Short summary of results from my master-thesis concerning the acoustical correlates of happiness in speech

Felix Burkhardt
felixbur@kgw.tu-berlin.de

June 7, 1999

Contents

Introduction

This is a short summary of the result from my master-thesis. The topic was "Simulation of the emotional speaker-state ‚happiness' with a speech-synthesizer"
I structured the task in several explorations and will follow in this summary according to the structure:

1. a short overwiev of results from earlier studies.

2. an analytical comparison between happy and neutral spoken sentences

3. a listening test to evaluate some of the results from the comparison
4. a self-programed "emotion-filter" and it's evaluation in another listening test

1 Results found in literatur

This is a (uncomplete) summary of results from literatur:

- **Pitch:**
  avr. pitch:
  According to Öster & Riesberg, 1986; Trojan, 1975; Tartter, 1980 happy speech is accompanied with a high avr. pitch.
  range:
  Bergmann et al., 1988 finds the range broader.
  variability:
  Scherer, 1981 finds variability (measured as standard deviation) is higher.
  segmental contours:
  Mozziconacci & Hermes, 1997 measured for happy utterances more rising contours on the first stresses parts of syllables.

- **Stresses:**
  According to Trojan, 1975 the differences between stressed and unstressed syllables are bigger
  Following Bergmann et al., 1988, phrase-stressed syllables should be of high pitch and short (!) duration

- **Speechrate:**
  According to Öster & Riesberg, 1986; Carlson et al., 1992 happy speech is slower. Fonagy, 1981; Davitz, 1964 in contrast say, that the tempo is "lively" and faster

- **Intensity:**
  Davitz, 1964; Carlson et al., 1992 say, intensity increases

- **Voice quality**
  formants
  According to Fant, 1957, S. 19; Tartter, 1980; Laver, 1991, S. 219 happy speech has higher formants. (because of spreading the lips while smiling) phonation
  According to Fonagy & Magdics, 1963 happy speech is breathy.
  Energy distribution in spectral bands
  Klasmeyer & Sendlmeier, 1999 measured more energy between 3-5 kHz

2 Acoustical analysis of 9 sentence-pairs

I measured 18 sentences (a database from another project at my institute). They were spoken by actors, 9 in a happy expression and the same 9 in neutral. All of them were evaluated in a listening test and had recogniten rates of at least 98 %. The sentences came from 5 different actors, two male and 3 female.

Most of the results are mean-values for all the sentences of each actor. Most of the results are represented as the deviation of the happy from the neutral utterances in %. Their significance was tested by a double-sided, paired t-test.

2.1 General duration of syllables

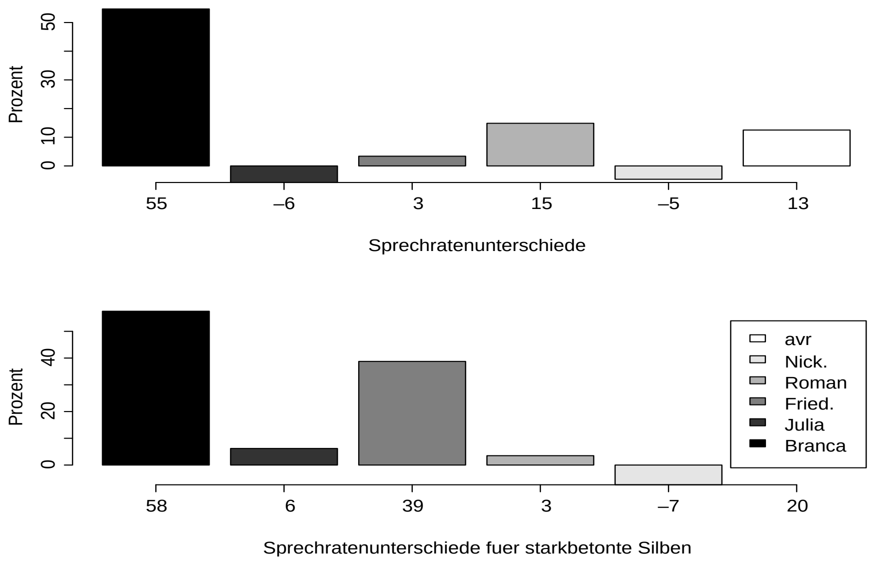

Figure 1: Differences of mean syllable duration

In figure 1 you can see the mean values for the differences between avr. syllable duration for happy vs. neutral speech. The upper figure is for all syllables, there's no significant difference, but a tendency (p-value was 0.079).

The lower figure shows differences for phrase-stressed syllables, the happy syls are avr. 20 % longer than the neutral ones.

2.2 Durations of certain sound-classes

In figure 2 you see the differences of mean durations of voiced fricatives (top) and strongly stressed vowels (bottom). These were the only significant differences for all sound-classes (short-vowels, long-vowels, nasals, fricatives, plosives, liquides, stressed vowels (phrase and word-stress)) I measured.

For the strongly stressed vowels there was only a tendency (p-value of 0.09). The voiced fricatives were highly significant longer (avr.: 37 %) in happy speech.

