Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Speech Communication 40 (2003) 61–70

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Affective encoding in the speech signal and in event-related brain potentials

Kai Alter [a,*], Erhard Rank [b], Sonja A. Kotz [a], Ulrike Toepel [c], Mireille Besson [d], Annett Schirmer [a], Angela D. Friederici [a]

[a] *Max Planck Institute of Cognitive Neuroscience, Stephanstraße 1a, D-04103 Leipzig, Germany*
[b] *Institute of Communications and Radio-Frequency Engineering, Technical University of Vienna, Gusshausstrasse 25/389, A-1040 Vienna, Austria*
[c] *University of Potsdam, Department of Linguistics, Karl Liebknecht St. 24-25, D-14476 Potsdam, Germany*
[d] *Language and Music Group, National Center for Scientific Research (CNRS), Center for Research in Cognitive Neurosciences (CRNC), 31 ch. Joseph Aiguer, F-13402 Marseille cedex 20, France*

Abstract

A number of perceptual features have been utilized for the characterization of the emotional state of a speaker. However, for automatic recognition suitable objective features are needed. We have examined several features of the speech signal in relation to accentuation and traces of event-related brain potentials (ERPs) during affective speech perception. Concerning the features of the speech signal we focus on measures related to breathiness and roughness. The objective measures used were an estimation of the harmonics-to-noise ratio, the glottal-to-noise excitation ratio, a measure for spectral flatness, as well as the maximum prediction gain for a speech production model computed by the mutual information function and the ERPs. Results indicate that in particular the maximum prediction gain shows a good differentiation between neutral and non-neutral emotional speaker state. This differentiation is partly comparable to the ERP results that show a differentiation of neutral, positive and negative affect. Other objective measures are more related to accentuation than to emotional state of the speaker.

Keywords: Acoustics of affective speech; Event-related brain potentials

1. Introduction

The assessment of emotional content of speech is a task of growing interest, both in the field of the analysis of pathological speech (e.g., Blanken et al., 1993) as well as in the field of man–machine communication for automatic speaker state recognition and as a pre-requisite for synthesis of emotional speech (Cahn, 1990).

In the present study the relation between several segmental acoustic features of the speech signal and affect (emotional state of the speaker, lexical content of the sentences) as well as noun and verb accentuation are explored with objective measures and event-related brain potentials (ERPs).

In many investigations concerned with the analysis of emotional speech the lexical content is neutral in order to isolate acoustic features independent of lexical content. Emotional content is

* Corresponding author. Tel.: +49-341-994-0119; fax: +49-341-994-0204.
E-mail address: alter@cns.mpg.de (K. Alter).

0167-6393/02/$ - see front matter
PII: S0167-6393(02)00075-4

categorized by types of the emotional state of the speaker. Here, we employed three emotional states of a speaker (neutral, happy, and cold anger) but considered the matching or mismatching lexical content (neutral, positive, or negative). Special attention was directed to the possibility that for mismatch conditions the encoding of the intended emotional state by the speaker could be stronger than for matching lexical content.

To explore this issue, we added ERPs on top of our acoustic analyses as the implicit on-line characteristics and high temporal resolution of this measure might add meaningful insight into the study of affective language processing (see Pihan et al., 1997, 2000 for an application of DC-potentials). ERPs are a transient change of electroencephalogram (EEG) voltages reflecting systematic brain activity which is triggered by a physical event. Accordingly, in a first ERP experiment using a prosodic judgment task different emotional states were realized in qualitatively different ERP traces (Kotz et al., 2000).

Our procedure for the present study was based on the following consideration: If the mismatch between emotional state and lexical content of a sentence is also acoustically encoded in the speech signal, i.e., happy emotion combined with a negative lexical content then we have to analyze match and mismatch conditions separately in both the objective measures and the ERPs.

Furthermore, often only one type of accentuation, namely the default accentuation, is explored. To investigate the influence of accentuation the position of the sentence accent was varied. First, the nominal phrase (NP) immediately preceding the sentence final verb was accentuated, indicating neutral/default accentuation in German for verb final sentences. Second, accentuation was on the sentence final verb. The use of different accent positions is based on the hypothesis that accented syllables including their nuclei, i.e., the vowels, are *hyper-articulated* (Lindblom, 1990). Unaccented syllables and their nuclei are *hypo-articulated*. Although most of the studies on the encoding of emotional states in the speech signal found a relation to global properties, i.e., speech rate, fundamental frequency, etc. (Ladd et al., 1985; Scherer et al., 1991) we wondered to what extent speakers arousal for non-neutral emotional states might be connected to hyper-articulation. Our questions were: If hyper-articulation is related to more vocal effort locally, what happens with the acoustic encoding of different emotional states in accented vowels, and is there a measurable acoustic difference between accented and unaccented vowels related to emotional states?

