Harmondsworth, UK: Penguin Books.

Bolinger, D. (1986). Intonation and emotion. *Quaderni di Semantica, 7,* 13–21.

Bourdon, B. (1892). *L'expression des émotions et des tendances dans le langage.* Paris: Alcan.

Bréal, M. (1897). *Essai de sémantique (science des significations).* Paris: Fayard.

Bühler, K. (1934). *Sprachtheorie. Die Darstellungsfunktion der Sprache* (2nd ed.). Stuttgart, Germany: Gustav Fischer.

Caelen-Haumont, G. (1978). *Structures prosodiques de la phrase énonciative simple et étendue.* Unpublished doctoral dissertation. Université de Toulouse-le-Mirail, Toulouse, France.

Caelen-Haumont, G. (1981) *Structures prosodiques de la phrase énonciative simple et étendue.* Doctoral dissertation, Hamburger Phonetische Beitrage, band 34. hamburg Buske.

Caelen-Haumont, G. (1991). Stratégies des locuteurs en réponse à des consignes de lecture d'un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodiques. Unpublished doctoral dissertation, Université de Provence, Aixen-Provence, France.

Caelen-Haumont, G. (à paraître). *Prosodie et sens: une approche expérimentale.* Doctoral dissertation (doctorat d'état). L'Harmattan-Marges Linguistiques.

Caelen-Haumont, G., & Auran, C. (2004). The phonology of melodic prominence: The structure of melisms. In B. Bel & I. Marlien

(Eds.), *Proceedings of Speech Prosody 2004*. Nara, Japan, 143–146. France: Université de Provence.

Caelen-Haumont, G., & Auran, C. (2005). Manuel d'utilisation de la procédure MOMEL-MELISM sous Praat, 1–57. Retrieved 2005 from http://www.lpl.univ-aix.fr/~lpldev/MELISM/

Caelen-Haumont, G., & Bel, B. (2000). Le caractère spontané dans la parole et le chant improvisés: De la structure intonative au mélisme. *Revue Parole, 15–16*, 251–302.

Chen, A., Gussenhoven, C., & Rietveld, A. (2002). Language-specific uses of the Effort code. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 215–218). France: Université de Provence, Aix-en-Provence

Crystal, D. (1975). *The English tone of voice*. London: Edward Arnold.

Crystal, D., & Quirk, R. (1964). *Systems of prosodic and paralinguistic features in English*. The Hague: Mouton and Co.

Dámaso, A., & Bousoño, C. (1951). Seis calas en la expresión literaria española, Madrid, Spain: Gredos.

Darwin, C. (1965). *The expression of the emotions in man and animals*. Chicago: University of Chicago Press. (Originally published 1872)

Davitz, J. R. (1964). Auditory correlates of vocal expressions of emotional meanings. In J. R. Davitz (Ed.), *The communication of emotional meaning* (pp. 101–112). New York: McGraw-Hill.

de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 97*, 491–504.

Delattre, P. (1966). Les dix intonations de base du français. *French Review, 40*, 1–14.

Ekman, P. (1999). Emotional and conversational nonverbal signals. In L. Messing & R. Campbell (Eds.), *Gesture, speech and* sign (pp. 45–57). Oxford, UK: University Press.

Fairbanks, G., & Hoaglin, L. W. (1941). An experimental study of the duration characteristics of the voice during the expression of emotion. *Speech Monographs, 8*, 85–90.

Fairbanks, G., & Pronovost, W. (1939). *Speech Monographs, 6*, 87–104.

Faure, G. (1970). Contribution à l'étude du statut phonologique des structures prosodématiques. *Studia Phonetica, 3*, 93–107.

Faure, G. (1973). Tendances et perspectives de la recherche intonologique. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 5–29.

Fónagy, I. (1971). Double coding in speech. *Semiotica, 3*, 189–222.

Fónagy, I. (1982). Variations et normes prosodiques. *Folia Linguistica, XVI, 1–4*, 17–38.

Fónagy, I. (1982). *Vive voix. Essais de psychophonétique*. Paris: Payot.

Fónagy, I., & Bérard, E. (1973). Questions totales et implicatives. *Studia Phonetica, 8*, 53–98.

Fónagy, I., Bérard, E., & Fónagy, J. (1982). Les clichés mélodiques. *Folia Linguistica, 161*(4), 153–185.

Fónagy, I., & Magdics, K. (1963). Emotional patterns in intonation and music. *Zeitschrift fur Phonetik 16*, 293–326.

Fónagy, I., & Sap, J. (1997). Traits prosodiques distinctifs de certaines attitudes intellectuelles et émotives. *Actes des 8e Journées d'Études sur la Parole*, 237–246. France: Université de Provence, Aix-en-Provence.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Hirst, D., & Espesser, R. (1993). Automatic labelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix, 15*, 71–85. France: Université de Provence, Aix-en Provence.

