2 

SIGNAL PROPERTIES

COMMUNICATION is characterized as a composite accomplishment that, besides matters of signal transfer and transmission, encompasses achievements of both a sender and a recipient of signals (Shannon & Weaver 1949; for further citations see Todt & Kipper 2003). With this well-proven model as a reference, it seemed expedient to prepare the main part of my study (i.e. chapters 3, 4 & 5) by a short inquiry into some fundamental signal properties of whispering. Therefore, this chapter will deal with some significant performance features of whispering and show, for example, how loudly or how softly the vocalisation can be expressed. In addition, the chapter will treat relationships between the physical properties of whispering and the influence of context variables, and demonstrate especially how environmental noise can constrain its transmission. Finally, the chapter will report experiments concerning the decoding of whispered speech. Inasmuch as speech can convey both verbal information and prosodic information, either of these two roles will be tackled separately.

Acoustical signals are well defined by parameter expressions in three dimensions: the frequency domain, the time domain and the power domain (reviews in Todt 1986; Todt & Kipper 2003). Features of the first two domains are often visualized by frequency spectrograms (see Figure 2.1) and in more detail addressed by different methods of signal analyses (review in Mundry & Sommer 2003). Quantitative approaches to the power domain, on the other hand, may require a separate procedure which usually includes an application of amplitude measures given in "dB" (review in Brumm 2003). In metrical terms, this measure has some advantages; that is, the value "0" dB stands for an amplitude which coincides to the hearing threshold of human beings. For comparison: the threshold of word recognition is normally given at 20 dB (Mrowinski & Scholz 2001).

Various laboratories of audiology and phoniatrics routinely take measures of both: the loudness of spoken language and the thresholds of auditory perception (Newby & Popelka 1992). Thus, there is already an abundance of data about such measures. Although their majority refers to data of normal speech, data of whispered speech are often collected as well; at least for purposes of control. When such data were checked for integration into the study reported here, however, they turned out to be not very appropriate. The rational of this judgement was based on two reasons: First, the data were sampled particularly for health matters, and thus given as mean values and standard deviations, only (Martin 1994). Second, I wanted to explore especially the limits of whispering performances, and such goal suggested some additional experiments that will be described in the following paragraphs.

2.1 Performance Features

Whispering is often distinguished as a low-volume signal. In order to elucidate this characterisation in more detail, I investigated the range of amplitude variation in unvoiced vocalizations compared it to the range of amplitude variation in normal speech. In addition, I examined whether the performance of whispering could be more arduous than normal speech. With respect to this, I expected that the number of syllables that can be produced per breath could be reduced during whispering. Such expectation was based on a study of Stathopoulos (1991) who found that the translaryngeal air-flow is larger during whispering than during a production of normal speech.

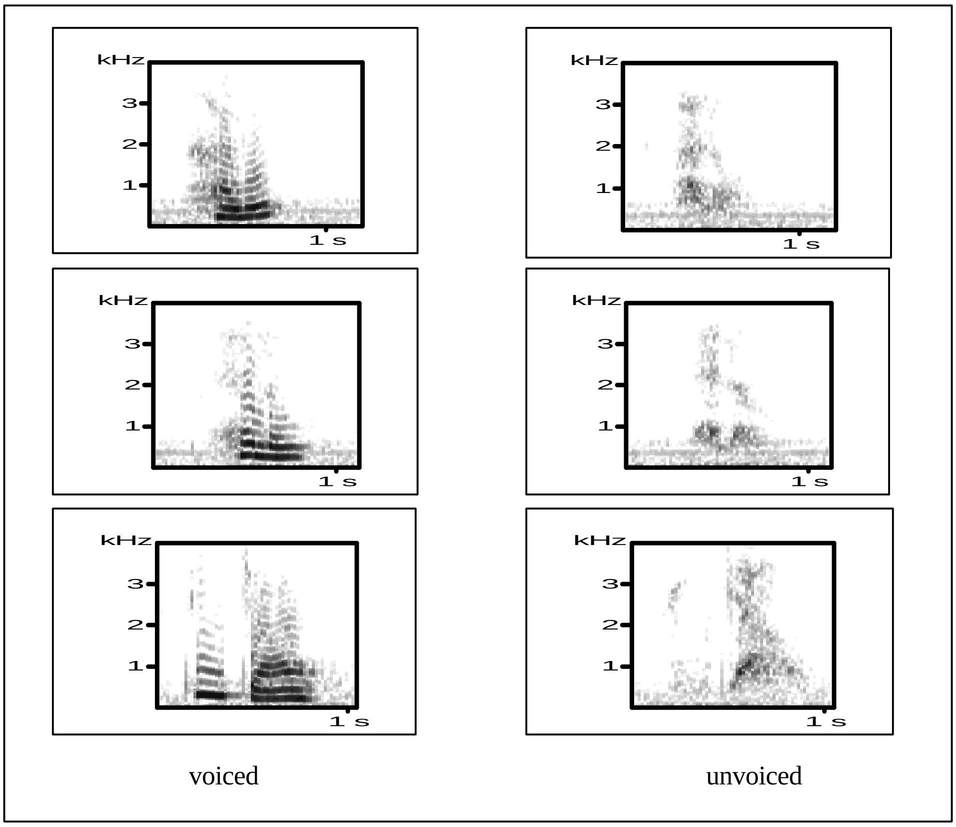

Figure 2.1 : Frequency spectrogams of words spoken either in a normal (=phonated) voice (left) or in a whispered (unvoiced) manner (right). Top: German expression of the word 'hallo'. Middle: English expression of the word 'hallo'. Bottom: Expression of ‘xin chao’, i.e. a Vietnamese equivalent of 'hallo'.

