

Cross Linguistic Interpretation of Emotional Prosody

*Åsa Abelin
Jens Allwood*

Abstract

This study has three purposes: the first is to study if there is consensus in the way listeners interpret different emotions and attitudes expressed by a Swedish speaker, the second is to see if this interpretation is dependent on the listeners' cultural and linguistic background, and the third is to ascertain whether there is any reoccurring relation between acoustic and semantic properties of the stimuli.

Recordings of a Swedish speaker uttering a phrase while expressing different emotions were interpreted by listeners with different first languages, Swedish, English, Finnish and Spanish, who were to judge the emotional contents of the expressions.

The results show that some emotions, e.g. anger, fear, sadness and surprise, were interpreted in accordance with intended emotion to a greater degree than others, which were interpreted as expected to a lesser degree. Emotions were interpreted with different degrees of success depending on the first languages of listeners; native listeners were the most successful. The results suggest that emotions with similar semantic features, e.g. anger and dominance or fear and shyness have similar acoustic features e.g. short duration and strong intensity (anger and dominance) or longer duration and weak intensity (fear and shyness).

Key words: Cross-linguistic interpretation, emotion, prosody

1. Introduction

This study had the purposes of:

- (i) Investigating if there is any consensus in the way in which listeners interpret different emotions and attitudes expressed by a Swedish speaker
- (ii) Investigating if this interpretation is dependent on the cultural and linguistic background of the listener

- (iii) Making acoustic analyses of the prosody of the stimuli used in the study, and making semantic analyses of the responses to the stimuli. Both these types of analyses have been made to investigate if there is any relation between acoustic and semantic properties.

In the study of interpretation of emotional prosody there is, we believe, a need for both acoustic and semantic analysis. The result of an acoustic analysis is not independent of how the semantic analysis is done. This is particularly true when dealing with cross cultural interpretation. Studies of languages other than Swedish have been made by e.g. Fónagy (1967) and Williams and Stevens (1972). The study by Williams and Stevens concerned acoustic analysis and the language is English. Fónagy has, among other things, studied articulation, glottal behavior. For articulation he claimed to find an iconic relation between the articulatory movements for emotions and the meanings of these emotions; Tenderness should, according to Fónagy, be characterized by slow, tender movements of the articulatory organs while wrath should be mirrored in violent articulatory movements and extreme articulatory positions. Fónagy also claimed that the emotional states influence the breathing patterns and thereby the sub glottal pressure, which has consequences for phonation frequency. Studies of German emotional prosody have been made by e.g. Paeschke, Kienast and Sendlmeier (1999). They showed among other things that the intensity of syllable stress distinguished the group of so called non-excited emotions (e.g. sadness) from excited emotions (e.g. anger), and that there were emotion specific F0 features. The question whether the expressions of emotions are universal or language specific also has importance for the question of how we interpret emotions. A recent cross-linguistic study of the interpretation of emotional prosody investigating the universality versus cultural relativity of emotions is Scherer, Banse and Wallbott (2001). This study, which was conducted in nine countries in Europe, the United States and Asia finds an overall recognition rate of 66% across all emotions and countries as well as a pattern of confusion which was very similar across all countries. It also found that accuracy decreased with increasing language dissimilarity from German. The approach of the present study of interpretation of Swedish emotional prosody is similar to the study of Scherer, Banse and Wallbott (2001), but is however conducted in a much smaller scale. Davitz and Davitz (1959) let eight subjects express ten different emotions while reciting parts of the alphabet. Each subject was recorded and the recordings were then played aloud to 30 judges, who were to identify the emotions they heard, from a list of the ten emotions. The results showed significant differences between the emotions, but also showed a great variation between the different judges, speakers and emotions. Hayashi (1999) studied how Japanese speakers interpret emotions, expressed by Japanese, in dialogue. The interpretations were mostly in accordance with the intended emotions. The results indicate that F0 conveys substantial information about emotional states. Chung (1999) investigated vocal expressions for emotions in Korean and the perception of emotions for Korean, American and French listeners. The results show that the listeners agreed on the global judgement of the emotions.

A basic problem in studies of how emotions and attitudes are manifested is the question of which emotions and attitudes should be chosen and how different emotions and attitudes can be delimited from one another. There are different theories of which are the primary emotions, e.g. Woodworth (1938) who distinguished love, joy, surprise, fear, suffering, anger, dominance, disgust, contempt; Izard (1971) who used the following dimensions: interest excitement, liking joy, surprise astonishment, despair anguish, disgust contempt, anger fury, shame humiliation, fear panic. Roseman (1979) used the emotions joy, relief, despair, sadness, hope, fear, frustration, liking, love, disapproval, hatred, anger, pride, shame and regret.

