

Pitch in emotional speech

Citation for published version (APA):

Berkvens, W. A. H. (1995). *Pitch in emotional speech*. (IPO rapport; Vol. 1059). Instituut voor Perceptie Onderzoek (IPO).

Document status and date:

Published: 11/07/1995

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Rapport no. 1059

Pitch in emotional speech

W.A.H. Berkvens

Pitch in emotional speech

Trainee: W.A.H. Berkvens

TUE Identity number: 300588

Faculty: Electronic and Electrotechnical Engineering
Eindhoven University of Technology

Date: 6 july 1995

At the Institute for Perception Research, Eindhoven
Research group: Hearing and Speech

Supervisor:

Sylvie Mozziconacci (Institute for Perception Research,
Eindhoven)

Pitch in emotional speech

Contents

1 INTRODUCTION	1
2 SPEECH MATERIAL	2
3 PITCH	3
3.1 Introduction	3
3.2 Choice of units	3
3.3 Measurements	3
3.4 Precision	5
3.5 Results	5
3.5.1 Intonation labelling	5
3.5.2 F_0 measurements	7
3.5.3 Various types of pitch information	10
3.6 Conclusion	11
4 DISCUSSION	12
Bibliography	13
Appendix I: Speech recordings	14
Appendix II: Standard deviation	15
Appendix III: Pitch movements according to the intonation grammar for Dutch	16
Appendix IV: Averaged values in Hertz, over all of the fifteen utterances per emotion	17
Appendix V: Averaged values in semitones, over all of the fifteen utterances per emotion	18
Appendix VI: Figures containing the results in Hertz of the tables of appendix IV and V	19
Appendix VII: Averaged values in Hertz, using only the utterances in the categories 5A/5&A, 1&A, 1A and 1EA	20
Appendix VIII: Averaged values in semitones, using only	

Pitch in emotional speech

	the utterances in the categories 5A/5&A, 1&A, 1A and 1EA	21
Appendix IX:	Figures containing the results in Hertz of the tables of appendix VII and VIII	22
Appendix X:	Various types of pitch information in semitones	23

1 INTRODUCTION

In the history of mankind, speaking was one of the late evolutions. It was a secondary function of the organs needed for eating and used for breathing (Vogten, 1988). In speaking, man uses, among others, pitch to express his emotions and attitudes.

The computer can be used to generate synthesized speech. Synthesized speech sounds monotonous. As stated by Granström and Nord (1991), emotional speech production introduced in speech synthesis increases naturalness and intelligibility of spoken text and also gives the possibility to include the important aspects of speaking style variation and the speaker's attitude.

Odé and Van Heuven (1994, p. 2) observed that 'Prosody comprises all properties of speech that cannot be understood directly from the linear sequence of segments (vowels and consonants)'. Prosody can be divided into two broad categories of phenomena: temporal structure and melodic structure. Pitch is one of the parts of the melodic structure (Odé, and Van Heuven, 1994).

This study is done as part of a greater study by Mozziconacci (1995). In this report acoustic analyses are performed for the purpose of getting some relevant aspects of pitch in emotions. These aspects can be used in speech synthesis.

Summarising, it is shown that the pitch contours for the emotions considered are different from each other. The results can be useful for the production of synthesized speech.

2 SPEECH MATERIAL

In this study, recordings of sentences spoken by two Dutch actors (a male and a female) are used. These recordings have been recorded by Vroomen. Seven emotions are considered. The seven emotions considered are: joy, neutrality, boredom, anger, sadness, fear and indignation. In each emotion the actor had to say three times an emotional sentence and three times five sentences. The content of these five sentences is attempted to be as neutral as possible (see appendix I). The emotional sentences are used to bring the speaker in the right mood and are not further used in this study.

From both speakers the recordings were digitized at a sampling frequency of 16kHz (16 bit precision). The program used for this purpose is GIPOS¹. Our database contains one hundred and twenty-six sentences (three tokens * six sentences * seven emotions) per speaker.

¹ : Ercan Gigi, Institute for Perception Research (IPO), Eindhoven.

3 PITCH

3.1 Introduction

Considering the fact that everyone uses pitch movement during speech production it is interesting to investigate pitch for the benefit of the generation of synthesized speech by a computer.

3.2 Choice of units

In this report, both Hertz and semitones are used in relation to pitch. Hertz is used because this is most often used for the representation of pitch. Semitones are interesting to compare the results between the male and female speaker.

Regarding the perception of pitch, we are interested in frequency distances and in the absolute frequencies. To calculate the frequency distances a formula is used to transpose Hertz into semitones. In the formula a distance D , in semitones, between any two frequencies f_1 (in Hz) and f_2 (in Hz) can be calculated ('t Hart, Collier, and Cohen, 1990, p. 24):

$$D=12*\log_2\left(\frac{f_1}{f_2}\right). \quad (1)$$

The arbitrary chosen base-line (f_2) is 100Hz.