Methods

Participants were students of biology (n=8), who - after instruction - voluntarily agreed to take part in the following two experiments. In one of them subjects had to deal with six different tasks: That is, they had to speak a given text as loudly or as softly as possible, or to vocalize at an intermediate level. And, each of these three versions had to be spoken with a phonated and also with a whispered voice. In the other experiment, subjects had to deal with these six different tasks, too. Here, however, they were instructed to use one single breath to articulate as many syllables as they could. This manoeuvre had to be performed as loudly or as softly as possible, or at an intermediate level, and also with both a phonated and a whispered voice. To standardize the design, subjects were asked to simply speak successions of different numbers.

The procedure was conducted on individuals seated in a sound-protected test chamber with walls shielded against echo-effects. To avoid serial effects, the order of experiments and also the order of tasks within experiments were changed across subjects. Amplitude measures were taken at a speaker-microphone distance of 1m by using a Rode & Schwarz EZ GA2 precision sound level meter. The successions of numbers produced by the subjects were recorded on tape. These recordings served for supplementary control**.**

Results

As shown in Figure 2.2, amplitude measures received for a given verbal category formed a coherent distribution. Amplitudes measures of phonated speech varied from 74 dB(A) (extremely loud), over 59 to 52 dB(A) (normal) to 42 dB(A) (extremely low); whereas measures of whispered speech ranged from 52 dB (extremely loud), over 40 to 35 dB(A) (normal) to 31 dB(A) (extremely low). All subjects showed significant systematic differences in the median amplitude values between the three speaking levels for both phonated and whispered speaking [Friedman ANOVA: ? ^{2} = 16, N = 8, p < .001]. Some overlap aside, the measures of whispering provided a nice continuation of measures assessed for normal speech. This distribution suggested that whispering, indeed, could make an appropriate supplement of the latter.

The results on syllables produced per breath are illustrated in Figure 2.3. Comparing the data received for phonated speech to data received for whispered speech uncovered several specificities. These were:

(1) The topmost amount of syllables spoken per breath was around 160. Individuals who reached this amount also by whispering did so only by using an extremely soft voice.

(2) The number of syllables whispered loudly was smaller than those of syllables spoken otherwise. This indicated that the production of such whispering can require more air or breathes, respectively, than normal speech.

(3) The distributions of syllable numbers measured for each performance category (i.e., either loud, or intermediate, or soft) were similar for normal speech [Friedman ANOVA, $?^2$ = 5.68 , N = 8, n.s.] but extremely different across categories when ascertained for whispered vocalizations [Friedman ANOVA, $?^2$ = 13.56 , N = 8, p < .001]

(4) The inter-individual variation of amplitude data ascertained within a given performance category (i.e., either loud, or intermediate, or soft) was remarkably small for whispered vocalisations, but larger for that of data belonging to the measures of normal speech.

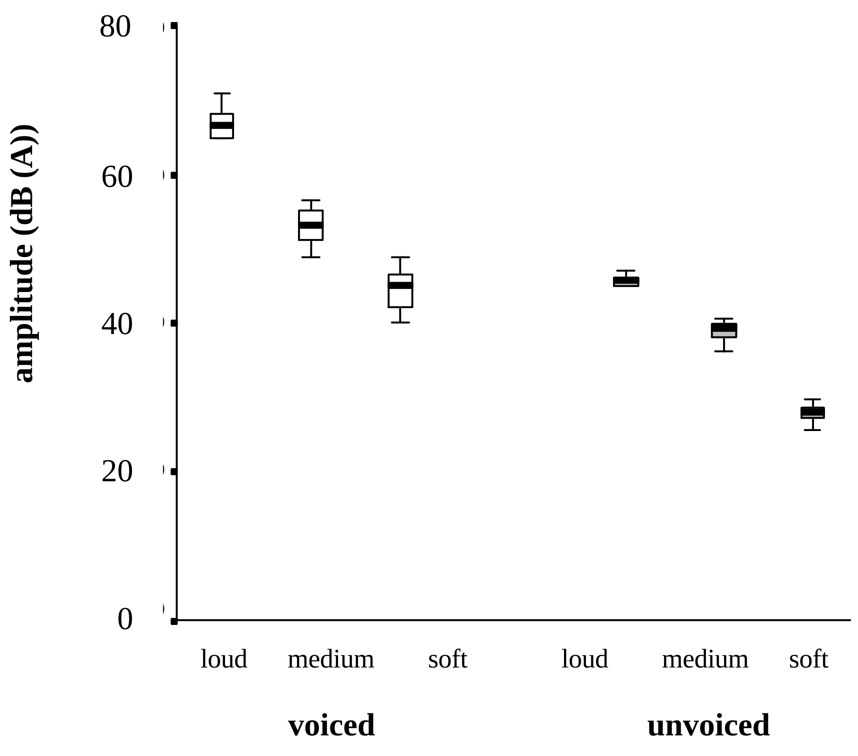

Figure 2.2 : Boxplots of amplitude measures ascertained for words (here numbers) spoken either in normally (left) or in a whispered manner (right). Pooled data. Data referred to a total of 400 words recorded from eight subjects. (see text.)

