7 Prosodic analysis of situational adaption in a spontaneous speech database

Anja Hübner

7.1 Introduction

A conversational situation is shaped by its various components, such as the interlocutors, the purpose or the topic of communication. These situational components lead to an adaption of prosody in daily communication. However, the question arises as to which situational aspects are decisive for the adaption of voice quality. Is this also detectable in linguistic units like phrases of courtesy? And which prosodic parameters are suitable to distinguish situational voice quality adaption?

7.2 Theory

Voice quality, as one of the five prosodies (Pfitzinger 2006), encodes paralinguistic information in social interaction. Pfitzinger (2008) states that the individual human voice emerges from static properties and the dynamic control of the glottal behaviour and of the shape of the vocal tract. Laver (1980) characterizes voice quality as the characteristic auditive colour of an individual voice which contains linguistic and paralinguistic information. Moreover, voice is influenced by further effects like emotions or time of day.

Although various methods to parametrise the glottal source for identifying voice quality exist (e.g. Mokhtari et al. 2003) the *Amplitude Quotient AQ* (Alku & Vilkman 1996) turned out to reliably cover a voice quality range from *pressed* to *normal* to *breathy voice*: First, the residual signal is calculated by means of inverse filtering to remove influences of the vocal tract. Then, the glottal flow is estimated from the residual signal by numerical integration. Finally, the amplitude-domain quotient is calculated by dividing the peak amplitude of the glottal flow waveform by the minimum of its first derivative (equation 7.1):

$$AQ = \frac{A_{AC}}{d_{min}} \tag{7.1}$$

Alku et al. (2002) introduce a further development of the AQ. By normalizing the AQ by the fundamental frequency F0 of the speaker's voice, comparable values of male and female voices are obtained. The *Normalized Amplitude Quotient NAQ* is given by equation 7.2:

$$NAQ = \frac{A_{AC}}{d_{min} \cdot T} = \frac{AQ}{T} \tag{7.2}$$

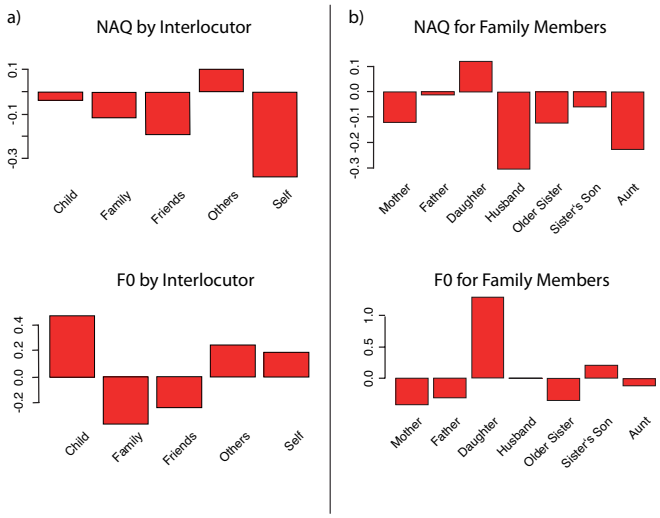

Fig. 7.1: Averages of NAQ and F0 for the interlocutors in a) all groups and b) in the family (Campbell & Mokhtari 2003).

Furthermore, Alku et al. (2002) calculate the NAQ for breathy, normal, and pressed phonation for ten speakers (5 male and 5 female). Their results show a considerable decrease of the NAQ values following the phonation types breathy – normal – pressed.

Campbell & Mokhtari (2003) use the NAQ to analyse a corpus of conversational speech. Within the scope of the JST/CREST-ESP project a female Japanese speaker was wearing a microphone to record her daily communication during five years. They extracted the NAQ and F0 in the categories *interlocutor, speaking-style* and *speech-act*.

Their normalization of the AQ was accomplished by the equation $NAQ = log(AQ) + log(F0)$.

Their database contains 13,604 utterances with labels of the following groups:

- interlocutor: child, family, friends, others, self-directed
- speaking style: polite, friendly, casual
- speech act: giving information, exclamations, requesting information, muttering, requesting repeats

Fig. 7.1 a) reveals that the speaker's voice is breathy in communication to others and her child but not in self-directed speech. F0 is highest in interaction with children, whereas very low in communication to friends and family. This figure illustrates the independent control of NAQ and F0 for each group of interlocutor. Inside the family F0 and NAQ are highest in child-directed speech (Fig. 7.1 b). The degree of breathiness shows the order daughter > father >