2.3 Pitch: global parameters

In figure 5 you can see the differences for (from top to bottom): avr. pitch, range and standard deviation. Except for the range all differences were significant. The fact, that the range-difference wasn't significant is mainly due to one sentence, which's pitch contours are shown in

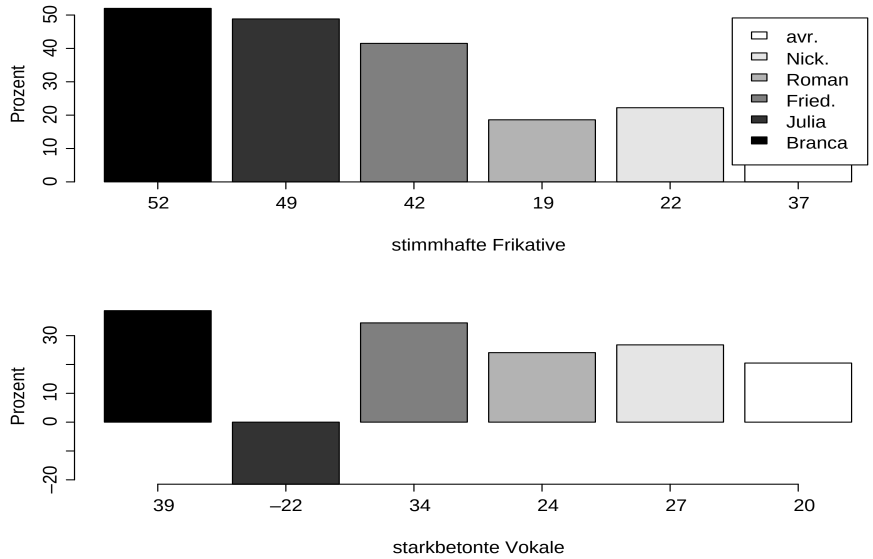

Figure 2: Differences of mean durations for voiced fricatives and phrase-stressed vowels

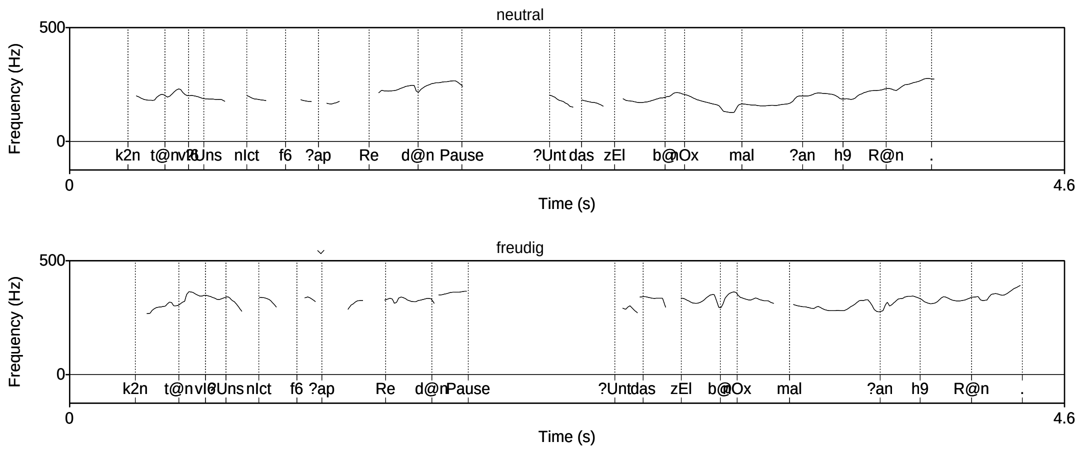

Figure 3: Pitch-contour of Friederike

figure 3. It was the only question of all 9 sentences, and I think question-sentences should have been viewed separately.

The pitch-contours of a typical sentence-pair is shown in figure 4.

2.4 Pitchheight of stressed syllables

I measured the deviation of pitch height at phrase-stressed syllables and the sentence's avr. pitch. The results are shown in figure 6. The deviation was significantly higher for the happy sentences.

