The Communicative Setting and Markers in Speech

Jo Verhoeven
Department of Germanic Languages • Universiteitsplein 1 • B-2610 Wilrijk
verhoeven@pandora.be

1. Strands of information
The communicative setting in which people meet and talk can be regarded as highly complex interaction between human beings with an equally complex exchange of information at various levels between the participants in a conversation. At the core of the communicative process is the communicative act or an attempt by a speaker to establish a communicative effect in a listener. In order to realise a communicative act, any speaker has tools available at three strands, i.e. the aural, visual and tactile strand.

The aim of this paper is to present a brief theoretical reflection on the communicative setting in order to provide a wider background against which the subject of phonetics can be situated.

The aural strand relates to speech as a medium of communication, produced by a speaker and perceived by one or more listeners. The visual strand has to do with all aspects of the speaker’s body language in interpersonal interactions and which are visually perceived by listeners. This strand covers e.g. aspects of posture, facial expressions, and gesture to support communicative acts. These aspects of interpersonal interaction have been studied by Kinesics (Birdwhistell, 1972). Signals on the visual strand may be a source of information about individuals’ backgrounds, social status and situational role. In addition, visual aspects may contribute to the organisation of turn taking in which gaze direction, head and face gestures, posture shifts have been found to play an important role. The tactile strand, finally, relates to the proximity between speaker and listener, and body contact. A typical example of exchanges on the tactile strand are handshakes, embraces and kisses in greeting and goodbye rituals. The frequency and form of handshakes e.g. is to a large extent culture-specific and may consequently be more or less appropriate in a particular linguistic communicative setting. In Western cultures handshakes are typically short and intense, whereas in e.g. African contexts handshakes may be given according to a specific pattern (shake, thumb taking, shake in South African Zulu or Xhosa culture) and may easily develop into holding hands during a substantial part of the introductory exchange of a conversation.

When the information on a specific strand is purposefully manipulated by a speaker to create a particular effect in a listener, this strand is used with a communicative effect in mind, while the other strands largely remain supportive of the message to be conveyed.
Selection of the main communicative strand may depend on the message to be communicated. In many instances, the aural strand is the main communicative strand. But often the speaker may feel that certain messages are communicated more efficiently by means of the glance of an eye rather than words. In efficient communication the information on all strands is supportive of one another to reach the desired communicative effect in that e.g. communicating commiseration with a smile does not seem very effective.

2. The aural strand and speech
On the aural strand **speech** is the medium by means of which information is exchanged. The speech signal on the aural strand contains various **types of information**.

2.1. Semantic information
**Semantic information** relates to what is traditionally regarded as the message of the utterance. Speech in this respect is the carrier of language and the organisation of speech sounds provides the listener with information necessary to decode the intended message. It should be noted that in most part the relationship between sound and meaning is arbitrary and language-specific. The connection between sound and meaning is at best indirect. The semantic content and the speech characteristics which carry meaning of utterances is clearly related to the communicative value of utterances.

2.2. Symbolic information
In contrast to semantic information where the relationship between sound and meaning is indirect, sound symbolism can be defined as “the direct linkage between sound and meaning” (Hinton, Nichols & Ohala, 1994: 1). Clear examples of such direct linkage between sound and meaning are involuntary expressions of pain or bodily states such as hiccups. Here the sounds have a specific meaning in that they directly reflect the internal state of the speaker’s body.
A somewhat more interesting form of sound symbolism can be described as synesthetic sound symbolism, which “is the process whereby certain vowels, consonants, and suprasegmentals are chosen to consistently represent visual, tactile, or proprioceptive properties of objects, such as size or shape” (Hilton et al 1994: 4). In order to achieve this type of sound symbolism, languages seem to operate on the principle of the Frequency Code (Ohala 1984): this principle refers to the phenomenon that “high tones, vowels with high second formants (notably /i/) and high-frequency consonants are associated with high-frequency sounds, small size, sharpness, and rapid movement; low tones, vowels with low second formants (notably /u/), and low frequency consonants are
associated with low-frequency sounds, large size, softness, and heavy, slow movements.” (Hinton et al 1994: 10).