To summarize, three general questions were addressed:

1. Are there relations between the acoustic parameters measured and the neural responses analyzed (objective measures and ERPs)?
2. Can we exclude possible interactions in the speech signal for the production of emotional states conflicting with the lexical content of speech match versus mismatch relative to 1?
3. Are there local acoustic interdependencies of accent placement related to hyper-articulation and affective encoding in the speech signal?

We therefore related a three-dimensional classification for the production of the speech signal to the outcome of the acoustic measures. This should allow to discriminate the influence of each dimension at first for the acoustic measurements—lexical content, emotional state and accentuation.

To anticipate our findings, on the basis of the acoustic analyses we performed an ERP study. Our findings are that some of the measures are indeed strongly correlated with accentuation and not with affect, like the harmonics-to-noise ratio (HNR) correlates with the accentuation type and lower glottal-to-noise excitation (GNE) ratio which is a characteristic of the accented word. Nonetheless, maximum prediction gain shows a basic differentiation of non-neutral affect in comparison to the neutral state. A similar discrimination between neutral and non-neutral emotion is visible in the temporal progression of ERPs of the subjects listening.

In the subsequent sections we give an illustration of the corpus recorded (Section 2), a description of the objective acoustic measures chosen and of their application to the corpus (Section 3), the assessment of the corpus by ERPs (Section 4), and conclusions in Section 5.

2. Recorded material

A corpus of emotional speech comprising 148 sentences with the same syntactic form (subject–auxiliary–NP–verb) with matching and non-matching lexical content was recorded for this investigation. The text for the sentences was chosen to cover the range of neutral, positive, and negative lexical content. The lexical content was rated by a group of subjects ($n = 20$) and classified into one of the three categories neutral, positive, or negative.

All sentences were spoken by a trained female speaker of Standard German. Each sentence was produced with two different forms of accentuation (on the NP and on the sentence final verb) and three different forms of emotional state (happiness, neutral, and cold anger) combined with semantically mismatch conditions resulting in a total of $2 \times 3 \times 148 = 888$ recorded utterances.

Example sentences with different lexical content are:

Positive:	Sie hat den Preis gewonnen. She has the prize won (literal translation).
Negative:	Er hat das Bein gebrochen. He has the leg broken (literal translation).
Neutral:	Sie hat die Tür geschlossen. She has the door closed (literal translation).

The variation of lexical content and of intended emotional state of the speaker resulted in match (e.g., positive lexical content spoken with a happy voice) and mismatch (e.g., positive lexical content spoken with an angry voice) conditions between these two factors as well as for the position of accentuation (see Table 1).

The complete crossover of emotional state and lexical content was evaluated once with a prosodic judgment task and once with a lexical judgment task, with two different subject groups.

In a first experiment 444 sentences with noun accentuation were tested in the ERP experiment. Twenty subjects each judged either the lexical content or the prosodic contour of the sentences on a five point scale (negative = 1, positive = 5). Here, only the results of the prosodic judgment task will be reported. Measurements of ERPs with verb accentuation are still ongoing, thus will not reported here.

Table 1
The dimensions analyzed acoustically in the present study

Lexical content	Emotional state	Accentuation
Positive ($n = 37$)	Happy	NP
Neutral ($n = 37$)	Neutral	Final verb
Negative ($n = 37$)	Cold anger	

Each of the lexically positive, neutral or negative sentences was completely crossed with all three emotional states resulting in $37 \times 3 \times 3 \times 2 = 846$ sentences analyzed acoustically. Note that also the unaccented vowels from both NP and verb accentuation i.e., the vowel from the unaccented verb in the NP accented condition and vice versa were analyzed (in total $2 \times 846 = 1692$ vowel samples).

For the acoustic analyses the vowels in the accented and unaccented categories (NP and verb, respectively) were extracted manually from 141 of the 148 original sentences. Thus a total of 1692 vowel samples were utilized in the acoustic analyses.