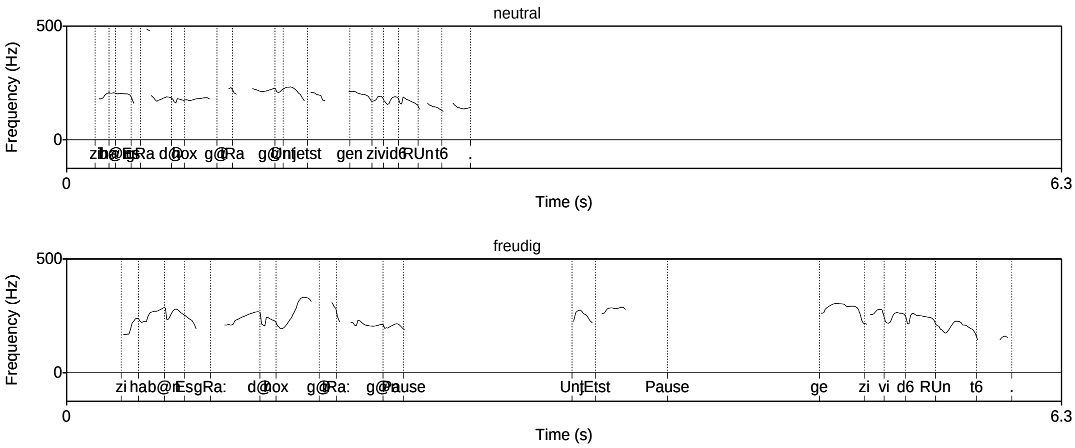

Figure 4: Pitch-contour of Branca

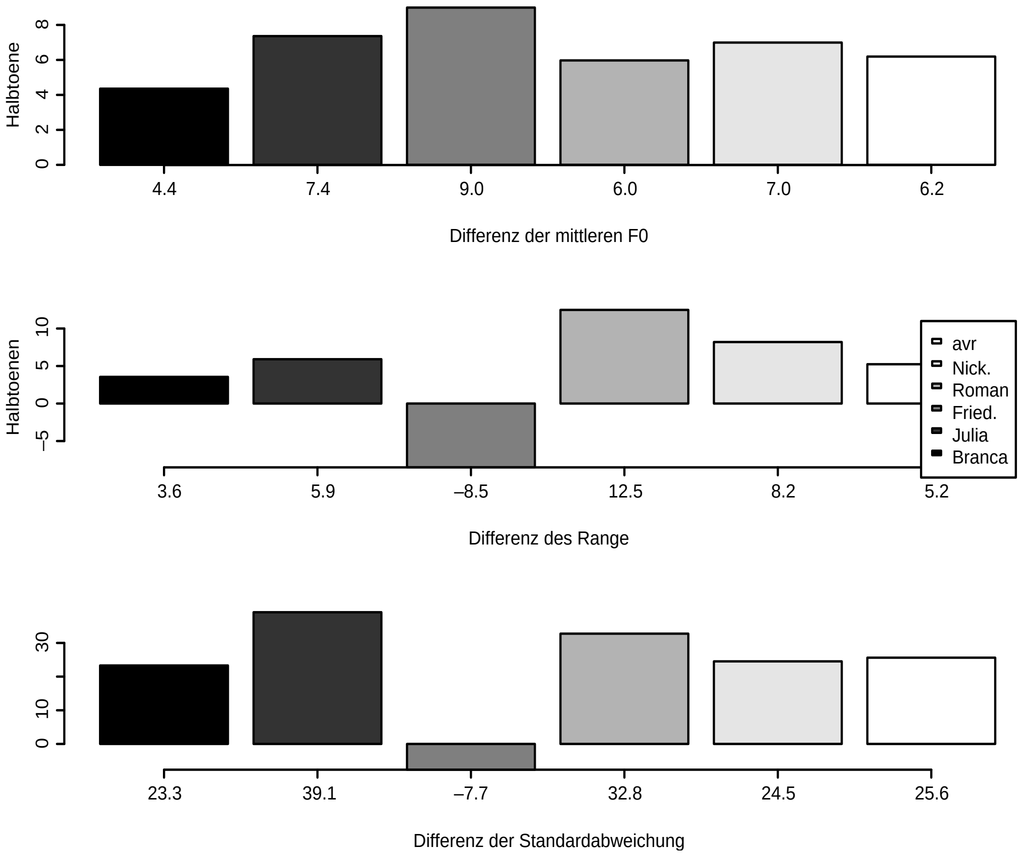

Figure 5: Differences of avr. pitch, range and std. deviation

2.5 Pitch-contours over syllables

I examined the pitch-contours over syllables using an algorithm similar to d'Alessandro & Mertens, 1995. I analysed 9 different types (combination of 2 halftones with 3 types (straight,

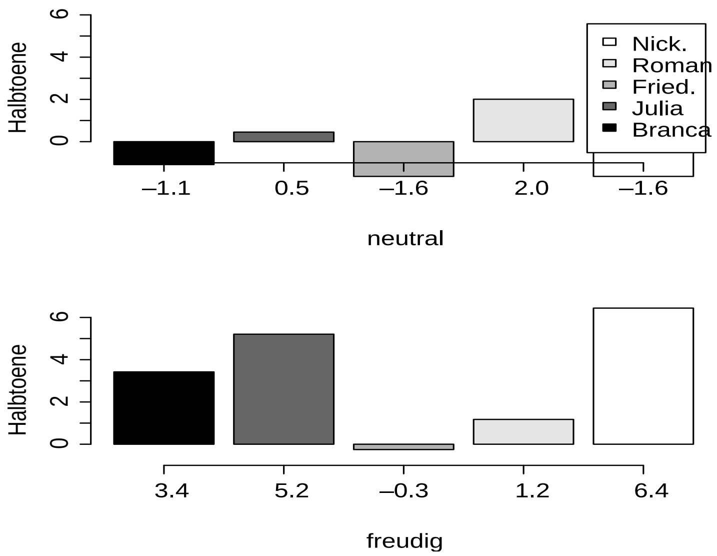

Figure 6: Pitch-deviation at phrase-stressed syllables

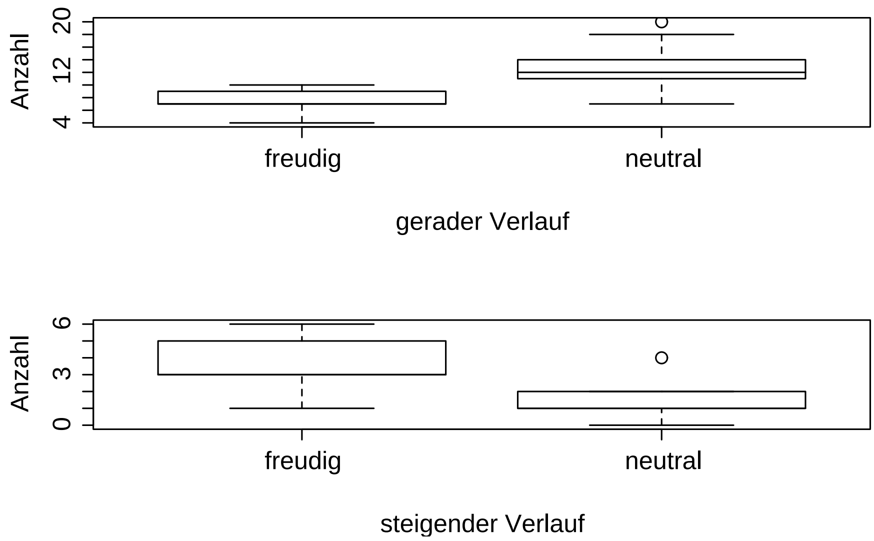

Figure 7: Boxplots for straight and rising pitch-contours over syllables

	rising [ST/s]	falling [ST/s]
happy	43	-47
neutral	19	-26
p-value	0.02*	0.012*

Table 1: Mean gradients for rising and falling pitch-contour types

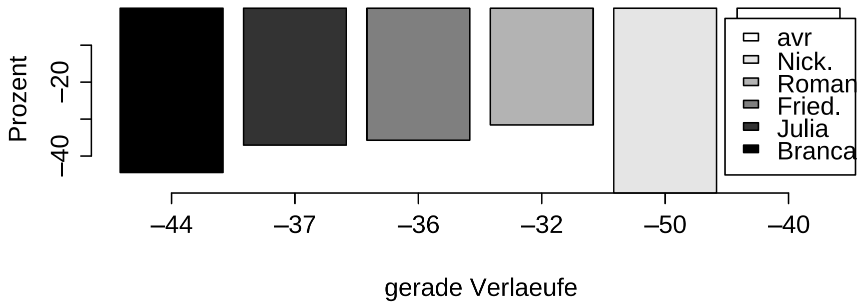

Figure 8: Differences of frequencies of straight (top) and rising (bottom) pitch-contours over syllables

