Perception of Intonation

JACQUELINE VAISSIERE

(préversion, version finale corrigée sous presse)

lundi 22 mars 2004

1 Introduction

This chapter provides an overview of the role of intonation in speech perception, with special focus on the perception of intonation contours.

All primates vocalize on an expiratory airflow and they make use of the oscillation of the vocal folds to generate sounds. The acoustic correlate of the rate of vibration is the *fundamental frequency* (Fo) of voice; its perceptual correlate is *pitch*. By manipulating the stiffness and length of the vocal folds, elevating or lowering the larynx and changing the sub-glottal pressure, humans can vary the periodicity of vocal fold vibration and control the temporal course of the modulation, Fo range, Fo height, the size and direction of Fo movements, the shape of the glottal-pulse waveform, their rate of change and their timing relative to the articulatory maneuvers for the realization of phonemes. All human languages exploit Fo modulation in a controlled way to convey meaning, i.e. *intonation*. As a first approximation, intonation is the use of Fo variation for conveying information at levels higher than the word, i.e. the phrase, the utterance, the paragraph, and discourse as a whole. Fo modulation contributes to the perception of the syntactic structuring of the sentence (it has a demarcative function), of its modality (it has a modal function), of the informational structuring (focus marking and topic delimitation), of the speaker attitudes and of his emotions, and of the dialog situation (the speaker's communicative intention, and his intention to give or to keep turn).

Nowadays, intonational aspects of speech are not neglected anymore. The current interest in research on intonation has been encouraged by several factors. First, there has been tremendous technical progress in the last decade. There is now wide access to inexpensive speech analysis and synthesis software, real-time Fo detection, large databases facilities, video techniques for multi-modal analysis, as well as neuro-imaging techniques. Furthermore, the hope of drawing on intonation to improve speech synthesis of texts, speech recognition and man-machine dialog systems, as well as language identification and speaker recognition, attracts many engineers. Second, the last two decades have witnessed a conceptual advance in the formal representation of pitch contours, and phonologists have been strongly encouraged to study intonation. Finally, the shift of interest from the purely syntactic aspects of language to the communication process as a whole has brought out the leading role of prosody in real-life situations and focused interest on the role of intonation in interaction. Many, if not all, of the communicative functions of intonation were not observable in laboratory speech, but surface in interactive, spontaneous speech. The perceptual (central and peripheral) and cognitive (innate and acquired) principles underlying the processing of intonation are not known, however. Despite conceptual advances, there is as yet no comprehensive model of intonation which includes the interaction between the various (often conflicting) functions of intonation.

The goal of this chapter is threefold. First, we present a number of facts which explain why there is as yet no complete theory concerning the perception of intonation, and why it is all so complex. Second, we review the findings that have nonetheless been made on the syntactical, informational, interactive, modal, attitudinal and emotional aspects of intonation (see Table 1). The contribution of Fo contours to speaker identity (sex, age, socio-cultural background, regional accent, Gussenhoven & Rietveld, 1998; Grabe, Post et al., 2000), to source separation (Brokx & Nooteboom, 1982; Darwin, 1975) and attention focusing (Cohen, Douaire et al., 2001), to intelligibility of speech (Swerts & Geluyens, 1993) and to acceleration of lexical access are not considered here, nor are the tactile perception of intonation (Auer, Bernstein et al., 1998), the perception of subjective pauses (Duez, 1993), the neural basis of intonation (Gandour, Wong et al., 2003), the processing of intonation by hearing-impaired subjects (Grant, 1987) and the effect of age (Most & Frank, 1994) or the contribution of intonation to the perception of a foreign accent (Van Els & DeObt, 1987) and the development of the comprehension of intonation patterns during language acquisition (Moore, Harris et al., 1993). Third, we present some tentative components of a *psychophonetic code* that seem to account for several striking cross-linguistic similarities in the interpretation of intonational features across languages, but that need further investigation. Some suggestions for further research will be presented in the conclusion.

Syntactic	***Segmentation of continuous speech into syntactic units of different size:*** **Prosodic** words, syntagma, propositions, utterances, paragraphs
Informational	***Segmentation of continuous speech into informational units*****:** theme/rheme, given/new, focus/parenthesis
Interactive	***Regulation of the speaker-listener interaction:*** Attraction of attention and arousal, turn taking/holding, topic end/continuation
Modal	***Communicative intent*** Assertion/question/order, etc**.**
Attitudinal	***Attitudes of the speaker toward what he says*****:** Doubt, disbelieve, etc. ***Attitudes of the speaker toward the listener:*** Politeness, irony, etc.
Emotional	***Speaker's arousal*** Joy, anger, etc.
Others	***Characteristics of the speaker:*** identity, sex, age, physiological state, regional varieties, stylistic variations, sociocultural background, etc. *Prosodic continuity, intelligibility, lexical access, memory and recall*

Table 1 : Some of the multiple functions of intonation (see text).

2 Why is intonation difficult to study?

2.1 Lack of a clear definition of intonation

There is currently no universally accepted definition of intonation. The term may be strictly restricted to the perceived Fo pattern, or include the perception of other prosodic parameters fulfilling the same functions: pauses, relative loudness, voice quality, duration, and segmental phenomena related to varying strengthening of the speech organs.

Furthermore, there is no broad consensus as to the *object* and *aim* of intonational studies. Pierrehumbert considers "that it is just the grammatical intonation distinctions which are properly of interest for linguists" (Pierrehumbert, 1980: 60), whereas other researchers emphasize that "the grammatical functions of intonation are secondary to the emotional one" (Bolinger, 1986: 260; see also Fonagy, 1983). Intonation is “a symptom of how we feel about what we say, and how you feel when you say it” (Bolinger, 1989). In this chapter, all the major functions of intonation will be considered because they interact in everyday conversation.

There is no general agreement either on the *representation* of intonation. Should one focus on pitch *levels*, pitch *movements*, or *configurations*? Phonologists these days generally prefer a pitch-level approach with only two levels (as in Pierrehumbert, 1980), while phoneticians generally favor more levels (4 to 6 or more), or expandable/compressible ranges or sloping grids of (near) parallel lines rather than levels. We do not tackle the problem of the representation of pitch contours, because it is a matter of too much controversy (Ladd, 1996).

2.2 Approaches to intonation and its perception

There are many theoretical approaches to intonation, reflecting deep theoretical and representational differences. First, for those researchers committed to a strictly linear system, the *symbolic representation of intonation* and the number of prosodic units play the primary role. In Pierrehumbert’s model for American English, intonation is essentially considered as the sum of atomistic local events: pitch accents, phrase accents and boundary tones (Pierrehumbert, 1980).

Second, the adherents of a superpositional representation of intonation suggest that the final Fo contour is best reconstructed as the sum of superimposed global baselines, semi-global phrase components and local word accent commands. Öhman and Fujisaki give physiological motivation for such a superimposition (Ohman, 1967; Fujisaki & Sudo, 1971).

Third, for the advocates of a morphological or pragmatic approach, function plays a key role. All the cues (whether pragmatic particles, or intonational, syntactic, facial and gestural cues) that fulfill the same function in a given language are therefore to be described together: *functional equivalence* is a major concern (Danes, 1960; Uldall, 1964; Rossi, 1999).

Fourth, the listener-oriented approach at IPO (Institute for Perception Research at Eindhoven) assumes that Fo contours should be described in terms of a number of *perceptually relevant* Fo patterns. Various Fo contours may be “perceptually equivalent", i.e. variants of the same Fo pattern (Cohen & t'Hart, 1967 and t'Hart, Collier et al., 1990).

Linear and superpositional approaches converge in their effort toward a compositional interpretation of intonational meanings (see part 3). More and more researchers consider it necessary to attempt to study intonational phenomena under several aspects (phonetic, phonological, physiological, functional, perceptual and neuronal).

2.3. Multiple cues, cue trading and cross-linguistic functional equivalence

Intonational cues can surface as temporally short-ranged *local* cues (such as juncture tone), as *semi-global* cues (such as resetting of the baseline), concerning a part of the utterance and as *global* cues (such as manipulation of declination tendency, pitch range, pitch register and rate of speech over an entire utterance). Pitch-accent and tone languages display more semi-global and global cues than languages with a very strong lexical stress (English), which favor more local cues mainly anchored relative to the lexically stressed syllable. Local cues may however be used in tonal and pitch accent languages: the final tone in tone languages may undergo change due to intonation; in Serbo-Croatian, lexical contrasts are neutralized by intonation in sentence initial and final position (Lehiste, 1970:101) (see Hirst & Di Cristo, 1998, for a survey on intonation in twenty languages).

All the parameters of speech melody, local and global, are perceived *in an integrated way*. Several properties of the pitch contour guide the interpretation of an utterance as a question or a statement (Gosy and Terken, 1994) and combine additively in producing finality judgments (Swerts, Bouwhuis et al., 1994 ; for affect, see Ladd, Silverman et al., 1985).

Fonagy made the hypothesis that the melody of sentences for which the listeners give multiple meanings (in free choice tests) is produced by the effective superimposition of several everyday simple intonation patterns and there should be considered as complex: the *complex intonation pattern* expresses simultaneously the messages conveyed usually by two or more simple intonation patterns. He found that the complex melodies were reproduced in a much less consistent way by French and Hungarian speakers than simple melodies (Fonagy & Fonagy, 1987).

Other parameters than pitch, such as pause, duration, intensity and voice quality may help to signal a prosodic contrast. The effects of duration pattern and pitch contour seem to be additive in phrasing (Streeter, 1978). More phonetic cues create a perception of stronger boundaries (De Pijper & Sanderman, 1994, see also Swerts, 1997). Intonation is a perceptually more important factor than pause for the clarification of the topical make-up of a text (Swerts & Geluytens, 1993). For native speakers of English listening to Czech or Slovakian Fo fall is a relatively more important cue to the perceptual segmentation of speech than is pause (Henderson & Nelms, 1980). The presence of a pre-focal hesitation pause strengthens the interpretation of a focal peak delay as signaling question intonation (House, 2003).