2.3. **Regulative information**
Regulative information relates to characteristics of the speech signal which indicate that the speaking turn of one participant in the conversation is over and that the floor is ready for another speaker to take over. Here the use of suprasegmental/prosodic speech phenomena such as intonation probably play a crucial role.

2.4. **Aesthetic information**
Aesthetic information in the speech signal are those characteristics which contribute to a sensation of beauty in the listener. Here reference is often made to technical devices to create an impression of beauty such as versification and metre in poetry (De Groot, 1968). But it should also be mentioned that people find some languages more attractive and more pleasant to listen to than others. This consideration no doubt plays a role for student-linguists to study one language rather than another.

2.5. **Evidential information**
Since speech signals are actively produced by a speaker it can be assumed that speech contains a large number of characteristics which indirectly provide information about the speaker, his speech community and the speaking situation.

2.5.1. **Speaker characteristics**
2.5.1.1. **Markers of identity**
Regarding speaker-characteristics, a distinction has to be made between those that are relatively permanent or long-term and those that are relatively transient or short-term. The former provide indirect evidence as to the identity of the listener in that they essentially refer to characteristics in the speech of individuals which make them stand out amongst other speakers of the speech community. These characteristics originate from the typical anatomical configuration of the speaker’s vocal tract and/or his habitual pronunciation characteristics. These characteristics can be labelled as **markers of identity**.

Markers of this kind are often actively used conversationally. This can be most clearly illustrated by telephone conversations which often open with a statement like “Hello, it’s me.” By means of this utterance, the speaker who opens the telephone conversation suggests that he belongs to a circle of people known to the listener and actively assumes that the listener will be able to identify him on the basis of his voice. In practice, this
identification process works quite well, identification is almost instantaneous and often occurs even before the speaker is able to proceed to the “it’s me” part of the utterance. This process obviously only works if the receiver is familiar with the voice of the caller, which suggests that such speaker characteristics are relatively permanent or long-term features of the speaker’s voice.

Markers of identity in telephone conversations are not only used to identify the speaker, but also to assess the characteristics of speakers whom we are not familiar with. There is substantial evidence which suggests that listeners are able to derive information about the speaker’s age, physique, weight, height and sex on the basis of his/her voice characteristics.

Vocal markers of age are discussed in Helfrich (1979), who suggests that there are at least four important dimensions on which listeners could base their judgements about speaker age. The first is pitch\(^1\) of the speaking voice. The results of measurements of mean pitch in several studies on men and women indicate that there is a substantial lowering of mean pitch from childhood (ca. 425 Hz) to young adulthood (ca. 175 Hz). The same studies suggest that the further development of average pitch from young adulthood onwards is different for men and women. For women there seem no further systematic changes in mean pitch as a function of progressing age. In men, however, mean pitch decreases slightly from early adulthood up to the age of 40-50 and a slight upward shift of pitch is observed from the age of 65 onwards.

The second factor is pitch range, which can be defined as the difference between the lowest and the highest pitch in utterances. Helfrich (1979) indicates that pitch range appears to remain constant during childhood and increases from adolescence to adulthood. In old age, pitch range seems to narrow again.

The third factor relates to the number of perturbations – or involuntary and irregular movements - in the pitch curve. This corresponds perceptually to a trembling of the speaking voice. Helfrich (1979) observes that advancing age goes hand in hand with an increase of the number of perturbations and explains this by impaired co-ordination in the nervous system.

Finally, Helfrich (1979) points to voice quality as an indicator of age. It is reported that voices of elderly people are typically characterised as more “hollow” and an increase of breathiness of the voice has been noted with increasing age.

In connection with speech as a marker of age Helfrich (1979) also points out that the speaker’s voice does not only reflect information about the speaker’s own age, but may

---

\(^1\) The term ‘pitch’ refers to the melody of speech or tone of voice, which is related to the speed of vibrations of the vocal folds. The higher the speed of vibration, the higher the tone of voice; the lower the speed, the lower the tone of voice.
also reflect certain assumptions made by the speaker about the listener’s age. It can often be observed that elderly listeners are addressed in a loud voice and at a lower tempo than younger listeners, whereby the speaker assumes that the listener may be hard of hearing and has difficulty in understanding.

