

Speaking styles and Phonetic variation

Term paper autumn 2004
Speech Technology level 1

Ulla Bjursäter (ullabj@ling.su.se)
Department of Linguistics, Stockholm university

ABSTRACT

Human speech is a very dynamic phenomenon with nearly endless forms of variations. Different factors affect the speech signal; this paper aims at giving a short overview of different speaking styles and phonetic variations that affects human speech production and perception. Various topics are mentioned, such as age and gender, sound symbolism, speaking styles, emotions, universal features, voice quality and prosodic aspects. A conclusion can be made on the importance of integrating all these aspects in the research concerning human-computer interaction.

Introduction

When we produce speech, the vocal organs do not move from sound to sound in a series of separate steps. Instead, speech is a continuously varying process, and sounds continually influence their neighbours (Crystal, 1997). We do not discern speech as discrete units, all the units are integrated and we perceive a holistic representation of the speech signal.

Speech has to be easy to produce but also easy to distinguish, to maintain distinctive linguistic/phonetic aspects. When we talk, the speech sound can be assimilated and reduced without greater misunderstandings because of the redundancy and context. Some aspects that are of importance are different prosodic aspects and paralinguistic phenomena. Prosody plays a very important role in human-computer interaction as well as in the communication between human beings (Hirschberg, 2002). The more we know of a language, the more we understand, despite great reductions in the speech signal. If we as listeners have limited knowledge of a language, we depend on the speaker leaving as much information as possible in the speech signal (Lundström-Holmberg & af Trampe, 1987).

Some of the difficulties in speech perception and subsequently also in automatic speech recognition, are the great variability of speech production and perception. Some important factors of the comprehensibility are the speaker's age, gender, anatomy and also the dialect, idiolect and sociolect of the speaker. The speech can be affected by the emotional status and health condition and also if the person is much stressed up. The speaking style is also important for the intelligibility. Speakers tend to vary their articulatory speech output from hyper- to hypospeech (Lindblom, 1990). Clear speech is produced in an effort to be highly intelligible, and is relatively easy perceived by the listener. Speech can also be produced on a spontaneous, conversational basis, with

reductions, non-grammatical sentences and hesitations that may affect the intelligibility of the utterance. A listener can have difficulties understanding what is said if they listen to a non-native language or if they don't know the person that is speaking. Bad hearing and different age problems can also affect the comprehension. When people grow older, the speaking habits are more set and it can be difficult to accept new lexical words and the change of meaning in familiar lexical words.

This paper aims at giving a short overview of different speaking styles and phonetic variations that affects our speech production and perception.

Speaking styles

The ways we express ourselves vary from one situation to next, depending on the context and speaker intentions. A speaker might vary his or hers speaking style depending on the listener and the situation, using different expressions, pronunciations and tone of voice. The manner of pronunciation is quite strict when a more formal speaking style is used; the talker makes an effort to be easily understood by modifying the articulation to make speech slower and acoustically more distinctive (Kent & Read, 1992). These are also the characteristics of clear speech. These forms are often used when reading out loud. There is, of course, a great difference between reading a formal text out loud and reading fairy-tales to children. In reading a formal text out loud in public, the need of comprehension influences both articulation and prosodic aspects, while reading a fairy-tale there is a different demand on the speaking style, as it addresses a child or a group of children and the read speech may often include passages of spontaneous speech.

There is quite a difference between public speech and spontaneous speech. When speaking in public, more precise articulations are produced in contrast to spontaneous speech. When the speaker is more comfortable in a conversation, the speaking style turns into a more casual one, with simplified words and phrases. This is when we start to produce reductions, assimilations and coarticulations. Speech rate is increased compared to clear speech and so the quantity of reductions also increases. The reductions are dependent on where the stress lies in the word. Stressed words or syllables are not reduced as much as an unstressed and they are usually not completely reduced, as might be the case with unstressed syllables or words (Lundström-Holmberg & af Trampe, 1987).

Humans tend to have a certain way of speaking to their infants. The syntax is simpler and the prosodic organization seems to be almost universal with a few culturally based variations, with higher pitch, slower tempo and enhanced intonation as a standard. Even speaking with a whispering voice is not uncommon when communicating with young infants (Fernald & Simon, 1984).

We gradually change our speaking style as the child grows and adjust the communication according to the child's linguistic/developmental level. When the children develop into youths, they often try to find new ways of expressing themselves as a form of differentiation from their elders and sometimes also from other groups of people in their own generation, thus forming their own idiolects and sociolect.

