COGNITION AND EMOTION, 2001, *15* (1), 57–79

The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample

Caterina Breitenstein

University of Southern California, Los Angeles, USA
and *University of Trier, Germany*

Diana Van Lancker

University of Southern California, Los Angeles, USA
and *NYU School of Education, New York, USA*

Irene Daum

Ruhr-University Bochum, Germany

The present study examined acoustic cue utilisation for perception of vocal emotions. Two sets of vocal-emotional stimuli were presented to 35 German and 30 American listeners: (1) sentences in German spoken with five different vocal emotions; and (2) systematically rate- or pitch-altered versions of the original emotional stimuli. In addition to response frequencies on emotional categories, activity ratings were obtained. For the systematically altered stimuli, slow rate was reliably associated with the ‘‘sad’’ label. In contrast, fast rate was classified as angry, frightened, or neutral. Manipulation of pitch variation was less potent than rate manipulation in influencing vocal emotional category choices. Reduced pitch variation was associated with perception as sad or neutral; greater pitch variation increased frightened, angry, and happy responses. Performance was highly similar for the two samples, although across tasks, German subjects perceived greater variability of activity in the emotional stimuli than did American participants.

Please send correspondence and requests for reprints to Dr Caterina Breitenstein, Dept. of Neurology, University of Muenster, Albert-Schweitzer Strasse 33, 48129 Muenster, Germany. Phone: +49-251-834 8195; Fax: +49-251-834 8181; e-mail: breitens@uni-muenster.de

This research was supported by a postdoctoral fellowship granted to Caterina Breitenstein by the Deutsche Forschungsgemeinschaft (Br 1753/2).

We thank JoAnn Olson, Daniel Kempler, Jody Kreiman, Barry Schein, and Maryellen MacDonald for their support in the recruitment of the American student sample, Andreas Stevens for his help with the German sample, and Ingo Hertrich for technical advice. We also express our appreciation to the listeners on both sides of the Atlantic Ocean.

© 2001 Psychology Press Ltd

http://www.tandf.co.uk/journals/pp/02699931.html DOI:10.1080/02699930042000114

Little is known about the contribution of single acoustic parameters to the perception of specific vocal emotions. Therefore, the primary goal of the present investigation was to examine the effects of systematic acoustic manipulation of vocal emotional stimuli on listeners' perception. In addition, even less is known about the cross-cultural processing of vocal emotions (emotional prosody), as the majority of cross-cultural studies have focused on cue utilisation in the perception of *facial expressions* (e.g., see Cornelius, 1996, for review; Ekman & Friesen, 1971; Massaro & Ellison, 1996) or vocal correlates of *personality dimensions* (e.g., Apple, Streeter, & Krauss, 1979; Brown, Strong, & Rencher, 1973; Kramer, 1963; Scherer, 1972, 1974, 1978). Thus, a second, more exploratory aim was to compare two different nations (American, German) on the same set of stimuli to address questions about sample differences in the perception of emotional tones of voice.

Acoustic correlates of vocal emotions

Different emotions are cued by different combinations of acoustic parameters, with speech rate and fundamental frequency (F0; the acoustic correlate of perceived pitch in a speaker's voice) presumably exerting the strongest effect (Banse & Scherer, 1996; Murray & Arnott, 1993, 1995; Pell, 1999; Protopapas & Lieberman, 1997; Scherer, 1982; 1995; Scherer & Oshinsky, 1977). It has also been shown that the use of F0 for linguistic distinctions is functionally separable from the role of F0 in vocal emotions (McRoberts, Studdert-Kennedy, & Shankweiler, 1995). Mean F0 and speech rate are generally highest for emotions associated with high sympathetic arousal (anger, fear; see Ellgring & Scherer, 1996); a slower speech rate is generally observed in passive, e.g., sad utterances (Johnson, Emde, Scherer, & Klinnert, 1986). Multiple regression analyses examining the association of specific acoustic parameters and recognition rates for different vocal emotions revealed significant correlations for most of the 14 studied emotions with a particular set of acoustic parameters (Banse & Scherer, 1996), with F0 and temporal parameters overall exerting the strongest effects.

Furthermore, not all emotional categories are recognised equally well. Raters' recognition scores are generally highest for (hot) anger and sadness, lower for happiness, fear (Banse & Scherer, 1996), and worst for disgust (Scherer, Banse, Wallbott, & Goldbeck, 1991).

To provide direct evidence for the specificity and independence of acoustic parameters to the judgement of vocal emotions, the systematic and independent variation of these parameters in natural speech (resynthesis) is required. Previous attempts to systematically vary acoustic parameters experienced technical constraints with regard to natural sound quality (Carlson, 1992; Frick, 1985; Kappas, Hess, & Scherer, 1991), especially for the manipulation of F0 variation (Brown et al., 1973). Previous studies that investigated the contribution of F0 and amplitude on listeners' judgements of synthesised *tone stimuli*, in which acoustic information was systematically manipulated, demonstrated that mean F0, F0 variation, and tempo were the most powerful cues for listeners' judgements (Lieberman & Michaels, 1962; Scherer & Oshinsky, 1977). However, the use of tone sequences instead of naturally spoken utterances limits the generalisability of these findings for interpersonal communication. Using naturally spoken utterances with *neutral intonation* for resynthesis showed that gradual increases in F0 range resulted in gradual changes in arousal ratings (Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985); utterances with a small F0 range were judged as sounding sad, and stimuli with a wide F0 range as sounding happy (Bergmann, Goldbeck, & Scherer, 1988). Furthermore, increased vowel duration was associated with higher sadness and lower anger ratings (Bergmann et al., 1988).

Finally, Bergmann et al. (1988) altered *emotionally intoned* sentences with respect to F0 range, intensity, and duration, trying to neutralise the emotional meaning of the stimuli (see also Pell, 1998, using a similar design in comparing stroke patients and healthy elderly controls). The most dramatic effects on listeners' ratings arose from alteration of speech rate: Ratings for sadness increased with an increase in speech duration for the sad stimulus; for the angry stimulus, ratings on the anger scale increased with a reduction in stimulus duration.

The present study aimed to further investigate the effects of independent variation of two acoustic parameters (F0 variation, speech rate) on listeners' judgements of vocal emotions, using a *wider span* of manipulation factors and emotional categories and using *qualitative* (emotional labels) as well as *quantitative* (activity) measures. The present investigation used utterances, which were actually spoken with *emotional tones* of voice by a professional actress (and not as neutral statements). Recent technical advances in speech analysis software (e.g., LPC Parameter Manipulation/Synthesis by Kay Elemetrics, USA), which take the importance of pitch perturbations into account (Lieberman, 1961; Lieberman & Michaels, 1962), make it possible to maintain a natural sounding voice quality while allowing for systematic variation of different acoustic parameters.

Hypotheses

Of special interest was whether the systematic manipulation of one acoustic parameter (F0 variation *or* speech duration) would result in a gradual shift in perception, such that, for example, an utterance with a frightened prosody would be perceived as sad when speech duration was systematically increased. A second goal was to examine whether persons from different *nations* use similar acoustic cues in the recognition of vocal emotions. We therefore compared native speakers of American English and German, using stimuli in the German language.[1]

[1] In many ways, English and German are similar in phonological structure, with respect to consonant-vowel patterns and syllabic shape (e.g., Scherer, 1974).

METHODS

Subjects

American sample. A total of 30 undergraduate and graduate students (22 female, 8 male) from the University of Southern California in Los Angeles participated in the experiment (all but 5 were right-handed). All participants (and their parents) were native speakers of American English, and none was fluent in German. The majority of subjects ($n = 21$) was brought up in Southern California.