All the cues seem to form coherent wholes: young infants (aged 0:9) react differently to normal (coincident) phrase boundaries and non-coincident boundaries (strings segmented within the predicate phrase) (Jusczyk, Hirsh Pasek et al., 1992).

There is evidence of *trading relations* between parameters (Nooteboom, Brokx et al., 1976; see Repp, 1982 ; Pisoni & Luce, 1987, on trading relationships and context effect in speech perception). Cues form a context-dependent hierarchy. In post-focus position, the range of Fo is smaller (see examples on Figure 3) and temporal cues may take over the leading role for phrasing (French), and for stress marking (Swedish) (and laryngealization for stress marking in Serbo-Croatian, Lehiste, 1970:101). Peak Fo height and slope also trade in determining the category boundary between contrastive and non-contrastive focus in English (Bartels & Kingston, 1994); a delayed peak often substitutes for a higher peak (House, 2003). There is little evidence of cue trading for children, for whom duration has a strong influence on identification of the phrasal units, whereas pitch has only a slight influence (Beach, Ganguly et al., 1991). The weight of each acoustic parameter depends on the position of the syllable in the word (McClean & Tiffany, 1973)and of the word in the prosodic contours (Nakatani & Schaffer, 1978).

Several cues may be *functionally equivalent* cross-linguistically. In mora-based Japanese, duration is constrained at the phonological level, and Fo at the lexical level. I suggest that (i) a semi-global resetting of the baseline at the beginning of each new prosodic phrase (Fujisaki & Sudo, 1971) in Japanese may be considered as a kind of substitute for phrase-final lengthening, as phrase boundary marker, that (ii) the final topic-marker "wa" may be considered as a substitute for a continuation rise, and that (iii) the expansion of the local Fo range is functionally equivalent to the displacement of phrase accent in English for the expression of focus. As a consequence, the description of intonation in a language cannot be done without considering whether or not other acoustic cues and other linguistic means fulfill similar functions.

2.4 Non-applicability of otherwise well-established (psycho-acoustic and linguistic) methods

The well-established experimental methods developed in *psychoacoustics* do not actually apply in the field of intonation. Fo perception in speech includes not only a psychoacoustic level but also higher-level cognitive and linguistic processing (Pisoni & Luce, 1987). While higher-order linguistic decision may determine auditory shape in some cases, the reverse can also be true (Studdert-Kennedy & Hadding-Koch, 1973).

Traditional *linguistic* paradigms, such as the criteria of distinctiveness and *semantic differentiation*, which have proved their validity in the discovery of phonemes, fail in the study of intonation. When the listener is ask to identify the same stimulus pattern as *yes-no question, echo question, call for confirmation, alternative question, rhetorical question, disbelieving question, a question to oneself*, an all-or-none decision is hardly possible.

One may reasonably consider intonational meaning as involving both all-or-none contrasts on the one hand, and dimensions of gradiency within categories, on the other hand (Ainsworth & Lindsay, 1986; Ladd & Morton, 1997). Differences in *F0 range* are commonly assumed to have continuous rather than categorical effects on affective judgments (). The coexistence of discrete and continuous dimensions of intonation makes perceptual experiments difficult to carry out and interpret.

How does one decide what is a linguistic function of intonation and what is not ? In a given language, a morpho-syntactic means, such as a modal particle, word order, an expression such as “I strongly believe that …”, may be entirely or not replaceable by intonation and intonation may be a reinforcing cue. In some languages, intonation may be the only means available for expressing yes-no questions or some attitudes. Intonation may be the only factor used to resolve certain syntactic ambiguities, but ambiguities are rare in every day speech.

Furthermore, intonation has no self-evident *units.* A unit such as a word tends to be acoustically marked by a lexically prominent syllable even in tone languages. Two or more words may be prosodically grouped into a single phrase, which is variously called *intermediate phrase*, *phonological phrase*, "*groupe rythmique*", "*groupe mineur*", “sense-group”, "*buntetsu*" (in Japanese). This low-level grouping is generally achieved by superimposing the prosodic characteristics of a single word on a sequence of several words. Two sense-groups may also be further grouped into a larger intonational group, or *“groupe majeur”* or *intonational phrase*. The end of an utterance non-final intonational group is typically marked by some kind of Fo raising and pre-boundary lengthening. Intonational phrases are then grouped into a *prosodic* (or *phonological*) *utterance*, typically ending in an Fo fall and low Fo value, low intensity and final lengthening. Utterances are also grouped into prosodic paragraphs (or *topic units*), with raised Fo values at its beginning, and lowest Fo value at its end. The various units are marked by both a *strength hierarchy of stress* and varying *strength of boundaries* (See Shattuck Hufnagel & Turk, 1996 for a summary of English facts). It is hardly feasible to find acoustically well-defined units: there are large inter-style, inter-rate, inter-speaker variations, and the final prosodic organization is also determined by the size of the constituents. Implementing five levels of boundary strength gives synthetic speech higher quality than rule sets with fewer levels (Sanderman & Collier, 1996b).

2.5 The lack of standardized methods

A large number of methods have been used to study the perception of intonation and the role of intonation in speech perception (they are listed in footnote[i]). Synthetic speech allows a researcher to change one parameter at a time, such as Fo, which represents a clear advantage over natural speech production for evaluating the contribution of each individual parameter. Intonational focusing, however, always involves an action of the respiratory muscles (see part 3.2.2). Such an increase has several “natural” acoustic consequences (unless compensatory adjustments are made at the glottal level): higher intensity, higher Fo, less steep spectral slope, stronger release of the obstruents, longer VOT, etc. Isolated Fo manipulation using synthetic speech stimuli therefore may be inappropriate for studying the perception of focus in everyday speech. Furthermore, the Fo-manipulated versions retain much of their original characteristics: the accompanying cues may substitute for Fo features.

2.6 Effect of phonetic context and discourse context on the perception of intonation

The intonational features are perceived according to context.

2.6.1 Intrinsic and cointrinsic context

First, perception of the Fo features of a vowel is by no means independent from its loudness and duration, its phonetic quality, the voicing quality of the subsequent consonant (see Lehiste, 1970 for an almost exhaustive review), the melodic context (Hadding-Koch & Studdert-Kennedy, 1964), the timing of the Fo movement relatively to the onset of the vowel (House, 1990), and the time course of intensity (Rossi, 1978; see also Fonagy, 2000: 137-148). Such results call for a configurational approach to Fo contour perception, or at least for a combination of atomistic cues and holistic patterns.

2.6.2 The discourse context

Second, there is often no agreement between the speaker’s intention and the listener’s interpretation when utterances are heard out of context in Danish, Dutch, French and Swedish (Uldall, 1964; Beun, 1990; Fonagy & Bérard, 1972; Hadding-Koch, 1961). In their judgments of speakers' intentions, like sarcasm, adults rely heavily on context as well as intonation, but children are less attuned to contextual information (Capelli, Nakagawa et al., 1990).

As a consequence of the influence of the pragmatic context, it is not possible to draw any permanent link between form and function. The perceived meaning of prosodic signals should be treated as a pragmatic implicature or a pragmatic inference (Wichmann, 2002).

2.7 Perception of intonation as a language-specific process and non-language-specific process

Tonal languages use Fo primarily to signal lexical contrasts. In Japanese, a pitch-accent language, the word tonal patterns are the most straightforward component of the *shape* of the Fo contour. In stress languages, the need to realize the lexical stresses strongly constrains Fo. In French, Fo, duration and intensity are tighted to the word boundaries and intonation (especially its demarcative function) is the main determinant of the *shape* of the Fo contour. There are highly language-specific characteristics, "semantic", "systemic", "realizational" and "phonotactic" distinctions in intonational structure across languages (Ladd, 1996), but all types of languages, tonal, pitch-accent, stress and boundary languages, use intonation and share intonational features.

It is generally hypothesized that the perceptual mapping between the acoustic signal and intonational categories is sensitive to the abstract structural properties of individual phonological systems. Every single perceptual experiment on speech-like stimuli involving listeners whose native languages differed shows differences in the identification and discrimination of basic intonational elements such as word *stress* (Berinstein, 1979; Watanabe, 1988), *prominence, modalities* (Makarova, 2001)*, attitudes* and *perceived emotion* (Abelin & Allwood, 2000;).

The same cues are generally used but their hierarchy may be different. In the perception of *prominence* in speech and non-speech signals, amplitude cues override duration cues for English-speaking listeners, whereas native Estonian listeners are more responsive to duration cues (Lehiste & Fox, 1992). French informants do not classify a syllable as accented when it has a falling pitch movement, whereas Swedish and Dutch listeners do. Also, the location of the onset of the pitch movement seems to have much less weight in French than in Dutch or Swedish for detecting accentuation (Beaugendre, House et al., 1997). When Japanese and Russian subjects were asked to identify two-syllable re-synthesized stimuli with modified rising-falling contours as exclamations, interrogatives or declaratives, the perception of stimuli as declarative was similar for both groups of subjects, while the perception of stimuli as interrogative and exclamatory was in some cases significantly different (Makarova, 2001).

3 Cross-linguistic similarities in intonational meanings and underlying psychophonetic code

Despite differences obtained in perception experiments involving native speakers of different languages, the literature on a large number of related and unrelated languages points to several universal tendencies in the relation between intonational form and intonational meaning. It seems natural to attempt to explain similarities in the interpretation of intonational contours by features shared by human speakers (the present proposal follows up on Vaissière, 1995, and Gussenhoven, 2002). The following table illustrates five elements of a proposed hypothetical *psychophonetic code*. The elements will be described and discussed below.

