ROLE OF INTONATION PATTERNS IN CONVEYING EMOTION IN SPEECH

S.J.L. Mozziconacci* and D.J. Hermes

IPO, Center for Research on User-System Interaction, the Netherlands

** Now: "Projet Plurifacultaire Prosodie", University of Geneva, Dept of Linguistics,2 Rue de Candolle, 1211 Genève 4, Switzerland*

ABSTRACT

In a production and perception study the relation was studied between the emotion or attitude expressed in an utterance, and the intonation pattern realized on that utterance. In the production study, the pitch curves of emotional utterances were labelled in terms of the IPO intonation grammar. One intonation pattern, the '1&A', was produced in all emotions studied. Some other patterns were specifically used in expressing some of these emotions. In the perception study, in which the perceptual relevance of these findings was checked, the communicative role of the intonation patterns found in the database was tested. This listening test provided converging evidence on the contribution of specific intonation patterns to the perception of some of the emotions and attitudes studied. Some intonation patterns, such as final '3C' and '12', which were specifically produced in some emotion, also introduced a perceptual bias towards that emotion. In that sense, the results from the perception study supported those from the production study.

1. INTRODUCTION

Speech not only conveys the strictly linguistic content of sentences but also the expression of attitudes and emotions of the speaker. Prosody plays a role in this, which may result in adding information to the linguistic content and/or in its modification. Pitch level, pitch range, and speech rate are known to be important prosodic cues for, among others, the expression of emotions and attitudes in speech [14, 1; 2; 8, 10]. Related studies were reviewed in [4, 13, 11]. These studies rarely related production data to perception data, and did not consider the role of intonation patterns. In spite of this, it is likely that information on the emotion expressed by the speaker is not only present in global prosodic properties of the utterances, such as pitch level and pitch range, but also in more local properties as represented in the intonation patterns.

The present study focussed on the role of these intonation patterns in conveying emotions and attitudes in the production and in the perception of speech. First, two experienced experts in intonation listened to the pitch curves in the speech produced by three speakers and labelled them in terms of the Dutch grammar of intonation by 't Hart, Collier and Cohen [7], and the distribution of the thus found intonation patterns over the emotions was investigated. Next, originally neutral utterances were resynthesized using pitch contours with the intonation patterns most regularly occurring in the production data. An experiment was conducted to investigate whether specific intonation patterns contribute to the perception of particular emotions.

2. PRODUCTION OF EMOTION: AN ANALYSIS OF SPEECH MATERIAL

2.1. Speech material

Three Dutch speakers, two male speakers, MR and RS, and one female speaker, LO, enacted seven emotions or attitudes: neutrality (as a reference), joy, boredom, anger, sadness, fear, and indignation. In the following, the term emotion will be used as short for both emotion and attitude. In order to elicit the emotions, the speakers first spoke a semantically emotional sentence, such as: "How nice to see you here!" for the expression of joy. Once in the aroused mood, each speaker realized five sentences with that same emotion. This procedure was repeated three times. The five sentences are presented below. They are divided into an *initial* and a *final* part separated by a vertical slash. Stressed syllables are underlined.

1. Zij hebben | een nieuwe auto gekocht (They have bought a new car)
2. Zijn vriendin | kwam met het vliegtuig (His girlfriend came by plane)
3. Jan | is naar de kapper geweest (John has been to the hairdresser)
4. De lamp | staat op het bureau (The lamp is on the desk)
5. Het is bijna | negen uur (It is almost nine o'clock)

Sentences 1 and 5 could carry two accents in the second part of the sentence.

2.2. Procedure

According to the Dutch grammar of intonation [7], intonation patterns are composed of basic pitch movements:
'1': an early prominence-lending rise;
'2': a very late non-prominence-lending rise;
'3': a late prominence-lending rise;
'4': a slow rise extending over various syllables;
'5': an overshoot after a rise;
'A': a late prominence-lending fall;
'B': an early non-prominence-lending fall;
'C': a very late non-prominence-lending fall;
'D': a slow fall extending over various syllables;
'E': a half fall that may be prominence-lending.
When two movements occur on a single syllable, the symbols are linked with an ampersand, e.g., '1&A'.