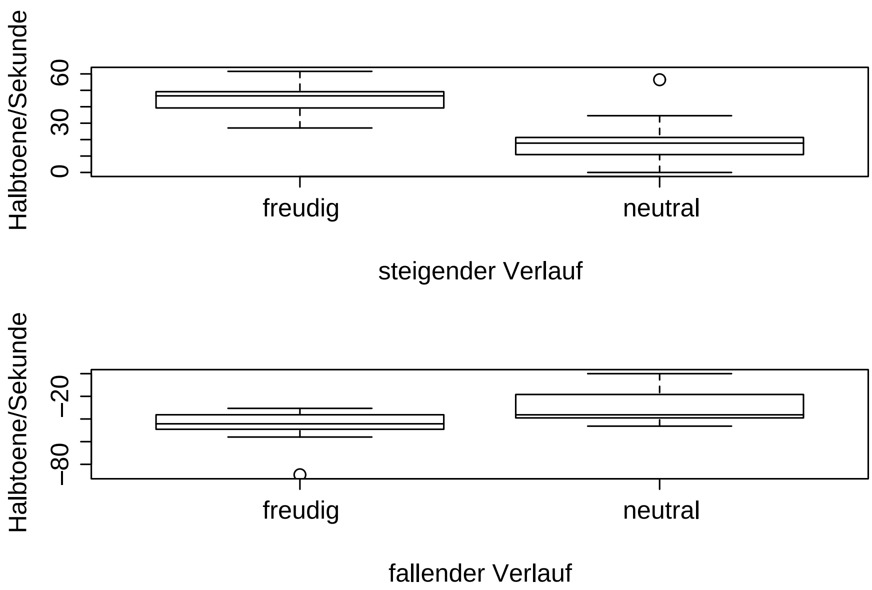

Figure 9: Boxplots of the gradients for rising and falling types

rising, falling)). There were significant more rising and less straight contours in happy speech (see figures 7 and 8).

The gradients of happy speech were higher both for rising and falling types (see figure 9 and table 1). The gradients are given in semitones/sec.

2.6 Formants

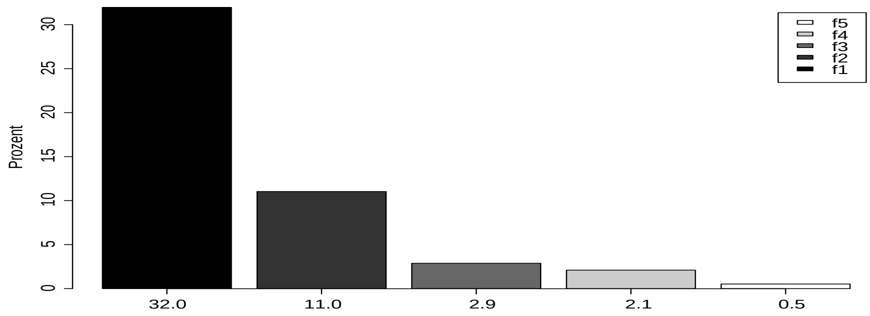

Figure 10: Differences of center-frequencies of formants 1-5

From each sentence I cut out 10-12 vowels of equal size and measured (automatically) the mean center-frequencies of formants 1-5. The first and second formants were at happy speech significantly higher than at neutral speech (see figure 10).

2.7 Energy in spectral bands

For the same vowels as above I measured the energy difference between the frequency-bands from 0-1 and 1-8 kHz and 0-4 and 4-8 kHz (see figure 11, a): 0-1 and 1-8 kHz, b): 0-4 and 4-8 kHz). In happy speech the spectral damping is significantly less than in neutral. The figures 12 and 13 show wideband-spectrograms for some of the vowels. You can see higher formants and more energy in higher frequency-areas.

2.8 Summary of results

- The avr. pitch is rising around 50 %.
- The F_0-range becomes rather broader.
- The utterances split in more phrases
- Phrase-stressed syllables are stronger (slower and higher pitch)
- The pitch-contours over syllables are more lively and are more often of the rising type
- The gradient of the rises and falls is bigger
- Especially voiced fricatives and strongly stressed vowels are longer
- The first 2 formants are higher.
- The higher frequency-bands contain decidedly more energy.