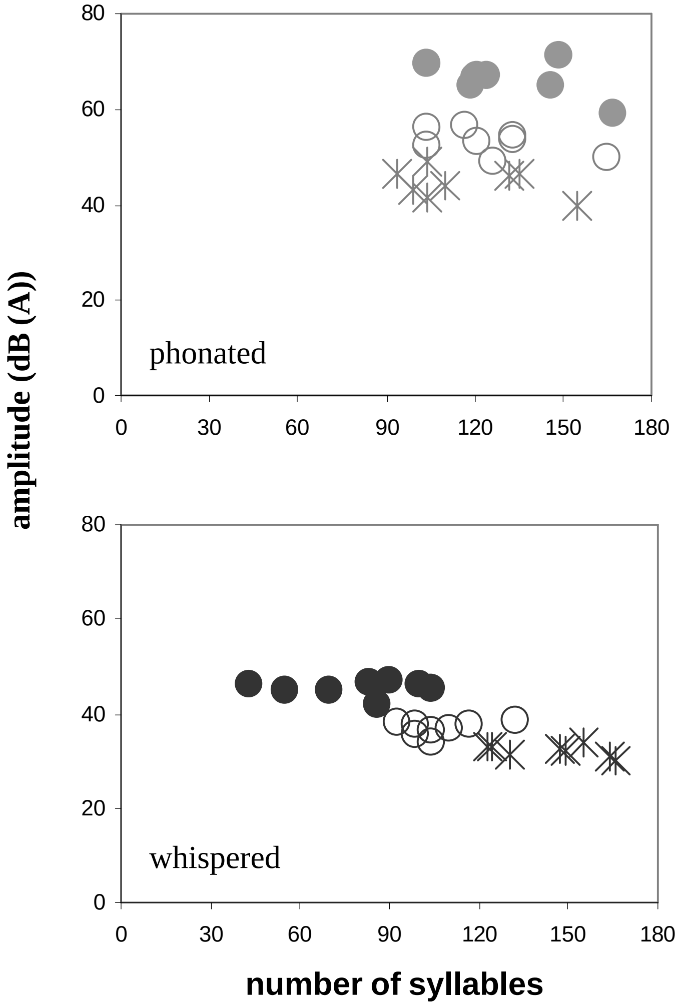

Figure 2.3 : Amounts of syllables that subjects were able to produce per breath, here plotted against their acoustical amplitude top: data received for phonated speech. bottom: data received for whispered speech. Filled circles: vocalisations that subjects uttered as loudly as possible. Circles: vocalisations that subjects uttered at an intermediate amplitude level. Crosses: vocalisations that subjects uttered as softly as possible.

Conclusions

The results invited some comments and conclusions. In the first place, they confirmed that whispering, indeed, can be specified as a low-volume signal which could be used especially in close-contact communications. Implications of the latter statement will be examined in the next chapter.

Second, it was noticeable how good subjects managed the task to separate their vocal performances so clearly according to the three categories 'loud', 'intermediate' and 'soft'. As shown, the range of amplitude data ascertained within a given category was relatively small, especially for whispered words. This finding pointed to a sophisticated system of performance control which here might have benefited from the instruction to achieve the two opposite extremes in loudness. Then, these might have served as an anchorage to correspondingly place the intermediately loud expressions.

Third, subjects differed widely in the amounts of syllables per breath, but only a little in their amplitude ranges. These differences suggest that the production of as much syllables as possible, or an economically optimal use of air, respectively, could be a matter of individual training. The amplitude measures, on the other hand, could reflect a skill that either may be independent of individual exercise, e.g. indicating inherent constraints of the system, or may be a consequence of training processes that had reached already an advanced stage of expertise. There is evidence that also the training of whispering can start early; that is, at an age of approximately two years (see chapter 5).

Finally, our experiment documented that the performance of whispering, aside from its very soft expressions, can require more air than normal speech. This finding confirmed my expectation that whispered speech could be explained as being more arduous as the latter, or, from a modern perspective of biocommunication, as being more 'costly'. This aspect contributes to a discussion of developmental and evolutionary aspects and, will be treated in chapter 5, as well.

2.2 Decoding Verbal Information

As mentioned in the 'General Introduction', whispered speech is considered to be as intelligible as a normal speech, at least, if applied in a short sender-recipient distance. In addition, it was shown that the main communicative effects of whispering are obviously independent of language specificity (Jensen 1958). Aside from these overall characteristics, however, some specific aspects of the use of unvoiced speech remained unclear. They concerned, for example, the limits of such use and its relationships to a drastically lowered loudness of whispered stimuli, or the impact of context variables, such potential masking effects caused by environmental noise. In the following, I describe a study which tested these issues by two different experiments.

Methods

Participants were students of biology (n= 104), who - in these experiments - were placed in a sound protected room and asked to decode auditory stimuli presented to them via headphones. This setting allowed us to test up to 12 subjects simultaneously. Nevertheless, we made sure that each subject treated her/his experimental task individually; that is, notated her/his decoding result immediately after each stimulus on a special list. Subjects with hearing deficits were excluded from the tests.

To prepare the experiments sets of ten different numbers had been recorded on tape in either a phonated or a whispered expression. Tapes were composed only by numbers that contained four syllables and were presented in a random succession. Playbacks of tapes differed in the amplitude of stimuli measured close to the headphone membrane (distance: 0.01 m). Again amplitude measurements were conducted by using a Rode & Schwarz EZ GA2 precision sound level meter.