Every speaker has a personal idiolect that differs from people with the same dialect. We use our voice and speaking style, intentionally or unintentionally, to mark our personality. Our sociolect is a form of social dialect, it tells people who and what we are

in more or less hierarchically ordered social groups. It gives the information of social position and educational level. There is a cultural variation in the norms and rules for accepted behaviour. People usually speak to higher-status people in the respectful way used when speaking to strangers, while lower-status people are addressed in the more intimate, first-name way similar to when speaking to friends. The way we address other people indicates our social distance and social status in relation to the other person (Myers, 2002).

Sound symbolism

There are some speech sounds that are conscious, non-arbitrary and iconized forms of speech. This can be termed “sound symbolism” and is a direct linking between speech sound and meaning (Hinton, Nichols & Ohala, 1994). This depicting phenomenon gives indirect associations and a universal pattern can be detected, with a common base that is realized in different language specific patterns. In “*mama*” and “*papa*”, imitations may originate from an infant’s spontaneous vocalization of CV-syllables (Jacobson, 1962). Humans produce imitative animal sounds in an onomatopoeic way (*pip pip*), imitations of sounds from nature (*swisch*), and imitations of sounds originating from the human being (*gurgle, mumble, babble*) (Hinton, Nichols & Ohala, 1994; Traunmüller, 2000).

Emotions

Our voice is greatly affected by our emotional reactions. The emotional status of a speaker can be revealed by several acoustical aspects. Mozziconacci (1995) studied pitch variations and different emotions (neutral, joy, boredom, anger, sadness, fear, indignation) in speech. She found that measurements of this single aspect were not enough to establish the speaker’s emotional status. There are other studies investigating different aspects in emotional speech, like spectral and temporal changes (Kienast & Sendlmeier, 2000). Table 1 contains a listing of the acoustical correlates of the four emotions “happy”, “sad”, “angry” and “afraid” from observations of some prosodic characteristics of speech, in vowels and fricatives in relation to neutral speech (Murray & Arnott, 1993).

Table 1: Acoustical correlates of the emotions happy, sad, angry and afraid.

	Speech rate	Variation of fundamental frequency	Intensity	Vowels	Fricatives
Happy	Varying speech rate	Big F ₀ variation	Elevated	Raised F ₁ , a little raised F ₂	Spectral balance, a lot of energy at high frequencies
Sad	(Low F ₀ -variation) + reductions and assimilations	Low F ₀ variation	Low	Less periphery formant frequencies	Lower spectral balance compared to neutral speech
Angry	Increased speech rate (often)	Elevated F ₀ variation	High (especially at stress)	Raised F ₁ , little raised F ₂	Spectral balance, a lot of energy at high frequencies
Afraid	Increased speech rate (often)	High pitch, lower F ₀ variation	High intensity (+jitter, at extreme fear)	Less periphery formant frequencies	Spectral balance, a lot of energy at high frequencies

Acoustic analyses indicate that smiling raises the fundamental and formant frequencies for all speakers, and also the amplitude and/or duration for some speakers. The elevated formant frequencies might be a side effect from altering the vocal tract by spreading the lips and drawing the corners of the mouth backwards, a procedure that shortens the vocal tract (Tartter, 1980).

According to Künzel (2000) the general increase of stress and nervousness, like for example during a crime, tend to raise his or hers F_0 values during the “crime situation”. It is important to mention that there are of course culture- and language-specific differences in the listener’s interpretation of the emotion (Scherer, Banse & Wallbott, 2001). It can be hard to correctly identify different emotions because a variety of interacting variables may manifest themselves in a complex way.

Prosodic aspects

Prosody has, as mentioned earlier, a very important role in human-computer interaction as well as in the communication between human beings (Hirschberg, 2002). Prosodic aspects of a language are a collection of linguistic/phonetic effects, like tonal, temporal and dynamic aspects. The most significant prosodic effects in a language’s intonation system are provided by the linguistic use of pitch. Different levels of pitch (tones) are used in particular sequences to convey a wide range of meanings. For instance, the difference between a falling and a rising pitch pattern can express the contrast between “stating” and “questioning”. Duration is another prosodic parameter. Variations in the temporal rate at which syllables, words and sentences are produced convey different kinds of meaning. In several languages, a sentence spoken with increased tempo conveys urgency; while slower speed conveys deliberation or emphasis (Crystal, 1997). Another significant prosodic aspect is intensity, which is used to convey differences of emotional aspects, such as the increased volume usually associated with anger. Intensity is also used to express divergence in lexical aspects as in terms of the contrast heard in the different syllables in a word. Syllabic intensity is usually referred to as stress, but the term “accent” is often used (accented vs. unaccented) referring to the way prominence manifests in loudness as well as pitch (Crystal, 1997).