German sample. Undergraduate and graduate students (23 women, 12 men) were recruited from the University of Bochum ($n = 30$) and the University of Tübingen ($n = 5$). There were 30 right-handed and 5 left-handed subjects in the group. All students were native speakers of German and brought up in Germany.

Stimulus material

Twenty sentences from a standardised set of emotional prosodic stimuli (Breitenstein, Daum, Ackermann, Lütgehetmann, & Mueller, 1996) were selected. The sentences were in German and neutral in content, but differed in meaning[2] (4 different meanings; see Appendix A) and emotional tone of voice (5 categories: happy, sad, angry, frightened, and neutral). All sentences had identical grammatical structures (subject, verb, object). The inclusion of a neutral category was considered methodologically crucial as a control category because confusion errors do not occur with the same likelihood across all emotional categories. With the inclusion of a neutral category, subjects are less likely to use exclusion rules, which may result in artificially high rates of correct guesses (cf., Scherer, 1986). All sentences were recorded by a professional German actress[3] in a sound-attenuated room using a portable digital tape player (SONY TCD-D7) and a microphone (SONY ECM-959A) placed at a mouth distance of 30 cm. The actress was instructed not to use additional vocal sounds, such as laughs or sobs, for the portrayal of vocal emotions.

It may be argued that spontaneous rather than posed productions are a more valid design to study vocal emotions. It has been pointed out in this context, however, that ‘‘it is rather unlikely that strategic, emotive messages use signal patterns that differ strongly from spontaneous, emotional expressions’’ because naturally occurring vocal emotional expressions are generally controlled by

---

[2] Prior to the main statistical analyses, ANOVAs were conducted to test any systematic effect of sentence content on response frequencies. As expected, neither main effects of sentence content nor interactions with nationality were significant. Data were therefore pooled across sentence contents.

[3] Previous research indicated that effects for acoustic variables are speaker-independent and generalisable across different speakers (Ladd et al., 1985).

sociocultural rules of emotional expression (cf., Banse & Scherer, 1996, p. 618). Furthermore, comparisons of voice samples obtained under real-life and simulated emotional conditions indicate comparable changes in acoustic parameters (Williams & Stevens, 1972) and a much larger database is available on acoustic correlates of portrayed emotions (cf., Scherer, 1995).

Speech signals were digitised at a sampling rate of 25 kHz[4] (after low-pass filtering at 9 kHz) and 16-bit quantisation using speech analysis software (''Computerized Speechlab 4300'', Kay Elemetrics Corp., USA). The F0 was determined using a pitch frame length of 20 ms and a pitch frame advance of 10 ms (maximum frequency of F0: 350 Hz; minimum F0: 50 Hz). Following the visual correction of artefacts in the intonation contour (acoustic data for the original sentences are presented in Appendix B), the data underwent systematic variation of two acoustic parameters (standard deviation of F0,[5] speech rate) using LPC Parameter Manipulation/Synthesis (Kay Elemetrics Corp., USA). The choice of these two acoustic parameters was based on prior studies indicating the crucial and independent role of pitch variation and temporal aspects in real vocal emotions (Williams & Stevens, 1972) and in judgements of vocal emotions portrayed by actors (Banse & Scherer, 1996; Bergmann et al., 1988; Ellgring & Scherer, 1996).

For each acoustic parameter (speech rate, F0 variation), six[6] sentences were resynthesised from each of the 20 original sentences (tokens): For half of the sentences, the values for the parameter of interest were *decreased* by multiplication with the factor levels .9, .7, and .5. For the other half, the values were *increased* by multiplication with the factors 1.1, 1.3, and 1.5. Following the recommendations by Bergmann et al. (1988), to ensure that the stimuli sounded as natural as possible, the new values for F0 (after the multiplication) were corrected by addition or subtraction of a constant, so that the speaker's baseline F0 of 120 Hz (which was the lowest F0 data point occurring in any of the 20 tokens) remained constant. For speech rate manipulation, both consonant and vowel information were linearly expanded or compressed.[7] Altogether, 240

[4] The speech analysis software required a sampling rate of 10 kHz (or less). For that reason, all sentences were downsampled at 10 kHz prior to setting of pitch markers.

[5] Mean F0 was not selected for the acoustic manipulation because global changes in mean F0, independent of F0 variation, produce voice quality differences which do not constitute clear emotional prosodic cues (cf., Bergmann et al., 1988).

[6] The final stimulus set was selected on the basis of pilot work involving a larger stimulus set with a finer-grained manipulation (10 sentences from each of the original sentences) of the two acoustic factors which yielded almost identical results.

[7] Vowel duration is usually slightly more affected by changes in speech rate than is the duration of the consonants. Because: (a) the exact ratio of vowel to consonant lengthening or shortening is unknown, and (b) alteration of total speech rate yielded stronger effects on emotion attributions compared to alteration of stressed syllables (Bergmann et al., 1988), we decided to linearly expand and compress each utterance (which resulted in a very natural-sounding speech quality).

stimuli (5 emotional categories times 4 different meanings times 6 factor levels) were resynthesised. All 240 stimuli were recorded in random order on a digital tape with an ISI of 4 seconds (the trial number was announced by a male voice in English). These are the ''rate-'' and ''pitch-altered'' stimuli.

A second tape was prepared and presented to all subjects. It contained the 20 original sentences, which were rated separately from the synthesised stimuli because of the detectable difference in sound quality.

Procedure

The procedures were kept very similar for the American and German samples. After subjects gave written informed consent, they filled out a demographic data sheet which contained questions about years and field of education, their foreign language abilities, time spent living abroad, and years of musical training. Subjects were tested in groups of two to three, and all stimuli were presented on portable digital tape recorders (SONY TCD-D7) and headphones (beyerdynamic DT211). Headphones were chosen instead of loudspeakers to minimise distraction effects and to allow each individual to proceed at his/her own pace. The subjects were told to set the volume to a comfortable setting. They were also informed that they could stop the tape at any time and listen to a sentence again if they wished to. All instructions were given in written form to optimise standardisation. The American subjects were given an English translation of the sentences before performing the listening task. For the 240 synthesised stimuli (rate- and pitch-altered tape) and the tape with the 20 original sentences, the participants were instructed to categorise and rate the emotional tone of voice of each sentence on two scales: (a) an emotional category which contained the five emotion labels, and (b) a 5-point activity scale (with the endpoints ''very active'' vs. ''very passive''). The rationale for including the activity scale was based on recent findings that: (1) differences in pitch range have continuous (rather than categorical) effects on listeners' judgements of vocal emotions (Ladd et al., 1985), particularly on attributions of arousal, and (2) the activity dimension is correlated with larger pitch variability and faster speech rate (see Frick, 1985).

Practice trials (three for the original stimuli, six for the acoustically altered stimuli) were included to ensure acquaintance with the speaker's pattern of vocal emotions (Cosmides, 1983). Subjects were not given feedback at any point during the experiment.

To control for effects of experience, fatigue, and carry-over, the sequence of the two tapes (synthesised stimuli, original stimuli) was counterbalanced across subjects. For the synthesised stimuli, the sequence of part 1 (trials 1–120) and part 2 (trials 121–240) was counterbalanced for the same reasons. The session lasted about 60 minutes, and all subjects were reimbursed for their participation or received course credit.