3.3 Measurements

Before any calculations are done, the utterances have been segmented at the beginning and the end of the speech.

The following pitch information is considered in this report:

- intonation pattern,
- mean pitch, standard deviation, minimal, maximal, begin and

end frequency,

- F_0 measurements on six points for each utterance.

From each utterance the intonation pattern is built from pitch movements as proposed in chapter four of the book by 't Hart, Collier and Cohen (1990). The pitch movements can be decomposed into perceptual parts. These parts can be subdivided in their direction, their timing with regard to syllable boundaries, their rate of change and their size. The parts for the different pitch movements of Dutch are mentioned in a table (see appendix III).

Every utterance is divided into little parts called frames. Each frame can either be voiced or unvoiced and can have a positive or negative gain. A frame that is both voiced and has a positive gain is called a frame with a positive F_0 value. In the programs MEANPITCH² and GEP³ only the frames that have a positive F_0 value are used. Voiced frames are frames that have a periodic waveform; unvoiced frames don't. A positive gain means that the amplitude of a frame lies above an arbitrarily chosen lower limit. This lower limit is chosen to overcome the problem of using frames with only noise.

The program MEANPITCH is used to calculate the mean pitch and the standard deviation and the program GEP is used to obtain the minimal, maximal, begin and end frequency.

The standard deviation is calculated by using the formula in appendix II.

For the F_0 measurements at six points for each utterance, six syllables are chosen. In these syllables an anchor point is determined in the stable part of the vowel.

² : Institute for Perception Research (IPO), Eindhoven.

³ : Leo Vogten, Institute for Perception Research (IPO), Eindhoven.

3.4 Precision

To get pitch information about an utterance a pitch detection algorithm (pdt-algorithm) is used. The pdt-algorithm had a upper frequency of 500Hz. That was not high enough for the utterances of the female speaker. The pdt-algorithm was changed so that its upper frequency became 1250Hz. The results presented in this chapter are made with this changed pdt-algorithm.

3.5 Results

3.5.1 Intonation labelling

J. Terken and R. Collier (Institute for Perception Research (IPO)), labelled all the utterances according to the intonation grammar for Dutch ('t Hart, Collier, and Cohen, 1990) (see appendix III).

For each utterance a category was chosen considering the labelling of the utterance. From every utterance that contains more than one label the most representative label is chosen. For every emotion the results are presented in table 1 for the male speaker and in table 2 for the female speaker. The following categories are considered: 12/1&2, 5A or 5&A, 1&A, 1A, 1EA and 3C/1C.

Table 1: Categories for the male speaker

	12 or 1&2	5A or 5&A	1&A	1A	1EA	3C or 1C
joy	0	2	13	0	0	0
neutrality	0	0	13	0	2	0
boredom	0	0	7	4	2	2
anger	0	2	12	1	0	0
sadness	0	0	9	1	0	2
fear	0	0	5	1	0	7
indignation	0	0	3	0	0	8
total	0	4	62	7	4	19

Table 2: Categories for the female speaker

	12 or 1&2	5A or 5&A	1&A	1A	1EA	3C or 1C
joy	0	0	7	2	1	5
neutrality	0	0	5	7	2	0
boredom	0	3	1	3	0	5
anger	0	2	8	3	0	2
sadness	0	0	6	6	0	3
fear	0	2	7	2	0	3
indignation	13	0	1	0	0	0
total	13	7	35	23	3	18

The total of every row is maximal fifteen, because there are fifteen sentences spoken. When the total of a row is less than fifteen the other sentences could not be placed in one of the considered categories.

The utterances in category 12/1&2 and 3C/1C are left out in the further study because they end not at the lower declination line.

3.5.2 F_0 measurements

Measurements are done at six anchor points. The begin and end frequency could complement the results to get a total of eight points. This is omitted because the first c.q. last frame with a positive F_0 value is sometimes existing later c.q. earlier than one of the anchor point.

First, for every emotion all of the fifteen utterances are used. The results of the F_0 measurements in semitones (see appendix VI for the same results in Hertz) are shown in figure 1 for the male speaker and figure 2 for the female speaker (see appendix IV for the values in Hertz and see appendix V for the values in semitones). For each anchor point the fifteen values are used. In this way the influence of the spoken utterances was averaged.

Second, the F_0 measurements are done only for those utterances classified as 5A/5&A, 1&A, 1A and 1EA. Because not all the utterances are used, we weight the utterances so that the influence of the sentences spoken are averaged. First we calculate the mean over the utterances of each of the five sentences and then take the mean over these results. The results in semitones (see appendix IX for the same results in Hertz) are presented in figure 3 for the male speaker and figure 4 for the female speaker (see appendix VII for the values in Hertz and see appendix VIII for the values in semitones).