3. Acoustic correlates for breathiness and roughness evoked by emotional speaking state and accentuation

Breathiness and roughness have been used as perceptual features for the assessment of emotional speech (Klasmeyer and Sendlmeier, 1995; Klasmeyer, 1997). Objective measures for these features have been applied, e.g., for the classification of pathological voices (Fröhlich et al., 1998). Compared to segmental measures based on the spectral envelope, like spectral slope, spectral balance, or a more general description of the spectral distribution, the measures chosen capture the spectral fine structure.

We wanted to investigate the relation of several objective measures for breathiness and roughness with the parameters of the recorded database (emotional state of speaker, lexical content, and accentuation). As an acoustic correlate for breathiness and roughness we use an estimation of the *HNR*, the *GNE ratio*, a measure of *spectral*

flatness, as well as the *maximum prediction gain* for a speech production model computed by the mutual information (MI) function.

The results indicate that the HNR estimation correlates with sentence accentuation, GNE ratio with word accentuation, whereas a low maximum prediction gain indicates arousal, i.e., positive or negative emotional state of the speaker in comparison to the neutral state.

3.1. Estimation algorithms

3.1.1. Harmonics-to-noise ratio

For the computation of HNR first the harmonic components are estimated in the cepstral domain by finding the peaks at the lag corresponding to fundamental frequency (f0) and its multiples and classifying the range of cepstral coefficients around each peak as corresponding to the harmonic components. These are subtracted from the original cepstrum. The computed noise cepstrum is transformed back into the spectral domain and aligned appropriately below the original spectrum, and the HNR is computed as the total spectral energy of the original signal in relation to the energy of the noise spectrum (de Krom, 1994).

Examples for estimated noise spectra in relation with the original spectra are shown in Fig. 1 for the vowel /a/ in the sentence final verb for neutral emotional state, neutral lexical content, and accentuation on the NP versus accentuation on the sentence final verb.

3.1.2. Glottal-to-noise excitation ratio

The GNE ratio measure is based on the correlation between the Hilbert envelopes of the linear prediction residual signal in different frequency bands (Michaelis et al., 1995). For a signal evoked by glottal oscillation the glottis closure impulse triggers a pulse of the Hilbert envelope in all frequency bands. Thus, the correlation between the envelopes is high, whereas for a noise signal the correlation between the envelopes in different non-overlapping frequency bands is low. The GNE ratio thus provides a measure for the relation between glottis evoked versus noise evoked signal parts. GNE ratio is to a high degree immune to variations in fundamental period (jitter) and amplitude (shimmer) of individual pitch cycles.

GNE ratio estimation is performed by applying linear prediction analysis and inverse filtering to the speech signal downsampled by a factor of 4 ($f_S = 11,025$ Hz). Then a fast Fourier transformation (FFT) is applied and the Hilbert envelopes are calculated by performing the inverse fast Fourier transformation on 10 non-overlapping frequency bands using only FFT points corresponding to positive frequencies. For each pair of envelopes the maximum cross correlation regarding time lags between -3 and $+3$ samples is computed, and in the original algorithm the maximum of these correlation values is used as the GNE parameter. Here we also used an average over the five highest correlation values (GNEm) as control parameter.

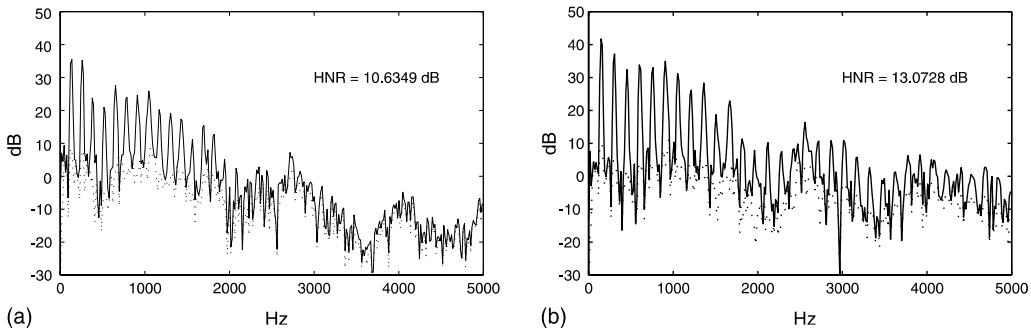

Fig. 1. Spectra of the original signal (—) and the estimated noise component (···) for the vowel /a/ in the sentence final verb for neutral emotional state and (a) accentuation on the NP, (b) accentuation on the sentence final verb, corresponding to *neutral case* '*s*' versus *neutral case* '*v*' in Fig. 2.