The evaluation of effects was done by first comparing the lists of notations to our lists of stimuli and then relating the percentage of correct notations to the amplitude of stimuli (20 dB(A), 25 dB(A), 35 dB(A)). Statistical significance of effects was tested by a one-way ANOVA.

Results

The analyses of experimental data yielded a number of clearly amplitude-related effects. In particular, I could ascertain the following four results:

(1) Stimuli presented with an amplitude of about 45 dB(A) did not induce any decoding errors. That is, at this level of stimulation, whispered numbers were as well understandable as numbers given in a normal voice.

(2) With a stimulus amplitude of about 35 dB(A), phonated numbers were decoded correctly, whereas a few decoding errors appeared towards whispered numbers. However, this effect was not significant.

(3) The amount of correctly decoded numbers was clearly reduced for either kind of stimulus version, if amplitude levels were lowered to about 25 dB(A). However, the reduction was especially drastically for the whispered stimuli. This effect was highly significant [$F (1, 103)= 139.37$, $p< .0001$]. The data are given in Figure 2.4.

(4) With a stimulus amplitude of about 20 dB(A), the amount of correctly decoded numbers showed a further decline for either kind of stimulus version. The difference between test results remained, nevertheless, significant [$F (1, 103) = 33.56$; $p< .0001$] . The data are given in Figure 2.4.

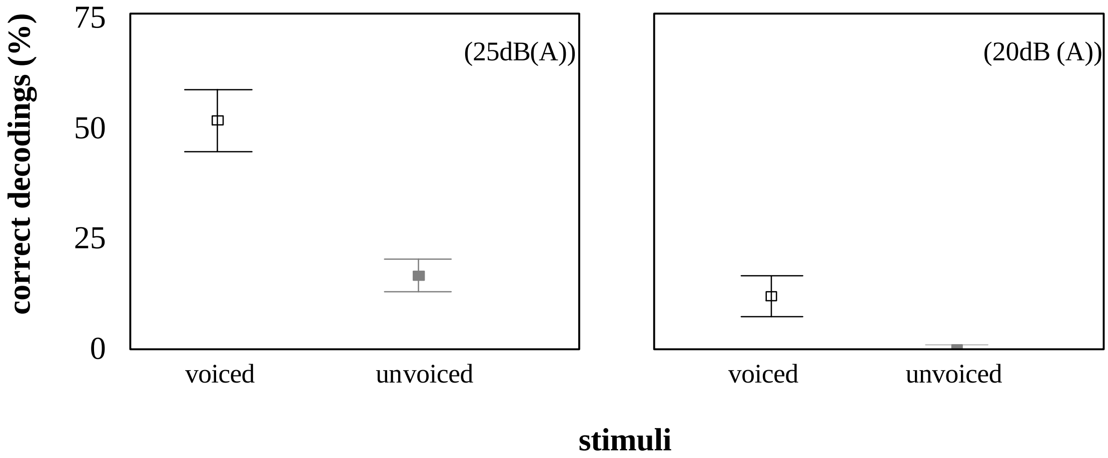

Figure 2.4 : Percentage of correctly decoded stimuli plotted against stimulus quality (median and interquartiles); i.e., voiced (=phonated) numbers (left), or unvoiced (=whispered) numbers (right). The two squares in the left box refer to stimuli presented with an amplitude of about 25 dB(A). The two squares in the right box refer to stimuli presented with an amplitude of about 20 dB(A).

In a subsequent study, I examined whether and how effectively environmental noise can impede the decoding of verbal information. Overall, I applied the same type of setting as described above. But, I implemented an additional device here. That is, besides being presented with sets of well-defined stimuli, subjects were also exposed to a playback of environmental noise. In contrast to the spoken numbers which subjects received via headphones only, the noise was given through a loudspeaker and with an overall amplitude of 45 dB(A). With this design, I wanted to simulate an almost normal situation. For the same reason, I did not use white noise, but background noise which I beforehand had recorded in the waiting hall of an airport. In other words, I resigned to apply a completely standardized test condition.

An evaluation of lists documenting the correct notations of decoded stimuli yielded effects that – at a stimulus amplitude of 35 dB(A) - clearly differed in relation to stimulus quality: When the background noise ranged around 45 dB(A), numbers presented in a phonated version received significantly less decoding errors than other numbers presented in a whispered version [$F(1, 95) = 375.75$, $p < .001$] The data are illustrated in Figure 2.5 .