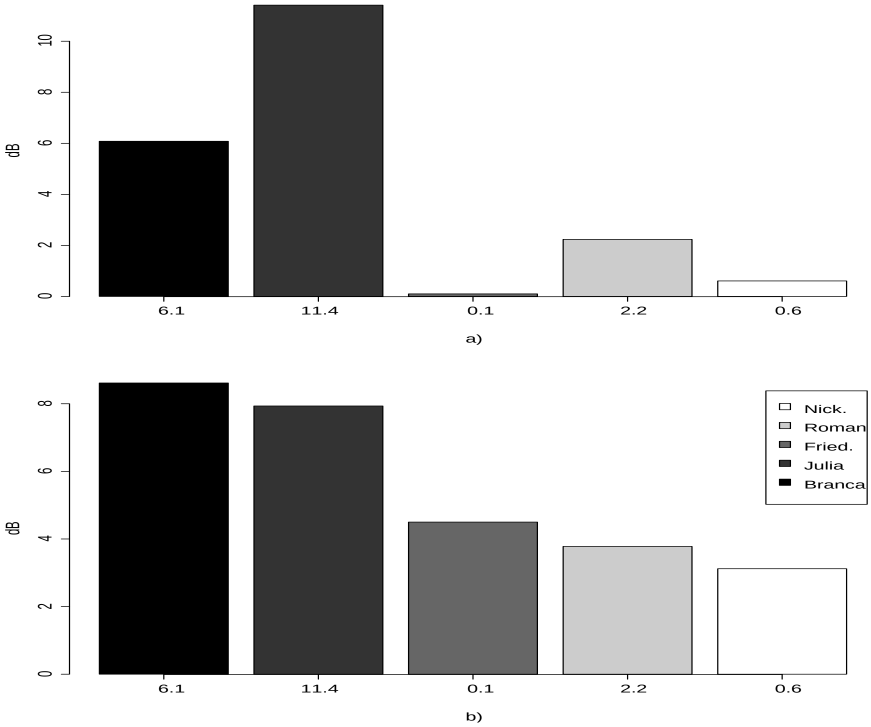

Figure 11: Differences of energy between frequency-bands

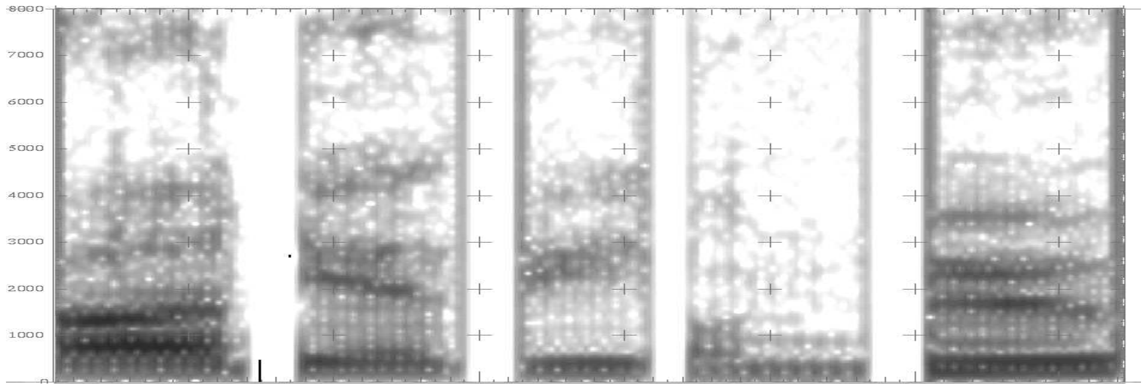

Figure 12: Wideband-spectrogram of the vowels a,e,i,ʊ, ø from a neutral sentence

3 Listening test

3.1 Testdesign

I did a listening-test concerning three factors: formants (2 steps), pitch-contour (3 steps) and durations (2 steps). I modified a neutral sentence systematically according to the factors. This was done by lpc-resynthesis and td-psola technique.

In addition I created one stimulus, where the pitch-contour and sound-durations of the neutral version were copied to the happy version (happy sentence with neutral prosody).

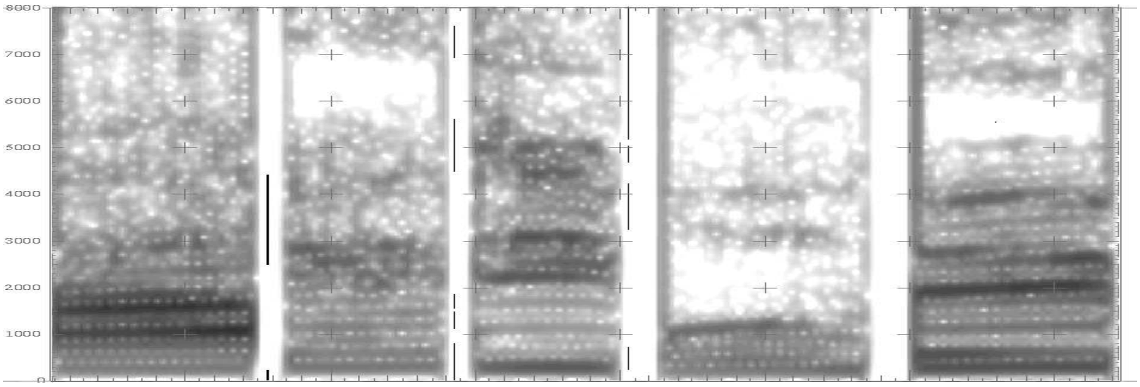

Figure 13: Wideband-spectrogram of the vowels a,e,i,ʊ, ø from a happy sentence

Formants

- Version 1: raised the first formant by 6 % and the second by 4 %.
- Version 2: raised formants 1-5 by 5 %.

Pitch-contours (see figure 14)

- Version 1: copy the pitch-contour of the happy variante of the sentence.
- Version 2: assign 3 phrase-accents to the sentence and raise the stresssed syllables by 25 %.
- Version 3: raise the stresssed syllables by 50 % and smooth the surrounding.

Durations

- Version 1: copy the sound-durations of the happy variante of the sentence.
- Version 2: prolong the 3 stressed syllables by 24 % and all shorten all other sounds by 20 %.

3.2 Results

17 native listeners judged the stimuli by answering the question "how happy sounds the speaker?" with a cross on a 7-step scale between "not happy at all" and "very happy". I analysed the results with an analysis of variance with complete repetition of measurments. All three main effects were significant. The interaction between durations and pitch-contours was also significant.

The mean values for the judgements of formants, pitch-contours, durations and durations/pitch-contours are shown in figures 15, 16, 17 and 18).

The p-values for the differences are shown in tables 2, 3 and 4.

The stimulus of the happy sentence with neutral prosody was still judged as the most happy one (!).