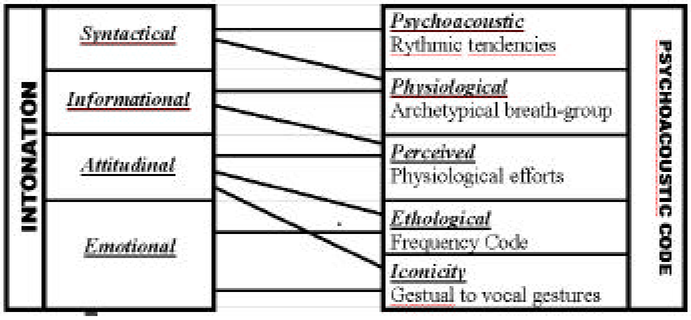

Table 2 : Some elements of the psychophonetic code (right) and their hypothesized relationship with intonational components (left). Lines indicate main relationships.

3.1 Psychoacoustic rhythms, initial strengthening and final lengthening.

The first part relates to basic *psychoacoustic* rhythmic tendencies. The similarities in the way of *segmenting speech* (demarcative function) in languages are tentatively explained by the inference of two non-linguistic rhythmic principles which associate lengthening with the notion of end (and relaxation) and strengthening with the notion of beginning.

3.1.1 The two basic rhythmic tendencies (Fraisse, 1956)

Through psychoacoustic experiments using non-speech stimuli, Fraisse (1956) distinguished two types of rhythmic organization: *"rythmitisation intensive"*, sensitive to *strengthening of the initial element*, and *"rythmitisation temporelle"*, building on the *lengthening of the final element* or pauses. Figure 1 illustrates the repetition of a hypothesized basic rhythmic unit, where the two basic rhythms are combined, with initial extra loudness and final lengthening.

.

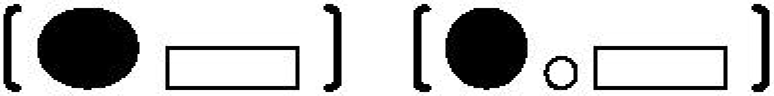

Figure 1 : Basic rhythmic unit composed of two (left) or three (right) elements. Filled circles and rectangles indicate perceived extra-loudness and final lengthening, respectively. Parentheses refer to the way listeners will most likely chunk the continuum.

3.1.2 Final lengthening as boundary marker

The two types of rhythm seem to be reflected in a large set of phenomena linked to segmentation, at the word and phrase level (see also Allen, 1975).

First, the basic tendencies are reflected in the manner languages mark *word stress*. The word-accent seems to come from intonation in exactly the same way that word-boundary phenomena come from utterance-boundary phenomena and intonation becomes grammaticalized as a word-stress when the suprasegmental features of pitch, duration and intensity that would have characterized a word in isolation (where it gets prominent intonation) are encoded with the word, and thus seem to function in words not in isolation (Hyman, 1977). As a general tendency, initial stress (or early stress, as in English) in a word tends to be marked by extra loudness, and late stress (French, Italian, Spanish) by extra lengthening[ii]. Even when they do not receive primary or secondary stress, word-initial consonant and syllable tend to be produced with greater strengthening, and word-final rhyme to be lengthened, at least in European languages.

Second, the basic rhythmic tendencies seem to have been phonologized in *boundary marking*:

a) **Final lengthening** and a decrease in speech rate convey *relaxation* and *ending*. Speakers tend to lengthen the final element in the unit: the final syllable in a word, the final stressed syllable (and the final syllable) in a phrase, the final phrase in an utterance, and the final utterance in a paragraph. English listeners expect the duration of the pre-boundary syllable to reflect the rank of the phonological boundary, whether or not it coincides with an intonation contour boundary (Gussenhoven & Rietveld, 1992; Streeter, 1978). When asked to judge the duration of pairs of vowels of equal duration, and with level Fo patterns, 68% of listeners judge the first syllable to be longer (Lehiste, 1975).

b) The **tendency to isochrony** has been observed in a number of languages (but not all): the average duration of syllables (words) in longer words (phrases) tends to be shorter than in shorter words (phrases), due to compression. It has two major consequences. First, shortening of syllables in long units and lengthening in short units contribute to perceptual integration into units. Preferred vowel duration depends on the number of following syllables in the word (). Second, isochronous intervals facilitate attention by guiding expectations as to when the next stressed syllable is likely to occur (Lehiste, 1980; see the chapter by Cutler, in this volume, and also Cutler, Dahan et al., 1997).

c) **Interstress** intervals tend to be perceived as more isochronous that they are acoustically. Disruption of the expected pattern is used to convey crucial information about syntactic structure: interstress interval lengthening is interpreted by listeners as indicating the presence of an underlying juncture ().

Third, the two types of rhythm coexist in languages. French uses mainly final lengthening for marking different degrees of syntactic boundaries, and initial strengthening (more extreme articulatory movements, with raised intensity and Fo) for informational structuring. In contrast, English makes greater use of "intensive rhythmitisation", by marking lexical stresses, but also uses final lengthening as a boundary marker (see Vaissière, 1991, and Vaissière, 2002 for the notion of dominant and regressive rhythms, and differences between English and French, respectively).

Fourth, most of the studies that compare the effects of Fo and duration show the importance of duration (in particular interstress lengthening) over Fo in the perception of phrase boundaries (Lehiste, Olive et al., 1976; Price, Ostendorf et al., 1991). In their production, children (age 5 and 7) use duration and not intonation to mark phrase boundaries in spontaneous speech (Verma, Mannering et al., 1994).

3.2 Breath-group and perceived effort.

The *structural and continuity markers* which have a cross-linguistic validity are interpreted as deviations from a physiologically-based Fo archetypal pattern (Lieberman, 1967). The same global archetypal Fo pattern seems to be cross-linguistically used for statements in a large number of languages (Bolinger, 1989). Deviations from that archetypical pattern carry meanings.

3.2.1 The archetypal breath-group and higher-level segmentation

a) Physiological aspects

Speech is superimposed on expiration. The basic Fo and intensity shape corresponding to phonation in a single expiration is a sharp rise followed by a fall, a pattern shared by non-human primates and infant cry (Lieberman, 1967). Between two successive breath-groups, there is inhalation and subsequent *resetting* of subglottal pressure and Fo values. Humans use sophisticated breath control, generally taking in more air when they intend to say more, which raises subglottal pressure and Fo.

An increasing muscular effort is required to maintain subglottal pressure constant as the volume of air in the lungs decreases; the unmarked Fo and intensity pattern is *declining*, which seems to indicate that speakers do not fully compensate for decreasing lung volume over the course of the utterance (Collier, 1975; for a summary of physiological explanations, see Vaissière, 1983).

Figure 2 (top) gives a schematic representation of the Fo pattern associated with the unmarked breath group (used for neutral statements). The Fo contour oscillates between two "abstract" declination lines, the *baseline,* connecting Fo valleys, and the *plateau* connecting Fo peaks. Fo range tends to decrease from the beginning of the sentence to the end. The plateau declines more rapidly than the baseline.

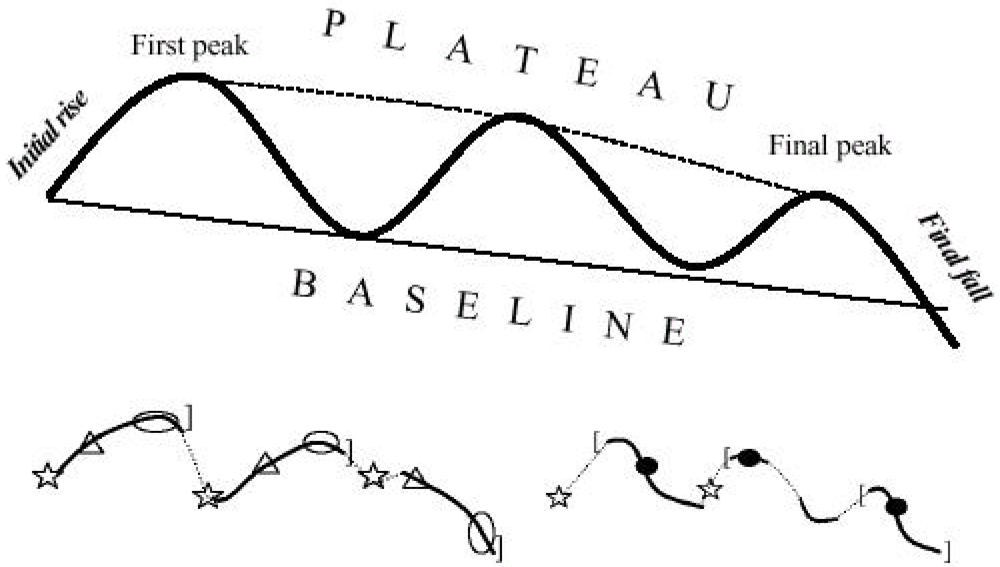

Figure 2: Top: the basic archetypal pattern found in many languages. The curve corresponds to the natural course of Fo and intensity. Bottom: highly abstract typical Fo contour in French (left) and English (right) declaratives. Bracketing indicates perceived boundaries. Empty black points represent word final syllables in French, and filled black points word stresses in English; stars represent function words and triangles word initial syllable in French.

b) Linguistic use of the characteristics of the archetypal pattern

Fo fluctuations within unmarked declarative utterances help delimit successive sense-groups or phrases. Figure 2 (bottom) represents typical Fo curves in declarative sentences (composed here of 4 and 3 prosodic phrases, for French and English, respectively; in the figure, each prosodic phrase comprises a single prosodic word). Each language seems to put emphasis on one aspect of the archetypal Fo pattern and English infants show very early preference for the predominant stress patterns of English words (Jusczyk, Cutler et al., 1993). For example, in French, the slowly *rising portion* ending by final lengthening is perceptually dominant. It contrasts with English, where a rapidly falling contour is associated with the word-stressed syllable (fig. 2) and the perception of *stresses* dominates. Danish favors the *low* Fo value, aligning it with the word-stressed syllable; in Japanese, what is mainly perceived is a contrast between successive stretches of baseline and plateau: an Fo jump to the baseline on the word's second mora is followed by a Fo plateau continuing until the word-accented syllable, and a low Fo value associated with the syllable (if any) following the accented one.