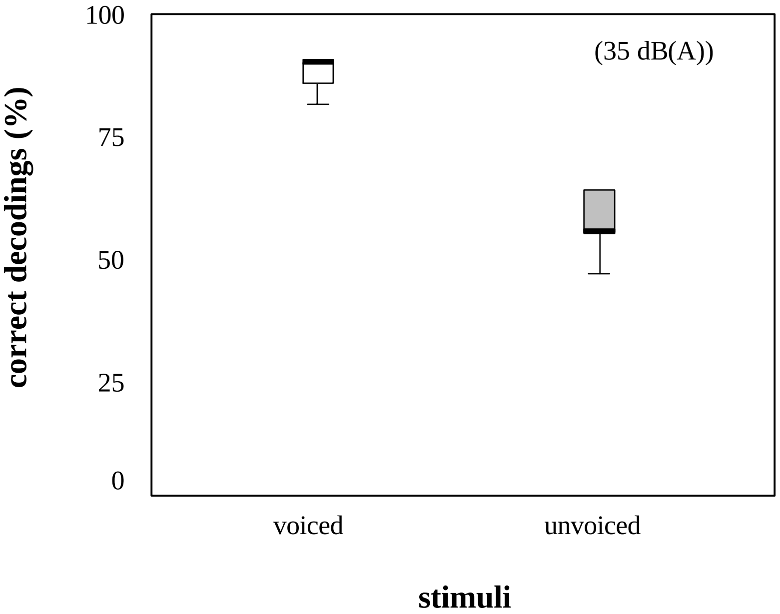

Figure 2.5 : Boxplots showing the proportion of correctly decoded stimuli plotted against stimulus quality, here: phonated numbers (left) or whispered numbers (right). Stimuli were presented with an amplitude of about 35 dB(A) in addition to environmental noise that was played-back with an amplitude of about 45 dB(A) and served to texts masking effects.

Conclusions

Which conclusions can be drawn from these results? Above all, the findings extend the knowledge about the pragmatic side of signal properties. As shown in Figure 2.2, whispered words usually were performed with an amplitude that at a speaker-microphone distance of 1 m ranged around 45 dB(A). Given the rule that in an acoustically ideal environment the loudness decreases by approximately 6 dB if the distance is duplicated, I conclude that hearing a whispered stimulus of about 35 dB(A) can symbolize a sender-distance of about 4 meters. As my results indicate, this distance can be regarded as that spatial range above which the use of a whispering voice may not be appropriate. Such distance is in line with a citation of other authors who declared 4 m as a boundary span for a communication by whispering (Hultsch et al 1992).

To avoid misunderstandings, my tests with stimulus amplitudes of 25 dB(A) or less served only to explore the limits of perceptional accomplishments, but not to study a recognition of whispered signals transmitted over a distance of about 10 m or more. Nevertheless, my results confirm an earlier conclusion saying that whispering is used best, if it is signalled to a nearby addressee (chapter 2.2.).

However, even when applied over a short distance only, whispering may be more ‘vulnerable’ than phonated speech. This can be concluded from the results of my second experiment which tested the impact of environmental noise. If such noise increases in volume it can have even more drastically effects: As I found in a pilot experiment conducted in the waiting hall of an airport, the masking effect of noise above 90 dB can cause that even the producer of a whispered voice himself is unable to recognize his own words (Cirillo, unpublished data). This effect can be explained as a consequence of the sound transmission via bones and tissues which can work fine for normal speech, but is constrained for whispered vocalisations with its missing fundamental frequency.

2.3 Decoding Emotional Information

The study of emotional expressions has a long tradition which can be traced-back even to Darwin's historical approach (Darwin 1872). As for prosodic information communicated by normal speech, the work of Scherer and his colleagues contributed most fundamental investigations to this issue (review in Scherer & Kappas 1988). As for the role that whispered vocalisations may play in this respects, the number of studies remained quite limited (Hultsch et al. 1992; Tartter & Braun 1994). Since these studies reported some discrimination problems for the deciphering of speech-encoded emotions, we reinvestigated this matter with a modified method. Our central questions were:

(1) How successful would subjects recognize and distinguish emotions, such as 'anger', 'fear', 'sadness', or 'joy', encoded in whispered speech? How would such decoding change, if they had been exposed to auditory stimuli which beforehand were filtered, e.g. by a telephone?

(2) Which acoustical parameters would encode a given emotion? And, which parameters would be used for its recognition?

(3) Which judgements would subjects announce in addition to their decoding result? Could the decoding task itself induce emotional feelings?

A first documentation of this investigation was recently published (Cirillo & Todt 2002). The following part of this subchapter provides an outline of the cited publication.

Methods

Participants (n=104) were students of biology who attended a course at the Free University of Berlin (Germany). The gender distribution of subjects was symmetrical, and their ages ranged between 22 and 32 years. Before being incorporated in the tests, individuals were asked whether they would like to take part, but they remained uninformed about the aim and other details of the experiments.

Our tests comprised two kinds of experimental manoeuvres: on dealing with a production and the other one with a perception of spoken language. In the tests, subjects were asked to speak the German phrase *"... ganze zwei-und-fuenfzig Wochen ..."* in 20 different versions (English translation: *"... all that fifty-two weeks long ..."*). The versions included four emotional expressions, namely (a) anger, (b) fear, (c) sadness (d) and joy, and for control, also an emotionally ' neutral' version. In addition, each version was spoken normally and also produced with a whispering voice. In pilot-experiments the test phrase was found to be extremely useful for this purpose. All 20 versions were recorded on tape, then marked by a secret label and finally incorporated in our pool of test material. Later, the material was submitted to detailed signal analyses (Figure 2.6) and/or used in the second experiment that focused on perceptional issues.

To prepare this setup, samples of the spoken phrase versions were drawn randomly from the pool of sound material and recorded on test-tapes which then were played-back to another group of subjects. Stimuli were presented via headphones and, within a given set of versions, in a random succession. At the same time, about six to eight subjects were gathered in the same room and then treated synchronously. Within the room, they were spaced and thus could neither interact nor influence each other otherwise.