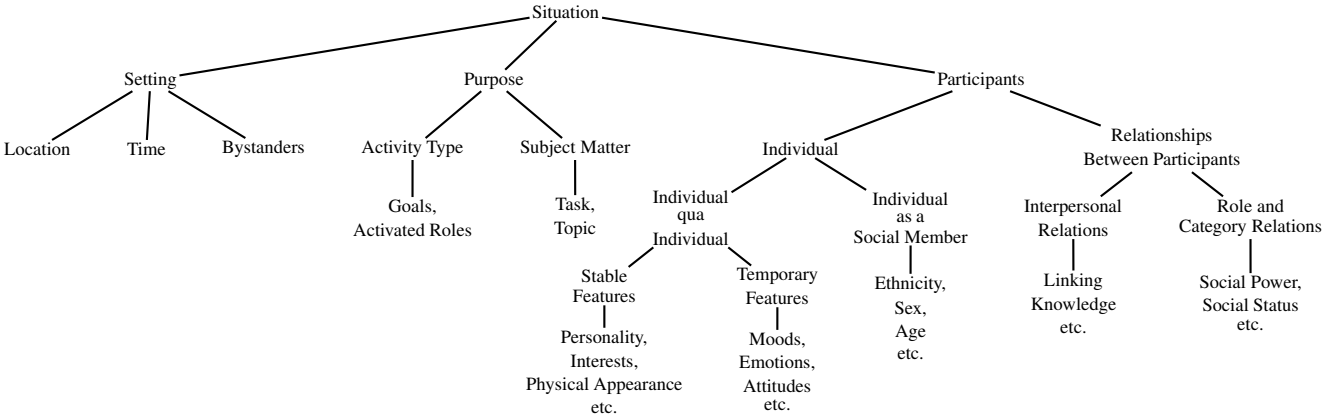

Fig. 7.2: According to Brown & Fraser (1979): Structure of a situation out of the main components *setting*, *purpose* and *participants.*

nephew > mother = older sister > aunt > husband (Campbell & Mokhtari 2003). Therein, Campbell & Mokhtari discern an order of the degree of care towards the speaker's family members, which is also valid in the analysis of the speaking styles. In this group breathiness is highest for careful speech followed by friendly, casual and last by self-directed speech. The analysis of F0 exhibits for careful and self-directed utterances high values, whereas friendly and casual speech was realized with a lower F0.

Also results of the analysis of speech acts are of interest. The category *giving information* shows low NAQ and F0 in contrast to the significantly higher values when the speaker is *requesting information*. *Requesting repeats* is realized with the highest NAQ. Obviously, the high F0 and breathiness in speech to children is a result of so-called child-directed speech, but the data shows that voice quality is consciously used to express interpersonal relationships.

Campbell (2004) establishes a connection to the *'park or ride'*-hypothesis (Falk 2003) which thematizes the human language development. Due to the fact that mothers could not take their infants along for foraging, they connect through voice instead of body contact to uphold the mother-child attachment and imply safety.

Humans also express with the tone of voice their personality, health and physical condition. The relationship between voice quality and emotions was examined by Klasmeyer & Sendlmeier (2000) by analysing their acoustic parameters. Happiness, anger and fear are realized by a rapid change of F0 and amplitude. Only small changes in F0 and amplitude show the contour of sadness and boredom. According to that they are expressed monotonously. Neutral spoken utterances were realized with modal voice, happiness with loud and anger with shouted modal voice.

A study of Gobl & Ní Chasaide (2003) documents an obviously contrasting usage of breathy and pressed voice. Relaxed, content, intimate and timid are emotional states which are mediated by a breathy or whispery voice. Conversely, a tense and harsh voice conveys an impression of a stressed, angry, formal or confident speaker.

However, to perceive the emotional content of an expression, the interaction of several factors is necessary. Thus, describing the prosody of emotions is multi-dimensional (Audibert et al. 2005).

An important part of social interaction are paralinguistic features which can be conveyed by voice. They are used to reveal the speaker's identity, personality and frame of mind. Abercrombie (1967) describes three classes of identity describing markers in speech. The markers signalize the membership of a group, characterize the individual or indicate the changing condition of the speaker.

Fundamentally, a speech event takes place in a situational context which consists of various components. According to Brown & Fraser (1979, p. 34) a situation is defined by the setting, the participants and their relationship, the purpose and the content of their conversation. Fig. 7.2 illustrates the individual components of communication.

A situation is composed of the components' setting, purpose and participants. The physical surrounding has little influence on the linguistic properties of the applied speech, however it requires an appropriate behaviour. Humans adapt the loudness of their utterance to the loudness of the location. This is called the *Lombard effect*, first documented in 1909 by Étienne Lombard (Silverman 2006, p. 191). Moreover, interaction is defined by the distance and orientation between speaker and listener. A professor in a lecture speaks in front of a large group in a big room, whereas guests of a cocktail-party stand in small groups facing one another. This form of communication is called *face-to-face*. Markers of time, e.g. greetings like "Good morning" open a situation. These deictic forms in communication mark physical and temporal aspects of the setting.

Also, the participants of a situation have to be considered in diverse perspectives (see Fig. 7.2). The communication is influenced by characteristics of the interlocutor as individual and as a social member. Thereby, speech is used as an expression of different speaker characteristics and as an indicator of the relationship between interlocutors. It consists of stable features and of the individual and temporary features like mood and attitudes. The relationship between interlocutors may be categorized by the interpersonal relations and the individual as a social member of a group (Brown & Fraser 1979, p. 50). Choice of words, syntax and register can

vary with the social relationship between interlocutors. An example for this fact is child-directed speech. A high pitch, an overplayed intonation contour and basic vocabulary and syntax characterize the communication to children.

Social position and setting define the degree of formality. Every human learns various ways of communicational behaviour in life. Often the degree of formality is defined by setting and purpose of a particular communication. Brown & Fraser (1979, p. 46) assume that formal and informal scenes can be distinguished by universal features in speech. Formal speech ("high") has characteristics like specific syntactic and phonological accuracy, corresponding choice of words and rhythm. Contrary to this, informal speech consists of e.g. elliptical constructions, repeating words, high speaking rate, cusses, etc.