The result for indignation in figure 4 (see also figure 8) is not very reliable because there are not enough points to calculate the average. Therefore it is presented using a dashed line.

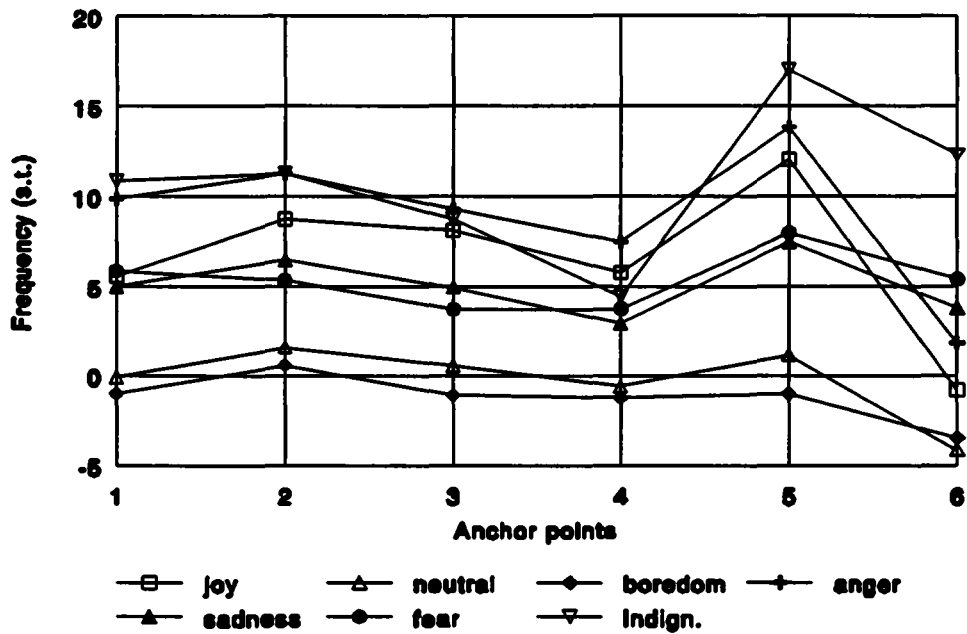


Figure 1: Results of the F₀ measurements over six points for the male speaker, over all of the fifteen utterances.

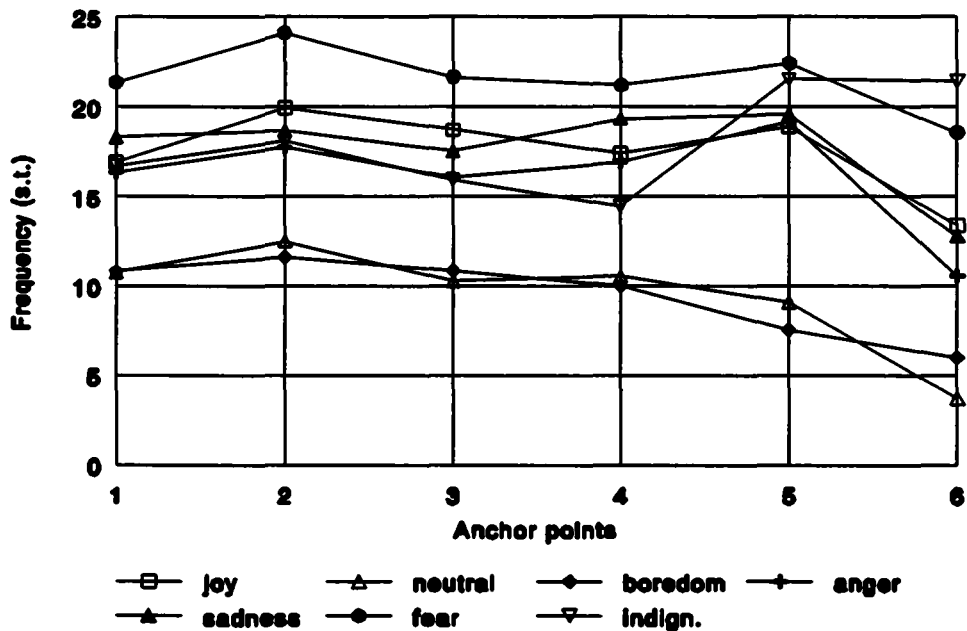


Figure 2: Results of the F₀ measurements over six points for the female speaker, over all of the fifteen utterances.