From this vantage point, one may suggest that in each language, the statistically most frequent way of combining Fo, duration, intensity and segmental characteristics (entailing a particular repartition of effort along the word syllables) becomes a prototype. Some of the impressionistic loudness and duration characteristics are represented in Figure 2.

The same basic components of the archetypal intonation pattern are used to contribute to the hierarchical organization of discourse, by marking beginnings and ends.

Initial rise - final fall in Fo summons up the notion of a *complete unit*, such as a definite statement.

Initial Fo rise is linked to the notion of *beginning*. In discourse, suppressing the initial Fo peak suggests to the listener that the utterance elaborates on the previous utterance(s) (Nakajima & Allen, 1993).

Final fall in Fo or lowered Fo contour and low intensity (and decreasing rate, as seen above) suggest *finality*, i.e. the end of a phrase (Streeter, 1978), an utterance, a paragraph (Lehiste, 1975), a topic unit in spontaneous discourse (Geluykens & Swerts, 1994), or a turn (Schaffer, 1984). A fall is judged to convey an impression of finality only if it restores the pitch to the level of the lower declination line (Collier, 1984). When topic and turn terminations interfere with each other in discourse, speakers avoid using low tones at the end of a topic, reserving them to signal turn finality. Listeners can reliably discriminate between turn-final and non-turn-final topic units (Geluykens & Swerts, 1994).

Continuing declination line is associated with the notion of *integration*, *continuity*, and *preplanning*. The rate of declination actually observed exceeds the rate of decline predicted by physiology, suggesting that declination has essentially been *phonologized* (Vaissière, 1995). Linguistically controlled declination, as in downstep and downdrift phenomena, indicates integration.

Fo fall-rise pattern at clause and phrase boundaries is associated with the notion of end-and-beginning, that is, of disjuncture between phrases, clauses, sentences (Lea, 1980), and syllables (Ainsworth, 1986).

c) Linguistic use of the characteristics of the deviations of the archetypal pattern

While prosodic word profiles differ greatly from one language to the next, they seem to be “deformed” by intonation in rather similar ways.

Final rising Fo and final non-low value are frequently associated with the notion of *non-assertiveness, incompleteness, continuity.* Raised tone seems to indicate points of "interest" within utterances and also indicate that more is to follow, as in questions (Bolinger, 1964). "yes-no" questions in Swedish display a terminal rise and an overall higher Fo than statements. Other Swedish utterances in which the speaker wants to draw the listener's special attention also display an overall high Fo and a terminal rise: in listening tests the labels "question", "surprise", "interest" have been found to be interchangeable (Hadding-Koch, 1961 : 126 ff.) An accent-lending rise followed by level pitch seems to indicate turn-keeping in Dutch, while accent-lending rise followed by a second rise is ambiguous between turn-keeping and turn-yielding (Caspers, 1998; for a review of recent research, Hirschberg & Swerts, 1998). When a language differentiates between question and continuation, higher Fo, less steep declination and/or steeper rise are associated with question marking.

Declination slope over the entire utterance carries meanings. More Fo declination than expected and higher articulation rate influence the listener’s perception of a "spontaneous" over a "read" speaking style (Lan, 1997), probably suggesting to the listener lack of preplanning. The most steeply falling intonation contours are identified as being declarative, the least falling ones as being interrogative, and contours in the middle of the continuum as being nonfinal (Thorsen, 1980). As a first approximation, topline is very sensitive to pragmatic factors, and the baseline to syntactic factors.

Non-continuity in the declination along the utterance and its partial Fo (and intensity) reset, with or without actual breathing-in and/or pausing, is interpreted by the listener as marking a new phrase (Streeter, 1978), an intonational clause, or topic. The amount of reset reflects the hierarchical structure: the higher the reset, the deeper the node. A lack of reset is used as a continuation mark: the unit is perceptually integrated with the preceding unit.

The same principles seem to hold for units larger and smaller than the utterance (for the notion of recursivity, see Vaissière, 1995). Utterance-highest Fo peaks decline from the beginning of a *paragraph* (or of a topic) to its end (Lehiste, 1975). In French, in an isolated *word*, a peak located at the word beginning, on the penultimate or at the very end of the word indicates an answer, a doubt and a question, respectively. Early Fo peak in the stressed *syllable* corresponds to *established fact*, medial and late peaks are perceived as implying a *new fact* (for German, Kohler, 1991; for Dutch, Caspers, 1999).

Concerning the perception of declination: the second accented syllable in a phrase should have *lower* pitch than the first one to be perceived as having *equally* strong stress. The listeners seem to normalize their perception of the overall contour in terms of expected declination (Pierrehumbert, 1979), in a very complex way (Terken, 1994; Gussenhoven, Repp et al., 1997).

3.2.2 Perceived expiratory effort and the expression of focus and arousal

The next element playing a role in the perception of intonation is the *effort code* (see also Gussenhoven, 2002). As well known, loudness judgments on speech are more closely related to the degree of *vocal effort in speech production* than to the speech signal's surface acoustic properties, such as intensity or Fo. The listener seems to perceive the amount of global effort made by the speaker and it seems advisable to separate the various types of perceived effort by taking physiology into account (for a first application to French with four physiologically-determined types of stress, Vaissière, 2001).

a) different degrees of strengthening of the *supraglottic speech organs* (the tongue, the lips and the velum) mainly affect *timber*. Supraglottic tensing results in higher position of the velum, longer and larger closure for obstruents, and more precise articulation. Initial stress in French should be considered as "supraglottic" stress and correspond to a localized hyperarticulation. Supraglottic strengthening may spread to the glottic level, but it need not be the case.

b) *laryngeal effort* affects both *Fo contours*, *voice quality* (i.e. the shape of the glottal-pulse waveform and spectral balance) and *glottal resistance*. Pitch-accent languages may be said to have a *laryngeal stress*. Spectral balance turned out to be a reliable cue in the differentiation between initial- and final-stressed words in Dutch, just behind duration, with the overall intensity and phonemic quality as the poorest cues (Sluijter & Van Heuven, 1997). Some attitudes and modalities are expressed solely by laryngeal maneuvers. Fo features, as well as voice quality, plays the key role here.

c) *sudden extra activity of the respiratory muscles*, mainly affecting subglottal pressure, leads directly to an increase in the speed of glottal opening, in amplitude of vocal folds displacement and in intensity, and indirectly to Fo raising and segmental changes through aerodynamic changes (e.g. VOT). Occurrences of *nuclear stress, sentence stress and emphatic stress* have always be found to correlate well with a burst of intercostal activity, increased subglottal pressure and increased loudness (Ladefoged, Draper et al., 1958, for English; Benguerel, 1973; for French), but not always with both intensity and pitch excursion. Expiratory stress (or flowinduced stress) seems to be perceived as *focus and emphasis*. Fo contour and height seem to be by-products of the local increase in subglottal pressure.

d) *sustained expiratory effort* leads to increased *Fo range* and *lack of declination,* and expresses *involvement and arousal*. Happiness and surprise are associated with large pitch variation, high pitch level and ascending scales. Music and speech seem to share similar interpretation (Collier & Hubbard, 2001).

e) *disturbance of the respiratory system* is frequently found in states of anxiety; the increase in respiration rate leads to increased subglottal pressure and higher Fo (). Listeners' stereotype of psychological stress includes elevated pitch and amplitude levels, as well as their increased variability (Streeter, MacDonald et al., 1983).

Focusing allows listeners to speed up the comprehension of words that convey new information, whereas given information is processed faster if it is not accented (Terken & Nooteboom, 1987). In a question-answering task, appropriately phrased utterances (quite predictably) produced faster reaction times (Sanderman & Collier, 1996a).

Figure 3 illustrates the typical Fo pattern in statements and in questions, with narrow focus on the initial, medial, and final word (this figure would be valid for French, Japanese, English, Danish and Chinese). The excess of physiological effort at one point in the utterance seems to be done at the expense of surrounding parts, particularly the following part. The long-range effects of the realization of focus call again for a multi-parametric and holistic approach to intonation in many languages.

Pitch, duration, and intensity were found to be relevant to the intonational-perceptual marking of focus (Batliner, 1991). Depending on language, speaker and style, acoustic cues include (i) a displacement of sentence stress onto the focused word, (ii) a further strengthening and lengthening of the lexically stressed syllable, hand in hand with the shortening and reduction of non-focused words, (iii) an increase in the magnitude of the underlying tonal movement, leading to a substantial pitch range expansion and top line modifications, followed by post-focal pitch range reduction, (iv) a sharp Fo fall between the emphasized word and the following word in the phrase) followed by a relatively flat and low Fo contour on post-focal words, (v) a supplementary focal tone, as in Swedish, or (vi) other “tonal” change. It has been proposed that the distinction between contrastive and non-contrastive focus in English can be conveyed by the difference between a *L+H** (*rise* on the stressed syllable) vs *H** (*high* on the stressed syllable) pitchaccent (Pierrehumbert & Hirschberg, 1990).