Humans use F0 cross-culturally to mark affect, intention and emotion according to Ohala (1983). This theory is called the *frequency code*. The central assumption is that a low F0 is an expression of aggression, dominance, assertiveness, confidence and intimidation, while a high F0 signals politeness, respect and social submission. Ohala (1983, p. 4) suggests that the function of the frequency code is used nearly cross-language.

7.3 Experiment

The study of Campbell & Mokhtari (2003) reveals relations between F0 and the degree of breathiness of voice in the categories *interlocutor, speaking style* and *speech act*. They use only a small, already labelled part of a large, over five years recorded database with only one female Japanese speaker. A follow-up experiment is required to replicate the results. Additionally, a different cultural background and first language of the speaker could be illuminative.

7.3.1 Hypotheses

According to Brown & Fraser (1979) a situation comprises several aspects which are leading to the adaption of voice quality. Part of the recorded data is telephone communication, whereas all other signals are in face-to-face interaction. Therefore, it is necessary to examine possible differences.

1. There are no significant differences in F0 and amplitude between the communication situation *telephone* and *face-to-face*.

A speaker adapts F0 and NAQ to the interlocutor (Campbell & Mokhtari 2003).

2. a) The comparison of communication groups shows a decrease in F0 in the order children, public, friends, family.
   b) Within the group *family* F0 decreases in the order children, cohabitee, father, friends, mother.
3. a) The comparison of communication groups shows a decrease in NAQ according to the order public, children, family, friends.
   b) Within the group *family* NAQ decreases in the order father, children, mother, sister, cohabitee.

Brown & Fraser (1979) and Levinson (2000) discuss the pragmatic significance of ritual behaviour patterns. Connected with the *frequency code* by Ohala (1983), the following hypothesis regarding the adaption to a situation can be expected.

4. Ritualized social behaviour patterns like greetings and farewell as an expression of courtesy are realized with a high F0.

The content of communication as part of a social situation induces an adaption of the voice (Brown & Fraser 1979).

5. In conversation about private, informal topics, the voice quality is more breathy and shows higher F0 unlike in work-related, formal topics.

7.4 Data Acquisition

The adaption of voice to different situations is of great interest. For this purpose, a German database solely consisting of spontaneous speech was generated. Therefore, a person had to wear a microphone over a longer period and had to feel up to record even private conversations in distinct situations with a large variety of interlocutors. Additionally, the private surrounding with which the person interacts on a daily basis should be composed of different groups like parents, brothers and sisters, infants, friends, co-workers and others. Using different communication media, such as the telephone, is enriching the database.

7.4.1 Recording of the speech database

The database has been recorded between August 20th and December 14th 2009. A head-mounted condenser microphone with cardioid polar pattern *AKG C-520* and a portable recording device *Fostex FR-2LE* were used in order not to limit the interaction. The microphone was adjusted with the help of a mirror before every recording session.

The choice of a head-mounted microphone is advantageous since it is neither restricting the mobility nor biasing the communication due to the fact that it can be worn without attracting attention. Practically, this means that the recorder was in a shoulder bag all the time and the cable was hidden underneath a jacket. Moreover, great care was taken to not alter the recordings by any means like alcohol or decreasing health. The signals were recorded with a sampling rate of 48 kHz and a resolution of 24 bit. The latter was chosen as e.g. emotions can produce amplitude variation of more than 27 dB (Pfitzinger & Kaernbach 2008).

In total, 174 files in *.wav*-format were created, which together sum up to 57.48 GB or about 119 hours. The size of the files varies between 128 kB and 1.56 GB. At the time of the recordings the female speaker is 25 years old.

7.4.2 Labeling System

The creation of a database requires a well-suited labeling system. Here, the focus was set to an as precise description of the discourse as possible as well as to the ability to analyse the data in an automated way. The labeling system (Fig. 7.4) consists of base and additional items as well as those labeling non-verbal contents of the communication. Base labels are marking three essential aspects describing the communication situation, namely the *interlocutor*, the *location* and the *type of communication* (Brown & Fraser 1979, p. 34) whereas additional items are labeling topics, emotions as well as the speaking style categorizing forms. Also, labels describing non-verbal aspects in communication are added. By marking non-verbal events one can exclude or consider those separately during analysis.

Basic labels include the group of interlocutors. These are members of the family, like the parents, the older sister and her own family and the grandparents. Regarding the possible adaption on speaking styles, the range of age is of interest. The niece is two months old and therefore an infant, the nephew is five years old whereas the grandparents are older than 80 years. Other interlocutors are the cohabitee and his family, friends, workmates and others which cannot be classified into one of the above groups. This database contains also several conversations with people in public e.g. the pharmacists, the baker or a delivery service. Furthermore, utterances to animals and the computer as well as self-directed speech are labelled. In total there are conversations to 70 interlocutors.

Equally, the category *location* belongs to the basic labels as well as *type of communication* which can cause differences in communication according to Brown & Fraser (1979, p. 44). Thus, a distinction of face-to-face vs. telephone is necessary in the category *type of communication*.