Tonal aspects

Intonation, variation in tone, present a variety of different functions. One obvious function is to express emotions. Intonation co-varies with other prosodic and paralinguistic aspects to mark all kinds of emotional expression (Crystal, 1997). An expressive intonation pattern can also be used in a synesthetical¹ way, like in the use of a deep voice and vowel lengthening in speaking of large objects as in “It was a bi-i-ig fish” (Hinton, Nichols & Ohala, 1994). Intonation also plays an important role in the marking of grammatical contrast. Pitch contours break up utterances, which facilitates comprehension. Statements and questions or positive or negative intentions may be signalled by intonation. Intonation conveys information structure with the intonation prominence; there is a big difference of meaning in the way we say “*I like fish*”, the prominence can land on either “*I*”, “*like*” or “*fish*”, meaning different things (Crystal, 1997). A language can also contain minimal pairs that contrast only in word accent. In

¹ Syntesthetic sound symbolism is the process whereby certain speech sounds are chosen to consistently represent visual, tactile, or properties of objects, such as size and shapes (Hinton, Nichols & Ohala, 1994).

Swedish, for example, the sentence “*Den här tomten är bra*” means either “*This site is fine*” or “*This goblin is fine*”, depending on the accentual pattern of “*tomten*” (Crystal, 1997). Monotonous intonation can be a sign of language retardation in children (Nettelbladt, 1997).

Intonation can also be used distinctively in read speech. Textual information is divided in larger stretches of paragraphs, when you read a text out loud a distinctive melodic shape may give information. When a new item is read, the pitch level rises only to gradually descend as you continue to read on. The use of intonation also can help organize language into units that are easier to recognize and memorize, like listening to a long sequence of numbers. This is an aspect that may be missing in some cases of language disorder. Intonation can have a significant function as a marker of personal identity in an “indexical” function, as it can help to identify people as belonging to different social groups and various occupations (Crystal, 1997).

Temporal aspects

Temporal aspects may also reflect various attitudes and emotions of the speaker (Lundström-Holmberg & af Trampe, 1987). Temporal aspects in prosody have two functions; quantity and juncture. Several kinds of meaning are conveyed by variations in the temporal rate at which syllables, words and sentences are produced. In *final lengthening*, the duration in last part of an utterance is extended as an indication of the utterance coming to an end. In many languages, variations of the length of the segment are used to make a difference in meaning, such as in Swedish where the use of quantity creates long and short vowels. There are also long and short consonants, depending on the quantity of the vowel. If the vowel is long - the consonant is short and vice versa (V:C / VC:). Another temporal function is juncture, that might manifest itself through audible pauses, but more often it is just short closures of the air flow or extensions of certain segments (I scream vs ice cream) (Crystal, 1997).

Intensity

Production, acoustics and perception

Intensity is dependent on variation in vocal effort controlled by the respiratory muscles. Syllabic and phrase intensity is usually referred to as stress, but the term accent is often used (Crystal, 1997). In Finnish, main stress is fixed on the first syllable, while in French main stress always fall on the last syllable. Other languages, like Swedish and English, might have stress that fall differently depending on whether the word is a noun (*‘import, ‘pervert*) or a verb (*im ‘port, per ‘vert*). Stress may also convey a difference of meaning on phrase level (*‘sleep in* or *sleep ‘in*).

Intra-speaker variations in vocal effort creates various degrees of loud and soft speech. This affects the production and, subsequently, the acoustics. Articulation changes when intensity is raised. In vowel production, the opening of the jaw increases and the lips and the tongues movements are compensating with necessary, extreme movements. The duration also increases in relation to the openness of the vowel; the more open the jaw- the longer the vowel duration. With consonants, the place and manner of articulation generally remains unchanged, but hypertension in the muscle tonus may occur. The vocal folds tense and a higher subglottal pressure occur. This also influences vocal fry, which decreases with increasing intensity (Shulman, 1989). Increased rate of articulation

produces shorter consonants and make stressed vowels longer, the total segmental duration to remains practically unchanged due to duration compensation. Adults take longer pauses as they need more air in this production of amplified intensity (Shulman, 1989).