Statistical analyses

From our previous experience with the stimulus material (pilot study), it was predicted that the effects of the acoustic manipulations (decrease and increase of speech rate or pitch variation as compared to the original prosodic stimulus parameters) could be described as linear or quadratic trends with respect to frequency of response categories. It was expected that specific responses would increase or decrease fairly linearly across the acoustic manipulation continuum (e.g., the slower the rate, the higher the number of sad responses and vice versa). For each acoustic parameter manipulation (speech rate, pitch variation) and each emotional category (happy, sad, angry, frightened, neutral), a 2 (*nationality*: German, American) by 5 *response category*[8]: happy, angry, sad, frightened, neutral) by 7 (*manipulation* factor: .5, .7, .9, 1.0, 1.1, 1.3, 1.5) design with repeated measures over the latter two factors was used for all statistical analyses. The original stimuli (manipulation factor: 1.0) were included in the same analysis with the acoustically altered stimuli. This was done to control for baseline group differences with respect to emotion detection and to achieve a purer measure of group differences due to the acoustic manipulations. For clarity of presentation, if higher order interactions achieved significance, lower order interactions are only reported when the lower order effect is of primary interest with respect to the study purposes.

Group differences on single variables were analysed by Bonferroni-adjusted independent *t*-tests (pooled or separate variance) or Mann–Whitney *U*-tests, as appropriate.

RESULTS

Background variables

Information on subjects' demographic data is presented in Table 1. The American and the German groups did not differ significantly with respect to age, years of education, years of musical training, or the distribution of sex or handedness (all ps > .05). The groups did, however, differ significantly with respect to years spent living abroad ($U = 439$, $p = .046$) as well as in the number of self-reported languages spoken fluently ($U = 179$, $p < .001$). This is probably due to a cultural difference in lifestyle and education, because it is becoming increasingly popular among German students to spend a portion of their university training in America. Furthermore, the German school system requires every student to learn English.

[8] If a response category lacked sufficient variance, it was dropped from analyses. ANOVAs were then conducted with 4 (or in some cases 3) response categories. Because not all response categories were observed for each of the emotional categories, the five emotional categories were not examined simultaneously in a 2 (nationality) by 5 (emotional category) by 5 (response category) by 7 (manipulation factor) design.

TABLE 1
Demographic data

	American sample	*German sample*
Sample size (*n*)	30	35
Female/Male (*n*)	22/8	23/12
Right/Left-handed (*n*)	25/5	30/5
Years of education (*M*/SD)	15.6+/ − 1.5	15.8+/ − 2.3
Age in years (*M*/SD)	22.1+/ − 4.2	24.0+/ − 3.6
Years of musical training (*M*/SD)	4.3+/ − 3.5	4.9+/ − 3.9
Years spent living abroad (*M*/Median)	0.0/0	0.2/0
Number of languages (*M*/Median)	1.1/1	1.9/2

M = mean; SD = standard deviation.

Test or item sequence

Neither the sequence of the tasks (original stimuli, rate- or pitch-altered stimuli) nor the sequence of items within the speech-manipulated stimuli yielded an effect on the groups' overall scores (all ps > .25). Therefore, the data were pooled for all analyses.

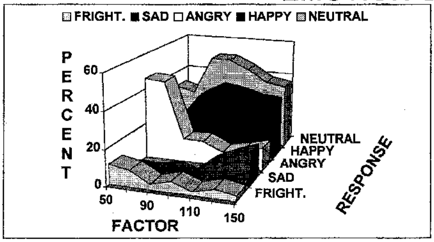
Rate-altered stimuli

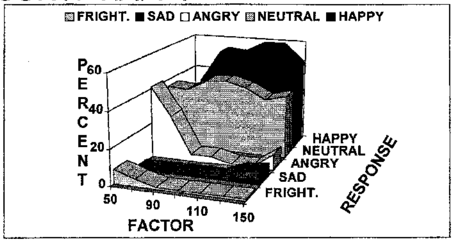
Happy category. Analyses of variance with the between-group factor *nationality* and the within-factors *response category* (happy, sad, angry, frightened, neutral) and *manipulation factor* yielded a significant interaction of the three factors, quadratic trend: $F(1, 63) = 3.76$, $p < .05$, as well as significant two-way interactions of response category by manipulation factor, linear trend: $F(1, 63) = 22.60$, $p < .001$), and nationality by response category, $F(4, 252) = 5.18$, $p = .001$. To explain the three-way interaction, group differences of trends across manipulation factors were analysed separately for each of the five response categories. The analyses yielded different quadratic trends of the two groups for the *happy* response category only, quadratic trend: $F(1, 63) = 3.85$, $p < .05$. For both groups, the frequency of happy responses was lowest at the extreme ends of manipulation, quadratic trends: both Fs(1, 29/1,34) > 20.04, ps < .001, but the trend was steeper in the German group. As can be seen in Figure 1, this was because at less extreme manipulation levels, the German subjects were much more likely to rate the stimuli as happy, thereby

Figure 1 (opposite). Frequencies of responses (in per cent; mean and standard error) for speech rate manipulation across the seven manipulation factors for the five different emotional target categories in the American and German groups (50 to 150 = original stimulus duration × 0.5 to 1.5; target emotions are displayed in black, confusion errors in grey).

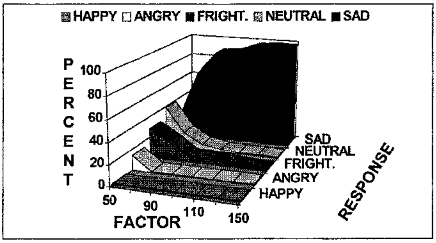
AMERICAN **GERMAN**

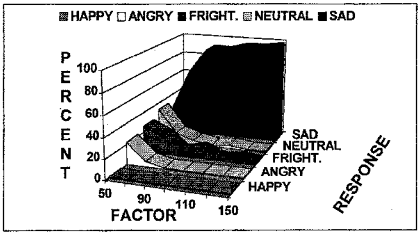
EMOTION CATEGORY: "HAPPY"



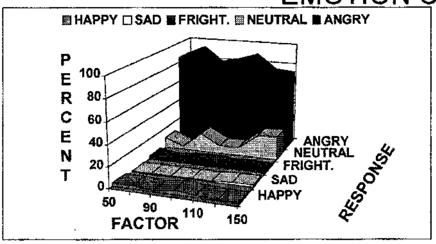


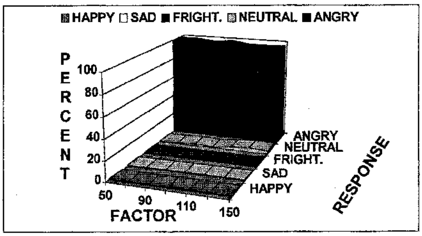
EMOTION CATEGORY: "SAD"



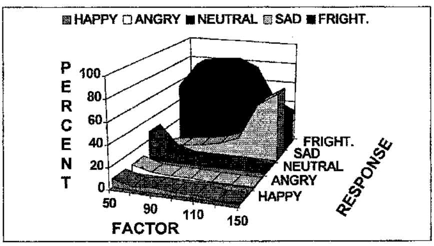


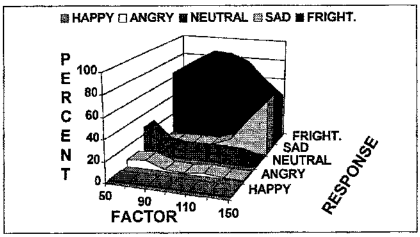
EMOTION CATEGORY: "ANGRY"



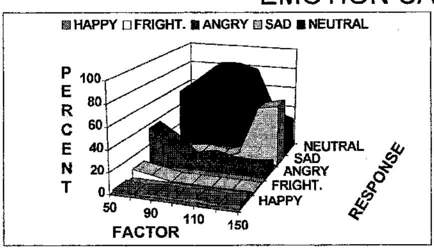


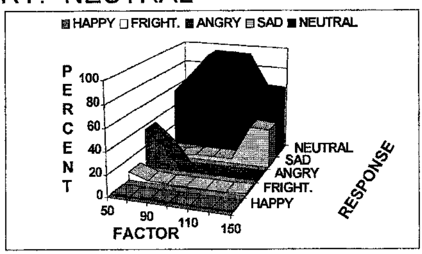
EMOTION CATEGORY: "FRIGHTENED"





EMOTION CATEGORY: "NEUTRAL"





producing a steeper drop-off for this group at the more extreme levels of manipulation.