The emotions we have chosen are joy, surprise, sadness, fear, shyness, anger, dominance and disgust. These are the so-called basic emotions (see, e.g. Ekman et al, 1987) plus dominance and shyness. Dominance, which can be described as a personality trait was included in the study in order to see its relation to anger. The reason for including shyness was that this emotion, or rather attitude, is thought to be highly dependent on social context. Therefore we considered it interesting to see if it was possible to correctly judge vocal expression of shyness without knowing the context in which the emotion was uttered. We are aware that the eight emotions we have chosen are not all of the same basic level. We are also aware that it is possible to express vocally, and to distinguish, more than the eight emotions we have tested in this study. But since there seems to be no agreement on how emotions and attitudes are manifested vocally, we believe the present study although not covering all emotions, is of general interest.

For further study of the expression of emotions in speech there is a need for empirical studies, but also a need for a theoretical clarification of the concept of emotion. For example, is it important to distinguish emotions from attitudes? Do words such as joy and anger stand for both emotions and attitudes? Other important questions to consider when determining the concept emotion concern the relation of emotions to physiological states, psychological states, outward behavior and situational factors. For an overview of different theories of emotion, see Cornelius (2000). According to our view this depends, at least partly, on the fact that emotional expressions in natural languages modify many different aspects of the processes, which are brought to life when emotions occur. Better investigations of the semantic content, and the relations between emotional words in semantic fields, is therefore probably a prerequisite for more insightful studies of how emotions are manifested and signalled in communicative behavior.

2. Method

Material and recordings

For constructing the speech material we used the method of acted samples. This method has its drawbacks, e.g. the restricted possibilities of making generalizations to the

expression and interpretation of emotions in natural communication, but is nevertheless often used. The advantage of this method is that it is easy to produce structured speech materials, e.g. the same carrier phrase with different emotional expressions.

One male speaker, from the west of Sweden, not a professional actor, was used for producing the utterances. Naturally, using only one speaker puts a restriction on a general interpretation of the results since there is most probably a variation in the expression of emotion depending on e.g. age, sex, and idiosyncrasies of the speaker, and this study must be seen as the first in a longer series. However, this study is of interest in the light of the absence of similar studies of the expression and interpretation of emotions in Swedish.

The voice samples were elicited for the emotions anger, fear, joy, sadness, surprise, disgust, dominance and shyness and were recorded on audio-tape. The actor was instructed to try to experience the different emotions and then express them. To avoid the influence of semantics on the performance of the actor, and later, on the interpretations of the listeners, we used a carrier phrase which would be as neutral as possible with respect to the different emotions. We also wanted a sentence with a content against which different emotions naturally could be felt. The carrier phrase contained words for different dishes "Salt sill, potatismos och pannkakor" (Salted herring, mashed potatoes and pancakes), and the idea was that many different emotions could be held towards food. This assumption might be correct within the Swedish culture but it is probably somewhat naïve within a cross-cultural perspective. The dishes of the carrier sentence are to a certain degree associated with Swedish or at least northwestern European culture.

The speaker produced the utterance with the eight different emotional expressions until he was himself satisfied with the expressions of emotions. These eight emotional expressions of the carrier phrase were then played back to listeners who were to identify the expressions of emotions.

The listener group and listener test

The listener group consisted of 35 native Swedish speakers, 23 native Spanish speakers, 23 native Finnish speakers and 12 native English speakers. The non-Swedish listeners were Swedish immigrants and all had knowledge of Swedish, of varying proficiency. The reason for having a much smaller number of English speaking listeners in this study was the scarcity of native English speaking immigrants in Sweden. The listeners were of varying ages and sexes. For practical reasons the listening tests took place in groups or individually. The stimuli were presented over loud-speakers and only presented once. When the listeners had written down their interpretation of a stimulus the next stimulus on the tape was presented. The listeners were told to write down his or her interpretation of each emotional expression immediately after listening to it. The listeners were instructed to write down any word that came into their mind and to write the interpretations in their own language.