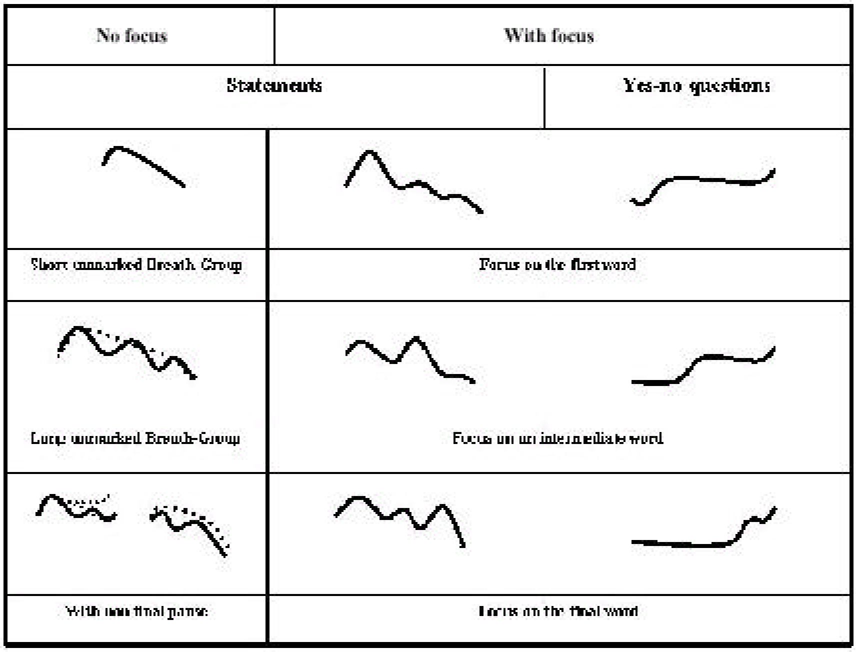

Figure 2 : Typical Fo contours in neutral statement (left) of different length, and the deformations they often undergo to mark modalities and focus (See text).

3.3 Iconicity and the Frequency Code

3.3.1. Expressive settings, facial gestures and intonation

Universals in intonational meaning may draw on an archaic *genetic code* where the motivation of the signs originates, this being especially salient for the expression of affect (Bolinger, 1989).

Expressive intonation often goes hand in hand with deviations superimposed on the (ideal) neutral *phonatory and articulatory setting* of the entire utterance. *Tenderness* is expressed by melodicity (intrasyllabic regularity of vocal folds vibration), relaxed vocal folds, breathiness as well as smooth articulatory transitions, and labialisation. If two fundamental frequency curves differed only in their angularity (in suddenness of change or direction) the French sentences with a more angular fundamental frequency curve were rated as being significantly more aggressive than their smoother counterparts (Fonagy, 1979). *Smallness* is suggested by high Fo as well as palatalization (a fronted position of the tongue mainly raises F2 frequency). The expression of *disgust* includes glottalization and pharyngealization. Fo irregularities, forceful innervation of the glottal muscles, narrow constriction of the glottal space as well as retracted lips and tongue retraction characterize anger and hostility (See Fonagy, 2000 drawing on tomographic and cineradiographic studies).

Intentions can often be recovered not only from vocal (especially intonation) gesture and settings, but also from *facial gestures*, the nominal content of the message then being redundant (Bolinger, 1989). In making judgments about intonation patterns, much of the listener's attention is directed to visual inspection of the upper part of the talker's face, which may play a greater role than words (Lansing & McConkie, 1999). Raised eyebrows often go hand in hand with an Fo rise in the expression of surprise, and labialisation with tenderness (see Fonagy, 2000 for a review). Loudness judgments were also found to be affected significantly by visual information even when subjects were instructed to base their judgments only on what they hear and not what they see (Rosenblum & Fowler, 1991).

There seem to be large cross-linguistic similarities in phonatory and articulatory deviations, facial gestures for the expression of attitudes and emotion. This is the basis for proposing the *Iconicity* Code, a gestural-to-lingual code. Iconicity involves ethology, and the development of more or less elaborate intonational "signifiés" from instinctive "signifiants" or signs that originally expressed uncontrolled primary emotion. These primitive signs were then conventionalized and integrated into the linguistic code; but there remains a resemblance between the spoken forms and the things they stand for. Motivation dominates in the expression of emotion, while the expression of attitude is more conventionalized; the expression of *moods* is part of the grammar. It is indeed a general observation that the least motivated intonation phenomena (the most arbitrary) are least recognized cross-linguistically. In foreign languages, it may be easier to recognize and render emotions than to distinguish a question from a statement. We refer to Fonagy, 2000 for a thorough account of these issues; see also Bolinger, 1989).

According to Fonagy, Bolinger and others, speech melody could be conceived in terms of *virtual bodily gesturing*. First, the degree of general excitement and tension (arousal) is reflected in the degree of tension of the vocal folds: higher pitch means greater *excitement*. Low Fo (and a slower speaking rate) characterize *passive* emotion and detachment, whereas high Fo and more rapid speech characterize *active* emotion. Second, melodicity characterizes *agreeable* emotion, and lack of melodicity *disagreeable* emotion (Fonagy, 1981)[iii]. Third, vocal gesturing is in *proportion* to expression and content: the pitch range is proportional to the degree of *involvement*. In synthetic speech experiments, small pitch variation is found to be associated with *disgust, anger, fear, boredom*, and large pitch variation with *happiness, pleasantness, activity, surprise* (Bolinger, 1989; Hirschberg, 1992 ;) and *benevolence* (Brown, Strong et al., 1973, see also Fonagy, 1983).

3.3.2. The Frequency Code

The well-known Frequency Code (Ohala, 1984; Bolinger, 1989; Morton, 1994) is one aspect of the Iconicity principle; it is the one best documented in the literature, and while there is no firm evidence for it, there seems to be no clear counter-example, and it tends to be accepted. First, by the laws of acoustics, a larger vocal tract has lower formant frequencies. Second, a slower rate of vibration of the vocal folds is physiologically related to larger vocal folds, and so to a larger vocalizer. Nonhuman terrestrial vertebrates use similar sounds in similar ways; it is often hypothesized that the fundamental, unifying principle of vocalizations used in *hostile* or *friendly, appeasing* contexts is that they convey an impression of the size of the vocalizer (Morton, 1994). Third, a majority of women have a posterior glottal chink: female voice tends to have a breathier voice quality than male voice.

Speakers can control up to a certain extend mean formant frequencies, Fo and degrees of breathiness in their voice. The natural links seem to provide an explanation for the cross-linguistic observation that high formant frequencies, high Fo and breathiness are all associated with the same primary meaning of *small vocalizer*, conveying secondary meanings such as *subordinate, submissive, non-threatening, desirous of the receiver's goodwill, polite, lack of threat by the subject, more femininity, hesitation, uncertainty* or *surprise*. The association of greater pitch range with *incredulity* can be accounted for by the previously noted tendency of listeners to associate larger pitch ranges with greater degree of speaker involvement; conversely, the association of smaller pitch range with *uncertainty* can be explained as a consequence of the perception of less speaker involvement (Hirschberg & Ward, 1992; Ward & Hirschberg, 1988). Also, the lower the degree of certainty, the higher the mean Fo value will be (Bolinger, 1989). Speech directed to infants typically exhibits a larger pitch range and higher average pitch, probably to empathize with the infant's smallness or to attract and sustain its attention. The intonation at the end of a sentence has a great impact on politeness judgments and speech rate plays an important role: raised Fo is heard as more polite (Ito, 2002; Ofuka, McKeown et al., 2000).

Lower formants, lower Fo and the creaky quality of the voice conveys the primary meaning of *large vocalizer*, with the secondary meanings *dominant, aggressive, threatening*, *definitive, more authoritative* (see also Bolinger, 1964).

4. Discussion, conclusion and suggestions for newcomers to research on the perception of intonation

The first part of the chapter considered the difficulties researchers are faced when starting intonational studies: lack of clear definitions, non-applicability of otherwise standardized experimental methods used in psychoacoustics and phonology, the effect of phonetic and melodic context and the speaker's native language on the perception of intonational phenomena. As a consequence, it appeared that the results obtained in one context cannot be extended and generalized to other prosodic contexts in any simple straightforward way (Verhoeven, 1994). In the second part, we attempted to put together a number of non-linguistic facts to explain cross-linguistic similarities in the interpretation of intonational phenomena: psychoacoustic rhythmic tendencies, physiological considerations, and an ethological iconicity code.

As illustrated in this chapter, intonation plays many different roles in speech perception. The studies on the use of prosody in parsing suggest a *supporting*, rather than a leading role for prosody *in the grouping of words into constituents* (Cutler, 1997). Not all phonetically similar sentences can be disambiguated by listeners on the basis of prosodic differences alone (Price, Ostendorf et al., 1991) and ambiguous sentences are very rare in every day conversation. In utterances in which prosody and syntax conflict, the localization of the click by listeners was determined by the syntactic structure assigned to the sentence, and not by prosody (Fodor & Bever, 1965). When presented with recordings of ambiguous constituent structures, listeners generally ignore prosodic features when other linguistic cues (semantic and pragmatic) are available (Berkovits, 1980). Syntax and semantics provide much stronger topic cues than does intonation (). In contrast, the *affective functions* of intonation are generally not redundant with the linguistic properties of an utterance. Under some circumstances, the way we say things may be more important that what we say. Intonation offers effective shortcuts: a simple word like "yes", "oui, "da" may express approbation, confirmation, doubt, impatience, joy, anger, irony, evidence or tenderness. Listeners seem to have no difficulties in differentiating all these, though many of the differences are so subtle that they are hard to identify acoustically.

Intonational studies are in vogue at present; indeed, intonation is an extremely lively field, but still much has to be done. First, perception (rather than acoustics) has strong claims to being the best starting-point for the study of intonation, a phenomenon that involves multiple acoustic cues. Only listening tests can provide reliable behavioral data on perceptual equivalence, on the one hand, and functional equivalence (within one language, and cross-linguistically) on the other hand. It is not reasonable to build intonational models from intuition and acoustic analysis alone, without perceptual testing.