Laver and Trudgill (1991) indicate that listeners are quite good at judging speakers’ **physique, weight and height** because of the good correlation between these factors and the dimensions of the speaker’s vocal system. In this claim Laver and Trudgill (1991) refer to a.o. research in Fay and Middleton (1940) who investigated the relationship between physique and voices transmitted over a public address system. Informants had to indicate which Kretschmerian² body types corresponded best to a series of voices. The results indicate that listeners judged 22% better than chance for pyknic³ types, 20% for leptosomes⁴ and only 1% for athletic⁵ types.

Lass, Beverly, Nicosia and Simpson (1978) report that listeners can judge weight to within ca. 1-2 kilo’s, while height judgements are accurate to within 4 centimetres.

The seemingly good correlation between speech characteristics and the physical dimensions of weight and height is, however, not supported by research in Künzel (1989) at least not as far as the parameter of pitch is concerned. Künzel (1989) measured the average pitch of 183 subjects reading a short passage of speech and was not able to establish any statistical relationship between pitch and weight/height of the informants:

> This conclusion is consistent with this author’s everyday experience with persons involved in criminal offences, who have been anonymous for some time before being identified: when these individuals’ voices are analysed using instrumental and auditory phonetic techniques strong discrepancies between parameters such as height and/or weight on the one hand and average F0, overall loudness and certain features of voice quality may become obvious in quite a number of cases when actually seeing the persons in court or at an arrest. (Künzel, 1989: 122)

Judgements of sex of the speaker seem usually accurate and based on pitch range. Hollien, Dew and Philips (1971) measured average minimum and maximum pitch for a

---

² Ernst Kretschmer (1888-1964) was German psychiatrist who attempted to correlate body build and physical constitution with personality characteristics. In his best-known work *Körperbau und Charakter*, he developed the theory that certain mental disorders were more common among people of specific physical type. He assumed three constitutional groups, i.e. the tall, thin asthenic type, the more muscular athletic type and the round pyknic type.

³ The pyknic type is characterised by shortness of stature, broadness of girth and powerful muscularity.

⁴ The leptosome type has a light body build.

⁵ The athletic type is characterised by a heavy frame, large chest and powerful muscular development.
group of 332 adult men and 202 adult women. The range for men was 78-698 Hz, while the female range amounted to 139-1108 Hz. The female range is thus considerably wider than for men. This difference can be attributed to the intrinsic length of the vocal folds, in that intrinsically shorter vocal folds in women give rise to higher pitch levels than longer vocal folds in men: average vocal fold length in adult women is typically 17 mm and 23 mm for men (Kaplan 1960).

Another factor which may play a role in speaker sex identification is suggested in Laver and Trudgill (1991), i.e. the difference in average frequency of the formants⁶ in male and female speech. Laver and Trudgill (1991) refer to Fant (1960) to suggest that formants of female speech are 17% higher than in men as a result of their generally shorter vocal tract.

The fact that speaker characteristics are relatively permanent and to a large extent determined by the anatomy and physiology of the speaker’s vocal organs can be clearly illustrated in this connection by speakers who have undergone a biological sex transformation. Although it is technically feasible to transform women into men and visa versa by means of operations and hormonal treatment, the vocal clues as to the original sex of the transformee remain present to a greater or lesser extent, even after intensive speech therapy. Often the original sex characteristics still present in the transformee’s speech are the only clues to the listener which make him wonder about the appearance of the speaker as a man or a woman.

As a last point in the discussion of individuating markers we would like to refer to personality markers which are extensively discussed in Scherer (1979). He presents evidence that listeners attribute certain personality characteristics to speakers on the basis of certain vocal clues in speech. He concludes that higher pitch levels in male American speech tend to be associated with dominant and competent personalities. In German speakers similar higher pitch levels are associated with discipline, responsibility and social conformity. Higher pitch levels in these types of personalities are accounted for by the fact that these personality characteristics are accompanied by a “habitually high degree of arousal or state of readiness of the organism” (Scherer 1979: 155). Higher arousal can be assumed to cause an increase in muscular tension, which may lead to higher pitch levels in speech. Scherer (1979) also points out that there may be a relationship between loudness of the speaking voice and extroversion, and that the characteristic of voice quality seems quite promising as an indicator of personality characteristics.