Acoustically, increased articulation effort affects the fundamental frequency and formant frequencies. The F_0 value increases as a function of increased vocal intensity, the formant frequencies (especially F_1) shifts upwards, thus facilitating a correct phonetic identity. Independent of the vowel's degree of opening, the phonetic vowel identity remains unchanged when the tonotopical distance between F_1 and F_0 (in bark) is constant. Perceptually, the formant positions are evaluated in relation to each other and to F_0 . The listener also catches the information of a voiced/voiceless consonant by information from F_1 . Increased intensity also gives increased spectral emphasis at the higher formant frequencies (Traunmüller, 1988).

Voice Quality

Paralinguistic features

Apart from the contrasts signalled by tone, tempo and intensity, languages make use of several distinctive vocal effects, using the range of articulatory possibilities of the vocal tract. The laryngeal, pharyngeal, oral and nasal cavities can all be used to produce “tone of voice” which may alter the meaning of the utterance. One of the clearest examples of paralinguistic aspects is whispered speech, which is used in many languages to add “conspiratorial” meaning to the utterance (Crystal, 1997).

There are different dimensions of voice quality. Important voice quality factors are the laryngeal conditions and articulation habits. Voice qualities originate from the larynx, where the character of the speech material is produced to render different laryngeal qualities such as vocal fry (creaky), strained voice and breathy voice (Lindblad, 1992). Vocal fry is caused by strong, irregular, relatively low vocal fold pulses. A breathy voice occurs when the edges of the vocal folds do not quite close when vibrating. In a falsetto voice, the sound is produced by long, thin and tense vocal folds. It is very hard to control intensity and pitch when using falsetto voice (Lindblad, 1992).

Table 2: Various combination possibilities of different phonation types according to Laver (1980)

	Modal voice	Falsetto	Breathy voice	Whisper	Creaky, Vocal fry	Rough
Modal voice	x	-	+	+	+	+
Falsetto	-	x	-	+	+	+
Breathy voice	+	-	x	-	-	-
Whisper	+	+	-	x	-	-
Creaky, Vocal fry	+	+	-	-	x	-
Rough	+	+	-	-	-	x

Additionally, the voice gets its characteristic quality from the special shape of the speakers' vocal tract and its adjacent cavities. A speaker usually have certain habitual settings and gestures when moving the lips, tongue, jaw and velum. An example on articulatory quality is nasal voice. Nasality can be used in different sociolect (e.g some parts in Stockholm), and also in different Swedish dialects (Elert, 1997).

Even though the range of combinations of different types of phonation is vast, there are certain types of phonation that are improbable or even impossible because of physical limitations. Table 2 contains a description of various combinations of different types of phonation. For example, rough and breathy voice usually exists in combination with other types of phonation. A rough, whispering voice usually occurs in combination with modal or falsetto voice. With a breathy voice, whisper or vocal fry can usually only occur in combination with a modal voice (but not in combination with a falsetto voice) (Laver, 1980).

Age and Gender

Physiology, acoustics and perception

A listener can (almost always) hear the difference between a male and a female voice, at least in adult voices. Even a very short utterance or even a cough or laugh contains enough information about the vocal tract and vocal folds for the listener to form an instantaneous impression. Organic variations (like in gender differences) caused by differences in the dimensions of the vocal tract affect all the formant frequencies and F_0 (Titze, 1989). This generates acoustical differences between male and female speech; men have a fundamental frequency of about 120 Hz, while women have a F_0 almost one octave higher – around 220 Hz, as a result of the anatomical difference of the vocal folds, mass and length, between men and women. Men have thicker vocal folds, but the length is more important as an acoustic gender-parting factor; a male have longer vocal folds which affects the fundamental frequency. The length of the vocal tract differs, men have about 1,5 cm longer vocal tract (ca.17 cm) than women (ca. 15,5 cm), which gives women higher formant frequencies (Diehl, Lindblom, Hoemke & Fahey, 1996).