Second, at this stage, we need to develop uncontroversial universal methods for describing intonation, prosodic transcriptions should include several sharply distinguished levels (one must be perceptual, another acoustic and another interpretative). At the highest level, transcription should capture the interpretation of the sentence by the listener, its syntactic structure, its mode, and its information structure as well as the attitude perceived. An intermediate level should include local intonational phenomena: the perceived strength of boundaries between words, jumps and glides in pitch, and local stresses, among others. Parallel listening experiments on music and (non-sense) speech material may help.

Third, intonational studies should include as many parameters as possible, including physiological ones. *Fo contour* has been privileged (see Pierrehumbert, 1980: 11-12; t'Hart & al, 1990); but it is not the only cue. Limits of perceptibility of pitch phenomena may change depending on accompanying cues and on the physiological interpretation of the deviations by the listener. It should be borne in mind that for want of appropriate tools, most research concentrates on an incomplete inventory of parameters, such as mean Fo and duration values (Murray & Arnott, 1993, for emotion). For example, in Moba, an African tone language, glottal stop and vowel shortening mark assertion, while breathiness and vowel lengthening mark interrogation (Rialland, 1985).

The role of *intensity* need wider empirical attention: the fact that intensity downtrend may not parallel Fo downtrend has been largely unexplored.

Intonation (and prosody in general) can barely be considered independently from *segmental characteristics* (glottalization, initial strengthening, etc.)

Also, *multimodal analysis* including body and facial gestures seems very promising for future studies in the perception of intonation, starting out from laboratory materials, and moving on towards a study of intonation in *spontaneous speech,* in *real-life situations*.

It is likely that continuing interest in emotions and attitudes will show the limitations inherent in a “Tone and Break Indices” description restricted to pitch levels.

Fourth, great care should be taken to ensure that stimuli presented to the listeners have physiological plausibility. Ways of obtaining physiological data (e.g., subglottal pressure, EMG) are often invasive, but glottographic and fibroscopic data are rather easy to obtain. Listeners interpret differently stimuli seem with very subtle nuances. In particular, perceptual experiments are needed to test whether the listener can distinguished between loudness from different physiological origins, an increase in respiratory effort resulting in increasing intensity and/or Fo, or to laryngeal maneuvers (see the chapter on loudness, by Lehiste, 1970), since at least in French, they assume a different function (Vaissière, 2001).

Fifth, statistical knowledge on the use of the different intonational patterns within and across languages, dialects, styles and speakers is an indispensable component of any study of the perception of intonation.

Within a given language, statistics are necessary for establishing the most common intonational pattern the listeners are exposed to and for building probabilistic models of the auditory processing of intonational cues by both infant and adult listeners.

Enormous amounts of speech are necessary to obtain statistical evidence on factor combinations in the study of the interaction between layers of meaning, in particular the interplay between discourse structure and grammatical and informational structure: systematic perceptual experiments with synthesized stimuli are a necessary complement. An intonational pattern, although rarely observed, may be easily interpreted by listeners.

Note that the statistical knowledge should be derived from both databases of spontaneous speech and from different kinds of well-controlled material: the syntactic role of intonation stands out in isolated sentences; in a text, the information structure becomes apparent; in dialog, it is interaction which will attract the researcher’s attention. None of these functions should be considered as more important than others: they all coexist in every day conversation and listeners as well as students of intonation have to cope with such a fact.

Sixth, more *cross-linguistic*, *cross-dialectal,* and cross-stylistic studies on differences in perception (and production) of intonation in infants, children and adults are needed. There should be more studies on the developmental aspects of intonation in different cultures: newborns are more sensitive to such prosodic phenomena (and phonotactic characteristics) than previously believed. No firm conclusion on the universality of part of intonational phenomena can be drawn until more languages have been studied and for the time being, we mainly rely on the lack of counter-evidence. Perceptual tests involving listeners of different languages are extremely promising. Nonetheless, great care should be taken in designing the task : while it seems a very natural task for a native of English, Dutch or German to decide which syllable in a word is more "stressed, it is very awkward for a native of French: a French word carries often more than one (supralexically determined); the coexisting (supralexical) stresses correspond to different physiological maneuvers in French, have different acoustic correlates, and therefore cannot be ranked; the notion of lexical "stress" does not correspond to a physiological reality (Vaissière, 2001). Cross-linguistic studies should involve linguists who master the phonological (segmental and prosodic) systems of their respective languages.

Seventh, the interaction between the different types of intonation inside a language calls for more detailed exploration. At present, very few studies venture into the field of multiple-parameter interactions between structural and affective aspects of intonation (see, however, Pell, 2001). Thorough investigations must be conducted to test out the elements of the psychophonetic code and their interactions.

An important assumption that runs throughout the present chapter is that language-specific traits of intonation can only be understood within the context of universal principles. Studies are necessary to explore the psycho-phonetic code and reveal the true *common denominators*. Collaboration between specialists of various domains becomes essential: psychoacousticians, psycholinguists, physiologists, specialists of ontogenesis and phylogenesis, and others. If one takes a narrow view of studies on intonation and its perception, it has to be acknowledged that most of the literature in this field actually consists of case studies. There is a long way to go before firm and solid conclusions can be reached. From what has been said about language-specificity, it is a natural conclusion that successful experiments in perception have to involve listeners whose native languages differ greatly to avoid unwarranted conclusions; investigations that confine themselves only to the languages of Europe are limited in scope from the outset. Languages from Africa and Asia that possess widely varied tonal systems should by all means be included. This appears as a major challenge for future research on the perception of intonation.

Of all dimensions of speech, is intonation the most difficult to study? Young children (and, in fact, dogs) know how to decode much of intonational meaning, and yet no existing speech synthesis system can be said to be able to reproduce attitudinal and emotional nuances carried by intonation. The problem of how intonation works is still very far from being solved. One has to remember Bolinger's word to the wise: intonation is a half-tamed savage (Bolinger, 1978: 475). If, as it seems, the complexity of intonation is typical of human complexity, then there is still a long way to go before perception of intonation yields all of its secrets.

Acknowledgements

Un grand merci to Ilse Lehiste and Ivan Fonagy for their careful reading and suggestions. Many thanks to Alexis Michaux, Klaus Kohler, Cécile Fougeron and Shinji Maeda for their useful suggestions on former versions of this chapter.

References

Abelin, Å. & J. Allwood. Cross-linguistic interpretation of emotional prosody. *ISCA Workshop on Speech and Emotion: A conceptual framework for research*, Belfast, (2000).

Ainsworth, W. A. "Pitch change as a cue to syllabification." Journal of Phonetics 14-2: 257-264, (1986).

Ainsworth, W. A. & D. Lindsay. "Perception of pitch movement on tonic syllables in British English." Journal of the Acoustical Society of America 79(2): 472-80, (1986).

Allen, G. D. "Speech rhythm: its relation to performance universals and articulatory timing." Journal of Phonetics 3: 75-86, (1975).

Auer, E., L. Bernstein, et al. "Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: an initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals." Journal of the Acousitic Society of America 104(4): 2477-89, (1998).

Bartels, C. & J. Kingston. "Salient pitch cues in the perception of contrastive focus." Journal of the Acoustical Society of America 95(5): 2973, (1994).

Batliner, A. "Deciding upon the Relevancy of Intonational Features for the Marking of Focus: A Statistical Approach." Journal of Semantics 8(3): 171-189, (1991).

Beach, C. M., A. Ganguly, et al. "Children's use of prosody to identify phrasal units."?? 90(4): 2297, (1991).

Beaugendre, F., D. House, et al. Accentuation boundaries in Dutch, French and Swedish. *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications*, Athens, Greece, (1997).

Benguerel, P. "Corrélats physiologiques de l'accent en francais." Phonetica 27(1): 21-35, (1973).

Berinstein, A. E. Cross-linguistic study on the perception and production of stress, UCLA Working Papers in Phonetics 47, Los Angeles, (1979).

Berkovits, R. "Perception of intonation in native and non-native speakers of English." Language and Speech 23(3): 271-280, (1980).

Beun, R. J. "The recognition of Dutch declarative questions." Journal of Pragmatics 14(1): 39-56, (1990).

Bolinger, D. Intonation as a universal. *Proc. 9th Int. Congr. Linguistics*, Mouton, The Hague, (1964).

Bolinger, D. *Intonation and its use: Melody in Grammar and Discourse*, Standford University Press, (1989).

Bolinger, D. L. *Intonation and its Parts: melody in spoken English*. Standford., Standford University Press, (1986).

Brokx, J. P. L. & S. G. Nooteboom. "Intonation and the perceptual separation of simultaneous voices." Journal of Phonetics 10(1): 23-36, (1982).

Brown, B. L., W. J. Strong, et al. "Perceptions of personality from speech: Effects of manipulations of acoustical parameters." Journal of the Acoustical Society of America 54 (1): 29-35, (1973).

Capelli, C. A., N. Nakagawa, et al. "How children understand sarcasm: the role of context and intonation." Child Development 61(6): 1824-1841, (1990).

Caspers, J. "Who's next? the melodic marking of question versus continuation in Dutch." Language and Speech 41(3-4): 375-398, (1998).

Caspers, J. "An experimental investigation of meaning differences between the 'early' and the 'late' accent-lending fall in Dutch." Linguistics in the Netherlands 16: 27-39, (1999).

Cohen, A. & J. t'Hart. "On the anatomy of intonation." Lingua 19, (1967).

Cohen, H., J. Douaire, et al. "The role of prosody in discourse processing." Brain Cogn 46(1-2): 73-82, (2001).

Collier, R. "Physiological correlates of intonation patterns." Journal of the Acoustical Society of America 58(1): 249-255, (1975).

Collier, R. Some physiological and perceptual constraints on tonal systems. *Explanations for Language Universals*. B. Butterworth, B. Connie and O. Dahl. Berlin, New York, Amsterdam., Mouton: 237-248, (1984).