---

⁶ Formants are an acoustic concept. In the acoustics of speech, formants on a spectrogram refer to stretches of high energy in the speech signal.
In conclusion, it can be said that the speech characteristics relating to the speaker’s identity provide a general articulatory setting for individuals. The concept of articulatory setting will be discussed in more detail later.

2.5.1.2. State markers

Besides revealing the identity of the speaker, the speech signal can also provide information about the physical and/or emotional state of the speaker. Such speech characteristics can be labelled state markers and they differ from individuating markers in that they tend to be transient, rather than permanent. They imply, however, that listeners have established a notion of what is the ‘normal’ articulatory setting for a given speaker and that they can detect deviations from this ‘normal’ setting, which may then be attributed to physical or emotional changes that have taken place. The range of state markers is quite varied and this section only highlights some of the more interesting ones.

In the first instance, we would like to refer to the more transient medical states relating to inflammations of the organs of speech viz. laryngitis, pharyngitis, nasal catarrh or cold. In the common cold, for instance, the glands in the nose are swollen and block the air passage through the nasal tract. This sounds as if the patient’s speech is heavily nasalised. In reality, however, this type of speech is entirely denasal since the air cannot escape via the nose as a result of the blockage. This absence of nasality in speech is easily noticeable and common colds are readily recognised perceptually in for instance telephone conversations.

Alcohol intoxication is another well-known example of a temporary change in the speech of individuals. The consumption of alcohol affects the transmission of neural commands to the speech organs and impacts on their smooth operation. As a result of this, articulation becomes less co-ordinated and precise, giving rise to a slurred impression which is easily recognised by any sober individual in a party of merry people well over the legal limit. A survey of speech features of intoxicated voices is given in Johnson, Pisoni & Bernacki (1990). At the phonetic level, they distinguish between segmental effects and suprasegmental effects of alcohol intoxication. Segmental effects relate to the misarticulation of segments. The most common errors are misarticulation of [r] and [l] and the misproduction of [s] which tends to be realised as [s]. Other processes which are commonly observed are the wordfinal devoicing of stops ([dEt] instead of [dEd]) and deaffrication ( [S´S] instead of [tS´tS] ). The most important suprasegmental effects mentioned in Johnson, Pisoni & Bernacki (1990) is a reduced speaking rate, which may be the result of the speaker’s attempt to compensate for the loss of motor co-ordination when intoxicated. Furthermore, there is
an increase in the unvoiced/voiced ratio in intoxicated speech, a decreased level of loudness and a lower pitch range.

Johnson, Pisoni & Bernacki (1990) also provide an interesting example of the forensic application of speech research in that they analyse the speech of the captain of the Exxon Valdez, the ship involved in the massive oil disaster before the coast of Alaska in 1989. They present measurements of certain target words and phrases in the captain’s radio communication with the Coast Guard at various times before and after the incident. It was shown that as the captain’s intoxication progressed, an increasing number of characteristics of intoxicated speech could be observed.

Abuse of the vocal apparatus generally as a result of excessive smoking, drinking or working in a noisy environment can result in a permanent harsh voice quality. It should be noted that this can have sociological implications in that in certain societies ‘harsh voice’ is associated with lower social status (cfr. Section 2.5.2.). Women with such harsh voice quality tend to be perceived as loose in behaviour, since in many societies smoking and drinking is regarded as inappropriate behaviour for women.

Fatigue is also marked by voice features in that the mode of phonation can become inefficient, resulting in a whispery voice or in a weak breathy voice (Laver & Trudgill, 1979).

Finally, reference should be made to the relationship between speech features and the psychological state of the speaker. Emotions such as happiness, anger, boredom and even sexual arousal can be associated with typical speech features. Anger in English is often conveyed by a harsh voice quality, with a wider pitch range and an increased loudness (Laver & Trudgill, 1979). When the speaker feels on top of the world, his voice characteristics will be different from a situation where he feels utterly bored: in these cases differences in speech rhythm (fast vs. slow) and pitch range (wide vs. small) can be observed. The speech features of sexual arousal are an interesting case. In situations of sexual intimacy between people, voices of both men and women become rather breathy and the fine control of pitch seems rather difficult. Laver and Trudgill (1979) attribute this to “changes in the copiousness and consistency of the supply of lubricating mucus in the larynx and in the characteristics of the mucal lining covering the vocal folds, affecting the efficiency of their vibration.” (Laver and Trudgill, 1979: 13). These changes seem to result from hormonal changes leading up to intercourse. Again it should be noted that in a sociological perspective people have come to associate ‘breathy’ voice with sexiness and it is often heard in radio and television commercials in which products are advertised with breathy voice. Thus the product is associated with all the pleasant feelings surrounding the situation of sexual intimacy.
Outside the commercial context, breathy voice in women is associated with the stereotype of ‘sexy’ and this again has sociological implications.