Additional labels denote common topics, emotions and different speaking styles from spontaneous speech in interlocution. The group of emotions turned out to be unratable since marking the on- and offset of emotions can be done arbitrarily. For that reason, the system in Fig. 7.4 for the classification of emotions (happiness, fear, rage, anger and disgust) with the rating of the intensity in slight and intense emotional degree, is merely theoretical.

The section of topics is subdivided in topics of formal and informal content, because of the situational difference according to Brown & Fraser (1979, p. 47). Informal topics are interlocutions about food, health, other persons as well as conversations about the niece (infant). Contrary to this, formal topics are about master's thesis, computer or work. Direct speech and read speech are marked as different from spontaneous speech. Additionally, pragmatics states that social deictic forms grammaticalize social relationships. Hence, the system includes a label for rituals of courtesy like greeting and leave-taking.

Non-verbal content of a communication, such as laughing, singing or other sounds are also social indicators of a situation, which are represented by labels in the system in Fig. 7.4. These were not considered during the analysis. Another marker was given for especially interesting contents like irony or whispering to be analysed in later studies.

The database contains 18625 labelled utterances. Every signal is marked by the three basic labels interlocutor, location and type of communication. If the signal was labelled with additional labels (e.g. topics) or non-verbal labels, they were attached to the elementary file name. A typical example is the signal name fg_13_FF_+_TBBY, marking a communication to the mother in the location Bad Kleinen, which occurred face-to-face, whispered and with the topic TBBY (the newborn niece).

7.5 Results

7.5.1 Type of Communication

The type of communication essentially determines the situation of a conversation (Brown & Fraser 1979). At this point only these groups are considered in which face-to-face and telephone conversations are marked. The prosodic parameters amplitude and F0 were analysed for members of the family, the cohabitee, his parents and people in public. The statistic evaluation of the amplitude by means of a univariate, two-factorial analysis of variance shows that signals of telephone conversation are with 65.0 dB significantly louder than face-to-face conversations which have a mean amplitude of 63.2 dB ($df_1 = 13; df_2 = 12075; F = 141,7; p = 0.001; \eta^2 = 0.012$).

F0 was statistically analysed as well. Fig. 7.3 displays the average and the standard deviation of F0 for interlocutors of different groups. The average of F0 in face-to-face situations amounts 27.0 st, whereas telephone conversations have a mean F0 of 27.7 st. This evaluation shows that F0 in conversations on the phone is significantly higher than face-to-face ($df_2 = 12075; F = 96.1; p = 0.001; \eta^2 = 0.08$).

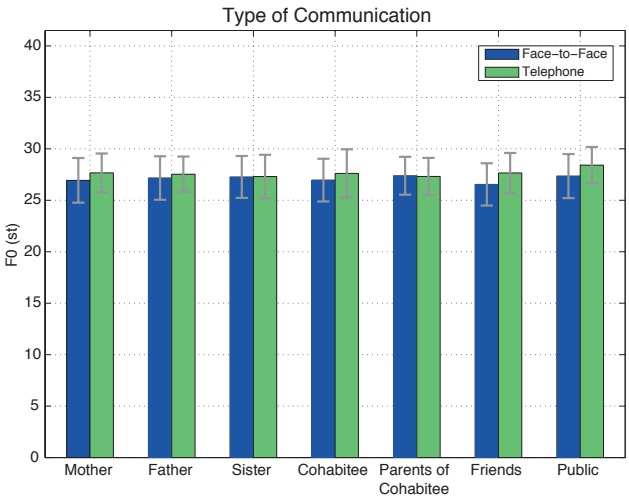

Fig. 7.3: Averages of F0 in face-to-face and telephone conversation for mother, father, sister, cohabitee, parents of cohabitee, friends and public persons.