3. Analysis

In those cases where the listeners did not have Swedish as a first language we made a translation into Swedish with the help of bilingual persons. When this had been done, all the interpretations of the emotions made by the listeners were classified into semantic fields. As an example of how the answers have been grouped we can mention that angry, mad, rage, wrath and resentment (in English translation) have been grouped into one semantic field, Anger, and that fear, afraid, timid, frightened, horror, fright, terror were grouped into the semantic field Fear. In some cases it is difficult to judge which semantic field a word or an expression belongs to, but this problem is unavoidable when letting the subjects use free choice as a method of categorization. The advantage of using free choice is that the listeners do not feel forced to choose an emotion, which he or she does not consider correct. In some cases, however, the listeners used more imaginative expressions to describe the emotion expressed in the recording, and in these cases the answers have not been considered, due to problems of classification. An example of this is when disgust is classified as "suburban kitchen without beer" or when shyness was classified as "said as a culinary footnote" or "said as a children's program". Other informants have not been able to overlook the lexical content of the carrier phrase and given interpretations, which are specific for attitudes and emotions towards food, e.g. "indigestion" for disgust or "he asks for food" for sadness. However, this occurred quite seldom. Now and then the informants focused on some other dimension, e.g. how natural the speaker sounded, for instance when answering theatrical for the emotion fear.

Something else, which is problematic in a cross-cultural study is of course that there is no one-to-one relation between word meanings in different languages. In Spanish there is for example the word *espantado* which expresses simultaneous surprise and fear which the words *asombrado* and *sorprendido* do not. It is a well-known fact the lexicons of different languages are differently structured.

There are certainly drawbacks in using free choice as a method for listeners to respond; the responses can be difficult to classify, and overall the number of responses for each stimulus can be quite large. However, using the method of forced choice may add a loss in validity and conceal the variety of interpretations, which do exist and which is manifested when using free choice.

4. Results

4.1 Interpretations

In the sections below first follows a presentation of the listeners' answers, in tables. The presentations of the interpretations are presented in a confusion matrix for five of the emotions. Figures are in percent of total number of answers. When listeners have given answers which are not the same as those of other listeners, or not classifiable into the

same semantic field as other answers, these answers are not counted. This means that the percent answers for a given emotion, by a given language, seldom adds up to 100%. The shortfall to 100% shows how great a proportion of the answers were isolated, indicating the difficulty the listeners found in making their interpretations. For reasons of space, the isolated answers are not listed here.

After the confusion matrix the remaining three emotions are presented separately. After this section a semantic analysis of the interpretations and thereafter an acoustic analysis of the stimuli are presented.

Stimulus		Anger	Fear	Joy	Sadness	Surprise	Total mean
Anger	sw	66					
	en	50					
	fi	87				9	
	sp	78				30	
Fear	sw		66				
	en		42				
	fi		9				
	sp		30			17	
Joy	sw			89		9	
	en			50			
	fi			35			
	sp			22			
Sadness	sw				69		
	en			8	100		
	fi		61	17	70	9	
	sp		48	35	91		
Surprise	sw					74	
	en					42	
	fi					65	
	sp					22	
Mean % expected answer for each emotion		70	37	49	83	51	58

Table 1. Confusion matrix for five of the eight emotions studied. Numbers are in % of total number of answers.

Table 1 shows the percentage expected answers, as well as non-expected answers, for each of the five emotions anger, fear, joy, sadness and surprise. The Swedish interpretations of joy, show quite a large correspondence with the intended emotion. Among the interpretations from the other language groups the Spanish listeners are the most deviant. They interpret joy as sadness in 35% of the cases. They interpret the emotion as joy in 22% of the answers and as contentment also in 22%. The Finnish group is also deviant in that 17% of the Finnish listeners interpret joy as sadness.

The different language groups all interpreted surprise quite successfully. Some of the Swedish interpretations were happy surprise in 17% of the 74% that interpreted this emotional expression as surprise. This could indicate that joy and surprise have similar

acoustic characteristics in Swedish. The most striking results are the Spanish listeners' interpretation: anger in 30% of the answers and fear in 17% of the answers. Again the Spanish listeners are the most deviant. The English listeners produced a large amount of isolated answers (which they did for most of the emotions), and these are not shown in the table; therefore the cumulated answering percent for the English listeners is quite low.

As can be seen in Table 1, the Swedish listeners identify sadness to a lower degree than listeners of the other languages. The English listeners interpreted the emotion correctly to 100%. Many of the Swedish listeners give the interpretation disappointed, and so did a similar proportion of the Finnish listeners (not shown in Table 1). The results of the Swedish listeners could indicate that the speaker did not really succeed in expressing pure (Swedish) sadness. Nevertheless, this is an emotion which most listeners, whatever their first language, succeeded in interpreting.