Male articulatory gestures consume more energy than female gestures because of the need of larger tongue movements to reach uvula, velum or the pharyngeal wall from a neutral position. These time-demanding, energy consuming articulatory gestures do not quite reach the intended articulation target and thus a smaller vowel space is generated. Female formants are more extreme than are male formants; women have a more expanded vowel space with larger distribution in the $F_1 - F_2$ - space, which gives an increased vowel contrast. This might possibly be some kind of compensation for the higher female fundamental frequency, which gives a slight reduction in distinction, but it can also depend on the flexibility of the female articulation organ, that renders them to easier reach the intended articulation target (Lindblom, 1983).

There are no acoustical differences between boys and girls speech before puberty; the difference between adult men and women mostly depends on the pharyngeal prolongation in boys during puberty. This pharyngeal prolongation is a result of a descent of the larynx in the vocal tract (Fitch & Giedd, 1999). Likewise, the gender differences diminish with old age. As male hormones increases in women and female hormones increases in men, the voices also change and sound more similar.

An important factor is the multimodal character of communicative behavior and speech and language processing. The influences of visual and auditory factors affect perception. An audiovisual integration of the speech signal occurs and visual images affect the expectations based on the listener's experience. Johnson, Strand and D'Imperio (1999) have examined the auditory/visual effect in vowel perception and their results indicate that listeners tend to integrate abstract information of gender with phonetic information in speech perception. The listener uses all available clues to what they can expect; if you see a female face you expect female formant frequencies.

Universal features

There are specific features that seem to be more or less universal. For instance, all languages do have consonants and vowels but they have different phoneme inventories. The languages' selected speech sounds are chosen to get enough dispersion to achieve lexical distinction. Most languages contain the three vowels /i,a,u/² because they give a maximal perceptual distance (Liljencrants & Lindblom, 1972).

Ohala (1983) analyzed languages' different sound patterns and looked for phonetic universal features in an attempt of understanding the production of speech. By observing the universal physical, phonetic characteristics in the speech mechanism, especially aerodynamic qualities, Ohala discusses how different languages build their phoneme inventories based on different physical and biological conditions. Voiceless stops seem to be more frequent than voiced ones throughout the languages of the world. Voiced bilabial stops, /b/, are more common than voiceless /p/, while there seem to be a preference for voiceless velar stops /k/. Also, voiceless fricatives are preferred to voiced.

Languages make use of different categorical distinctions. Lisker and Abramson (1964) measured VOT (voice onset time) in a cross-language study and found that languages tend to use VOT as a distinctive contrast in categorizing voiced/voiceless stops. Yet another distinction is provided by aspiration. Languages have different ways of using aspiration as a distinctive contrast in the production of stops. This can be noticeable in second language production, where an incorrect production of aspiration either sound somewhat unfamiliar to the native or even can be a source of misunderstanding.

Naturally, the speech sounds have to fall within limitations of the human speech apparatus. Lindblom (1983) points out that normal speech only uses a small part of the potentially available gestures of articulation. Humans try to minimize the energy consumption of the speech gestures and speakers have a universal tendency to more hypo- than hyper-articulate, which results in coarticulations and vowel reductions (Lindblom, 1963). Humans seem to avoid extreme articulations in speech production, but speech is only economized to the limit of being perceptually appropriate. The speaker strives for articulatory relief while the listener demands perceptual distance; the sounds must be different enough for the listener to be able to separate them. Languages tend to develop sound patterns that adapts to the biological constraints of speech production (Lindblom, 1983). "Easy way sounds OK" seem to be a functional way of maintaining a balance between production and perception (Lindblom, 2000).

² The International Phonetic Alphabet (IPA) SIL Doulos93

Concluding remarks

The aim of this paper was to give a short overview of different speaking styles and phonetic variations that affect our speech production and perception. A conclusion can be made on the importance and problems of integrating all these aspects in the research concerning human-computer interaction. People tend to use various forms of speech depending on situation, from a more formal, hyper-articulate way of speaking to a more casual, spontaneous form of hypo-articulation. The study of how different emotions affect our voice is important in designing human-computer interactive software, as the simulation of emotions in a synthetic voice can be used to indicate “personality”, which could influence the intelligibility of the speech and the intended message (Murray & Arnott, 1993). Prosodic aspects play a very important role in the human-computer interaction, though software technologies have to provide more sophisticated abilities in both the recognition and the generation of prosodic variation to further the development of current research (Hirschberg, 2002).