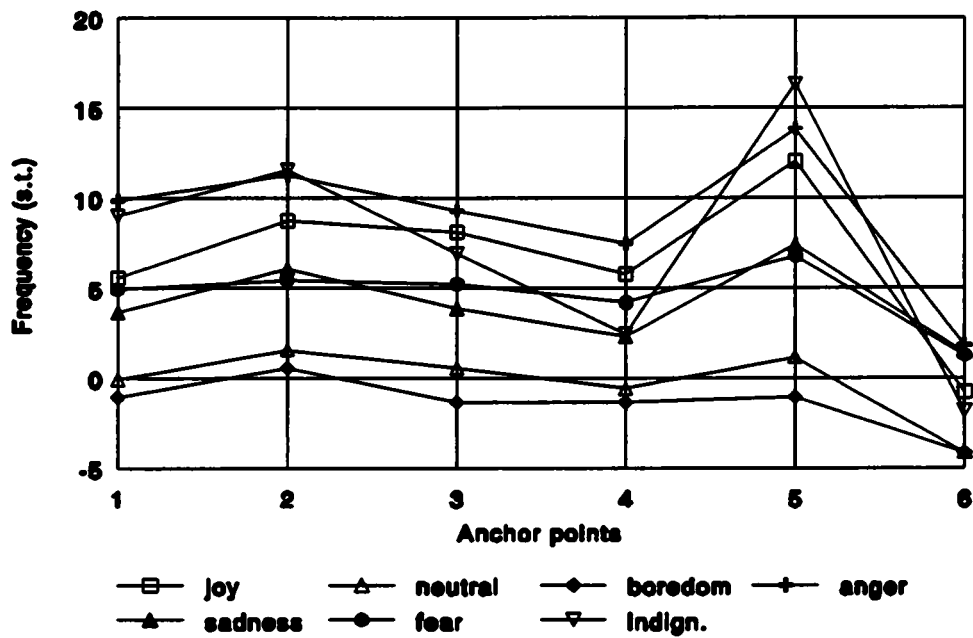


Figure 3: Results of the F₀ measurement over six points for the male speaker, over the utterances in the categories 5A/5&A, 1&A, 1A and 1EA.

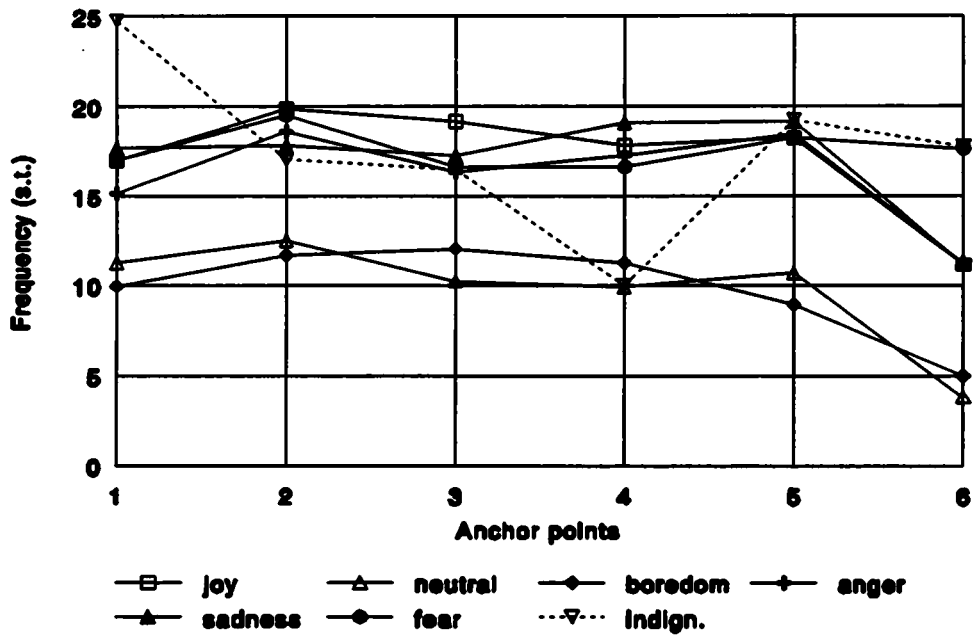


Figure 4: Results of the F₀ measurement over six points for the female speaker, over the utterances in the categories 5A/5&A, 1&A, 1A and 1EA.

3.5.3 Various types of pitch information

Various pitch information has been gathered. These include mean pitch, standard deviation, minimal, maximal, begin and end frequency. In table 3 this is done for the male speaker and in table 4 for the female speaker (see appendix X for the same information in semitones).