Collier, W. & T. Hubbard. "Musical scales and evaluations of happiness and awkwardness: effects of pitch, direction, and scale mode." American Journal of Psychology 114(3): 355-75, (2001).

Cutler, A. Prosody and the structure of message. *Computing prosody : Computational models for processing spontaneous speech*. Y. Sagisaka, N. Campbell and N. Higuchi, Springer Verlag: 63-66, (1997).

Cutler, A., D. Dahan, et al. "Prosody in the comprehension of spoken language : A litterature review." Language and Speech 40(2): 141-201, (1997).

Danes, F. "Sentence intonation from a functional point of view." Word 16(1): 34-54, (1960).

Darwin, C. On the dynamic use of prosody in speech perception. *Structure and Process in Speech Perception*. A. Cohen and S. G. Nooteboom. Berlin, Springer Verlag: 178-195, (1975).

De Pijper, J. R. & A. A. Sanderman. "On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues." Journal of the Acoustical Society of America 96(4): 2037-2048, (1994).

Duez, D. "Acoustic Correlates of Subjective Pauses." Journal of Psycholinguistic Research 22(1): 21-39, (1993).

Fonagy, I. Emotions, voice and music. *Research aspects on singing*. J.Sundberg. Stockholm. Research aspects on Singing: 51-79, (1981).

Fonagy, I. *La vive voix: essais de psycho-phonétique*. Paris, Payot, (1983).

Fonagy, I. *Languages within languages: an evolutive approach*. Amsterdam, Philadelphia, John Benjamins, (2000).

Fonagy, I. & E. Bérard. ""Il est huit heures": contribution à l'analyse sémantique de la vive voix." Phonetica 26: 157-192., (1972).

Fonagy, I. & J. Fonagy. Analysis of Complex (Integrated) Melodic Patterns. *In Honor of Ilse Lehiste*. R. Channon and L. Shockey. Dordrecht - Holland, Foris Publications**:** 75-97, (1987).

Fujisaki, H. & H. Sudo. "A generative model for the prosody of connected speech in Japanese." Ann. Res. Eng. Inst. Logopedics Phoniatrics, Tokyo 30: 75-80, (1971).

Gandour, J., D. Wong, et al. "A cross-linguistic fMRI study of perception of intonation and emotion in Chinese." Human Brain Mapping 18: 149 –157, http://www.anatomy.iupui.edu/~wong/Chinese3_HBM_03.pdf, (2003).

Geluykens, R. & M. Swerts. "Prosodic cues to discourse boundaries in experimental dalogues." Speech Communication 15(1-2): 69-77, (1994).

Gosy, M. & J. Terken. "Question Marking in Hungarian: Timing and Height of Pitch Peaks." Journal of Phonetics 22(3): 269-281, (1994).

Grabe, E., B. Post, et al. "Pitch Accent Realization in Four Varieties of British English." Journal of Phonetics 28(2): 161-185, (2000).

Grant, K. W. "Identification of intonation contours by normally hearing and profoundly hearing-impaired listeners." Journal of the Acousitic Society of America 62(4): 1172-&&è8, (1987).

Gussenhoven, C. Intonation and Interpretation: Phonetics and Phonology. *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, France, http://www.lpl.univ-aix.fr/sp2002/pdf/gussenhoven.pdf, (2002).

Gussenhoven, C., B. H. Repp, et al. "The perceptual prominence of fundamental frequency peaks." Journal of the Acoustical Society of America 102(5): 3009-22, (1997).

Gussenhoven, C. & A. C. M. Rietveld. "Intonation Contours, Prosodic Structure and Preboundary Lengthening." Journal of Phonetics 20(3): 283-303, (1992).

Gussenhoven, C. & T. Rietveld. "On the speaker-dependence of the perceived prominence of F0 peaks." Journal of Phonetics 26(4): 371-380, (1998).

Hadding-Koch, K. Acoustic-phonetic studies in the intonation of Southern Swedish. Lund, Travaux de l'Institut de Phonétique de Lund III, C.W.P. Gleerup, (1961).

Hadding-Koch, K. & M. Studdert-Kennedy. "An experimental study of some intonation contours." Phonetica 11: 175-185, (1964).

Henderson, A. I. & S. Nelms. "Relative salience of intonation fall and pause as cues to the perceptual segmentation of speech in an unfamiliar language." Journal of Psycholinguistic Research 9(2): 147-159, (1980).

Hirschberg, J. & M. Swerts. "Prosody and conversation." Language and Speech (special issue) 41(3-4), (1998).

Hirschberg, J. & G. Ward. "The influence of pitch, range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise pattern intonation contour in English." Journal of Phonetics 20: 241-251, (1992).

Hirst, D. & A. Di Cristo. *Intonation Systems: A Survey of Twenty Languages*. AI&SI, England, Cambridge University Press, U.K., (1998).

House, D. *Tonal Perception in Speech*. Lund, Sweden, Lund University Press, (1990).

House, D. Perceiving question intonation: the role of pre-focal pause and delayed focal peak. *Int. Cong. of Phon. Sc.*, Barcelona, (2003).

Hyman, L. On the nature of linguistic stress. *Studies in stress and accent*. L. Hyman. University of Southern California, Southern California Occasional Papers in Linguistics. 4, (1977).

Ito, M. Japanese Politeness and Suprasegmentals: A Study based on Natural Speech Materials. *Speech Prosody 2002*, Aix-en-Provence, France, (2002).

Jusczyk, P., K. Hirsh Pasek, et al. "Perception of acoustic correlates of major phrasal units by young infants." Cognitive Psychology 24(2): 252-293, (1992).

Jusczyk, P. W., A. Cutler, et al. "Infants' preference for the predominant stress patterns of English words." Child Development 64(3): 675-687, (1993).

Kim, S., E. Curtis, et al. "Emphatic Koreans and neutral Americans?" Journal of the Acoustical Society of America 109(5): 2474-2475, (2001).

Kohler, K. "Prosody in speech synthesis: the interplay between basic reserach and TSS application." Journal of Phonetics 19: 121-138, (1991).

Ladd, D. R. *Intonational phonology*, Cambridge University Press, (1996).

Ladd, D. R. & R. Morton. "The perception of intonational emphasis: continuous or categorical?" Journal of Phonetics 25(3): 313-342, (1997).

Ladd, D. R., K. Silverman, et al. "Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect." Journal of the Acoustical Society of America 78(2): 435-444, (1985).

Ladefoged, P., M. Draper, et al. "Syllables and stress." Miscellanea Phonetica 3: 1-14, (1958).

Lan, G. P. M. "The Contribution of Intonation, Segmental Durations, and Spectral Features to the Perception of a Spontaneous and a Read Speaking Style." Speech Communication 22(1): 43-65, (1997).

Lansing, C. & G. McConkie. "Attention to facial regions in segmental and prosodic visual speech perception tasks." J Speech Lang Hear Res 42(3): 529-39, (1999).

Lea, W., Ed. *Prosodic Aids to Speech Recognition*. Trends in Speech Recognition. Englewood Cliffs, NJ, Prentice-Hall, (1980).

Lehiste, I. *Suprasegmentals*. Cambridge, (1970).

Lehiste, I. The Phonetic Structure of Paragraphs. *Structure and Process in Speech Perception*. A. Cohen and S. G. Nooteboom. Berlin, Springer Verlag: 195-206, (1975).

Lehiste, I. "Phonetic manifestation of syntactic structure in English." Ann. Bull. RILP 14: 1-27, (1980).

Lehiste, I. & R. A. Fox. "Perception of prominence by Estonian and English listeners." Language and Speech 35(4): 419-434, (1992).

Lehiste, I., J. Olive, et al. "Role of duration in disambiguating syntactically ambiguous sentences." Journal of the Acoustical Society of America 60-5: 1199 - 1202, (1976).

Lieberman, P. *Intonation, perception and language*. Cambridge, MIT Press, (1967).

Makarova, V. "Perceptual Correlates of Sentence-Type Intonation in Russian and Japanese." Journal of Phonetics 29(2): 137-154, (2001).

McClean, M. D. & W. R. Tiffany. "The acoustic parameters of stress in relation to syllable position, speech looudness and rate." Language and Speech 16(3): 283-290, (1973).

Moore, C., L. Harris, et al. "Lexical and prosodic cues in the comprehension of relative certainty." J Child Lang 1993(20): 1, (1993).

Morton, E. S. Sound symbolism and its role in non-human vertebrate communication. *Sound symbolism*. L. Hinton, J. Nichols and J. Ohala. Cambridge, England, Cambridge University Press: 348-365, (1994).

Most, T. & Y. Frank. "The effects of age and hearing loss on tasks of perception and production of intonation." Volta Review 96(2): 137-149, (1994).

Murray, J. R. & J. L. Arnott. "Towards 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, (1993).

Nakajima, S. & J. F. Allen. "A study on prosody and discourse structure in cooperative dialogues." Phonetica 50: 197-210, (1993).

Nakatani, L. H. & J. A. Schaffer. "Hearing 'words' without words: prosodic cues for word perception." Journal of the Acousitic Society of America 63(1): 234-245, (1978).

Nooteboom, S. G. Production and perception of vowel duration. Doctoral dissertation, Utrecht, (1972).

Nooteboom, S. G., J. P. L. Brokx, et al. "Contributions of Prosody to Speech Perception." IPO Annual Progress Report 11: 34-55, (1976).

Ofuka, E., J. D. McKeown, et al. "Prosodic Cues for Rated Politeness in Japanese Speech." Speech Communication 32(3): 199-217, (2000).

Ohala, J. J. "An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice." Phonetica 41(1): 1-16, (1984).

Ohman, S. E. G. Word and sentence intonation. A quantitative model. Quarterly Progress and Status Report. Stockholm, Speech Trans. Lab.: 20-54, (1967).