3.1.3. Spectral flatness

Spectral flatness is the ratio of the geometric to the arithmetic mean of the spectral energy distribution (Markel and Gray, 1976). As such, it is limited to a range between zero and one, and equal one only for a perfectly flat spectrum. If we express the spectral flatness in dB the resulting range of values is $-\infty$ to zero. The spectral energy distribution is computed by a FFT. Signal and FFT length were chosen according to the propositions by Markel and Gray (1976) with the signal downsampled to 11,025 Hz windowed to 128 points and applied to a 256 point FFT.

3.1.4. Maximum prediction gain

The maximum prediction gain has been chosen as a measure for the amount of glottis oscillator evoked—and thus predictable—in relation to the noisy–unpredictable–signal components. The maximum prediction gain estimated by the MI function regards the non-linear characteristics in the signal production system (Bernhard, 1997, 1998), and, as the results show, achieves a measure clearly distinct from the other methods used.

The application of the MI function relies on the embedding of the speech signal in a low dimensional pseudo phase space with a dimension according to the underlying production system. It has been shown that voiced phonemes—and particularly vowels—can be considered as signals produced by a low dimensional system ($d \sim 3$). So for the estimation of the maximum prediction gain a three-dimensional pseudo phase space reconstruction by time delay embedding ($T = 0.7$ ms) was used for the speech signal and the maximum prediction gain was computed for a one sample ahead prediction (Bernhard, 1997).

It is noteworthy that three of the measures used here, the HNR estimation, spectral flatness, and the maximum prediction gain computed by use of the MI function, cannot provide a measure for the noisy signal components alone, but—since they rely on stationarity—are also affected by other attributes like frequency variations (jitter) or amplitude variations (shimmer) (Pinto, 1990). Moreover, the estimation algorithms for both HNR and maximum prediction gain also depend on the length of the signal. Thus, although the measures were chosen as possible correlates for the perceptual features breathiness and roughness they constitute features of the speech signal and hence have to be classified as acoustic rather than as perceptual features.

3.2. Analysis of results

Analysis was performed on the signals corresponding to vowels in the NP and the sentence final verb. The signal was sampled at 44,100 Hz with 16 bit resolution. All analyses were performed for the verbs in the NP and in the sentence final verb for 141 sentences (comprising different lexical content) uttered with three different emotional states and two different accentuation types, summing up to 1692 vowel samples.

Generally, no explicit dependence of either HNR, GNE ratio, spectral flatness, or maximum prediction gain on the lexical contents of a sentence was found. Also the mismatch conditions between lexical content and affect yield *no effect* in the speech signal. Thus, in the following the results presented are restricted to the parameters *emotional state of speaker* and *accentuation*.

The results for the HNR analysis show a slightly higher value for the cold anger emotional state than for neutral or happy emotional state. A more explicit effect is exemplified in Fig. 2 for the vowels in the sentence final verb: if sentence accent is on the NP (case 's') the HNR is generally lower for the vowel in the verb than if the verb is accented (case 'v').

Also for the vowels in the NP (Fig. 3) higher values (at least for cold anger and happy state) are indicated for accentuation of the sentence final verb (case 'v') then for accentuation of the NP itself.

So, for both the vowel in the NP and in the sentence final verb the HNR value is higher if the sentence accent is placed on the sentence final verb. HNR values thus correlate with the *accentuation type of the whole sentence*.

Analysis by the GNE ratio on the other hand yields distinctive values depending on the *word accentuation* for non-neutral emotional state. Figs. 4 and 5 again show the statistics for vowels in the NP and the sentence final verb. Note the difference

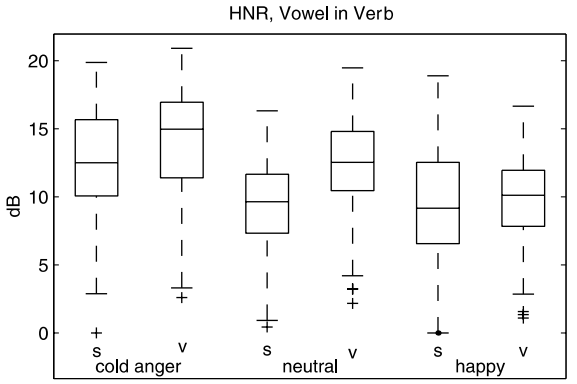

Fig. 2. Results of the statistical analysis of the estimated HNR for vowels from the sentence final verb and the three emotional states. For each case the box ranges from the lower to the upper quartile with the median value indicated by a line, and the total range indicated by the whiskers. Outliers are indicated by crosses. For each emotional state higher HNR is found when the verb is accented (case 'v').