Kant, E. (2001). *Critique de la raison pure*. Paris: Garnier Flammarion.

Kehrein, R. (2002). The prosody of authentic emotions. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 423–426). France: Université de Provence, Aix-en-Provence,

Ladd, D. R., Silverman, K. E. A., Tolkmitt, F., Bergman, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F_0 range in signalling speaker effect. *Journal of the Acoustical Society of America, 78*, 435-444.

Laver, J. (1968). Voice quality and indexical information. *British Journal of Disorders of Communication, 3*, 43-54.

Léon, P. R. (1970). Systématique des fonctions expressives de l'intonation, Analyse des faits prosodiques. *Studia Phonetica, 3*, 56-71.

Léon, P. R. (1971). Essais de Phonostylistique. *Studia Phonetica, 4.*

Léon, P. R. (1976). De l'analyse psychologique à la catégorisation auditive et acoustique des émotions dans la parole. *Journal de Psychologie, 3-4*, 305-324.

Marty, A. (1908). *Untersuchungen zur Grundlegung der allegemeinen Grammatik und Sprachphilosophie* (Vol. 1). Niemeyer, Austria: Halle.

Murray, L. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*(2), 1097-1108.

O'Connor, J. D., & Arnold, G. (1973). *Intonation of colloquial English*. London: Longman.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica, 40*, 1-18.

Piaget, J. (1970). *Epistémologie génétique.* Paris: PUF.

Piot, O. (2001). Ignorance, empathie et motivation: Une évaluation de leurs expressions dans la prosodie du français. In V. Aubergé, A. Lacheret-Dujour & H. Loevenbruck (Eds.). *Actes des Journées Prosodie*, 139-143. France: Grenoble.

Roach, P. (2000). Techniques for the phonetic description of emotional speech. *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast. Proceedings on line. Retrieved from http://www.qbc.ac.uk/en/isca /proceedings

Russel, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145-172.

Russel & Winterbottom (Eds.) (1972). Ancient literary criticism. Oxford, UK: Clarendon Press.

St. Augustino (1969). Le confessioni, Libro X. In C. Carena (trans.),. Rome, Italy: Città Nuova.

Sapir, E. (1927). Speech as a personality trait. *American Journal of Sociology, 32*, 892-905.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 293-318). Hillsdale, NJ: Erlbaum.

Scherer K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*,(3), 239-251.

Scherer, K. R. (2005). Unconscious processes in emotion: The bulk of the iceberg. In P. Niedenthal, L. Feldman-Barrett, & P. Winkielman (Eds.), *The unconscious in emotion* (pp. 312-334). New York: Guilford.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. Goldsmith (Eds.), *Handbook of the affective sciences* (pp. 433-456). New York: Oxford University Press.

Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*(5), 1346-1356.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*, 331-346.

Seashore, C. E. (1927). Phonophotography in the measurement of the expression of emotion in music and speech. *Scientific Monthly, 24*, 463-471.

Sebastiani, (1562). *Medicorvum doctrinam, Biblioteca medica storica della Ducale Università Coruzzi*. No 493.

Spitzer, L. (1961). *Stilstudien. Erster Teil: Sprachstile/Zweiter Teil: Stilsprachen* (Vols. 1-2). Munich, Germany: Max Hueber Verlag.

Steele, J. (1974). *An essay towards establishing the melody and measure of speech to be expressed and perpetuated by peculiar symbols.* (Also listed as *Prosodia Rationalis.*) Microfiche reproduction published by London: Scolar Press. (Originally published 1775)

Troubetzkoy, N. S. (1939). *Grundzüge der Phonologie*: *TCLP* (Vol. VII). Prague, Czech Republic: Travaux du Cercle Linguistique de Prague.

Wharton, T. (2003). Interjections, language and the 'showing-saying' continuum. *Pragmatics and Cognition 11*, 39-91.

Wichmann, A. (2002). Attitudinal intonation and the inferential process. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 11-16). Aix-en-Provence, France: Université de Provence, Aix-en-Provence.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, *52*(4/2), 1238-1250.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics, 38*(10), 1559-1579.

Zei, B., & Archinard, M. (2001). Acoustic patterns of emotions. In E. Keller, G. Bailly, A. Monagham, J. Terken, & M. Huckvale (Eds.), *Improvements in speech synthesis* (pp. 237-245). Chichester, UK: Wiley.

Zei-Pollermann, B. (2002). A place for prosody in a unified model of cognition and emotion. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference* (pp. 17-22). Aix-en-Provence, France: Université de Provence, Aix-en-Provence.