To introduce subjects to a test session, questionnaires that in pilot-experiments had proven to be well appropriate for this purpose, were handed out. In addition, subjects were exposed to an initial playback that served to provide them with a kind of baseline for the next step of their task (Wiedenmann & Todt 1992). However, no further information was given beforehand. Instead, subjects were asked the judge the following playbacks and to give their opinions within a questionnaire after each trial. The structure of the questionnaires has been published elsewhere (Cirillo 2001).

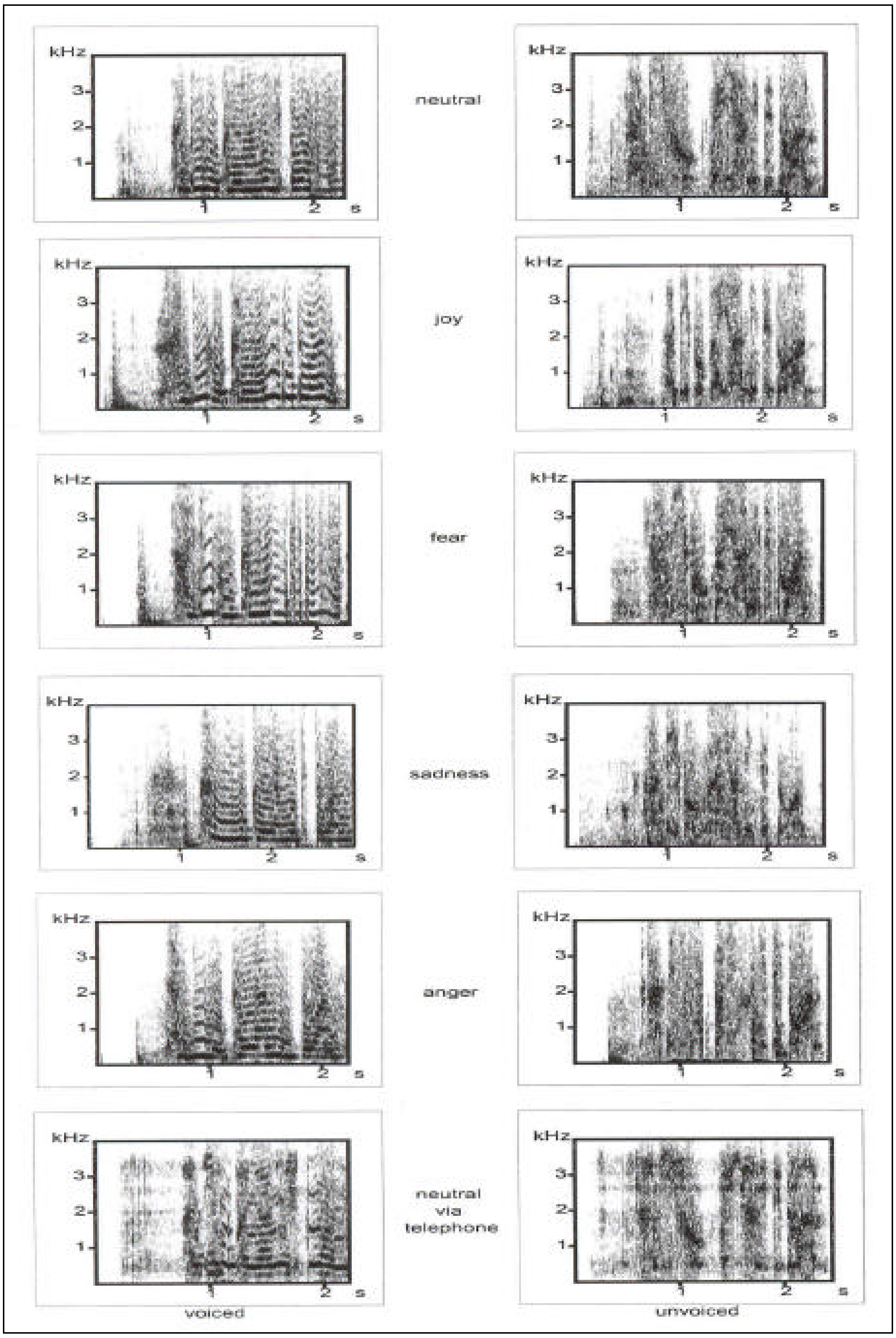

Figure 2.6 : Frequency spectrograms of verbally equal phrases. They were spoken normally (left) or with a whispering voice (right), and with different emotions. After their analysis, they were used as stimuli that subjects had to decode in terms of emotional cues.

In an additional line of experiments subjects were presented with a modified set of auditory stimuli. These were produced according to a method of Tartter and Braun (1994) who asked people to speak without trying to express a distinct emotion, but with a specific facial expression, e.g. a smiling face. Recordings of this material were incorporated in pool of stimuli. Finally, we extended this pool by filtering a large set of stimuli and adding this set together with the original material to the pool again. Filtering was done by sending the stimuli through a conventional telephone. The aim of this procedure was to examine whether and how a telephone would constrain a decoding of prosodic information.

Recordings of spoken phrases were analysed by sound spectrography using a Nicolet UA 500A spectrum analyzer that was connected to a Tönnies film camera for hard copy production of spectrograms. In addition, phrases were sampled and printed as spectrograms using the commercial program *Avisoft* or investigated parameter-wise. Finally, spectrograms of different phrase versions were compared subject-wise independently by two persons who were neither informed about the producer nor about the emotions that he/she intended to encode in his/her speech.