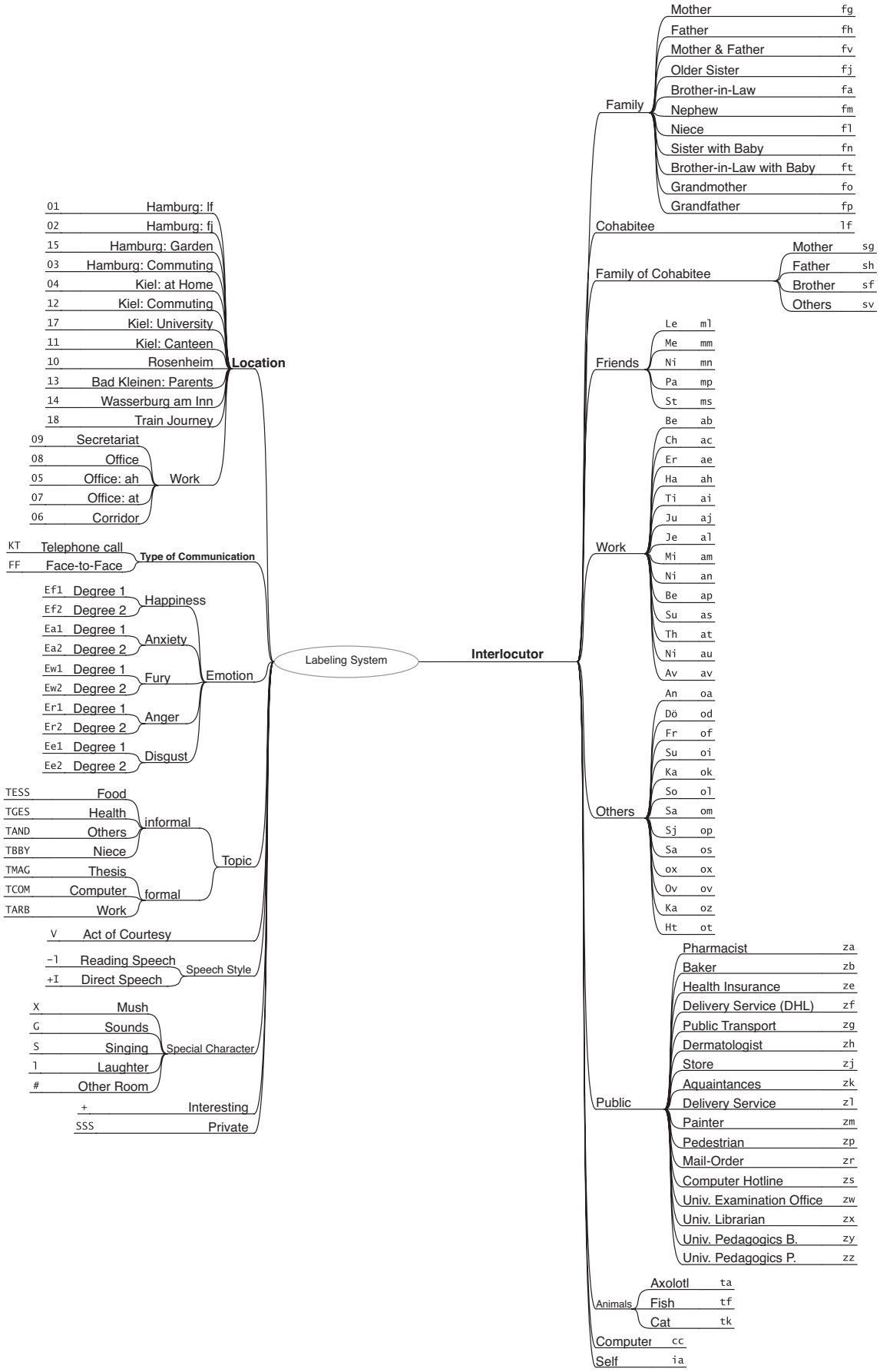

Fig. 7.4: Tree diagram of the labeling system. Beginning in the center, the system consists of basic labels (bold typed), additional labels and those which describe the non-verbal interaction.

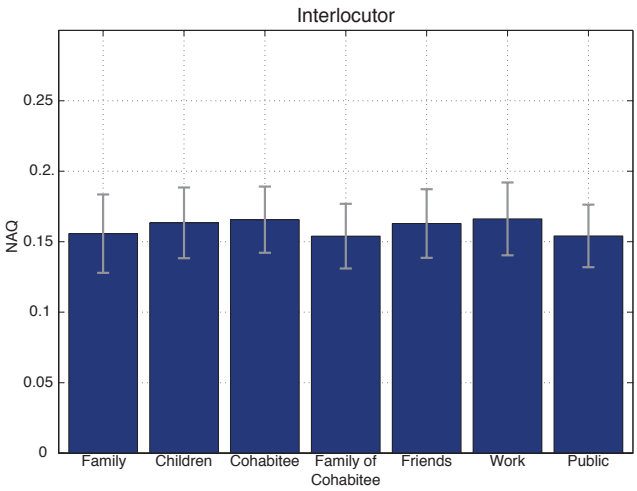

Fig. 7.5: Averages of NAQ for the interlocutors family, children, cohabitee, family of cohabitee, friends, work and public.

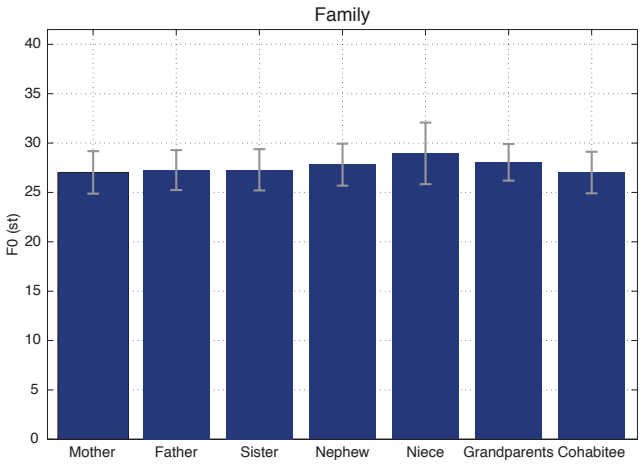

Fig. 7.6: Averages of F0 for family members: mother, father, sister, nephew, niece, grandparents and the cohabitee.

7.5.2 Interlocutor

Analyses of F0 and NAQ were conducted among the particular groups of communication family, children, cohabitee, family of cohabitee, work and public. By using a univariate, one-factorial analysis of variance the differences in the averages between the interlocutors were evaluated.