Table 3: General pitch information of the male speaker

	mean pitch (s.t)	SD	min. freq. (s.t)	max. freq. (s.t)	begin freq. (s.t)	end freq. (s.t)
joy	5.56	4.88	-6.11	14.42	1.56	-3.73
neutrality	-0.57	2.71	-7.20	4.22	-1.29	-4.32
boredom	-1.05	2.01	-5.82	3.64	0.29	-3.02
anger	7.76	4.99	-4.31	16.58	6.19	-1.47
sadness	4.66	3.47	-3.74	11.46	4.24	1.29
fear	4.44	3.44	-2.67	11.52	5.80	3.03
indignation	9.40	5.03	-0.63	18.67	9.56	10.87

Table 4: General pitch information of the female speaker

	mean pitch (s.t)	SD	min. freq. (s.t)	max. freq. (s.t)	begin freq. (s.t)	end freq. (s.t)
joy	15.97	4.94	1.28	23.31	12.76	1.90
neutrality	8.41	4.15	-3.37	14.24	8.46	-1.25
boredom	8.61	3.61	-0.29	14.02	9.39	2.01
anger	15.14	4.90	0.35	22.32	11.62	1.06
sadness	17.27	4.06	7.04	23.43	17.97	8.63
fear	22.80	4.09	10.81	28.73	18.89	13.54
indignation	15.90	6.07	2.66	29.14	11.44	19.69

3.6 Conclusion

The results of the F_0 measurements for each emotion (see figure 1, figure 2 and appendix VI), calculated by taking the mean over the fifteen utterances, show the differences and similarities between each emotion.

For all the emotions there is an accent on the second anchor point. For all the emotions, except boredom for the female speaker, there is an accent on the fifth anchor point.

The second accent is mostly larger for the sentences in the higher frequencies for the male speakers.

The sentences spoken with fear are the highest in pitch for the female speaker followed by the sentences spoken with sadness and joy. For the male speaker indignation and anger are the highest in pitch. The sentences spoken with boredom and neutrality are the lowest in pitch for both speakers, they are much lower than the sentences spoken with the other emotions. The sentences spoken with boredom and neutrality are followed by the sentences spoken with sadness and fear for the male speaker and anger for the female speaker.

The results of the F_0 measurements for each emotion (see figure 3, figure 4 and appendix IX), for the sentences classified as 5A/5&A, 1&A, 1A and 1EA, show also the differences and similarities between the emotions. The emotion indignation for the female speaker is not used because the results are not reliable.

The results for the male speaker are not changed very much because most utterances belong to one of the categories considered. There are some changes for the female speaker. The sentences spoken with joy and sadness are now higher in pitch than the sentences spoken with fear.

4 DISCUSSION

As mentioned in the introduction, other aspects besides pitch are also important in the expression of emotions. Because the pitch information is different for every emotion, it seems reasonable that the results can be used for the generation of synthesized speech. With these pitch results especially the monotonous way of speech production can be avoided and it is expected that the spoken sentences will be considered by listeners as more natural.

It would be interesting to investigate the speech of more speakers to see if the information gathered in this report will hold out.

Bibliography

Granström, B., & Nord, L. (1991). Ways of exploring speaker characteristics and speaking styles. *Proceedings of the XIIth international congress of phonetic sciences, Aix-en-Provence, volume 4, page 278-281.*

Hart, J. 't, Collier, R., & Cohen, A. (1990). *A perceptual study of intonation; an experimental-phonetic approach to speech melody.* Cambridge: Cambridge University Press.

Mozziconacci, S. (1995). Pitch variations and emotions in speech. *Proceedings of the XIIIth international congress of phonetic sciences, Stockholm.*

Odé, C., & Heuven, V.J. van (1994). Experimental studies of Indonesian prosody. Rijksuniversiteit Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië.

Vogten, L.L.M. (1988). Syllabus van het college Spraaktechnologie 0H050. IPO Report no. 649. IPO, Eindhoven.

Appendix I: Speech recordings

The emotions:

Joy - *blijheid*
Neutrality - *neutraliteit*
Boredom - *verveling*
Anger - *boosheid*
Sadness - *verdriet*
Fear - *angst*
Indignation - *verontwaardigd*

The sentences:

- * An emotional sentence, different for each emotion.
- * *Zij hebben een nieuwe auto gekocht.* - They have bought a new car.
- * *Het is bijna negen uur.* - It is almost nine o'clock.
- * *Jan is naar de kapper geweest.* - Jan has been to the hairdresser.
- * *De lamp staat op het bureau.* - The lamp is on the desk.
- * *Zijn vriendin kwam met het vliegtuig.* - His girlfriend came by plane.

Appendix II: Standard deviation

The standard deviation (SD) is

$$SD = \sqrt{s^2}, \quad (2)$$

where

$$s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - x_{mean})^2, \quad (3)$$

and

$$x_{mean} = \frac{1}{n} \sum_{j=1}^n x_j. \quad (4)$$

Appendix III: Pitch movements according to the intonation grammar for Dutch

The table with the pitch movements of Dutch, is copied from 't Hart, Collier and Cohen (1990).