Furthermore, the interaction of nationality by response category was due to significant group differences of happy, angry, and frightened responses, all Fs(1, 63) > 7.05, ps < .01, with higher frequencies of both angry and frightened responses in the American than the German group (both $ts_{(44.4/47.8)} > 2.58$, ps < .05), whereas happy responses were more frequent in the German as compared to the American group ($t_{(61.9)} = -2.94$, $p = .02$). The two groups did not differ with respect to overall frequency of sad or neutral responses.

The above-mentioned interaction of response category by manipulation factor could be explained with significant increases of angry and frightened responses with faster speech rate, linear/quadratic trends: all Fs(1, 63) > 8.59, ps < .01), and significant increases of happy and sad responses when speech rate was reduced, linear/quadratic trends: all Fs(1, 63) > 15.93, ps < .001, in both groups.

Sad category. Analyses of variance were conducted with the between-group factor *nationality* and the within-factors *response category* (sad, frightened, neutral) and *manipulation factor*. Of all three- and two-way interactions, only the interaction of response category by manipulation factor was significant, linear/quadratic trends: both Fs(1, 63) > 269.93, ps < .001. As depicted in Figure 1, sad responses were significantly more frequent at slower speech rates, linear/quadratic trends: both Fs(1, 63) > 167.72, ps < .001, whereas frequencies of frightened and neutral responses increased with faster speech rate, linear/quadratic trends: all Fs(1, 63) > 21.14, ps < .001. No group differences in response frequencies were observed in this emotion category.

Angry category. The three-way interaction of *nationality* by *response category* (happy, angry, neutral) by *manipulation factor* yielded significance, linear/quadratic trends: both Fs(1, 63) > 4.20, ps < .05. In subsequent analyses, group differences of trends across manipulation factors were analysed separately for each of the three response categories and revealed significant group differences in trends of the response categories angry and neutral, linear/quadratic trends: both Fs(1, 63) > 3.94, ps $\leq$ .05. As shown in Figure 1, German subjects responded almost exclusively with angry responses when listening to stimuli of this emotional category, independent of the degree of the acoustic manipulation. Americans subjects, however, decreased the frequency of angry responses, linear trend: $F(1, 29) = 5.88$, $p = .02$, when speech rate was slowed while tentatively increasing the frequency of neutral responses, quadratic trend: $F(1, 29) = 3.71$, $p = .06$.

Frightened category. A significant three-way interaction of *nationality* by *response category* (sad, angry, frightened, neutral) by *manipulation factor* was noted, linear trend: $F(1, 63) = 4.66$, $p = .04$. Analyses were followed up by a series of ANOVAs analysing group differences across manipulation factors separately for the four response categories. Only for sad responses, an interaction of nationality by manipulation factor, linear trend: $F(1, 63) = 5.00$, $p = .03$, was observed. As can be seen in Figure 1, American subjects showed a steeper increase of sad responses than German subjects when speech rate was slowed, linear trends: $Fs(1, 29/1, 34) > 106.15$, $ps < .001$.

A significant two-way interaction of response category by manipulation factor, linear/quadratic trends: both $Fs(1, 63) > 47.76$, $ps < .001$, could be explained with significant increases in frequency of frightened, neutral, and angry responses with faster speech rate, whereas the frequency for sad responses increased with slowing of speech rate, linear trends: all $Fs(1, 63) > 15.38$, $ps < .001$.

Neutral category. The three-way interaction of *nationality* by *response category* (happy, sad, angry, neutral) by *manipulation factor* yielded significance, linear trend: $F(1, 63) = 20.41$, $p < .001$. Separate ANOVAs for each of the four response categories revealed significant two-way interactions of nationality by manipulation factor of sad and neutral responses, linear trends: both $Fs(1, 63) > 19.17$, $ps < .01$. As depicted in Figure 1, both groups decreased the frequency of neutral responses when speech rate was slowed, linear trends: American subjects, $F(1, 29) = 30.46$, $p < .001$; German subjects, not significant, whereas responses on the sad category increased for both groups, linear trends: both $Fs(1, 29/1, 34) > 71.60$, $ps < .001$. The effects of rate manipulation, however, were significantly steeper in the American as compared to the German sample.

A significant two-way interaction of response category by manipulation factor was also noted, linear/quadratic trends: both $Fs(1, 63) > 100.51$, $ps < .001$, and could be explained with a significant increase of angry responses at faster speech rates and higher frequency of sad responses at slower speech rates, linear/quadratic trends: all $Fs(1, 63) > 48.80$, $ps < .001$. Neutral responses were lowest at both extreme levels of acoustic manipulation, quadratic trend: $F(1, 63) = 139.61$, $p < .001$.

In summary, *slower* speech rate resulted frequently in classifications as sad; *higher* speech rate was associated with greater angry, frightened, or neutral responses. As can be seen in Figure 1, both groups used the happy category least frequently and scarcely confused any of the other emotional categories with happy.

Activity ratings. A significant three-way interaction of *nationality* by *emotional category* by *manipulation factor* was found, quadratic trend: $F(1, 63) = 4.15$, $p < .05$. The interaction could be explained with significantly different trends of the two groups for stimuli of the happy, sad, and neutral target categories, quadratic trends: all $Fs(1, 63) > 5.79$, $ps < .02$. As can be seen in Figure 2, the slower the *happy* and *sad* stimuli, the lower the activity rating