Two experts in intonation research listened to the utterances, and to labelled the realized pitch curves of these utterances according to this grammar of intonation. They were instructed to discuss their judgment until they could agree on a single label.

2.3 Results and discussion

An overview of the occurrence of the various intonation patterns is presented in Table 1 for the initial part of the sentences and in Table 2 for the final part. Results are pooled over speakers.

emotion	intonation pattern					
	1&A	1	1B	1D	1&2(B)	rest
neutrality	7	12	15	3	5	3
joy	23	5	8	5	-	4
boredom	8	15	16	2	1	3
anger	18	6	16	1	2	2
sadness	19	3	16	1	1	5
fear	20	6	14	2	-	3
indignation	20	9	4	7	-	5
total	115	56	89	21	9	25

Table 1. Frequencies of occurrence of the INITIAL intonation patterns pooled over the 3 speakers.

emotion	intonation pattern						
	1&A	A	5A/ 5&A	EA	12/ 1&2	C	rest
neutrality	27	10	2	5	-	-	1
joy	32	2	2	-	-	9	-
boredom	15	12	3	5	1	7	2
anger	31	5	4	2	-	2	1
sadness	29	8	1	-	-	7	-
fear	27	3	2	-	-	11	2
indignation	9	-	9	-	13	10	4
total	170	40	23	12	14	46	10

Table 2. Frequencies of occurrence of the FINAL intonation patterns pooled over the 3 speakers.

The 'rest' category consists of utterances lacking fluency or including a pause, which could have affected the intonation pattern, utterances whose label could not unambiguously be assigned to one of the categories selected above, and utterances with a legitimate intonation pattern other than the selected ones if this pattern did not occur more than 5 times in the database. The rest category in Table 1 also included 5 cases in which no basic pitch movement was realized in the initial part of the utterance.

From Table 1 and Table 2, it appears that there is no direct coupling between single emotions and single intonation patterns. An emotion can be realized with different patterns, and a pattern can be used in the expression of different emotions. Consequently, there is no one-to-one relation between intonation pattern and emotion. This is not to say that intonation patterns are evenly distributed over all emotions. The '1&A' pattern is the pattern realized most frequently in emotional speech as well as in non-emotional speech. It can apparently be used in the expression of all emotions under study. This does not mean that this pattern is the best choice for the realization of any emotion, but that it is possible, using this pattern, to express each of the emotions studied here. With the exception of one speaker for indignation, the '1&A' pattern is the only one that was used by all speakers in expressing each emotion, both in initial and in final position.

For the expression of indignation, each speaker most often used a specific final intonation pattern different from the presumably more 'standard' '1&A' pattern; speaker MR usually used the '5&A' pattern, speaker RS the 'C', and speaker LO the '12' or the '1&2'. This suggests that the choice of intonation patterns may be of particular relevance for the expression of some emotions. Moreover, the patterns '12', '1&2', and 'C' were not used even once by any of the speakers in final position in the neutral utterances. A final '2' was only used by the female speaker LO, mainly in the expression of indignation. In final position, the 'C' pattern is the second most often used pattern in emotional speech, just after '1&A'. This pitch movement might be a good choice for signalling emotion in speech.

3. PERCEPTION OF EMOTION: AN EXPERIMENT

3.1. Aim

The previous analysis of emotional speech shows that in this recorded speech database, the '1&A' pattern was produced in the expression of all emotions and that patterns ending with '2' or 'C' were produced in emotional speech but not in neutral speech. This suggests that emotions can specifically be conveyed to the listener by using these patterns. As for the possibility to convey the seven emotions using the '1&A' pattern, the listening experiment reported in [9] already corroborates the hypothesis that, if one wants to keep the intonational structure constant, these emotions can be conveyed using the '1&A' pattern.