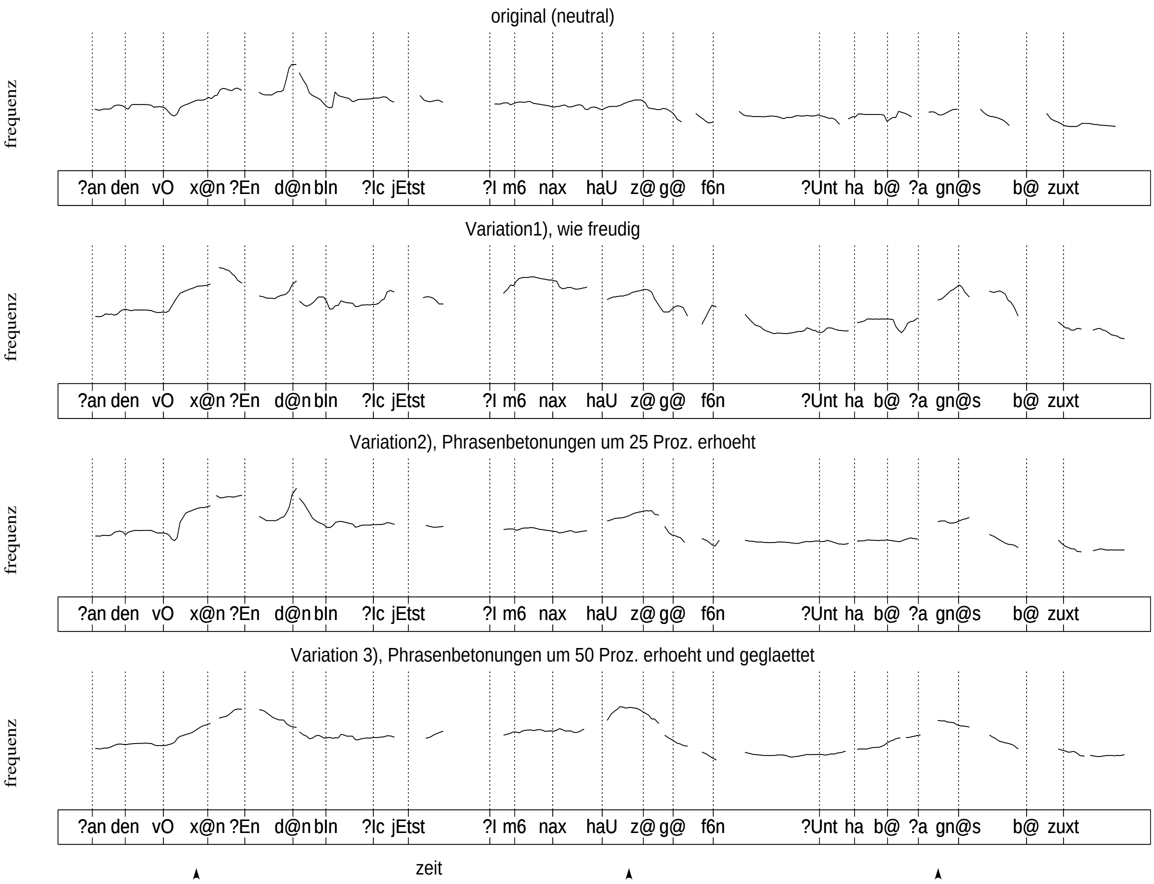

Figure 14: The modified pitch-contours

	orig.	Var. 1	Var. 2
orig.		0.04*	0.00**
Var. 1	0.04*		0.16
Var. 2	0.00**	0.16	

Table 2: p-values for the difference of formant-manipulations

	orig.	Var. 1	Var. 2	Var. 3
orig.		0,472	0,00**	0,02*
Var. 1	0,472		0,009**	0,153
Var. 2	0,00**	0,009**		0,026*
Var. 3	0,02*	0,153	0,026*	

Table 3: p-values for the difference of pitch-contour manipulations

4 A "Happiness by Rule"-system

4.1 Testdesign

I wrote a little program to manipulate the input-format of the TTS-system MBROLA, 1999. Similar approaches are described in Cahn, 1989 and Murray & Arnott, 1995

Table 4: p-values for the difference of the durations

	orig.	Var. 1	Var. 2
orig.		0,749	0,00**
Var. 1	0,749		0,00**
Var. 2	0,00**	0,00**	

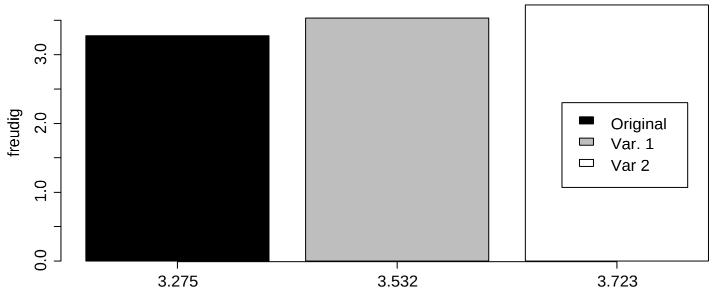

Figure 15: mean judgements for the formant-manipulations

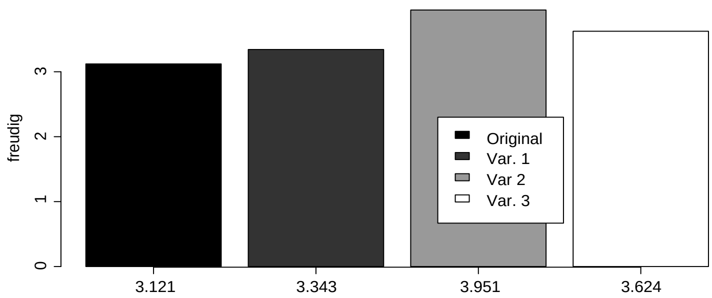

Figure 16: mean judgements for pitch-contour manipulations

To test the output, I did another listening test. I synthesized 5 one-phrased and 5 two-phrased sentences in three ways: neutral, happy and sad. The sad versions were used as distractors, so that the listeners wouldn't just judge everything strange sounding as being happy.