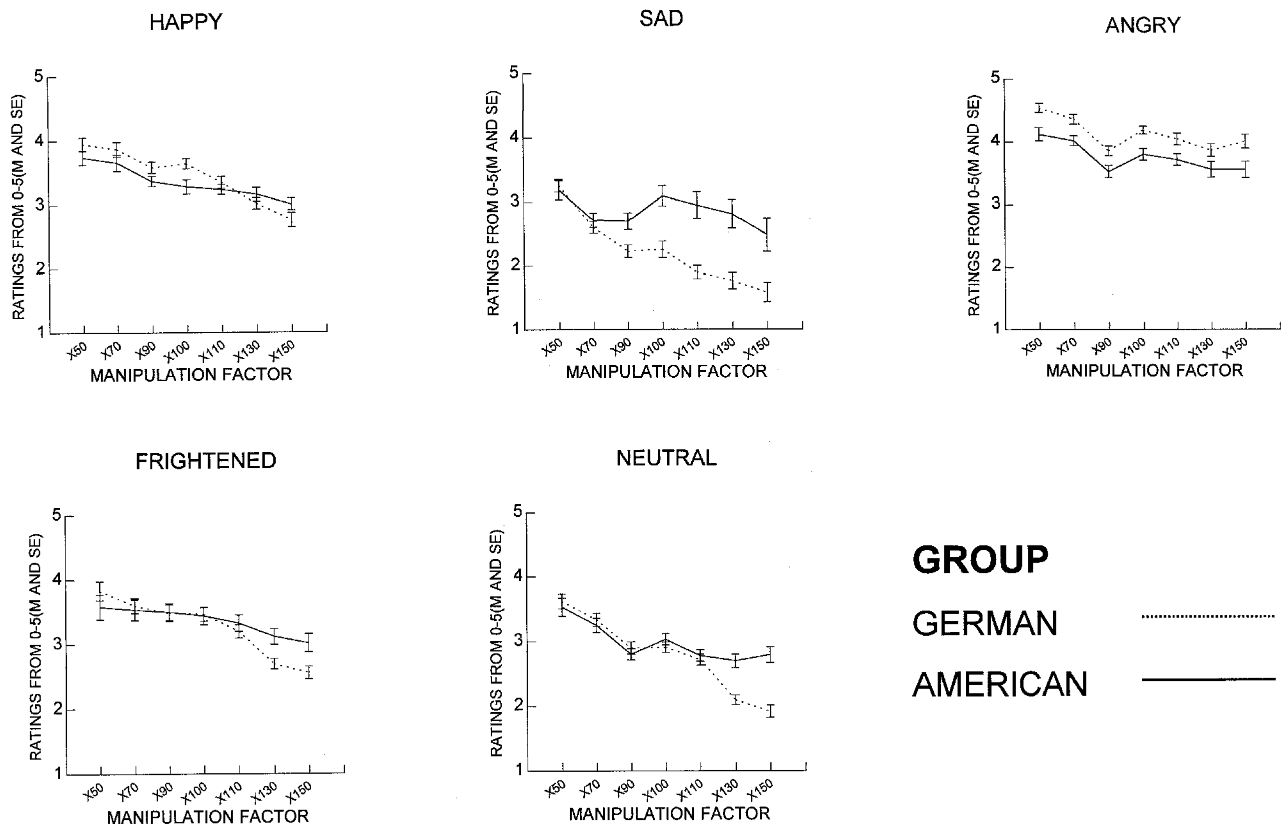


Figure 2. Activity ratings (mean and standard error) for speech rate manipulation across the seven manipulation factors for the five different emotional categories in the American and German groups ($\times 50$ to $\times 150$ = original stimulus duration $\times 0.5$ to 1.5).

given by the German subjects, quadratic trends: both Fs$(1, 34) = 4.77$, ps $< .04$. The American group was less affected in activity judgements across manipulation factors (quadratic trends: ps $> .05$). For *neutral* stimuli, both groups judged faster stimuli as significantly more active than slower stimuli, quadratic trend: $F(1, 63) = 10.40$, $p = .002$. The trend was, however, steeper in the German subjects as compared to the American sample, both Fs$(1, 29/1, 34) > 15.06$, ps $< .01$, shown in Figure 2. In summary, German subjects' activity ratings were apparently more ''sensitive'' to the decrease in rate for happy, sad, and neutral stimuli.

No group differences were noted for angry and frightened stimuli. Both groups, however, rated faster frightened stimuli as more active than slower stimuli, quadratic trend: $F(1, 63) = 8.21$, $p < .01$. Activity ratings of angry stimuli were not significantly affected by the rate manipulation.

Pitch-altered stimuli

Happy category. Analysis of variance revealed a significant interaction of *response category* (happy, sad, angry, frightened, neutral) by *nationality*, $F(4, 252) = 9.17$, $p < .001$, and a significant interaction of *response category* by *manipulation factor*, linear trend: $F(1, 63) = 40.49$, $p < .001$, but no significant three-way interaction.

Subsequent analyses of group differences in response categories yielded significant higher frequencies of angry and frightened responses of the American as compared to the German sample (both ts$_{(44.7/30.2)} > 3.83$, ps $< .01$). German subjects, on the other hand, responded more often with happy than the American group ($t_{(59.3)} = -3.91$, $p = .001$). No group differences were observed in frequencies of sad or neutral responses.

To explain the interaction of response category by manipulation factor, trends across manipulation factors were analysed separately for the five response categories, pooled for American and German subjects. For all response categories, significant linear trends across manipulation factors emerged, linear trends: all Fs$(1, 64) > 5.61$, ps $< .02$). As can be seen in Figure 3, neutral and, to a lesser extent, sad responses became more frequent with a lowering of pitch variation and were less likely when pitch variation increased. For happy, angry, and frightened responses, the opposite pattern was observed: Response frequencies increased with greater pitch variation.

Sad category. Of all two- and three-way interactions and main effects, only the main effect *response category* (sad, frightened, neutral) yielded significance, $F(2, 126) = 3087.40$, $p < .001$. *Post-hoc* analyses using paired t-tests with Bonferroni-adjusted probabilities revealed significantly higher response frequencies of sad as compared to frightened and neutral responses, as well as significantly higher frequencies of frightened as compared to neutral responses (all $ts_{(64)} > 2.51$, $ps < .05$). As can be seen in Figure 3, the frequency of sad responses was at ceiling in both groups.

Angry category. The three-way interaction of *nationality* by *response category* (happy, angry, neutral) by *manipulation factor* achieved significance, linear/quadratic trends: both $Fs(1, 63) > 5.50$, $ps < .03$. To explain the interaction, group differences in trends across manipulation factors were analysed separately for the three response categories. These ANOVAs indicated group differences of angry and neutral responses across manipulation factors, linear/quadratic trends: all $Fs(1, 63) > 7.44$, all $ps < .01$, but not of the happy response category (which was infrequently chosen in both groups). For *angry* responses, American subjects increased response frequencies at both extreme levels of manipulation, quadratic trend: $F(1, 29) = 6.27$, $p = .02$, whereas German subjects slightly decreased angry responses at the highest pitch variation levels, linear trend: $F(1, 34) = 5.57$, $p = .02$. For *neutral* responses, American subjects increased response frequencies when pitch variation was lowered, linear trend: $F(1, 29) = 16.46$, $p < .001$. German subjects, on the other hand, most frequently used the neutral response category at the extreme high ends of pitch variation manipulation, linear trend: $F(1, 34) = 38.44$, $p < .001$.

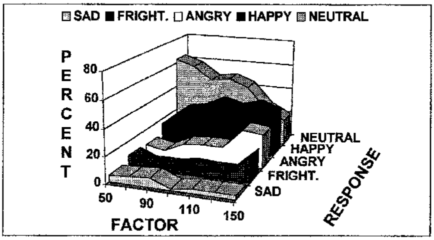
A two-way interaction of response category by nationality was also noted, $F(2, 126) = 36.54$, $p < .001$, which could be explained with overall higher frequencies of angry responses of the German as compared to the American sample ($t_{(32.0)} = -5.68$, $p < .001$) and more neutral responses of the American as compared to the German subjects ($t_{(32.5)} = 5.67$, $p < .001$).

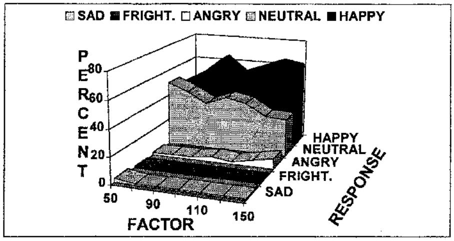
Frightened category. Analysis of variance with the factors *nationality*, *response category* (happy, sad, angry, frightened, neutral), and *manipulation factor* yielded a significant interaction of the three factors, quadratic trend: $F(1, 63) = 4.24$, $p = .04$, and a significant two-way interaction of response category by manipulation factor, linear trend: $F(1, 63) = 8.80$, $p = .004$.