The emotion fear has been interpreted mainly as fear or sadness. The semantic contents of these emotions are related and to some extent also their acoustic manifestations. The expression of this emotion could be more language dependent than sadness, since most Swedish listeners interpreted fear as expected, while the majority of the Spanish and Finnish listeners interpreted the emotion as sadness.

In summary, sadness and anger were the two emotions which were most often interpreted accurately. Joy was in many cases interpreted as sadness by the Finnish and Spanish listeners. Fear was confused with sadness also by the Finnish and Spanish listeners. Surprise was confused with all the other four emotions, by different language groups.

The result that the interpretation of sadness is the most successful could be due to a tendency to interpret all of these emotional expressions as sadness; both fear and joy have been interpreted as sadness by a large number of informants.

Furthermore the table shows the mean percent of expected answers calculated over the four language groups, and finally it shows the total mean for both the four language groups and the four emotions: 58%. If only four of the emotions are calculated, anger, fear, joy and sadness the average is 60%. This result can be compared with the result of Scherer, Banse and Walbott (2001) where the emotions anger, fear, joy, sadness and neutral were successfully recognized, across a large number of countries, to a mean of 66%. In Scherer, Banse and Walbott (2001) the averages across all countries ranged from 76% for anger to 42% for joy; in the present study the recognition rates for the emotions anger and joy were 70% and 49% respectively. The averages ranged from 83% for sadness to 37% for fear, in the present study.

None of the listeners identified the emotion shyness, see Table 2. This result is expected, since shyness was supposed to depend on the social context. It is difficult to judge this expression of emotion only from the voice quality. A large proportion of the

Swedish listeners judge this emotional expression as fear, and quite a large amount judge it as sadness. The emotional content of fear and sadness are quite similar to that of shyness. In contrast to this, nobody interpreted fear or sadness as shyness, probably because shyness is a less basic emotion, and also depending on context.

Shyness	Swedish	English	Finnish	Spanish
Sadness	11	8	30	30
Fear	34	8	0	17
Surprise	0	0	9	0

Table 2. Interpretations of shyness by listeners from four different languages. Numbers represent % of total number of answers.

The majority of all listeners interpreted the expression for dominance, cf. Table 3, as anger and as dominance. Some listeners interpreted it as the related emotions determination or irritation.

Dominance	Swedish	English	Finnish	Spanish
Anger	4	0	35	57
Dominance	71	67	43	9
Irritation	0	0	0	9

Table 3. Interpretations of dominance by listeners from four different languages. Numbers represent % of total number of answers.

The Spanish and Finnish listeners interpret the emotional expression for anger as anger to a higher degree than the Swedish and English listeners, see Table 1. Dominance and anger are semantically and acoustically similar, but there may be a cultural difference in the interpretation of these emotions.

The emotion disgust, cf. Table 4, was difficult to interpret, for all listeners. This could be because the speaker did not in fact succeed in expressing this emotion, or it could be because disgust is a more complex or less basic emotion.

Disgust	Swedish	English	Finnish	Spanish
Sadness	0	0	17	30
Surprise	6	0	9	0
Happiness	6	0	0	9
Irritation	0	0	9	0

Table 4. Interpretations of disgust by listeners from four different languages. Numbers represent % of total number of answers.

4.2 Semantic analysis

As we have seen above, all emotions were not interpreted as expected, but sometimes they were interpreted as an emotion which is, in some aspect, semantically close to the intended one, as when shyness is interpreted as sadness or fear (but not as anger or

dominance). For this reason we also examined the intended and interpreted content of emotions from a semantic point of view.

Our semantic analysis has the purpose of showing some of the relations between the emotional terms used, both in the listeners' answers and in our characterizations of the emotions that the speaker intended to express. The basis for the analysis is a comparison of similarities and differences between the terms in question, which could be relevant from an emotional perspective. This entails that our analysis of the meaning of the terms is only partial. For more complete analyses of some of the terms we study, see e.g. Abelin (1980), Boyd (1980), Hirsch (1980).

A convenient way of comparing terms within a semantic field is to try to find some of the properties or dimensions, which both unite and distinguish the terms in question. In the literature on emotions there are plenty of suggestions of such dimensions, but there is no agreement on this, nor is there any agreement on the question of which emotions are the basic ones. Some of the suggestions which have been given are the following: Wundt (1896) suggested three basic dimensions: lust – non-lust, relaxation – tension and calm – excitation. Schlosberg (1954) suggested almost the same dimensions: lust – non-lust, attention – rejection and sleep – tension. Osgood (1966) suggested degree of lust, degree of activation and degree of control which are similar to what he suggested in his semantic differential: evaluation, activity and strength. Frijda (1970) adds a dimension to the ones already mentioned, self-assured – insecure. Roseman (1979) takes a partly different path and suggests the following five dimensions need – wish, occurrence of a certain state, probability, type of cause and legitimacy.