32 naive listeners answered the question "does the speaker sound happy?" with a cross at either "rather not" or "rather yes". The parameters for my emotion filter for happy speech were the following[1]:

- Raise avr. pitch by 40 %.
- set Range parameter to 140 % (this doesn't mean, that the range was 40 % broader !)

[1] to understand, what the parameters do, see the programms doc under `http://www.kgw.tu-berlin.de/~felixbur/emofilt.html`

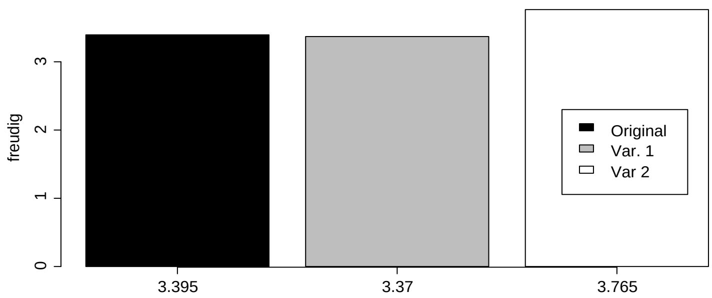

Figure 17: Mean judgements for durations

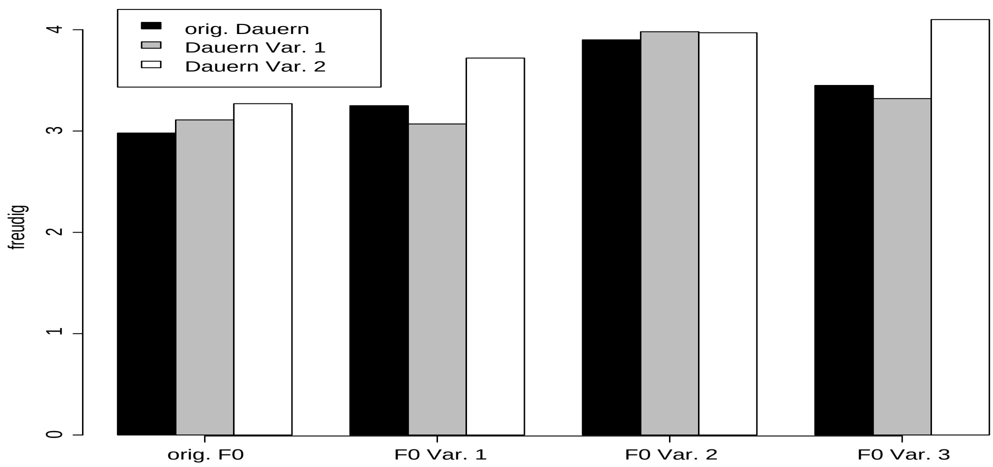

Figure 18: mean judgements for durations in dependance of pitch-contour

- set variability to 200 %
- raise phrase stressed syllables by 10 %
- use linear rising contour (by 20 ST/s) for phrase stressed syllables
- prolong phrase stressed syllables by 20 %
- shorten all other sounds by 20 %
- prolong voiced fricatives by 30 %

In figure 19 you see the neutral and modified pitch-contours of one of the test-sentences.

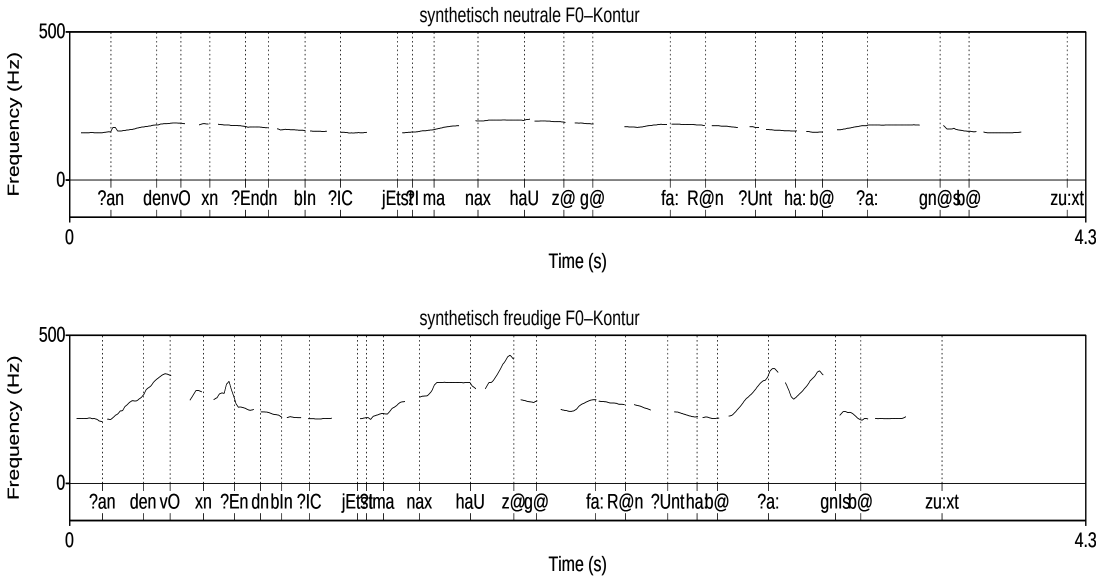

Figure 19: neutral and happy contour a test-sentence

4.2 Results

The results were analysed by three pairwise t-tests. The mean judgements for the four groups are shown in figure 20. There were all significant.

The pitch-contours of the sentences, who was judged as most happy (98 % of listeners), are shown in figure 21.