Table 5: The pitch movement of Dutch

	Transcription symbol									
	1	2	3	4	5	A	B	C	D	E
Direction										
rise	X	X	X	X	X					
fall						X	X	X	X	X
Timing										
early	X				X		X			X
late			X			X				
very late		X						X		
Rate of change										
fast	X	X	X		X	X	X	X		X
slow				X					X	
Size										
full	X	X	X	X		X	X	X	X	
half					X					X

Appendix IV: Averaged values in Hertz, over all of the fifteen utterances per emotion

Table 6: Pitch as measured at the anchor points for the male speaker (in Hz)

anchor point	1	2	3	4	5	6
joy	142.3	173.1	166.4	145.4	204.6	96.7
neutrality	100.1	110.4	103.9	97.5	107.7	78.8
boredom	95.0	104.1	94.6	93.7	94.9	82.7
anger	183.0	201.5	177.9	162.3	229.5	117.1
sadness	136.2	147.7	137.0	124.9	158.9	129.7
fear	142.4	139.5	128.6	126.1	164.6	145.0
indignation	190.1	197.6	172.9	139.2	269.4	220.7

Table 7: Pitch as measured at the anchor points for the female speaker (in Hz)

anchor point	1	2	3	4	5	6
joy	271.6	320.9	299.8	276.3	305.6	224.0
neutrality	186.9	206.2	181.3	184.5	171.2	124.7
boredom	189.0	198.7	191.0	181.9	157.9	143.4
anger	261.8	283.4	257.1	274.0	310.2	194.8
sadness	294.9	299.1	278.0	310.4	316.9	219.9
fear	351.7	411.1	357.4	352.8	378.2	303.5
indignation	268.9	290.0	257.3	243.0	354.4	363.1

Appendix V: Averaged values in semitones, over all of the fifteen utterances per emotion

Table 8: Pitch as measured at the anchor points for the male speaker (in semitones)

anchor point	1	2	3	4	5	6
joy	5.58	8.75	8.10	5.78	12.03	-0.80
neutrality	-0.09	1.58	0.55	-0.60	1.15	-4.13
boredom	-0.98	0.59	-1.09	-1.24	-1.02	-3.47
anger	9.86	11.30	9.33	7.45	13.85	1.81
sadness	5.01	6.50	4.96	2.93	7.46	3.81
fear	5.86	5.33	3.73	3.76	7.98	5.40
indignation	10.86	11.28	8.73	4.45	16.96	12.30

Table 9: Pitch as measured at the anchor points for the female speaker (in semitones)

anchor point	1	2	3	4	5	6
joy	16.88	19.93	18.71	17.40	18.88	13.36
neutrality	10.78	12.50	10.25	10.53	9.10	3.74
boredom	10.81	11.60	10.86	10.00	7.53	5.98
anger	16.31	17.73	16.03	16.91	19.16	10.55
sadness	18.30	18.68	17.53	19.33	19.58	12.78
fear	21.36	24.10	21.66	21.19	22.43	18.56
indignation	16.70	18.08	15.90	14.46	21.53	21.40

Appendix VI: Figures containing the results in Hertz of the tables of appendix IV and V

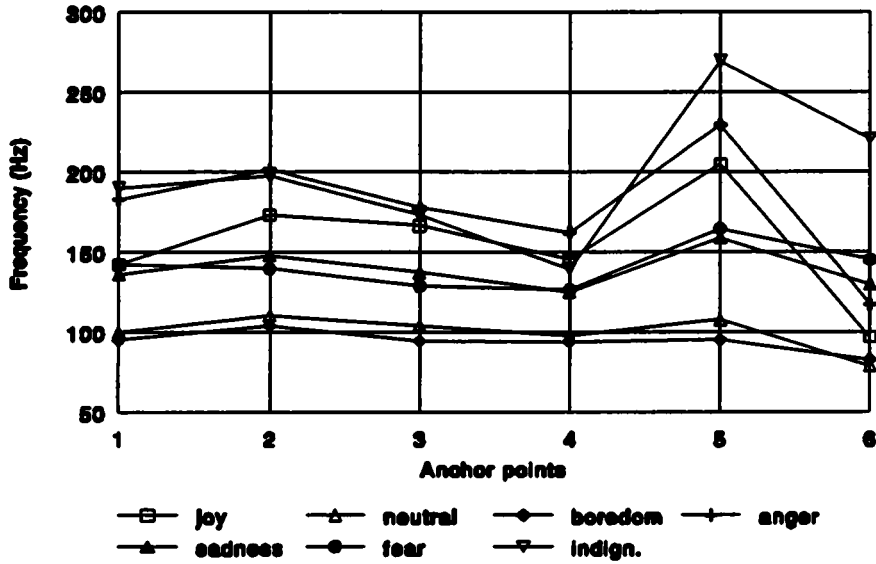


Figure 5: Results of the F_0 measurements over six points for the male speaker, over all of the fifteen utterances.