Significant differences appear in the use of F0 ($df_1 = 6; df_2 = 17247; F = 171; p = 0.001; \eta^2 = 0.056$). A clear variation in F0 can be observed in communication to workmates compared to public interaction. At work the subject speaks 1.38 st lower than in public ($df_1 = 369; T = 11.6; p = 0.001$). The voice to children with an F0 of 28.4 st is 0.67 st higher than to public persons ($df_1 = 662; T = 4.8; p = 0.001$). F0 in conversation to family members is with a value of 0.65 st significantly lower than to public people ($df_1 = 359; T = 5; p = 0.001$). The difference in F0 between the interlocutors children and family shows statistically significant evidence of child-directed speech. F0 is higher in speech to children with a mean difference of 1.27 st ($df_1 = 1517; T = -15.4; p = 0.001$).

Ordering F0 from high to low yields: **children > public > family of cohabitee > family > cohabitee = friends > work**.

The NAQ as a parameter indicating the degree of breathiness was yet analysed rarely. The statistical analysis of the NAQ shows significant differences between the averages of the groups of communication ($df_1 = 6; df_2 = 17247; F = 107; p = 0.001; \eta^2 = 0.036$) (see Fig. 7.5).

Between the cohabitee and the workmates no significant NAQ change can be detected ($df_1 = 6765; T = 0.9; p = 0.33$). However, the comparison of NAQ mean values of workmates and children reveals a significantly lower breathiness to children ($df_1 = 2125; T = 3.8; p = 0.001$). The degree of breathiness in conversation to children and to friends is equally breathy ($df_1 = 2350; T = 0.5; p = 0.6$). Moreover, the statistical analysis shows a lower NAQ in speech to family members than to children ($df_1 = 1842; T = -9.6; p = 0.001$).

The evaluation of NAQ shows this order from breathy to pressed: **cohabitee = work > children = friends > family = family of cohabitee = public**.

In the following the differences in the usage of F0 and NAQ in the group family are focussed. This group consists of mother, father, the older sister and her children as well as the grandparents. The longtime cohabitee was also considered as a member of the family.

Due to the different ages of the children (two month and five years) a separate consideration of F0 is necessary. Talking to the niece shows a mean F0 of 29.0 st, whereas a significantly lower voice of 27.8 st is used in speaking to the nephew ($df_1 = 1164; T = 8; p = 0.001$). On the other hand there is no significant difference between nephew and grandparents ($df_1 = 197; T = 2; p = 0.057$). F0 in conversation to the father is with a mean difference of 0.55 st significantly lower than to the nephew. A comparison of speech to the older sister and to the father shows no significant differences in F0 ($df_1 = 2635; T = 0.4; p = 0.68$), besides there is no difference in F0 between the mother and the cohabitee ($df_1 = 6411; T = 0.2; p = 0.85$).

With regard to the differences of F0 (see Fig. 7.6) in conversation to family members the following order can be determined: **niece > nephew = grandparents > sister = father = mother = cohabitee**.

The statistical analysis of the NAQ reveals a high breathiness in interaction with the niece, whereas the breathiness to the nephew is significantly lower ($df_1 = 1230; T = 11; p = 0.001$). Talking to the cohabitee the NAQ was significantly lower than to the niece ($df_1 = 930; T = 4.7; p = 0.001$), but higher than to the sister ($df_1 = 4250; T = 9.5; p = 0.001$). Also the NAQ between talking to mother and sister is significantly different ($df_1 = 4940; T = 4.5; p = 0.001$), whereas there is no significant difference between speech to mother and nephew ($df_1 = 941; T = 0.7; p = 0.5$). The communication with the father and the grandparents is more pressed than with the nephew ($df_1 = 1315; T = 4.6; p = 0.001$).

Based on the statistical analysis of the NAQ the order from breathy to pressed for family members is **niece > co-**

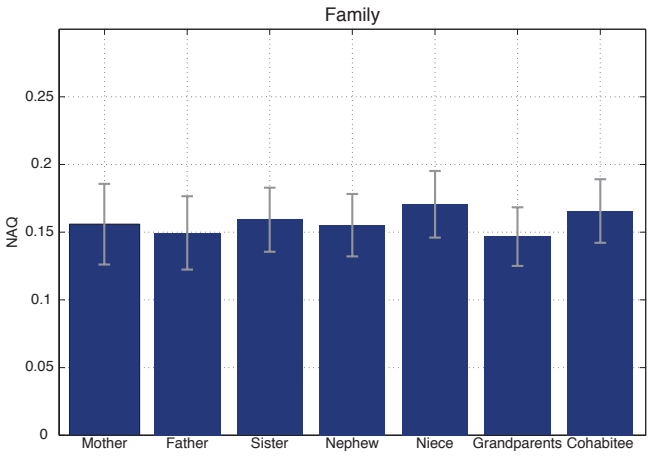

Fig. 7.7: Averages of NAQ for the family members mother, father, sister, nephew, niece, grandparents and the cohabitee.

habitee > sister > mother = nephew > father = grandparents. Fig. 7.7 illustrates the evaluated averages and standard deviations of the NAQ for all family members.

7.5.3 Acts of Courtesy

As acts of courtesy the pragmatics denotes ritualised social patterns of behaviour (Levinson 2000). One form of these linguistic signs are greetings and leave-taking.