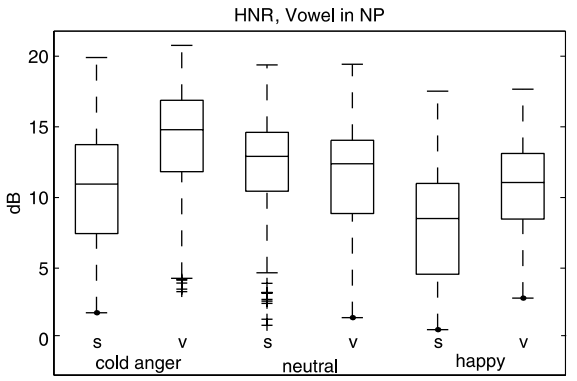

Fig. 3. Results of the statistical analysis of the estimated HNR for the three emotional states and vowels from the NP. Like in the analysis for the sentence final verb (Fig. 2) the HNR for cold anger and happy emotional state is higher when the verb is accented (case 'v') than when the NP is accented.

in Fig. 4 compared to Fig. 2 concerning the values for non-neutral emotional state. Both in the vowel of the NP and of the sentence final verb a lower GNE ratio—corresponding to a higher amount of noise excitation—is observed if the word bearing the vowel is accented (case 'v' for the vowel in the verb and case 's' for the vowel in the NP). The GNE ratio can thus be used as an indicator for word accentuation.

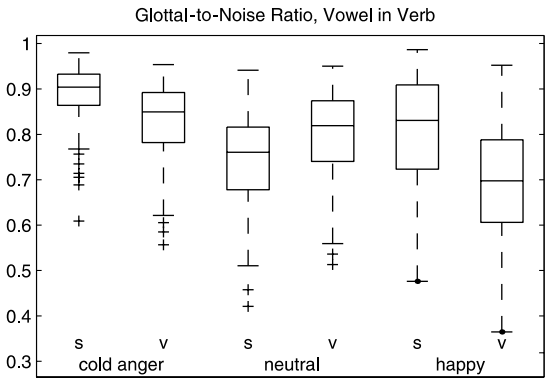

Fig. 4. Results of the statistical analysis of GNE ratio for the three emotional states and vowels from the sentence final verb.

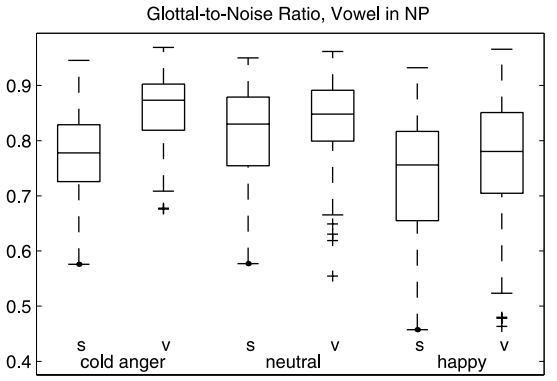

Fig. 5. Results of the statistical analysis of GNE ratio for the three emotional states and vowels in the NP.

Examination of spectral flatness shows no distinct values except for the case of unaccented vowels in the sentence final verb in neutral emotional state, which exhibit somewhat lower spectral flatness than all other cases (Fig. 6).

In contrast to the other measures, maximum prediction gain provides a distinction between the neutral emotional state (higher prediction gain) and cold anger/happy state (lower prediction gain) and is merely independent of accentuation, as shown in Fig. 7. Thus, arousal—i.e., a non-neutral affect—seems to result in a less predictable speech waveform.