The choice of dimensions probably to some extent depends on the purpose of the particular study. The dimensions which we have chosen should therefore be related to the purpose of our semantic analysis, namely to clarify semantic relations between the terms which have been used to characterize emotional qualities in the voice. The dimensions chosen are those which have been considered to play a role for expressive behavior. The dimensions that we have chosen are the following:

1. lust – non-lust
2. active – passive
3. secure – insecure

We treat these dimensions binary, while remaining well aware that a more thorough analysis would demand a finer grading of the dimensions.

Furthermore, in the use of the dimensions we leave a great many things vague – which ought to be distinguished in a finer analysis. As an example, we do not distinguish between if the non-lust is combined with the emotional state itself, or if it is rather connected with the cause of the state or with the object, which the emotion is directed towards. The three dimensions, which we have chosen, have not been utilized maximally for each term but only that or those dimensions that were judged appropriate.

In this way e.g. joy gets the values [+lust, +activity] while contentness only gets the value [+lust].

The use of the three dimensions gives the following results if we consider the answers given by the listeners, emotion for emotion.

joy	content	surprise	eagerness	intoxication	sadness
+lust	+lust	+lust	+lust	+lust	-lust
+active		+active	+active	+active	-active
		-secure	+secure		

Table 5. Semantic features of the interpretations of the emotion joy.

Following the three dimensions chosen, one could say that the relevant content of the emotion joy is lust-filled activity. This characterization unites joy with a larger number of the answers given by the listeners: surprised, eager and intoxicated. If we only look at the component lust, there is an association also with contentness.

If we accept this analysis the following proportions of the answers are completely or partly associated with joy: for Swedish 92%, English 66%, Finnish 65% and Spanish 79%. The conclusion is that using this analysis with semantic dimensions, a greater number of the listeners' interpretations of joy are successful. This is a result, which is replicated for all the emotions in this investigation; the classification of the interpretations with the aid of the three semantic dimensions gives a higher (or the same) percent answer in accordance with intended emotion.

Our analysis gives surprise the content lust filled active insecurity, as shown in Table 6. As we can see the variation of the listeners' interpretations is greater for surprise than for joy. But with the same kind of analysis as above it is possible to assert that all the interpretations, except for angry and sad, are possible to relate to surprise, either by the properties lust-filled activity or by the property insecurity. The following proportions of the answers in this way become related to surprise: for Swedish 83%, English 42%, Finnish 74% and Spanish 61%.

surprise	joy	enthusiasm	hesitance	content	anger	fear	sadness
+lust	+lust	+lust		+lust	-lust	-lust	-lust
+active	+active	+active	-active		+active	+active	-active
-secure		+secure	-secure		+secure	-secure	

Table 6. Semantic features of the interpretations of the emotion surprise.

The meaning of sadness, shown in Table 7, becomes non-lust filled inactivity.

sadness	disappointment	disgust	displeas	tiredness
-lust	-lust	-lust	-lust	
-active				-active

Table 7. Semantic features of the interpretations of the emotion sadness.

All the interpretations of this emotional expression can, through this analysis, be semantically associated with sadness, either via non-lust or via inactivity. The proportion of answers semantically associated with sadness is the same when using this analysis.

The meaning of fear becomes non-lust filled active insecurity, see Table 8.

fear	despair	upset	sadness	bitterness	love
-lust	-lust		-lust	-lust	+lust
	+active	+active	-active		+active
-secure					

Table 8. Semantic features of the interpretations of the emotion fear.

Again all the answers reported by the listeners can be semantically associated with fear, either by non-lust or by activity. The proportion semantically associated answers are: for Swedish 66%, English 42%, Finnish 70% and Spanish 78%.