Here, we investigate whether specific intonation patterns affect the perception of emotion in speech. Therefore, the identification of intended emotions encoded in speech is investigated by systematically varying the intonation patterns. If the distribution of the subjects' responses differs significantly for different patterns, the relevance of the choice of intonation patterns for conveying emotion in speech can be established.

3.2. Speech material

Neutral utterances of both sentence 1 'Zij hebben een nieuwe auto gekocht' and sentence 2 'Zijn vriendin kwam met het vliegtuig' were manipulated by analysis and resynthesis methods based on the PSOLA technique [12]. By means of these manipulations, the seven pitch levels and pitch ranges previously found optimal for a specific emotion [10] were independently and systematically varied with various intonation patterns regularly occurring in the database. The values found to be optimal for each of the seven emotions are reported in Table 3. Among all combinations of intonation patterns occurring in the speech database, only those patterns occurring more than a couple of times and considered 'legal' according to the intonation grammar for Dutch [7] were retained for testing in the listening experiment. They comprised 11 patterns: '1&A 1&A', '1B 1&A', '1D 1&A', '12 1&A', '1A', '1EA', '1 5&A', '1&A 3C', '1B 3C', '1D 3C', and '1&A 12'. Vowel onsets and end of voicing served as points of reference in synthesizing the pitch movements. The timing and the duration of the pitch movements in the synthesis of the '1&A' and the 'A' patterns were derived from [3]. The realization of the other patterns was partly based on a description of representative pitch movements given in [6]. For some pitch movements, this description was not exhaustive for timing, duration and excursion size. In these cases, the F_0

variations of the natural realizations by the actors were taken as point of departure to find appropriate specifications. The detailed results are presented in [10]. Segmental durations were left unchanged. The resulting 154 variants (11 patterns × 2 sentences × 7 combinations of pitch level and pitch range) served as stimuli.

	neutr.	joy	bored.	anger	sad.	fear	indign.
Pitch level (Hz)	65	155	65	110	102	200	170
Pitch range (s.t.)	5	10	4	10	7	8	10

Table 3. Pitch level and pitch range found optimal in previous study.

3.3. Design and procedure

Each sentence was presented in a separate block. The order of presentation of the two blocks was counterbalanced across the subjects. Within a block, the stimuli were presented to each subject in a different random order. In total, 24 subjects (12 female, 12 male) participated in the listening experiment. Half of them were either students or staff members at IPO, the rest came from outside. None of them had any particular knowledge of phonetics. Each subject took the test individually, using headphones and an interactive computer program. Subjects could listen only once to each stimulus, and did not get any feedback concerning their performance. They had a short break before the second block splitting the test into two periods of about 10 to 15 minutes each. Their task was to assign one of the seven emotion labels to the utterance they heard.

3.4. Results

The responses of the subjects were pooled into a three-dimensional table with 'combination of pitch level and pitch range' as one dimension, 'intonation pattern' as a second dimension, and 'response of the subjects' as a third one. These data were subjected to a three-way log-linear analysis [5]. In the simplest log-linear model into which the data can be fitted, there were significant interactions between 'pitch level - pitch range combination' and 'response', and between 'intonation pattern' and 'response', but not between 'pitch level - pitch range combination' and 'intonation pattern' ($\chi^2_{420} = 459.7, p > .9$). The implication is that, within each response class, 'pitch level - pitch range combination' and 'intonation pattern' can be assumed to be independent factors. It shows, for example, that one specific intonation pattern does not exclude a particular response of the subject and neither uniquely determines his or her response, whatever the combination of pitch level and pitch range. Also, the results can be collapsed over each of the independent variables. Doing so over the 'pitch level - pitch range combination' factor generates a table featuring the response distribution per intonation pattern. The resulting table forms the basis of a cluster analysis performed in order to find out whether intonation patterns can be combined into groups. The members of such groups give rise to similar responses. The composition of these clusters may provide us with information about which properties of the intonation patterns are essential in conveying a certain emotion. Therefore, for each possible combination of two rows, a (log-linear) test of independence was applied. Indeed, if the two intonation patterns corresponding with two rows induce a different distribution of the responses over the emotions, significant differences will be observed between the two rows. Otherwise, the two rows will be the same except for statistical fluctuations, and, hence, within one cluster, the composition of the rows will be independent of the intonation pattern. The computation of chi-squares (χ^2's) was assumed to provide a measure of association for the two intonation patterns corresponding with each two rows. Therefore, for each combination of two rows, these χ^2's were calculated and the corresponding p-values with six degrees of freedom were obtained. The smaller the χ^2, the more the two intonation patterns are associated, and inversely for the p-values. In Table 4, the p-values larger than .05 are presented. This analysis yields three clusters of intonation patterns. The first cluster is composed of