To conclude this section on state markers, it should be emphasised that the speech characteristics that have been discussed so far are clearly associated with the speaker and two groups can be distinguished on the basis of whether they are relatively permanent or rather transient. Both aspects, however, imply a notion of normal or typical articulatory setting and it seems that listeners operate perceptually against the background of this articulatory setting with which he may be extremely familiar (in the case of a speaker he knows very well) or operating on certain assumptions he makes about the speaker’s voice on the basis of typical articulatory settings that he has observed in other speakers and with which he has learnt to associate certain physical, emotional or personality characteristics.

2.5.2. Speech community characteristics

Since every individual is part of a speech community, it can be assumed that the speech characteristics of this community are to a greater or lesser extent reflected in every individual’s speech. These are speech characteristics from a whole group which have been acquired over time by individuals who are part of that speech community. In this respect these speech features differ from those in the previous section, since those were largely determined by factors within individual speakers themselves. The speech characteristics that are discussed here are the result of exposure to examples of the speech community in which the speaker is born and has grown up: they are the result of imitation or a learning process that has taken place.

The social background or status of speakers is often reflected in their voice characteristics. This has been shown by Trudgill (1974) who compared aspects of speech from working class speakers to those of middle class speakers in the Norwich region. He found that the working class speakers can be characterised by amongst others creaky voice, a high pitch range, generally louder voices and a substantial degree of nasality. Esling (1978) also found a clear correlation between voice quality and social class in the region of Edinburgh. Higher social status is related to more frequent use of creaky voice. Lower social status, however, correlated more strongly with whisper and harshness in the speaking voice.

A second reflection of the speech community in individual speakers’ speech relates to their regional affiliation. Leaving apart phonological, lexical and grammatical differences between various accents and language varieties, differences in the realisation of certain speech sounds may reveal the region where the speaker resides. This is documented in considerable detail in Wells (1995) and is also clearly illustrated
in standard Dutch with its two varieties spoken in The Netherlands and Flanders (Belgium) respectively. In a phonological perspective, both standard varieties have an identical vowel inventory. Whereas the vowels in the Belgian variety are without exception realised as monophthongs, Dutch speakers in The Netherlands have a strong tendency to diphthongise certain vowels, particularly [o] and [e].

A phenomenon which in many ways closely resembles regional affiliation is so-called **foreigner accent**. This is the phenomenon where non-native speakers of a language are recognised by mother tongue speakers as foreign. Just as regional affiliation is apparent from a regional accent, foreign accent informs about the mother tongue of the non-native speaker. In both instances, the accent can be quite variable. A regional accent can be very slight or extremely outspoken. The same holds for foreigner accent which may be hardly noticeable or so strong that the speaker’s utterances become difficult to understand. Both regional and foreign accent may have social implications in that some regional accents can be regarded as prestigious such as the Antwerp accent in the Dutch-speaking part of Belgium, whereas others may be regarded inferior. This is also true for foreign accent: due to historical reasons a French accent in the Dutch-speaking part of Belgium is socially less acceptable than a Scandinavian accent. In addition, both types of accents may have communicative implications. Depending on the strength of the accent speech may become less intelligible and this may have important consequences for communicative success. Furthermore it can be said that both accents refer to a pronunciation standard which is essentially supra-regional.

Apart from these similarities, there is at least one important difference in that regional accent is typically acquired by imitation of other speakers of the same (regional) speech community, while foreign accent is not acquired but rather represents the lack of successful acquisition of the target language. As to the causes of foreign accent, a vast body of research mainly focuses on phonological differences between the mother tongue of the learner and the target language. Although phonological interference certainly plays a role in foreign accent, it is likely that foreign accent also has a purely phonetic basis. Again reference can be made to the concept of articulatory setting, which has been said to differ between languages (Honikman, 1964).