In Table 9 is shown the meaning of shyness, which is non-lust filled inactive insecurity. Of the answers reported here, four ones (sad, afraid, depressed and nervous) are associated with shyness via two dimensions. The remaining ones are associated with shyness via one dimension. The percentage expected interpretations becomes somewhat higher if we take the semantically related answers into the analysis: for Swedish 45%, English 8%, Finnish 39% and Spanish 47%. If we take into consideration all the isolated answers, e.g. low-spirited, hesitant, unobtrusive, with pious wish, doubting, pleading, which can be associated with shyness, meaning non-lust filled inactive insecurity, the proportion of expected answers will be much higher.

shyness	sadness	fear	depression	nervousness	softness	thinking	surprise	disgust
-lust	-lust	-lust	-lust	-lust	+lust		+lust	-lust
-active	-active	+active	-active	+active		-active	+active	
-secure		-secure		-secure	-secure		-secure	+secure

Table 9. Semantic features of the interpretations of the emotion shyness.

Anger is non-lust filled active security, as shown in Table 10. As we can see, the answers given have a clear semantic association to anger. The results are: for Swedish 83%, English 75%, Finnish 87% and Spanish 96%.

anger	determination	irritation
-lust		-lust
+active	+active	+active
+secure	+secure	

Table 10. Semantic features of the interpretations of the emotion anger.

Dominance has been interpreted as determination, and its meaning becomes, with our dimensions, active security, see Table 11.

determination	anger	irritation
	-lust	-lust
+active	+active	+active
+secure	+secure	

Table 11. Semantic features of the interpretations of the emotion dominance.

We can see that the main variation between answers is the same for dominance and anger. The results for answers within the semantic field for dominance are: Swedish 81%, English 67%, Finnish 78% and Spanish 79%.

The meaning of disgust, in Table 12, can, with our dimensions, be said to be non-lust filled security.

disgust	Dis- appoint- ment	Displea- sure	Bore- dom	irrita- tion	bitter- ness	nervous- ness	sad- ness	tiredness	happi- ness	surprise
-lust	-lust	-lust	-lust	-lust	-lust	-lust	+lust		+lust	+lust
			-active	+active		+active	-active	-active	+activ e	+active
+secure						-secure				-secure

Table 12. Semantic features of the interpretations of the emotion disgust.

4.3 Discussion of the interpretations of the different language groups

The different language groups did not show the same success in interpreting the intended emotion of the speaker. The Swedish listeners performed best according to expectations for most of the emotions. An interesting result is that sadness and anger, in spite of getting approximately as high percent expected interpretation among the Swedes as the other emotions, are those emotions that are interpreted better by the other language groups (the only exception being the English interpretations of anger). The implications of this are unclear. Another question arising from the results is why not all the Swedes interpreted the emotions correctly. A possible answer to this is that all emotions are, more or less, dependent on context for their interpretation. Other possible answers are that the informants were stressed by the test situation, or they have difficulties finding the right word for their interpretation of the emotional expression or that the emotions were not unambiguously expressed. However, it must be pointed out that the Swedes, in most cases, had chosen semantically related emotions when not answering as expected. It is also reasonable to suppose an interaction between prosody and non-verbal signals such as facial expressions (cf. Massaro, 2000); this is however not investigated in the present study. It is not to be expected that all the information about the emotion is present in the acoustic signal. The proportion between the information conveyed by the acoustic signal and the information in the visual signal might even differ between different emotions, and of course the proportions can also vary between different languages. On the other hand, a part of the non-verbal signal,

namely the smile, is also coded in the acoustic signal (cf. Aubergé and Lemaître, 2000). Yet another type of interaction to be considered is that between prosodic and lexical information. Dubost and Tzu-ting (1999) showed that the perception of emotions in Mandarin was more difficult if a specific lexical marker was missing. In French, however, lexical markers seemed to be optional while intonation was decisive for interpretation of emotional expressions.

4.4 Acoustic analysis

The recordings of the ten emotions were studied in oscillograms. The length of the utterances was measured in two ways: with and without the pauses occurring between the nominal phrases. The intensity of the expressions was graded in relation to the intensity of the other emotions. The duration of the intensity peaks was not measured but the height of the dominating peak was measured. The intensity measures are not completely reliable, since no headworn microphone was used. The F0 curve was measured and classified with respect to if F0 was high or low, if it was level, increasing or decreasing.

F0 curve

Joy, fear, shyness and to some extent sadness show similarities. The F0 curve is even and quite high. Surprise, anger and dominance have an F0, which varies strongly up and down. Anger and dominance are distinguished from surprise mainly by being shorter. Paeschke, Kienast och Sendlmeier (1999), on English, showed that sadness, and boredom, were spoken with a lower average F0, compared with a neutral statement, whereas disgust, anger, fear and joy were spoken with a higher average F0. In the present study we did not look at average F0.

Intensity

The emotions that have the strongest intensity peaks are anger, surprise, disgust and dominance. The weakest are sadness and shyness.