To explain the three-way interaction, group differences were analysed separately for the five response categories. The two-way interaction of manipulation factor by nationality was significant for angry responses only, quadratic trend: $F(1, 63) = 4.54$, $p = .04$. Subsequent analyses of trends across manipulation factors separately for each group were not significant for either group, quadratic trends: both $ps > .10$. As displayed in Figure 3, response frequencies of angry were fairly small in both groups and data presumably lacked sufficient variance to detect group differences.

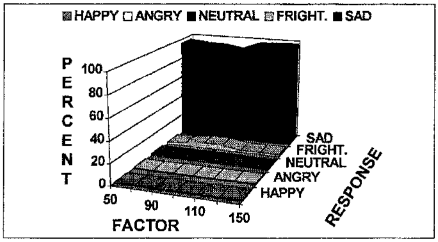
Figure 3 (opposite). Frequencies of responses (in per cent; mean and standard error) for manipulation of pitch variation across the seven manipulation factors for the five different emotional target categories in the American and German groups (50 to 150 = original stimulus duration × 0.5 to 1.5; target emotions are displayed in black, confusion errors in grey).

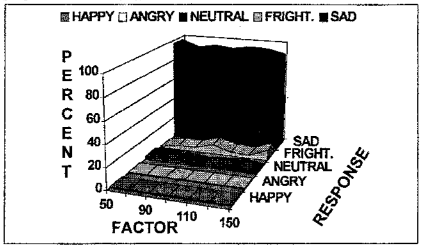
AMERICAN EMOTION CATEGORY: “HAPPY” **GERMAN**



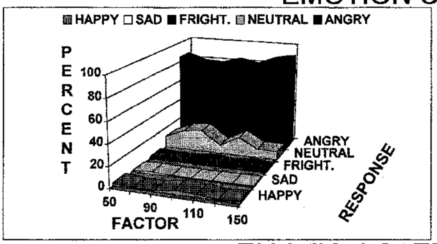


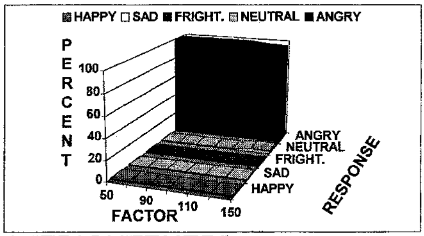
EMOTION CATEGORY: “SAD”



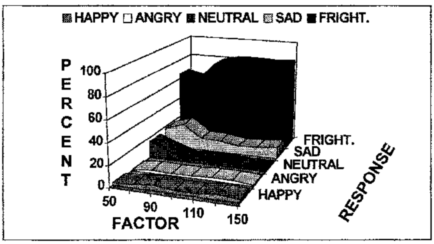


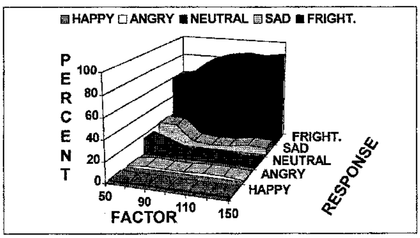
EMOTION CATEGORY: “ANGRY”



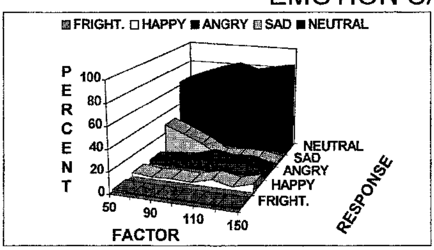


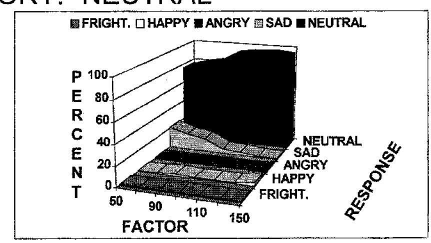
EMOTION CATEGORY: “FRIGHTENED”





EMOTION CATEGORY: “NEUTRAL”





The interaction of response category by manipulation factor could be explained with significant frequency changes of sad, angry, frightened, and neutral responses across manipulation factors, linear trends: all $Fs(1, 63) > 6.01$, $ps < .02$. As shown in Figure 3, both groups responded more frequently with sad, neutral, and angry with a lowering in pitch variation; frightened responses were more frequent with greater pitch variation.

Neutral category. Analysis of variance was conducted with the factors *nationality*, *response category* (happy, sad, angry, neutral), *and manipulation factor*. Of all interactions, only the two-way interactions of response category by nationality, $F(3, 189) = 8.76$, $p < .001$, and response category by manipulation factor, linear/quadratic trends: both $Fs(1, 63) > 13.39$, $ps < .001$, achieved significance.

The interaction of response category by manipulation factor was followed up by separate ANOVAs for the four different response categories, examining trends across manipulation factors. As shown in Figure 3, neutral, linear trend: $F(1, 64) = 51.20$, $p < .001$, and, to a lesser extent, angry and happy, quadratic/linear trends: both $Fs(1, 64) > 3.90$, $ps < .05$, responses became more frequent with an increase in pitch variation. Sad responses, however, were more frequent at the low extreme of pitch variation, linear trend: $F(1, 64) = 52.97$, $p < .001$.

To explain the interaction of response category by nationality, group differences were examined for each of the four response categories separately. Groups differed on angry and neutral response frequencies with American as compared to German subjects responding overall more frequently with angry ($t_{(30.9)} = 4.23$, $p = .001$). German subjects chose neutral responses more frequently than the American sample ($t_{(46.6)} = -3.22$, $p = .01$).

In summary, subjects' responses were altogether less affected by manipulation of pitch variation as compared to speech rate manipulation. In general, sad and neutral responses were more frequent when pitch variation was lowered whereas frightened, angry, and happy responses increased with greater pitch variation of the stimuli.

Activity ratings. Of all two- and three-way interactions, only the two-way interaction of *emotional category* by *nationality* yielded significance, $F(4, 248) = 14.13$, $p < .001$. Subsequent paired comparisons revealed significantly higher activity ratings of the German subjects for the happy and angry stimulus categories (both $ts_{(63)} > |-3.21|$, $ps < .01$) and significantly higher activity ratings of the American subjects on sad stimuli ($t_{(63)} = 4.60$, $p < .001$; see Figure 4). Significant main effects of *manipulation factor*, linear trend: $F(1, 63) = 110.57$, $p < .001$, and *emotional category*, $F(4, 248) = 98.30$, $p < .001$, were also found. Independently of emotional context, both groups significantly increased ratings on the activity scale when pitch variation of the stimuli became greater. Furthermore, overall activity ratings of angry stimuli were significantly higher

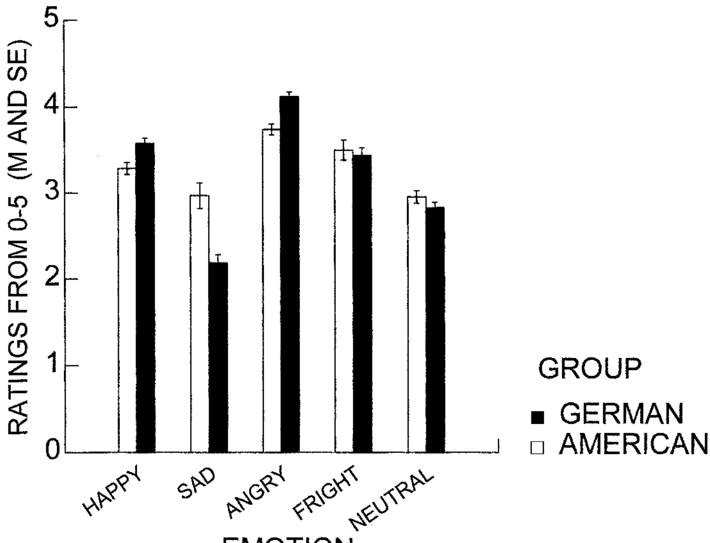


Figure 4. Activity ratings (mean and standard error) for manipulation of pitch variation (pooled across the seven manipulation factors) for the five different emotional categories in the American and German groups.