**Gender peer group** affiliation refers to pronunciation differences between men and women as a group. Smith (1979) indicates that there is substantial evidence that women tend to pronounce the sounds of their mother tongue more “correctly” than their male counterparts.

As far as the relationship between **ethnic origin** and speech parameters is concerned, it is suggested that pitch seems quite indicative. Yamazawa & Hollien (1992) have compared the average pitch of Japanese and American women. It was found that
Japanese women had a somewhat higher average pitch than American women, i.e. 223 Hz vs. 205 Hz. Although they argue that this difference probably has to do with structural differences between the two languages, an explanation in terms of ethnic origin is not ruled out completely. Hudson and Holbrook (1981, 1982) and Mack (1984) (referred to in Yamazawa & Hollien) compared the average pitch of black people with that of white people of similar age. They found that black people generally spoke at lower pitch levels.

2.4.3. Situational characteristics

Besides markers of the individual and the speech community, speech may carry characteristics imposed by the situation in which the communicative exchange takes place. In this context, the participants in the conversation adjust their speech or speaking style to the situational requirements. It shall be clear that most markers of situation relate to the lexicon, grammar in for instance formal vs. informal situations where speakers adapt their speaking style to what seems appropriate to the requirements imposed by the level of formality. “High form of language share certain properties, such as elaboration, syntax and lexicon, phonological precision and rhythmicity, whereas ‘low’ forms share other properties, including ellipsis, repetition, speed and slurring” (Brown & Fraser, 1979: 46).

In other instances a direct correlation can be established between speech features and situation. An example of this are situations of secrecy in which the communicated message is intended to remain hidden to people excluded from the conversation. In this context, speech is typically whispered in order to make it more difficult to overhear the message.

Another example of speech adjustment to the situation relates to the level of loudness at which speech is delivered. In a noisy environment such as the London underground, people speak with louder voices to make themselves understood. A very quiet environment such as a library invites people to whisper.

A last example is one with socio-linguistic implications and is reported in Woods (1992). In her research into the pitch differences between men and women, a research factor was situational level of formality. It was shown that the average pitch of women in formal situations was systematically lower than in informal situations, whereas this was not the case for men. Woods concludes: “(…) that because of the pervasive stereotype of high pitched voices, in certain formal speech encounters where serious issues are the topic of discussion, women suppress the high pitches which they use in more casual conversational styles” (Woods 1992: 93). The fact that women lower their voices in formal situations has also been observed in meetings of local councils in
Denmark (Gomard: personal communication). In essence, this lowering of the voice thus reflects a situational adjustment to the other party involved in the situation, i.e. men, probably in order to give more weight or authority to the statements made. In a society such as Denmark where women’s lib is high on the agenda, a general tendency can be observed for women to have lower voices in comparison to other languages. My personal impression is that in order to keep a clear speech distinction between the two sexes men lower their voices even more. Although this has never been investigated systematically, it fits nicely in the framework of the French Philosopher Baudrillard (1970), who argues that the consumer society is based on patterns of imitation and distinction. In this view, the higher classes set a the trend by driving certain cars (BMW), wearing certain designers clothes etc. This is imitated by the ambitious amongst the lower classes so that gradually the trend-setting value of these products is eroded. The consequence of this is a shift towards the consumption of other goods by the higher classes in order to distinguish themselves again. It would be an interesting point of research to investigate to what extent these patterns of imitation and distinction also apply to other phenomena such as language and speech.

It is probably true that situational characteristics of the kind that have been discussed are much more dynamic than markers of state and speech community markers. As soon as the situation changes, the typical speech characteristics associated with this situation disappear.

3. Conclusion
From the discussion, it should be clear that the phenomenon of speech is a highly complex, dynamic interaction of communicative as well as informative characteristics. Some speech characteristics can be manipulated as a system of segments to communicate messages on the aural strand. Other characteristics are simply there, they accompany the message and convey information about the speaker’s identity, his physical and/or emotional state, the speech community to which he belongs as well as the situation. It is against this general background that phonetics has to be situated as the discipline interested in the study of speech.

References