For both HNR and maximum prediction gain the analyses were also performed for the speech signal at 11,025 Hz sampling rate and qualitatively

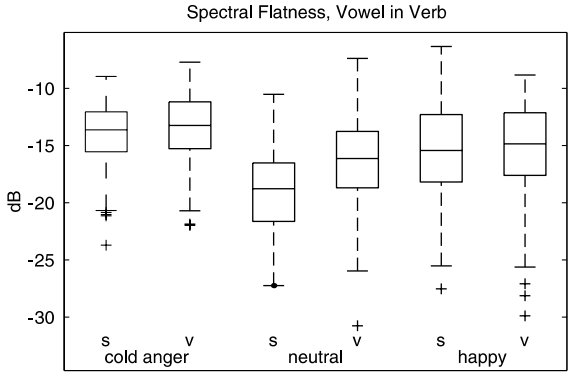

Fig. 6. Results of the statistical analysis of spectral flatness for the three emotional states and vowels from the sentence final verb.

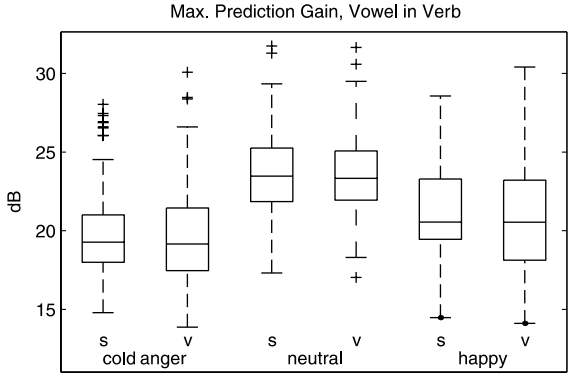

Fig. 7. Results of the statistical analysis of the maximum prediction gain computed by means of the MI function (graphic presentation like in Fig. 2). Non-neutral emotional state results in a generally lower maximum prediction gain than emotionally neutral speech; accentuation ('v' versus 's') has a remarkable low influence on maximum prediction gain.

the same results are achieved, with generally slightly higher HNR values and lower prediction gain.

Closer examination reveals that a more distinctive distribution of the analysis results can be achieved by taking into account the identity of the vowel. In Fig. 8 the vowel identity (SAMPA notation) is printed in the statistics for the maximum prediction gain.

Some vowels tend to generally yield lower prediction gain (/a/, /E/, ...) whereas others generally yield higher values (/o/, /U/, ...). Hence, using the

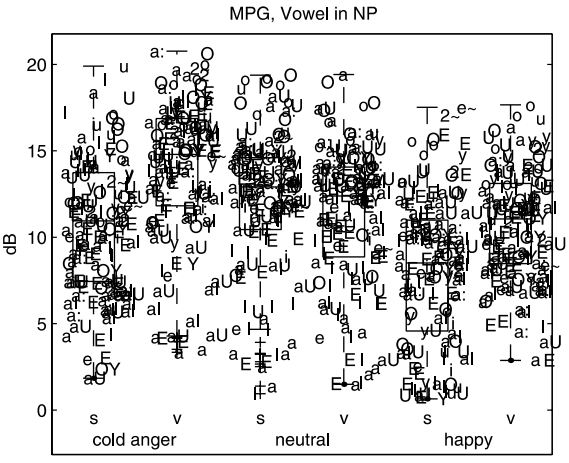

Fig. 8. Position of individual vowels in the distribution of the maximum prediction gain analysis.

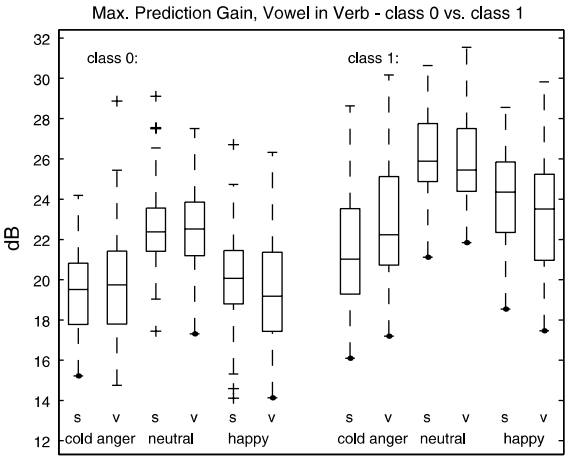

Fig. 9. Maximum prediction gain for two vowel classes.

vowel identity as a parameter may help to find a clearer distinction between neutral and emotional speaker state. This is exemplified in Fig. 9 where the maximum prediction gain is plotted for two different vowel classes (class 1: /a/, /a:/, /E/, /e/, /I/, /i/, /aI/ and /aU/, and class 2: /O/, /o/, /U/, /u/, /OY/, /Y/, /y/ and /2/).