Duration

On the duration parameter the emotions group themselves somewhat differently depending on whether or not the length of the pauses is counted. Table 13 shows the durations in milliseconds with and without pauses.

with pause	ms	without pause	ms	difference	ms
disgust	7 400	joy	5 670	shyness	2 290
sadness	6 850	disgust	5 610	sadness	2 240
surprise	5 860	surprise	5 170	disgust	1 790
shyness	5 750	sadness	4 610	dominance	1 550
joy	5 740	fear	4 300	fear	1 340
fear	5 640	shyness	3 460	anger	860
anger	4 210	anger	3 350	surprise	690
dominance	4 090	dominance	2 540	joy	10

Table 13. Durations, in milliseconds, of the eight different emotional expressions, with and without pauses. Emotions ordered from the longest to the shortest. To the right are shown differences of length when including and not including the pause; emotions are ordered from those with the longest length difference to the ones with the shortest length difference.

The emotions with the longest duration, without pause, are joy, disgust and surprise. Thereafter come sadness, fear and shyness. Shortest are anger and dominance. These results are partly in accordance with the results of Kienast, Paeschke, Sendlmeier (1999), which showed that sadness correlated with slow speech and fear with a speaking rate higher than the average. The greatest length difference between measures with and without pauses are found for shyness and sadness; these emotions then make use of pauses. They consist in part of long pauses between the nominal phrases.

5. Conclusion and discussion

The acoustic aspects of the emotional expressions are summarized in Table 14.

emotion	duration (including pause)	intensity	F0 curve	F0 range	F0 peaks
joy	medium (5 000–6 000 ms)	medium	monotonous	160–250	2 peaks
surprise	medium	strong	big variation	110–360	6 peaks
sadness	long (<6 000 ms)	weak	monotonous	125–275	3 peaks
fear	medium	medium	variation	125–375	4 peaks
shyness	medium	weak	variation	160–275	5 peaks
anger	short (>5 000 ms)	strong	big variation	160–260	6 peaks
dominance	short	strong	variation	95–250	3 peaks
disgust	long	strong	variation	100–330	5 peaks

Table 14. Summary of the acoustic aspects of the eight emotional expressions.

There are similarities between some emotional expressions. Anger and dominance are similar in having short duration and strong intensity. Fear and shyness have medium duration, weak or medium intensity and variation in F0. Sadness and joy are similar by having monotonous F0 and being weak or medium in intensity.

The results of the F0 range can be compared with Piot (1999), who showed that low F0 is related to dominance and that high F0 is related to submissiveness. In our material dominance had the lowest F0 values, while fear, surprise and disgust had the highest F0 values.

Another aspect on how acoustics correlate with emotions is brought up by e.g. Montero, Gutiérrez-Arriola, Colás, Enríquez and Pardo (1999), who showed that, for Spanish, the contribution of prosody to the recognition of the uttered emotion greatly varies, with sadness and surprise being more supra segmental, and joy and cold anger being more marked by segmental voice characteristics.

It is interesting to see which emotions have been confused and which consequently are similar in not only in expression but also in meaning. Anger was in many cases interpreted as dominance, or irritation. Furthermore, fear was interpreted as heartbroken or sad; on the other hand, sad was mostly interpreted as sad and not as afraid. Shy was interpreted as sad or afraid.

The Spanish listeners were good at interpreting sadness. This should, however, be seen in the light of them interpreting many of the emotions as sad. The most odd result is the Spanish listeners' interpretations of joy as sadness. The acoustic characteristics of these two emotions show a monotonous F0 curve.

Finnish and Spanish are the two languages whose listeners most often agree in their interpretations. This is most obvious for the emotions joy, fear, dominance and disgust and to a certain extent also for shyness and surprise. The interpretations of the Finnish and especially the Spanish listeners deviate considerably from the Swedish interpretations. However, there is no evidence to draw the conclusion that Finnish and Spanish listeners express emotions in a way which deviates from Swedish listeners and that emotional expression is not universal. It could be the case that the non-Swedish listeners do not distinguish between what is Swedish emotional prosody and what is Swedish sentence intonation, word accent, temporal characteristics etc. Also, research on a larger number of both speakers and listeners is necessary.

The conclusions are, that the Swedish listeners generally performed better than the listeners of the other language groups. Furthermore, the listeners with different first languages showed varied success in their interpretations of the different emotional expressions. No one language showed a consistently poorer performance for all the emotions, which could be the case if it had been the Swedish grammatical prosody that the listeners did not master.