than those of the other four emotional categories; happy and frightened stimuli were rated as significantly more active than neutral stimuli; sad stimuli obtained significantly smaller activity ratings than the other emotional categories (all $ts_{(64)} > |3.20|$, $ps < .01$).[9]

DISCUSSION

Synthesised stimuli based on the same original recordings were presented to native German and native American subjects in order to examine the effects of manipulation of single acoustic parameters (speech rate, pitch variation) on the perception of vocal emotions. The design was also intended to probe the respective roles of speech rate and pitch variation in emotion perception in these two samples. In addition to frequency of responses on five emotional categories (happy, sad, angry, frightened, neutral), an activity rating scale was included to examine associations with subjects' choice of emotions and to further compare patterns of recognition between the two samples. Rate was clearly the most potent cue, with pitch variation under matched manipulation conditions contributing less, which is in agreement with previous work (Murray & Arnott, 1993; Scherer & Oshinsky, 1977; but see also Protopapas & Lieberman, 1997).

[9] Paired t-tests using Bonferroni-adjusted probability levels.

Rate-altered stimuli

For *rate-altered* stimuli, both groups were affected in a similar way. Overall, despite the acoustic manipulations, subjects chose most frequently the emotion label which corresponded to the emotion category in which the actress had originally spoken the utterance. The only exception was for the American group in the happy category where subjects responded most frequently with neutral. The effects of the acoustic manipulation were most obvious at one or both extreme levels of the manipulation: Recognition of the happy and sad target emotions was mostly affected by increased speech rate whereas detection of the frightened and the neutral target emotions was reduced both at extreme fast and extreme slow levels of manipulation. The recognition of utterances originally spoken with angry intonation, however, was hardly affected by the rate manipulation, indicating that other vocal cues (e.g., tense/harsh voice quality) may have had a greater impact on subjects' responses than the rate manipulation. This was also evident in the activity ratings for this emotional category.

Both groups had strikingly similar response patterns, such that stimuli of the *happy* target category were frequently perceived as angry and, to a lesser extent, frightened or neutral (fast rate) and sad (slow rate); fast *sad* tokens were often recognised as frightened or neutral and slowing of the stimuli increased the frequency of sad responses; sentences originally spoken with *frightened* intonation were mostly classified as frightened and, to a lesser extent, neutral or angry (fast rate) and sad (slow rate); stimuli of the *neutral* target category were frequently perceived as neutral or angry (fast rate) and sad (slow rate). Across all emotional target categories except angry (see Figure 1), the frequency of sad responses increased when sentences became slower. Response frequencies on frightened, angry, and neutral, however, decreased with slowing of the utterances. The opposite pattern (less sad responses and more frightened, angry, and neutral responses) was observed when speed of the utterances was increased. As hypothesised, systematic manipulation of the acoustic parameter speech rate resulted in gradual shifts of subjects' emotion perception. The findings are consistent with previous correlational approaches, in that fast tempo (high sympathetic arousal) is a powerful cue in the perception of angry and frightened vocal emotions, whereas slower tempo is related to the perception of sad prosody (Banse & Scherer, 1996; Scherer, 1986; Scherer & Oshinsky, 1977). The findings are also in agreement with those of another study on systematic manipulation of acoustic cues using only two target emotions, in that a reduction in speech rate is associated with a perception of anger (Bergmann et al., 1988).

As predicted (Ellgring & Scherer, 1996), increases in speech rate were associated with higher activity ratings for both groups and all target emotions (except anger). Conversely, a decrease in speech rate led to lower activity ratings.

Although more similarities than differences were noted, the German and American samples differed in response frequencies across manipulation factors for some of the emotional categories. For the happy stimulus category, German subjects chose the happy response category more frequently than American subjects at faster rates, whereas American subjects responded more often with angry or frightened responses. Unlike German subjects, American participants chose angry less often when an angry stimulus was slowed, and responded with neutral more frequently. For the frightened and neutral stimulus categories, American subjects chose sad responses to a greater extent than German subjects when speech rate was slowed. American subjects were thus more affected by rate manipulation than the German sample.

Activity ratings of the two groups differed for happy, sad, and neutral target emotions with the Germans subjects showing greater differentiation between slow and fast exemplars of the same tokens. For the happy and neutral category, group differences in activity perception may partially explain group differences in response frequencies.

Overall, the data are consistent with those of previous cross-cultural studies on vocal emotions (Albas, McCluskey, & Albas, 1976; Beier & Zautra, 1972; Van Bezooijen, Otto, & Heenan, 1983) in finding overall higher intra-cultural compared to cross-cultural scores and finding highest recognition scores for angry, sad, and neutral target emotions.

Pitch-altered stimuli

For manipulation of *pitch variation*, response frequencies of the emotion label, which corresponded to the original emotional tone of voice of the stimuli, increased linearly with an increase in pitch variation for happy, frightened, and neutral stimuli. Perception of sad and angry stimuli was hardly affected by the manipulation of pitch variation, which may be partially due to a ceiling effect. Both groups tended to classify stimuli of the frightened and neutral categories as sad when pitch variation was lowered. For the happy category, happy, angry, and frightened responses increased with higher pitch variation; response frequencies of neutral and sad increased towards the low extreme of the manipulation. In summary, consistent with earlier reports (Bergmann et al., 1988), small pitch variation served as a cue for sad and neutral percepts. The overall effect of pitch variation on activity ratings was much weaker than expected (cf., Ladd et al., 1985).

As predicted (see Ladd et al., 1985), activity ratings increased with greater pitch variation of the stimuli. Independent of the acoustic manipulation, activity ratings were highest for stimuli of the angry and happy stimulus categories and lowest for the sad category.

As for the rate-altered stimuli, the distribution of response frequencies across manipulation factors was similar in the German and American samples, supporting the notion of similar pitch functions across cultures (Ohala, 1983, 1984). Group differences in response trends were noted for the angry and frightened stimulus categories with American subjects more frequently responding with angry at higher levels of pitch variation than the German subjects. As can be seen in Figure 3, however, group differences are very subtle.

Activity ratings differed between groups for stimuli of the happy, angry, and sad stimulus category and were not clearly related to the differences in response pattern.

Conclusions, study limitations, and future outlook

As mentioned earlier, recognition of the underlying target emotions is possible under conditions of acoustic manipulations. This indicates that: (a) the synthesised stimuli were of very good sound quality, and (b) additional vocal cues are important for listeners' judgements. Future studies should focus on the effects of systematic manipulation of additional acoustic parameters (e.g., intensity) on classifications of vocal emotions.

Consistent with previous studies on recognition of vocal emotions in non-manipulated stimuli, Figures 1 and 3 show that recognition of the target emotion was highest for the emotional categories anger and sadness and lowest for happiness (Banse & Scherer, 1996; Scherer et al., 1991).

Overall, two interesting differences to previous studies are that: (1) response patterns were not clearly related to the activity dimension (cf., Albas et al., 1976; Van Bezooijen et al., 1983), and (2) subjects in our study frequently chose the neutral category when incorrect. Because some of the earlier cross-cultural studies (Albas et al., 1976; Kramer, 1964) had not included a neutral response category, comparability of findings is limited. As in previous studies, response frequencies were not randomly distributed over emotions (Banse & Scherer, 1996).