4. Event-related brain potentials

ERPs allow the differentiation of language subprocesses as reflected in language-related

components. For example specific syntactic violations result in a biphasic pattern of early negative and late positive voltage changes, while semantic violations elicit a late negative change (see Friederici, 1995; Kutas and Hillyard, 1980). Thus, ERPs seem to be an appropriate tool to differentiate language-related characteristics. While there is ample evidence in the clinical literature that the processing of affective prosody might vary as a function of valence (positive versus negative affect; Davidson and Tomarken, 1989), there is very little evidence from online measures. However, some seminal DC-potential work by Pihan et al. (1997, 2000) reports that the discrimination of sentences with happy, sad or neutral intonation results in DC-potential patterns that vary as a function of fundamental frequency or the duration of syllable stress. However, the valence of prosodic affect does not vary by any of the acoustic manipulations.

The EEG was recorded from 32 cap-mounted tin electrodes with a sampling rate of 250 Hz/12 bits and with 40 Hz low-pass filtering. The left mastoid electrode served as the reference. A total of 20 subjects (10 female, mean age 23 years) were tested in the prosodic judgment task. Trials containing eye blinks or movement artefacts were rejected. Averages were first computed for each single subject. These averages then entered the grand averages. ERP components were quantified as amplitude means of specified time windows.

Concerning the different prosodic emotional states, there were differences between all conditions, as shown by repeated measures ANOVA. Fig. 10 shows the main pattern: there was a significant difference between the positive state and both neutral and negative states as reflected in a P200 component. At around 400 ms post-stimulus onset there was a stronger differentiation between neutral and both emotional states that persisted over the course of the sentence. However, between 400 and 700 ms positive and negative emotional states differed significantly.

The pattern for mismatch conditions was less clear, but broadly similar. As the comparison between match and mismatch conditions was always based on the same emotional content but mismatching prosody, the similarity with respect to match and mismatch conditions suggests that the differences between contours could not be attributed to lexical effects alone as Pihan's work suggests might be the case.

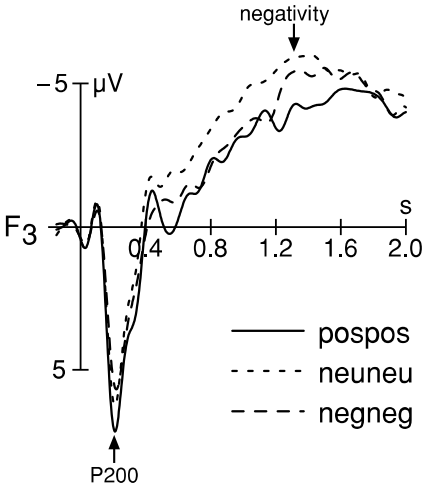

Fig. 10. The differentiation in the ERP between neutral (dotted line) and happy/cold anger (straight/dashed line) emotional speaker state. Waveforms illustrate the averages for all three conditions from 150 ms prior to sentence onset to 2000 ms at a selected frontal electrode site.

One of the attractions of the ERP technique is that it may illuminate the time course of affect processing. Fig. 10 suggests that distinctive brain reaction extends over most of the sentence rather than being highly localized. The traces begin to diverge after 200 ms and furthermore around the onset of the noun (about 400 ms after stimulus onset). The distinction is maintained as the sentence continues, and in some respects it is enhanced when the sentence final verb occurs (about 1400 ms after stimulus onset). These interpretations are tentative, because further work is needed to eliminate alternative explanations of the patterns. But the data indicate why the technique is promising.

5. Conclusions

In our study acoustic measures of the speech signal chosen as correlates for breathiness and roughness, as well as traces of ERPs were analyzed regarding their relation to the emotional state of the speaker (affect), lexical content, and accentua-

tion. It has been found that emotional state (match versus mismatch) could only be differentiated in the maximum prediction gain. The dimensions of *accentuation* and *affect* are almost mutually independently captured by distinct acoustic measures. They seem to be encoded in different features of the speech signal, and as the assessment of listener perception with EPR traces shows, also to trigger different perceptual events in the listener.

There is a correlation between the maximum prediction gain and the differentiation of ERP traces related to different emotional states. Utterances comprising neutral emotional state are characterized by a higher maximum prediction gain and a more negative ERP trace than utterances with a happy emotional state. Thus, those features could be an indicator of arousal, i.e., to distinguish between a non-neutral emotional state of the speaker and the neutral state.