The analysis of ratings extracted from the questionnaires was done according to methods recently described by Kipper and Todt (2001). In addition, transition matrices were used to assess and calculate relationships between the experimental variables; e.g. the quality of an emotion which a speaker intended to express and the classification provided by subjects who tried to decode the expressions. For this purpose, both sets of data were cast into transition matrices. In these matrices, categories of antecedent/consequent pairs were ordered. Statistical significance was tested by χ^2-methods.

encoded		***decoded***									
		anger		**fear**		**sadness**		**joy**		**neutral**	
		n	**w**	**n**	**w**	**n**	**w**	**n**	**w**	**n**	**w**
anger	**n**	84								10	
	w		84						10		6
fear	**n**			81		6		3			
	w				94		3				3
sadness	**n**					97				3	
	w				6		94				
joy	**n**					3		97			
	w				10				71		19

encoded		***decoded***									
		anger		**fear**		**sadness**		**joy**		**neutral**	
		n	**w**	**n**	**w**	**n**	**w**	**n**	**w**	**n**	**w**
anger	**n**	100									
	w		70						20		10
fear	**n**	10		60				17		13	
	w		7		63		7				23
sadness	**n**			17		83					
	w				20		73				7
joy	**n**	3				3		80		14	
	w				47		3		17		33

Fig. 2.7 : Matrices documenting how good specific emotional expressions were decoded by subjects (N=100) who heard the stimuli via headphone. Beforehand, emotions were encoded in single phrases recorded from other subjects (N=20) who had produced the phrases in either their normal (n) or a whispered (w) voice. Cells in diagonal give correct judgements; other cells show misjudgements. Top: matrix referring to judgments of unfiltered stimuli. Bottom: matrix referring to judgments of stimuli filtered by a commercial telephone. Note: whispered joy suffered most from a transmission via phone and was often mistaken for fear then (see text).

Results

Decoding success

To recall, in the main experiments our subjects were asked to estimate which emotional expression a given auditory stimulus would encode, and they were invited to address their vote to one of the following five categories: 'anger', 'fear', 'sadness', 'joy' or 'neutral'. However, as the subjects were not exposed to stimulus versions encoding 'neutral', they often used this category when having problems to give a clear decision. The decoding success was measured by calculating the portion of judgments matching exactly *that* emotion which the respective producer had been intended to express. In general, this success was high enough to provide significant evidence for both correct encoding and decoding manoeuvres [χ^2= 15.87, df 1, p< .01]. A more detailed analysis yielded the following results: First, the best record was received for stimuli taken from our pool of ordinarily generated phrases. Here, whispered stimuli were not significantly inferior to stimuli spoken with a normal voice. Nevertheless, there was one exception, and this concerned whispered 'joy' that often was decoded as either 'fear' or 'neutral'. Analyses of the spectrograms suggested that these failures resulted from the reduced tonal quality of whispered phrases (Fig. 2.6). Second, very good records were received for stimuli produced during a specific facial expression, but without a voluntary intention to express a specific emotion (see Methods). However, here most correct matches were found for only two facial displays, namely 'smile' and 'frown'. This was true for both kinds of vocalizations: whispered and normal speech. Thus, these results supplemented the findings of Tartter and Braun.

Finally, a fairly good record was received for the stimuli filtered through a commercial telephone. Here, however, voiced 'anger' and voiced 'sadness' was almost not impaired; whereas whispered 'joy' was clearly not decoded as intended, but either classed as 'fear' or as unclear or 'neutral', respectively.

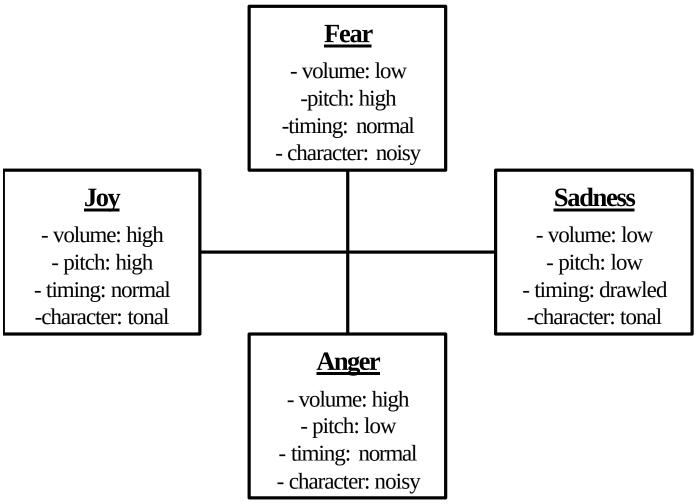

Figure 2.8. Schema listing typical expressions of four selected signal properties of vocalizations that encoded four studied emotions.

Encoding success

With the records listed above taken as a reference, we investigated the factors that could have caused the differences in stimulus judgment. The encoding success was measured by calculating the portion of phrases of a given speaker, which received correct judgments in terms of *that* emotion which he/she had been intended to express. In general, this success was extremely different across subjects. Thus, some speakers were excellent in signalling their emotions, whereas others appeared rather poor in such respect. Interestingly enough, this difference was consistent for both whispered phrases and phrases spoken normally.