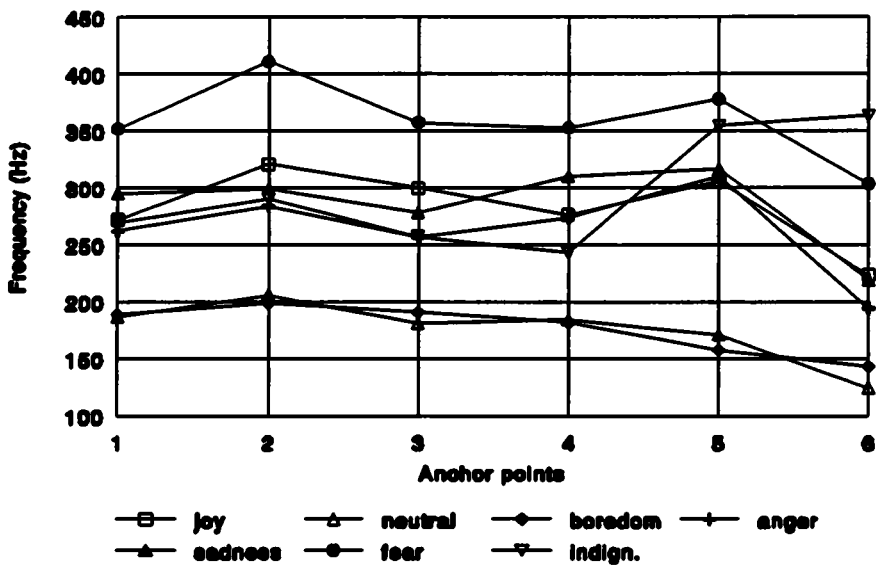


Figure 6: Results of the F_0 measurements over six points for the female speaker, over all of the fifteen utterances.

Appendix VII: Averaged values in Hertz, using only the utterances in the categories 5A/5&A, 1&A, 1A and 1EA

Table 10: Pitch as measured at the anchor points for the male speaker (in Hz)

anchor point	1	2	3	4	5	6
joy	142.3	173.1	166.4	145.4	204.6	96.7
neutrality	100.1	110.4	103.9	97.5	107.7	78.8
boredom	94.5	103.6	93.1	92.6	94.4	78.8
anger	183.0	201.5	177.9	162.3	229.5	117.1
sadness	130.5	148.8	133.2	125.1	161.2	109.5
fear	135.4	136.5	139.9	127.7	153.2	110.0
indignation	168.4	195.3	149.4	119.6	258.3	90.0

Table 11: Pitch as measured at the anchor points for the female speaker (in Hz)

anchor point	1	2	3	4	5	6
joy	300.6	357.4	341.2	315.2	331.0	217.1
neutrality	185.6	207.6	182.3	186.3	170.3	122.8
boredom	179.6	200.8	204.0	195.1	172.9	134.2
anger	257.0	290.9	264.2	283.8	312.3	181.9
sadness	286.5	283.3	274.1	307.1	309.1	198.2
fear	355.2	402.2	343.2	337.0	357.7	282.1
indignation	418.0	267.2	259.6	178.3	304.2	279.0

Appendix VIII: Averaged values in semitones, using only the utterances in the categories 5A/5&A, 1&A, 1A and 1EA

Table 12: Pitch as measured at the anchor points for the male speaker (in semitones)

anchor point	1	2	3	4	5	6
joy	5.58	8.75	8.10	5.78	12.03	-0.80
neutrality	-0.09	1.58	0.55	-0.60	1.15	-4.13
boredom	-1.07	0.54	-1.32	-1.37	-1.08	-4.17
anger	9.86	11.30	9.33	7.45	13.85	1.81
sadness	3.66	6.09	3.89	2.29	7.41	1.44
fear	4.92	5.45	5.23	4.20	6.76	1.26
indignation	9.01	11.58	6.88	2.45	16.32	-1.86

Table 13: Pitch as measured at the anchor points for the female speaker (in semitones)

anchor point	1	2	3	4	5	6
joy	16.95	19.85	19.18	17.82	18.23	11.17
neutrality	11.26	12.54	10.24	9.90	10.73	3.82
boredom	9.96	11.71	12.07	11.29	8.93	4.98
anger	15.15	18.62	16.35	17.23	18.43	11.27
sadness	17.66	17.80	17.28	19.08	19.18	11.22
fear	16.98	19.51	16.62	16.63	18.24	17.62
indignation	24.76	17.02	16.52	10.01	19.26	17.76

Appendix IX: Figures containing the results in Hertz of the tables of appendix VII and VIII