Having both groups also listen to American emotionally intoned sentences would have been desirable, except for the above-stated German-American asymmetry in familiarity of the companion language. Because of the considerable demands of the design (each subject rated 260 acoustic stimuli and each stimulus was judged on two scales), subjects were not presented with a second analogous stimulus set in the English language. This clearly constrains the interpretation of our findings for universality of vocal emotions because German subjects may have been more sensitive to subtle changes in the acoustic stimuli (cf., Albas et al., 1976).

Our long-term goal is the development of a ‘‘content masking procedure’’ by using stimuli in a foreign language. This is highly desirable for studying emotional dysprosody in neurologically impaired subjects who often present with (additional) attention and memory problems and may be much more distractible by the sentence content than neurologically healthy subjects (Breitenstein, Van

Lancker, Kempler, Daum, & Waters, 1998). Some groups have used content-filtering procedures, but the ecological validity of that approach remains questionable (Baum & Nowicki, 1998). Because we observed more similarities than differences between the two cultural groups, we feel confident using this design in future studies with non-native patient populations.

Using more than one actress/actor for emotion portrayals would have been desirable as well with respect to greater generalisability of our findings. As well as adding length to the subjects' task, it is doubtful that we would have obtained smooth linear/quadratic trends across the manipulation factors of our acoustic variations by the introduction of this additional source of variance.

Manuscript received 13 July 1998
Revised manuscript received 27 June 2000

REFERENCES

Albas, D.C., McCluskey, K.W., & Albas, C.A. (1976). Perception of the emotional content of speech—A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology*, *7*, 481–489.

Apple, W., Streeter, L.A., & Krauss, R.M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, *37*, 715–727.

Banse, R., & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636.

Baum, K.M., & Nowicki, S. Jr. (1998). Perception of emotion: Measuring decoding accuracy of adult prosodic cues varying in intensity. *Journal of Nonverbal Behavior*, *22*, 89–107.

Beier, E.G., & Zautra, A.J. (1972). Identification of vocal emotions across cultures. *Journal of Consulting and Clinical Psychology*, *39*, 166.

Bergmann, G., Goldbeck, T., & Scherer, K.R. (1988). Emotionale Eindruckswirkung von prosodischen Sprechmerkmalen [The inference of speaker emotions based on prosodic cues in speech]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *35*, 167–200.

Breitenstein, C., Daum, I., Ackermann, H., Lütgehetmann, R., & Mueller, E. (1996). Erfassung der Emotionswahrnehmung bei zentralnervösen Läsionen und Erkrankungen: Psychometrische Gütekriterien der ''Tübinger Affekt Batterie'' [Assessment of deficits in emotional perception following cortical and subcortical brain damage: Psychometric properties of the ''Tübingen Affect Battery'']. *Neurologie & Rehabilitation*, *2*, 93–101.

Breitenstein, C., Van Lancker, D., Kempler, D., Daum, I., & Waters, C.H. (1998). The contribution of working memory to the perception of emotional prosody in Parkinson's disease [Special Issue]. *Brain and Language*, *65*, 243–246.

Brown, B.L., Strong, W.J., & Rencher, A.C. (1973). Perceptions of personality from speech: effects of manipulation of acoustical parameters. *Journal of the Acoustical Society of America*, *54*, 29–35.

Carlson, R. (1992). Synthesis: Modeling variability and constraints. *Speech Communication*, *11*, 159–166.

Cornelius, R.R. (1996). *The science of emotion*. Upper Saddle River, NJ: Prentice Hall.

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 864–881.

Ekman, P., & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Ellgring, H., & Scherer, K.R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, *20*, 83–110.

Frick, R.W. (1985). Communicating emotion: the role of prosodic features. *Psychological Bulletin*, *97*, 412–429.

Johnson, W.F., Emde, R.N., Scherer, K.R., & Klinnert, M.D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, *43*, 280–283.

Kappas, A., Hess, U., & Scherer, K.R. (1991). Voice and emotion. In R.S. Feldman & B. Rime (Eds.), *Fundamentals of nonverbal behavior* (pp. 200–238). Cambridge, UK: Cambridge University Press.

Kramer, E. (1963). Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin*, *60*, 408–420.

Kramer, E. (1964). Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology*, *68*, 390–396.

Ladd, D.R., Silverman, K.E.A., Tolkmitt, F., Bergmann, G., & Scherer, K.R. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, *78*, 435–444.

Lieberman, P. (1961). Perturbations in vocal pitch. *Journal of the Acoustical Society of America*, *33*, 597–603.

Lieberman, P., & Michaels, S.B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America*, *34*, 922–927.

Massaro, D.W., & Ellison, J.W. (1996). Perceptual recognition of facial affect: cross-cultural comparisons. *Memory and Cognition*, *24*, 812–822.

McRoberts, G.W., Studdert-Kennedy, M., & Shankweiler, D.P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: evidence for a dissociation. *Perception and Psychophysics*, *57*, 159–174.

Murray, I.R., & Arnott, J.L. (1993). Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, *93*, 1097–1108.

Murray, I.R. & Arnott, J.L. (1995). Implementation and testing of a system for producing emotion-by-rule in synthetic speech. *Speech Communication*, *16*, 369–390.

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

Ohala, J.J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.

Pell, M. (1998). Recognition of prosody following unilateral brain lesion: influence of functional and structural attributes of prosodic contours. *Neuropsychologia*, *36*, 701–715.

Pell, M. (1999). Fundamental frequency encoding of linguistic and emotional prosody by right hemisphere-damaged speakers. *Brain and Language*, *69*, 161–192.

Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America*, *101*, 2267–2277.

Scherer, K.R. (1972). Judging personality from voice: a cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, *40*, 191–210.

Scherer, K.R. (1974). Voice quality analysis of American and German speakers. *Journal of Psycholinguistic Research*, *3*, 281–298.

Scherer, K.R. (1978). Personality inference from voice quality: the loud voice of extraversion. *European Journal of Social Psychology*, *8*, 467–487.

Scherer, K.R. (1982). Methods of research on vocal communication: paradigms and parameters. In K.R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (pp. 136–198). Cambridge, UK: Cambridge University Press.

Scherer, K.R. (1986). Vocal affect expression: a review and model for future research. *Psychological Bulletin*, *99*, 143–165.

Scherer, K.R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*, 235–248.
Scherer, K.R., Banse, R., Wallbott, H.G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, *15*, 123–148.
Scherer, K.R., & Oshinsky, J.S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *4*, 331–346.
Van Bezooijen, R., Otto, S.A., & Heenan, T.A. (1983). Recognition of vocal expressions of emotions. A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology*, *14*, 387–406.
Williams, C.E., & Stevens, K.N. (1972). Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, *52*, 1238–1250.

APPENDIX A

German sentences used in the emotional prosodic tasks (and the English translations in parentheses)

(1) Der Schal ist in der Truhe (The scarf is in the closet)
(2) Das Kind ging in den Zoo (The child went to the zoo)
(3) Die Stühle sind aus Holz (The chairs are made of wood)
(4) Die Uhr ist im Regal (The clock is on the shelf)

APPENDIX B

Acoustic data of the original sentences (pooled across the four sentence meanings)

	Mean pitch (in Hz)	*Pitch variation* (F0-SD/mF0)	*Duration* (in s)
Happy	209	.19	1.4
Sad	158	.08	1.6
Angry	193	.17	1.6
Frightened	239	.07	0.9
Neutral	162	.15	1.3

F0-SD = standard deviation of F0; mF0 = mean F0.

View publication stats