Amplitude and F0 of these language signs are investigated via univariate, two-factorial analyses of variance. The first factor consists of the levels mother, father, sister, cohabitee, parents of cohabitee, female friend and public persons. The second factor has two levels, on the one hand ritual utterances like greetings and leave-taking, on the other hand signals without phrases of courtesy.

Statistical analysis of amplitude reveals no significance ($df_1 = 13; df_2 = 11939; F = 0.71; p = 0.398; \eta^2 = 0.00$).

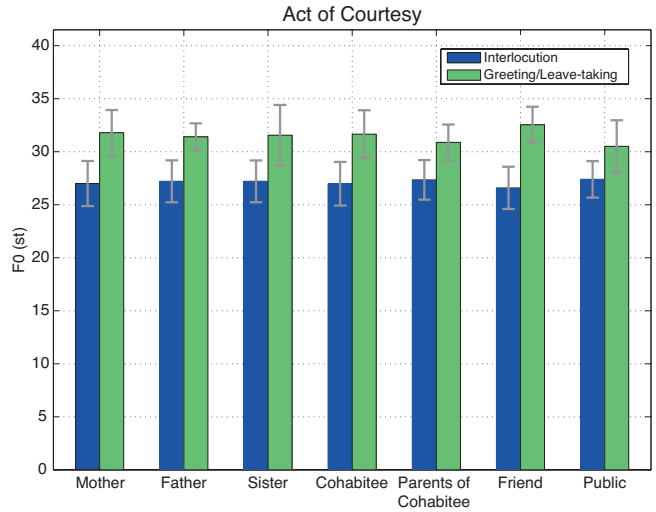

Fig. 7.8: Averages of F0 for utterances with and without courtesy in conversation with the interlocutors father, mother, sister, cohabitee, family of cohabitee, female friend and publicity.

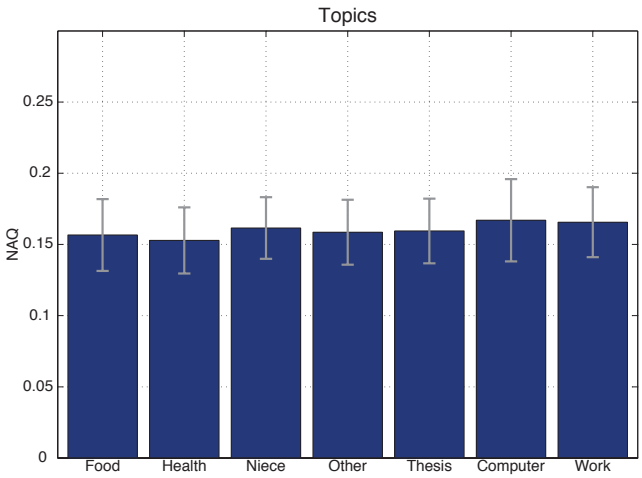

Fig. 7.9: NAQ for the informal topics food, health, niece and other persons and the formal topics master's thesis, computer and work.

In contrast, the evaluation of F0 shows significant values ($df_1 = 13; df_2 = 11939; F = 451; p = 0.001; \eta^2 = 0.036$). An illustration of the averages and standard deviations is given in Fig. 7.8. The F0 average of the phrases of courtesy is 31.4 st whereas in normal conversation without these ritual utterances F0 is 27.2 st. This means, that there is a difference of 4.2 st between the normal interlocution and the greeting or leave-taking.

7.5.4 Topics

The issue of an interlocution basically determines the choice of words in the interaction. But is there also a difference in prosody when changing topics? Formal and informal utterances are compared since language features of communication can be differed due to the degree of formality according to Brown & Fraser (1979, p. 46). Formal topics are the master's thesis, the computer and the work. Informal topics contain food, health, other persons and the niece.

The analysis of all formal and informal topics reveals significant differences in F0 ($df_1 = 6; df_2 = 5504; F = 30; p = 0.001; \eta^2 = 0.032$) as well as in the NAQ ($df_1 = 6; df_2 = 5504; F = 29; p = 0.001; \eta^2 = 0.031$). Using t-tests revealed that conversations about informal topics have 0.59 st higher F0 ($df_1 = 1582; T = 10; p = 0.001$). The degree of breathiness for formal and informal topics is shown in Fig. 7.9.

7.6 Discussion

Nowadays, verbal communication takes place non-exclusive from face to face. Other types of communication, for example video-telephony over the internet, are spreading out quickly. Nevertheless, in this study the speaker only interacts face-to-face or via telephone. The statistical analysis reveals higher values of amplitude and F0 for the communication over the telephone. Therefore, hypothesis 1 (p. 45) is refuted. A possible reason of the higher amplitude on the telephone can be the settings of the telephone. Is the signal too quiet the speaker will compensate this and will speak

louder. Alternatively, the interlocutor on the phone could have spoken louder which results in an adaption of the amplitude of the speaker, as predicted by the Lombard effect (Silverman 2006, p. 191). Also F0 was higher on the telephone, which could be explained with the non-existence of facial expression and gestural information which leads to a compensative reaction. F0 becomes higher without the support of facial expression and gesticulation because solely the voice can transport the message. Here a connection to the *'park or ride'*-hypothesis (Falk 2003) can be found.