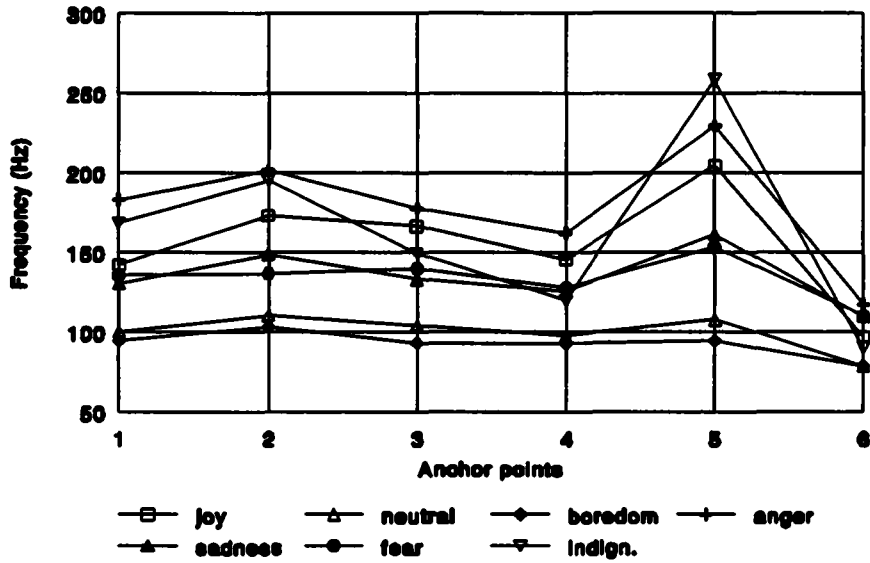


Figure 7: Results of the F₀ measurement over six points for the male speaker, over the utterances in the categories 5A/5&A, 1&A, 1A and 1EA.

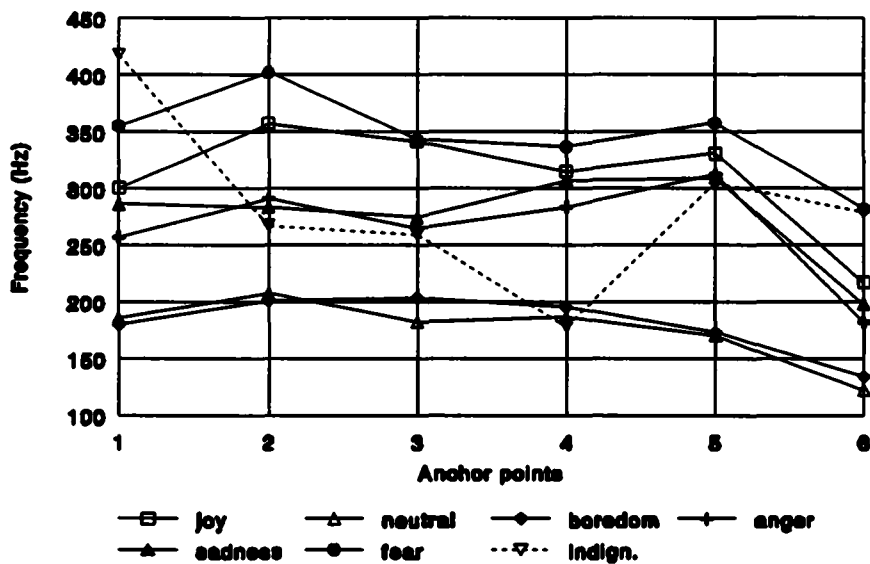


Figure 8: Results of the F₀ measurement over six points for the female speaker, over the utterances in the categories 5A/5&A, 1&A, 1A and 1EA.

Appendix X: Various types of pitch information in semitones

Table 14: General pitch information of the male speaker

	mean pitch (Hz)	SD	min. freq. (Hz)	max. freq. (Hz)	begin freq. (Hz)	end freq. (Hz)
joy	145.5	40.6	71.1	233.8	112.6	82.2
neutrality	98.2	15.1	66.3	128.1	95.6	78.5
boredom	95.2	11.3	71.7	124.7	103.6	84.5
anger	179.5	45.3	81.7	256.1	148.5	96.1
sadness	137.0	26.3	83.1	199.5	133.5	112.4
fear	134.1	27.1	88.4	197.4	143.1	127.3
indignation	184.4	50.1	102.9	296.9	179.1	197.7

Table 15: General pitch information of the female speaker

	mean pitch (Hz)	SD	min. freq. (Hz)	max. freq. (Hz)	begin freq. (Hz)	end freq. (Hz)
joy	263.5	66.3	114.3	389.8	214.3	119.0
neutrality	167.2	35.2	84.7	228.5	164.0	95.1
boredom	169.7	33.6	101.0	227.8	173.7	114.8
anger	252.6	62.7	105.8	367.5	202.0	110.1
sadness	282.7	58.5	165.4	393.8	289.4	184.3
fear	388.7	79.6	199.2	530.6	310.8	246.7
indignation	269.9	100.7	120.7	551.0	198.9	385.9