Pell, M. "Influence of emotion and focus location on prosody in matched statements and questions." Journal of the Acoustical Society of America 109(4): 1668-1680, (2001).

Pierrehumbert, J. "The perception of fundamental frequency declinaison." Journal of the Acoustical Society of America 66: 363-369, (1979).

Pierrehumbert, J. The phonology and phonetics of English intonation. Cambridge, USA, MIT, distributed by Indiana Linguistic Club, (1980).

Pierrehumbert, J. & J. Hirschberg, Eds. *The Meaning of Intonation in the Interpretation of Discourse*. Intentions in Communication. Cambridge MA, MIT Press, (1990).

Pisoni, D. B. & P. A. Luce. "Acoustic-Phonetic Representations in Word Recognition." Cognition 25(1-2): 21-52, (1987).

Pisoni, D. B. & P. A. Luce. Trading relations, acoustic cue integration, and context effect in speech perception. *The psychophysics of speech perception*. M. E. H. Schouten. Dordrecht, Boston, Lancaster, Martinus Nijhoff Publishers, (1987).

Price, P. J., M. Ostendorf, et al. "The use of prosody in syntactic disambiguation." Journal of the Acoustical Society of America 90-6: 2956-2970, (1991).

Repp, B. H. "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception." Psychological Bulletin 92(1): 81-110, (1982).

Rialland, A. "Le fini/infini ou l'affirmation/l'interrogation en moba (langue voltaïque parlée au Nord-togo." Studies in African Linguistics Supplement 9: 258-261, (1985).

Rosenblum, L. & C. Fowler. "Audiovisual investigation of the loudness-effort effect for speech and nonspeech events." J Exp Psychol Hum Percept Perform 17(4): 976-85, (1991).

Rossi, M. "Interactions of Intensity Glides and Frequency Glissandos." Language and Speech 21(4): 384-396, (1978).

Rossi, M. *L'intonation, le système du français: description et modélisation*, Orphrys, (1999).

Sanderman, A. A. & R. Collier. "Good prosody facilitates comprehension." Journal of the Acoustical Society of America 100(4): 2823, (1996a).

Sanderman, A. A. & R. Collier. "Prosodic rules for the implementation of phrase boundaries in synthetic speech." Journal of the Acoustical Society of America 100(5): 3390-3397, (1996b).

Schaffer, D. B. "The role of intonation as a cue to topic management in conversation." Journal of Phonetics 12(4): 327-344, (1984).

Scherer, K. "Vocal affect expression : a review and a model for future research." Psychological bulletin 92(2): 143-165, (1994).

Shattuck Hufnagel, S. & A. E. Turk. "A Prosody Tutorial for Investigators of Auditory Sentence Processing." Journal of Psycholinguistic Research 25(2): 193-247, (1996).

Sluijter, A. & V. Van Heuven. "Spectral balance as an acoustic correlate of linguistic stress." Journal of the Acoustical Society of America 100(4): 2471-2485, (1997).

Streeter, L. A. "Acoustic Determinants of Phrase Boundary Perception." Journal of the Acoustical Society of America 64-6: 1582-1592, (1978).

Streeter, L. A., N. H. MacDonald, et al. "Acoustic and perceptual indicators of emotional stress." Journal of the Acoustical Society of America 73(4): 1354-1361, (1983).

Swerts, M. "Prosodic features at discourse boundaries of different strength." Journal of the Acoustical Society of America 101(1): 514-521, (1997).

Swerts, M., D. G. Bouwhuis, et al. "Melodic cues to the perceived "finality" of utterances." Journal of the Acoustical Society of America 96(4): 2064-2075, (1994).

Swerts, M. & R. Geluyens. "The prosody of information units in spontaneous monologue." Phonetica 50: 189-196, (1993).

Terken, J. "Fundamental frequency and perceived prominence of accented syllables." Journal of the Acousitic Society of America 95(6): 3662-3665, (1994).

Terken, J. & S. G. Nooteboom. "Opposite effects of accentuation and desaccentuation on verification latencies for given and new information." Language and Cognitive processes 2(3/4): 145-163, (1987).

t'Hart, J., R. Collier, et al. *A Perceptual Study of Intonation*, Cambridge University Press, (1990).

Thorsen, N. "A study of the perception of sentence intonation- evidence from danish." Journal of the Acoustical Society of America 67: 1014-1030, (1980).

Uldall, E. T. Dimensions of meaning in intonation. *In Honour of Daniel Jones.* D. Abercrobie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott and J. L. M. Trim. London, Longmans**:** 271-279, (1964).

Vaissière, J. Language-Independent Prosodic Features. *Prosody: Models and Measurements.* A. Cutler and D. R. Ladd. Berlin, Springer-Verlag**:** 53-66, (1983).

Vaissière, J. Rhythm, accentuation and final lengthening in French. *Music, Language, Speech and Brain.* J. Sundberg, L. Nord, R. Carlson. and W.-G. I. S. Series, Macmillan Press. 59**:** 108-120, (1991).

Vaissière, J. "Phonetic explanations for cross-linguistic similarities." Phonetica 52: 123-130, (1995).

Vaissière, J. "Changements de sons et changements prosodiques : du latin au français." Parole 15/16(53-88), (2001).

Vaissière, J. Cross-linguistic prosodic transcription: French versus English. *Problemy i metody eksperimental'no-foneticheskih issledovanij, In honour of the 70th anniversary of Prof. L.V. Bondarko,.* N. B. Volslkaya, N. D. Svetozarova and P. A. Skrelin. St.-Petersburg, St.-Petersburg State University**:** 147-164, (2002).

Van Els, T. J. M. & K. DeObt. "The role of intonation in foreign accent." Modern Language Journal 71(2): 147-155, (1987).

Verhoeven, J. "The discrimination of pitch alignment in Dutch." Journal of Phonetics 22: 65-85, (1994).

Verma, S., A. M. Mannering, et al. "Prosodic cues for phrasal boundaries in productions by children and adults." Journal of the Acoustical Society of America 96(5): 3308, (1994).

Ward, G. & J. Hirschberg. "Intonation and propositional attitude: the pragmatics of L*+H L H%." Proceedings Eastern States Conference on Linguistics (ESCOL) 5: 512-522, (1988).

Watanabe, K. "Sentence Stress Perception by Japanese Students." Journal of Phonetics 16(2): 181-186, (1988).

Wichmann, A. Attitudinal intonation and the inferential process. *Speech Prosody 2002*, Aix-en-Provence, http://www.lpl.univ-aix.fr/sp2002/pdf/wichmann.pdf, (2002).

Williams, C. E. & K. N. Stevens. "Emotion and speech: some acoustical correlates." Journal of the Acoustical Society of America 52(2): 1238-1250, (1972).

[i] Due to the lack of a standardized method, a **large array of methods** are used in the study of the specific role of intonation in speech perception.

a) For the construction of the stimuli, manipulation can be done by building a corpus of semantically unpredictable sentences. Speakers (some favor actors) can be asked to vary phrasing, accentuation, modalities, social attitudes or emotions on one and the same sentence, to hum the utterances, or to mimic them by using reiterant speech and nonsense syllables (Liberman and Streeter, 1978). Cross-splicing allows to move part of an utterance into a new context. Degradation and delexicalisation of stimuli are done by band-pass filtering, rotating speech, spectral scrambling, noise addition or time-compression or using laryngographic recordings (content-filtered speech). There are now extensive possibilities for well controlled manipulation by computer of one prosodic parameter at a time, and to copy an intonation contour from one utterance onto another.

b) The listener may also be asked to perform a wide range of tasks: to discriminate stimuli; to estimate the perceptual distance between two contours; to assess the excursion size of a pitch movement in a comparison test; to perform by computer perceptual close-copy stylization of Fo contours (in the IPO style); to shadow the utterance; to transcribe prosodically the utterances (taken in or out of context), with or without access to the actual Fo curves; to localize clicks in sentences where intonation has been manipulated; to formulate judgments of prosodic appropriateness, felicity and naturalness on anomalous, ill-formed, cross-spliced prosodic contours or to give a pragmatic interpretation. Reaction times may be measured in all cases.

c) One of the main difficulties is to infer the level of listening used by the listeners. At the lowest, most concrete, psychoacoustic level, the naïve ear can be trained to perform an analytic transcription of each successive syllable (as high, mid, low, rising, flat or falling, etc.). As the size of the window analysis increases, the listener can locate intersyllabic discontinuous tone-steps (jumps), and intersyllabic distances. When it encompasses one or several words, he may detect (lexical) stresses and pitch accent, narrow focus, estimate the strength of the perceived boundary between the successive words, and derive the local syntactic structure (right or left branched phrase). When the subject listens to the whole utterance in a broad fashion, he is able to recognize the speaker's communicative intention from the linearized speech signal (t'Hart, J. & R. Collier. "Integrating different levels of intonation analysis." Journal of Phonetics 3: 235-255, (1975): (i) the structural aspects of the utterance (its syntactic *phrasing* and its informational structure)*,* (ii) its *modalities,* (iii) its affective aspects (*social attitudes* and *emotion)* and (iv) discursive aspects (*turn-taking* and change of topics). An important reminder is that in modifying one parameter only, experiments leave all the other traits of the original utterance unchanged, and some traits may entertain a trading relationship with Fo**.** The listeners may also be disturbed by the lack of the usual accompanying cues.

[ii] For a number of reasons linked to perception (an a higher sensitivity to durational difference on the penultimate (Lehiste, I. "The perception of duration within sequences of four intervals." Journal of Phonetics 7: 313-316, (1979)) and/or the necessity of a tail for Fo contrasts (Bolinger, 1978), the penultimate syllable is a favored position for stress (after the initial and final position) (see also Hyman, 1977).

[iii] Note that contrasts in melodicity are just suppressed in synthesized Fo !!!