It is not excluded that certain emotions, e.g. sadness and anger can be understood (and expressed) in similar ways in different languages while other emotions, e.g. joy are, at least in some manifestations of these emotions, more language specific. The acoustic and semantic analysis shows that the emotions, which are similar in expression and in meaning, are those which are confused in interpretation, the only exception being joy and sadness. The acoustic dimensions are analytically independent from each other, but also seem to co-occur, which could be a result of e.g. general excitation of the speech apparatus in certain emotional states. This shows that the connection between the expression of emotion and its content might not be arbitrary, a fact, which is compatible with universality. Our intention is to continue studying the expression and interpretation of Swedish emotional prosody, including more Swedish speakers and listeners from a greater variety of languages, as well as using a larger number of test phrases. Doing so we hope to be able to contribute to the growing area of cross linguistic studies of emotional prosody.

References

- Abelin, Å., 1980. A semantic field for anger, unpublished paper, Department of Linguistics, Göteborg.
- Aubergé, V., Lemaître, L., 2000. The prosody of smile. In: Proceedings of the ISCA Workshop on Speech and Emotion, pp. 122–126.
- Boyd, S., 1980. The semantic field of Swedish friendship terms, PAL 5, Göteborg.
- Chung, S.-J., 1999. Vocal expression and perception of emotion in Korean. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 969–972
- Cornelius, R. R., 2000. Theoretical approaches to emotion, In: Proceedings of the ISCA Workshop on Speech and Emotion, pp. 3–10.
- Davitz & Davitz, 1959. The communication of feelings by content free speech. *Journal of communication* 9, 6–13.
- Dubost, J.-M., Su, T., 1999. Prosodic differences and similarities between Mandarin and French in declarative, interrogative, surprise and doubt expressions. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 1561–1564
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K. R., Tomita, M., Tavaras, A., 1987. Universal and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53 (4), 712–717.
- Fónagy, J., 1967. Hörbare mimik. *Phonetica* 16, 42–51.
- Frijda, N.H., 1970. Emotion and recognition of emotion. In: Arnold, M. (Ed.) *Feelings and emotions*. Academic Press, New York.
- Hayashi, Y., 1999. Recognition of vocal expression of emotions in Japanese: using the interjection eh. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 2355–2358
- Hirsch, R., 1980. A study in Swedish fear vocabulary, PAL 4, Göteborg.
- Izard, C., 1971. *The face of emotion*. Appleton Century Crofts, New York.

- Kienast, M., Paeschke, A., Sendlmeier, W., 1999. Articulatory reduction in emotional speech. In: Proceedings of Eurospeech 1999, pp. 117–120.
- Massaro, D.W., 2000. Multimodal emotion perception: Analogous to speech processes. In: Proceedings of the ISCA Workshop on Speech and Emotion, pp. 114–121.
- Montero, J.M., Gutiérrez-Arriola, J., Colás, J., Enríquez, E. and Pardo, J.M., 1999. Analysis and modeling of emotional speech in Spanish. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 957–960
- Ohala, J. J., 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: Hinton, L., Nichols, J., Ohala, J. J. (eds.) Sound symbolism, Cambridge University Press, Cambridge.
- Osgood, C. E., 1966. The semantic differential technique in the comparative study of culture. *American Anthropologist*, 66, 171–200.
- Paeschke, Kienast, A., M., Sendlmeier W.F., 1999. F0-contours in emotional speech. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 929–932
- Piot, O., 1999. Experimental study of the expression of emotions and attitudes in four languages. In: Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 369–370
- Roseman, I., 1979. Cognitive aspects of emotion and emotional behavior. Paper read at the 87th Annual Convention of the American Psychological Association. Sept. 4. 1979. New York, unpublished manuscript from the Department of Psychology, Yale University.
- Scherer, K. R., Banse. R., Wallbott, H. G. (2001) Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76–92.
- Schlosberg, H., 1954. Three dimensions of emotion. *Psychological Review*, 61, 81–88.
- Tickle, A., 2000. English and Japanese speakers' emotion vocalisation and recognition: A comparison highlighting vowel quality. In: Proceedings of the ISCA Workshop on Speech and Emotion, pp. 104–109.
- Williams, U., & Stevens, K.N., 1972. Emotions and speech: some acoustical correlates. *JASA* 52, 1238–1250.
- Woodworth, R.S., 1938. *Experimental psychology*. Holt, New York.
- Wundt, W., 1896. *Grundriss der psychologie*. W. Engelmann, Tyskland.