	intonation pattern										
	1&A 1&A	1B 1&A	1D 1&A	12 1&A	1 5&A	1A	1EA	1&A 3C	1B 3C	1D 3C	1&A 12
1&A 1&A	-	.920	.843	.680							
1B 1&A	.920	-	.733	.800							
1D 1&A	.843	.733	-	.226							
12 1&A	.680	.800	.226	-							
15&A					-						
1A						-	.333				
1EA						.333	-				
1&A 3C								-	.437	.948	
1B 3C								.437	-	.280	
1D 3C								.948	.280	-	
1&A 12											-

Table 4. p-values associated with the chi-squares used for measuring dissimilarities between patterns

cluster	emotion							
	neutr.	joy	bored.	anger	sad.	fear	indign.	total
1...1&A	524↑	229↑	170	95↓	133	78	115↓	1344
15&A	90	69↑	31↓	52↑	20↓	16	58	336
1...A	179	52↓	83	138↑	65	26↓	129	672
1...3C	214↓	149	172↑	67↓	152↑	66	188	1008
1&A 12	30↓	27↓	29↓	29	25	48↑	148↑	336
total	1037	526	485	381	395	234	638	3696

Table 5. Response distribution per cluster of intonation patterns

'1&A 1&A', '1B 1&A', '1D 1&A', and '12 1&A', and will further be indicated with '1...1&A'. The second cluster is composed of '1A' and '1EA' and will be referred to as '1...A'. The third cluster, composed of '1&A 3C', '1B 3C', and '1D 3C', will be referred to as '1...3C'. The remaining intonation patterns, '1&A 12' and '1 5&A', differ from each other and from all the other patterns. Note that the members of these clusters correspond, largely , to the intonation pattern occurring in final position in the utterance. All intonation patterns ending in '1&A' form a cluster, as well as the three intonation patterns ending in '3C'.

Finally, the responses collapsed over the combinations of pitch level and pitch range were summed over all members of each cluster of intonation patterns. The result is presented in Table 5. Deviations from a log-linear model in which perceived

emotion and intonation pattern are independent, are marked with ↑ if the obtained value is higher than expected ($\chi^2 > 3.84$, $p < .05$), and with ↓ if the obtained value is lower than expected. The results show, for instance, that in the response class 'neutrality', stimuli with intonation patterns of the '1...1&A' cluster are significantly over-represented, while stimuli of the '1...3C' cluster and the '1&A 12' intonation pattern are under-represented. Another example is that within 'indignation' '1&A 12' is over-represented and cluster '1...1&A' is under-represented.

Additionally, if one considers an emotion to be 'correctly identified' when an utterance that had received the pitch level and pitch range adequate for that particular emotion was labelled with that same emotion by a subject, independently of the intonation pattern used in the utterance, then emotions were correctly identified in 22.1% of all cases. Although low, this percentage of correct identification is higher than a chance level of 14.3%. This percentage can naturally not be compared with the usual percentage of correct identifications, because, in the present study, the stimuli were not prepared to instantiate a specific emotion. Each of the eleven intonation patterns was indeed imposed on each of the seven pitch curves synthesized with optimal pitch level and pitch range. In other words, the choice of intonation pattern could very well provide information conflicting with the information provided by pitch level and pitch range. Another reason that can be given to account for the low proportion of correct identification in the experiment is that speaking rate and voice quality, that are known to be important determinants of the emotion conveyed, were kept constant.