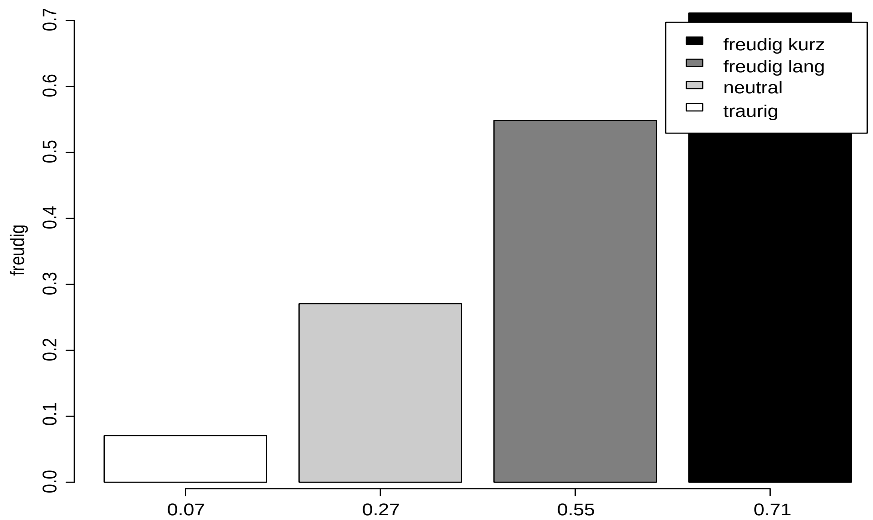

Figure 20: mean judgements of (left to right) sad, neutral, happy long and happy short stimuli

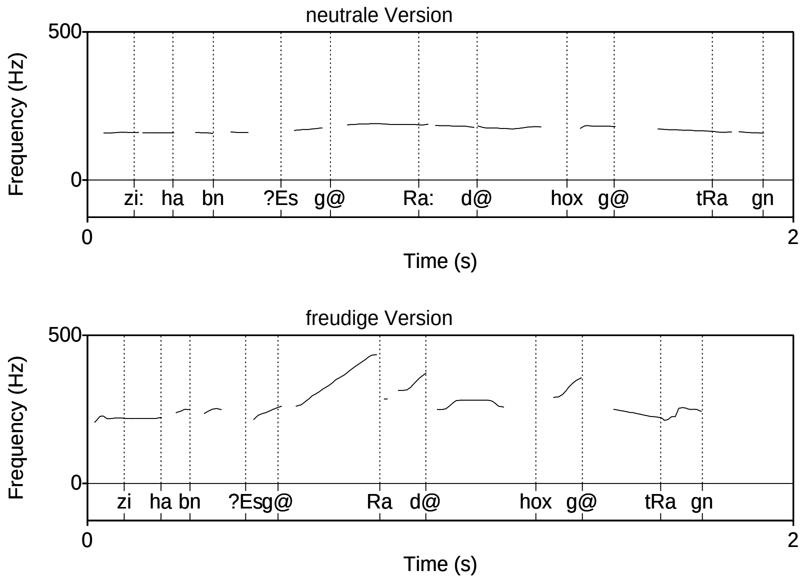

Figure 21: Neutral and happy pitch-contour of the most happy test-sentence

References

G. Bergmann, T. Goldbeck & K. R. Scherer. Emotionale Eindruckswirkung von prosodischen Sprachmerkmalen. *Zeitschrift f"ur experimentelle und angewandte Psychologie*, 35:167–200, 1988.

J. E. Cahn. The Affect Editor. *Journal of the American Voice I/O Society*, 8:1–19, 1989.

R. Carlson, G. Granstr"om & L. Nord. Experiments With Emotive Speech, Acted Utterances And Syntesized Replicas. *Speech Communication*, 2:347–355, 1992.

C. d'Alessandro & P. Mertens. Automatic Pitch Contour Stylization Using a Model of Tonal Perception. *Computer Speech and Language*, 9:257–288, 1995.

J. R. Davitz. Personality, perceptual and cognitive correlates of emotional sensitivity. In *The Communication of Emotional Meaning*. Mc-Graw-Hill, New York, 1964.

G. Fant. Modern Instruments and Methods for Acoustic Studies of Speech. *Proc. of the 8th International Congress of Linguists*, Seite 282–388, 1957.

I. Fonagy & K. Magdics. Emotional Patterns in Intonation and Music. *Zeitschrift f"ur Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 16:293–326, 1963.

I. Fonagy. Emotions, voice and music. In *Research Aspects on Singing*, Band 33, Seite 51–79. Royal Swedish Academy of Music, 1981.

G. Klasmeyer & W. F. Sendlmeier. Voice and emotional states. In R. Kent & M. Ball, editors, *The Handbook of Voice Quality Measurement*. Singular Publishibng Group, San Diego, 1999. im Druck.

J. Laver. *The Gift Of Speech.* Edinburgh University Press, 1991.

MBROLA, 1999. http://tcts.fpms.ac.be/synthesis.

S. J. L. Mozziconacci & D. J. Hermes. A Study of Intonation Patterns in Speech Expressing Emotion or Attitude: Production and Perception. *IPO Annual Progress Report*, 32:154–160, 1997.

I. R. Murray & J. L. Arnott. Implementation and Testing of a System for Producing Emotion-by-rule in Synthetic Speech. *Speech Communication*, 16:369–390, 1995.

A. Öster & A. Riesberg. The Identification of the Mood of a Speaker by Hearing Impaired Listeners. *Speech Transmiss. Lab. - Q.Prog.Stat.Rep*, Seite 79–90, 1986.

K. R. Scherer. Speech and emotional states. In J. K. Darby, editor, *Speech Evaluation in Psychiatry*, Seite 189–218. Grune & Stratton, 1981.

V. C. Tartter. Happy Talk: Perceptual and Acoustic Effects of Smiling on Speech. *Percept. Psychophys.*, 27:24–27, 1980.

F. Trojan. *Biophonetik.* Wissenschaftsverlag, 1975.