The adaption of the prosody of voice was regarded for all groups of interlocutors and within the family. The highest F0 was applied in interaction with children and public people. But the results show other orders of communication groups than Campbell & Mokhtari (2003) have evaluated. Hence, hypothesis 2 (p. 45) has to be refuted.

The very high F0 in interlocution with the grandparents shows the same properties like child-directed speech, but also the type of communication has to be considered. As mentioned before, the communication on the telephone was realized with a higher F0 than face-to-face. This can be the reason for the higher value due to the fact that the interaction with the grandparents only took place on the telephone.

An obvious explanation for the lower F0 with friends, workmates, and the cohabitee as well as mother and sister is the *frequency code* (Ohala 1983). It implies that a low F0 is used to signalize dominance, self-confidence or aggression whereas a high F0 exhibits subservience and courtesy. So, with the mentioned persons the speaker wants to prevail herself and show dominance.

Additionally, the results for NAQ in the different communication groups as well as in the family are not in the order Campbell & Mokhtari discovered which leads to the refutation of hypothesis 3. Unlike in their analysis the voice quality is pressed in public and breathy in communication to the cohabitee. Considering the order according to Campbell & Mokhtari (2003), this could be a societal effect. In several countries parents were adored and their children show subservience also in adulthood. Moreover, there is often social pressure due to hierarchical structures. This could be the case in the Japanese more than in the German culture.

The statistical analysis affirms hypothesis 4, because F0 is significantly higher in ritual social patterns of behaviour such as greeting and leave-taking. This could corroborate the fact that roles between interlocutors are also signaled on the basis of prosodic parameters. Also, a high F0 as an expression of courtesy corresponds with the *frequency code*.

A significantly higher F0 in informal conversation affirms hypothesis 5 only partially, which states that formal topics are discussed with lower F0. It also states a lower NAQ, but the results reveal a significantly higher NAQ for formal talk.

In this study, essential observations about the adaption of voice were evaluated. Most results show significance but some hypotheses, based on the literature, could not be affirmed. As already Brown & Fraser (1979) described, an interaction is influenced by several factors, e.g. the age, but also the stage of life and family status of the speaker.

Bibliography

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Alku, P., Bäckström, T. & Vilkman, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *J. of the Acoustical Society of America 112*(2), 701–710.

Alku, P. & Vilkman, E. (1996). Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering. *Speech Communication 18*, 131–138.

Audibert, N., Aubergé, V. & Rilliard, A. (2005). The prosodic dimensions of emotion in speech: The relative weights of parameters. In: *Proc. of Interspeech*. Lisbon, 525–528.

Brown, P. & Fraser, C. (1979). Speech as a marker of situation. In: Scherer, K. R. & Giles, H. (Eds.), *Social markers in speech*. Cambridge University Press, Ch. 2, 33–62.

Campbell, N. (2004). Accounting for voice-quality variation. In: *Proc. of Int. Conf. on Speech Prosody*. Nara, Japan, 217–220.

Campbell, N. & Mokhtari, P. (2003). Voice quality: the 4th prosodic dimension. In: *Proc. of the XVth ICPhS*, vol. 3. Barcelona, 2417–2420.

Falk, D. (2003). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences 27*(4), 491–503.

Gobl, C. & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication 40*(1–2), 189–212.

Höge, H., Kotnik, B., Kačič, Z. & Pfitzinger, H. R. (2006). Evaluation of pitch marking algorithms. In: *ITG-Fachbericht 192: Sprachkommunikation*. Berlin, Offenbach: VDE Verlag.

Klasmeyer, G. & Sendlmeier, W. E. (2000). Voice and emotional states. In: Kent, R. D. & Ball, M. J. (Eds.), *Voice quality measurement*. Singular Publishing Group, 339–357.

Laver, J. (1980). *The phonetic description of voice quality* (1. ed.). Cambridge University Press.

Levinson, S. C. (2000). *Pragmatik* (3. ed.). Max Niemeyer Verlag Tübingen.

Mokhtari, P., Pfitzinger, H. R. & Ishi, C. T. (2003). Principal components of glottal waveforms: Towards parameterisation and manipulation of laryngeal voice-quality. In: *Proc. of the ISCA Workshop on Voice Quality (Voqual'03)*. Geneva, 133–138.

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

Pfitzinger, H. R. (2006). Five dimensions of prosody: Intensity, intonation, timing, voice quality, and degree of reduction. In: Hoffmann, R. & Mixdorff, H. (Eds.), *Speech Prosody Abstract Book. Studientexte zur Sprachkommunikation*, vol. 40. Dresden: TUDpress, 6–9.

Pfitzinger, H. R. (2008). Segmental effects on the prosody of voice quality. In: *Proc. of Acoustics'08*. Paris, 5309–5314.

Pfitzinger, H. R. & Kaernbach, C. (2008). Amplitude and amplitude variation of emotional speech. In: *Proc. of Interspeech*. Brisbane, 1036–1039.

Poyatos, F. (1991). Paralinguistic qualifiers: our many voices. *Language & Communication 11*, 181–195.

Shrivastav, R. (2003). The use of an auditory model in predicting perceptual ratings of breathy voice quality. *J. of Voice 17*(4), 502–512.

Silverman, D. (2006). *A critical introduction to phonology*. Continuum.