4. GENERAL DISCUSSION

The first important conclusion is that the intonation pattern realized on an utterance is one of the determinants of the emotion conveyed in speech. Particular patterns seem to be better suited for conveying some specific emotions and less suitable for others. Although, in the production study, no clear-cut, one-to-one relationship between intended emotion and intonation pattern was found, some clear relationships could be distinguished. The pattern of pitch movements '1&A' occurred most often, both in the initial and in the final position of the sentences used in this study, and was produced in all emotions. Furthermore, some intonation patterns ending in 'C' or '2', were not used in neutral utterances in the database. These patterns could serve for specifically signalling some emotions.

The perception experiment confirmed that the intonation pattern is a relevant cue in signalling an emotion. Furthermore, the cluster analysis showed that it was predominantly the final part of the intonation pattern which affected the listener's response. The suggestions, based on the production study, that the '12' pattern was associated with indignation, was strongly confirmed. Another suggestion that the '12' and the '3C' patterns were negatively associated with neutral speech was also confirmed. The suggestion that '1&A' would lead to a reasonable identification of all emotions was confirmed too. Therefore, if one requires different intonation patterns not to contribute to experimental variability, a sequence of '1&A' patterns is found to be best suited for controlling this variability. The patterns frequently used in the expression of an emotion in the production study were found to introduce a perceptual bias towards that emotion in the perception experiment, so that both the production and the perception study support the same conclusions. Moreover, the perception study revealed many more interdependencies between intonation pattern and conveyed emotion. This can undoubtedly be attributed to the fact that in the perception experiment the two factors, 'pitch level-pitch range combination' and 'intonation pattern', can be varied independently.

A second conclusion concerns the description of the intonation of the utterances conveying emotion. In almost all cases, the labelling yielded patterns that were grammatically correct within the IPO grammar for Dutch intonation [7]. Hence, these results indicate that this intonation grammar is adequate for the description of the intonation patterns found speech conveying emotion.

ACKNOWLEDGEMENTS

This research was supported by the Cooperation Unit of Brabant Universities (SOBU).

REFERENCES

[1] Cahn, J.E. 1990. Generating expression in synthesized speech. *Technical Report*. Boston: MIT Media Lab.

[2] Carlson, R., Granström, B. and Nord, L. 1992. Experiments with emotive speech: Acted utterances and synthesized replicas. In: J.J. Ohala, T.M. Nearey, B.L. Derwing, M.M. Hodge and G.E. Wiebe (Eds.): *Proceedings of the International Conference on Spoken Language Processing, ICSLP-92, Banff, Alberta, Canada*, 671-674.

[3] Collier, R. 1991. Multi-language intonation synthesis. *Journal of Phonetics*, *19*, 61-73.

[4] Frick, R.W. 1985. Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*, 412-429.

[5] Fienberg, S. E. 1980. *The analysis of cross-classified categorical data, second edition*. The MIT Press, Cambridge, Massachusetts.

[6] Hart, J. 't and Collier, R. 1975. Integrating different levels of intonation analysis. *Journal of Phonetics*, *3*, 235-255.

[7] Hart, J. 't, Collier, R. and Cohen, A. 1990. *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.

[8] Kitahara, Y. and Tohkura, Y. 1992. Prosodic control to express emotions for man-machine interaction. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 75*, 155-163.

[9] Mozziconacci, S.J.L. 1995. Pitch variations and emotions in speech. *Proceedings of the 13th International Congress of Phonetic Sciences, ICPhS-95, Stockholm, Sweden, August 13-19, 1995, 1,* 178-181.

[10] Mozziconacci, S.J.L. 1998. *Speech variability and emotion: Production and perception*, Technical university of Eindhoven, The Netherlands.

[11] Murray, I.R. and 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.

[12] Moulines, E. and Laroche, J. 1995. Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication, 16*, 175-205.

[13] Scherer, K. R. 1986. Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, *99*, 143-165.

[14] Williams, C.E. and Stevens, K.N. 1972. Emotions and speech: Some acoustical factors. *Journal of the Acoustical Society of America*, *52*, 1238-1250.