A low GNE ratio of a vowel signal was found to go with accentuation of the word, whereas the HNR estimation correlates with accentuation type of the sentence, i.e., accentuation of the sentence final verb versus default accentuation on the NP.

A strong differentiation between ERP traces for neutral and non-neutral emotional state of the speaker is independent of lexical content and accentuation. Presumably, the hearers perception system seems to use signal properties at all stages of the incoming signal in order to process affective meaning. Accented and thus hyper-articulated signal portions are overlooked by the system during the processing of affective meaning. Both the acoustic analyses and the ERP data suggest that accentuation and the encoding of affect are two separate prosodic entities. The former seems to be a local quality of prosodic encoding, the latter seems to be realized globally.

Acknowledgements

The work was supported by the Leibniz Science Prize awarded to Angela D. Friederici. We thank Erdmut Pfeifer for constructive discussion of the corpus manipulations, and three anonymous reviewers for helpful comments of an earlier version of the manuscript.

References

Bernhard, H.-P., 1997. The mutual information function and its application to signal processing. Ph.D. thesis, University of Technology, Vienna.

Bernhard, H.-P., 1998. A tight upper bound on the gain of linear and nonlinear predictors for stationary stochastic processes. IEEE Trans. Signal Process. 43, 2909–2917.

Blanken, G., Dittmann, J., Grimm, H., Marshall, J.C., Wallesch, C.W. (Eds.), 1993. Linguistic Disorders and Pathologies. An International Handbook. de Gruyter, Berlin, New York.

Cahn, J.E., 1990. The generation of affect in synthesized speech. J. Amer. Voice I/O Soc. 8, 1–19.

Davidson, R.J., Tomarken, A.J., 1989. Laterality and emotion: an electrophysiological approach. In: Boller, F., Grafman, J. (Eds.), Handbook of Neuropsychology, Vol. 3. Elsevier, Biomedical Division, pp. 419–441.

de Krom, G., 1994. Acoustic correlates of breathiness and roughness. Ph.D. thesis, Utrecht. LEd.

Friederici, A.D., 1995. The time course of syntactic activation during language processing. A model based on neuropsychological and neurophysiological data. Brain Lang. 50, 259–281.

Fröhlich, M., Michaelis, D., Strube, H.W., 1998. Acoustic 'breathiness measures' in the description of pathologic voices. In: Proc. ICASSP'98, 2, Seattle, WA, pp. 937–940.

Klasmeyer, G., 1997. The perceptual importance of selected voice quality parameters. In: Proc. ICASSP'97, Munich, Germany, pp. 1615–1618.

Klasmeyer, G., Sendlmeier, W.F., 1995. Objective voice parameters to characterize the emotional content in speech. In: Proc. ICPhS'95, Vol. 1, Stockholm, Sweden, pp. 181–185.

Kotz, S.A., Alter, K., Besson, M., Schirmer, A., Friederici, A.D., 2000. The interface between prosodic and semantic processes: an ERP study. J. Cog. Neurosci. (Suppl. 123).

Kutas, M., Hillyard, St.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.

Ladd, D., Silverman, K., Tolkmitt, F., Bergmann, G., Scherer, K., 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. J. Acoust. Soc. Amer. 78, 435–444.

Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H & H theory. In: Hardcastle, W., Marchal, A. (Eds.), Speech Production and Speech Modelling. Kluwer, Dordrecht, pp. 403–439.

Markel, J.D., Gray, A.H., 1976. Linear Prediction of Speech. Springer, Berlin–Heidelberg–New York.

Michaelis, D., Gramss, T., Strube, H.W., 1995. Glottal-to-noise excitation ratio – A new measure for describing pathological voices. Acta Ac. 81, 700–706.

Pihan, H., Ackermann, H., Altenmüller, E., 1997. The cortical processing of perceived emotion: A DC potential study on affective prosody. Neuro. Rep. 8, 623–627.

Pihan, H., Altenmüller, E., Hertrich, I., Ackermann, H., 2000. Cortical activation patterns of affective speech processing depend on concurrent demands on the subvocal rehearsal system. A DC-potential study. Brain 123, 2338–2349.

Pinto, N.B., 1990. Unification of perturbation measures in speech signals. J. Acoust. Soc. Amer. 87 (3), 1278–1289.

Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T., 1991. Vocal cues in emotion encoding and decoding. Motiv. Emot. 15, 123–148.