References

- Crystal, D. (1997) *The Cambridge Encyclopedia of Language*. (2nd Edt) Cambridge University Press.
- Diehl, R.L., Lindblom, B., Hoemeke, K.A. & Fahey, R.P. (1996) On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics* 24, 187-208.
- Elert, C-C. (1997) *Allmän och svensk fonetik*. (7th Edt) Nordstedts Förlag, Stockholm.
- Fernald, A. & Simon, T. (1984) Explained Intonation Contours in Mothers' Speech to Newborns. *Developmental Psychology* 20 (1), 104-113.
- Fitch, W.T. & Giedd, J. (1999) Morphology and Development of the Human Vocal Tract: A Study Using Magnetic Resonance Imaging. *Journal of the Acoustical Society of America* 106 (3), 1511-1522.
- Hinton, L., Nichols, J. & Ohala, J. (1994) Introduction: Sound Symbolic Processes. *Sound Symbolism*. Hinton, Nichols & Ohala (ed). Cambridge University Press.
- Hirschberg, J. (2002) Communication and prosody: Functional aspects of prosody. *Speech Communication* 36, 31-43.
- Jacobson, R. (1962) “Why Mama and Papa?” in *Selected Writings, (1) Phonological Studies*, 538-545. The Hague: Mouton.
- Johnson, K. Strand, E.A. & D’Imperio, M. (1999) Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27, 359-384.
- Kent, R. & Read, C. (1992) *The Acoustic Analysis of Speech*. Singular Publishing Group Inc. San Diego, California.
- Kienast, M. & Sendlmeier, W.F. (2000) Acoustical analyses of spectral and temporal changes in emotional speech. *Proceedings of ISCA Workshop on Speech and Emotion*. Queen’s University, Belfast.
- Künzel, H. (2000) Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics* 7, 149-179
- Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge.

- Liljencrants, J. & Lindblom, B. (1972) Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 28 (4), 839-862.
- Lindblad, P. (1992) *Rösten*. Studentlitteratur, Lund.
- Lindblom, B. (1963) Spectrographic Study of Vowel Reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.
- Lindblom, B. (1983) Economy of Speech Gestures. *The Production of Speech*. P. MacNeilage (ed) Springer, New York.
- Lindblom, B. (1990) Explaining Phonetic Variation: A Sketch of the H&H Theory. *Speech Production and Speech Modelling*. W.J. Hardcastle & A. Marchal (eds). 403-439.
- Lindblom, B. (2000) Developmental Origins of Adult Phonology: The Interplay Between Phonetic Emergents and the Evolutionary Adaptations of Sound Patterns. *Phonetica* 57, 297-314.
- Lisker, L. & Abramson, A. (1964) A Cross-Language Study of Voicing in Initial Stops: Acoustic Measurements. *Word* 20 (3), 384-422.
- Lundström-Holmberg, E. & af Trampe, P. (1987) *Elementär Fonetik*. Studentlitteratur.
- Mozziconacci, S., (1995) Pitch variations and emotions in speech. *ICPhS 95 vol. 1*, 178 – 182.
- Murray, I.R. & Arnott, J.L. (1993) Toward the simulation of emotion in synthetic speech: A review of the literature on vocal emotion. *Journal of the Acoustical Society of America* 93 (2), 1097-1107.
- Myers, D.G. (2002) *Social Psychology*. (7th ed.) McGraw-Hill Higher Education, New York.
- Nettelbladt, U. (1997) De svårförståeliga barnen – aktuell forskning om specifik språkstörning. *Från Joller till Läsning och Skrivning*. R. Söderbergh (ed). Gleerups, Malmö.
- Ohala, J. (1983) The Origin of Sound Patterns in Vocal Tract Constraints. *The Production of Speech*. P. MacNeilage (ed) Springer, New York.
- Scherer, K., Banse, R. & Wallbott, H. (2001) Emotion Inferences From Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, 32 (1) 76-92.
- Shulman, R. (1989) Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America* 85, 295-310.
- Tartter, V.C. (1980) Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics* 27 (1) 24-27.
- Titze, I.R. (1989) Physiologic and Acoustic Differences between Male and Female Voices. *Journal of the Acoustical Society of America* 85, (4) 1699-1707.
- Traunmüller, H. (1988) Paralinguistic Variation and Invariance in the Characteristic Frequencies of Vowels. *Phonetica* 45, 1-29.
- Traunmüller, H. (2000) Sound Symbolism in Deictic Words. *Tongues and Texts Unlimited*. Aili, H. & af Trampe, P. (ed) Stockholm. 213-234.