In a further line of research we selected phrases that received high scores in encoding success and compared their acoustical signal parameters to the signal parameters of others that had received poor scores. This comparison allowed us to list a number of properties that were positively correlated to the four emotional expressions studied here (see Figures 2.6, 2.7, 2.8). The details of these results will be published elsewhere, but the most typical traits shall be given already. In brief, they are:

Typical 'anger' and 'fear' shared that their vocal expression did not require a tonal voice, but could be given in a noisy structure. And they differed clearly in pitch and often also in volume. Typical 'sadness' and 'joy', on the other hand, required a tonal voice, and this was especially important for 'joy', together with a high pitch. 'Sadness', in contrast, appeared encoded in a low-pitched voice, and also in a timing which concurrently appeared signalled a kind of slow-down (Figure 2.8).

Whispering specials

Aside of the role which whispering played in mediating messages and/or prosodically encoded emotion, we found that its application can encompass a further and equally interesting aspect. This was uncovered by evaluating a bipolar adjective scale which we had incorporated in our questionnaires in order to invite subjects to indicate how they personally liked an auditorily experienced phrase. Votes on this opinion were collected by two items, namely *'sympathetic'* and *'unsympathetic'*. Self-report data addressed to these items were collected by a scale of seven cells, namely *'sss', 'ss', 's', 'o', 'u', 'uu','uuu'*. Data in cell *'sss'* referred to votes for *'very sympathetic'* , data in cell *'o'* to votes for *'open',* or *'unclear'* or *'neutral',* and data in cell *'uuu'* to votes for *'very unsympathetic'*. Analysis of the ratings revealed that the votes were differently distributed for the tests with whispered and normally spoken phrases, and this difference was statistically significant [$\chi^2 = 12.48$, df 1, $p < 0.01$]. Whereas the votes given to normally spoken phrases showed a clear unimodal distribution with a frequency peak at cell *'ss'* , votes addressed to the whispered phrases were distributed in bimodal manner with a main peak at cell *'ss'* and a second peak at cell *'uu'*. This shape remained consistent, if the data were split randomly into three samples, each containing 40 votes. Since the second peak was statistically significant, we took it as relevant evidence for votes of individuals who obviously did not really like to listen to a whispered voice. With this as a reference we analyzed data collected by bipolar scales of other adjectives, and this allowed detecting the reason for the negative judgment of whispering: Subjects whose votes established the peak at cell *'uu'*, reported to feel uncomfortable during the exposure to whispering, because they had the impression to be socially excluded. The implications of these reports were investigated in a study reported in chapter 3.

Conclusions

Several investigators have studied the prosodic properties of spoken language that mediate emotions (Scherer & Kappas 1988; Hultsch et al. 1992; Tartter & Braun 1994; Sobin & Alpert 1999). However, the study reported here has provided the first comprehensive comparative approach to this issue, which also included the role of whispered speech. Some exceptions aside (see e.g. 'joy'), the portion of correctly decoded emotional expressions was remarkably large and exceeded that one reported for other studies. We interpreted this effect as a consequence of our method which was found to be very expedient already in other investigations (Wiedenmann & Todt 1992). The method allowed subjects to first extract a reference for discrimination, and then to decide among a few (here: five) alternatives, only. Based on these methodological aspects and also on our results, we concluded that the decoding of emotional expressions may be more likely a discrimination process than a process of instant identification.

Our study also suggested that, aside from the role which whispering plays in mediating prosodically encoded emotion, its social role clearly merits more attention than it received in the past. In particular it seemed interesting to further investigate whether and why some individuals obviously feel socially excluded, or at least uncomfortable, when being exposed to an unfamiliar whispering voice (see chapter 3).

2.3 Summary

In order to establish a solid foundation of my thesis, I investigated some essential signal properties of whispered speech and yielded the following results. (a) Volume: Whispering is a low-volume signal; i.e. its amplitudes (measurement distance: 1m) ranged from 52 dB(A) (extremely loud), over 40 to 35 dB(A) (normal) down to 31 dB (extremely low). Some overlap aside, these measures of provided a nice continuation of measures assessed for normal speech (extremely loud: 74 dB(A); normal: 59 to 52 dB(A); extremely low 42 dB(A)). (b) Economy: A use of whispering may not be economical; i.e. normal and loud whispering requires much more air than phonated speech. Only a very soft whispering may be economical as well, and allow to utter up to 166 syllables per breath. (c) Transmission range: Whispering is appropriate only for short-distance communication. Namely below an amplitude of 35 dB(A) it is clearly less useful than phonated speech. In addition, whispered words are especially vulnerable by masking effects of environmental noise. These qualities suggested that whispered speech is applied best if directly spoken into an addresses ear. Such suggestion, however, raised questions about its social aspects which will be treated in the next chapter.

The study on the prosodic properties of spoken language and the role of whispered speech in mediating emotional information provided a first comprehensive approach to such matters. Some exceptions aside, the portion of correctly decoded emotional expressions was larger than reported in other studies. This effect was interpreted as a consequence of our method which allowed subjects to first extract a reference for discrimination, and then to decide among five alternatives. Based on this and our results we hypothesized that the process of decoding emotional expressions is a discrimination rather than an instant identification, and thus benefits from access to one or more reference cues.

Finally, there was evidence that whispering is not only a useful tool to communicate emotional expressions, but can have also both social and psychobiological side-effects. This became clear, for example, when we analysed ratings that subjects had addressed to items such as *'sympathetic'* or *'unsympathetic'*. The results have stimulated already a further investigation of unvoiced speech, which will be reported in the next parts of this thesis (see chapters 3 & 4).