ENGLISH AND JAPANESE SPEAKERS' EMOTION VOCALISATION AND RECOGNITION: A COMPARISON HIGHLIGHTING VOWEL QUALITY

Alison Tickle

Department of Speech, University of Newcastle, UK

ABSTRACT

A major question in work on phonetic correlates of emotion is to what extent vocalisation of emotion is due to psycho-biological response mechanisms and is therefore quasi-universal and to what extent it is due to social convention. Cross-cultural research gives an angle in on this question. This paper describes the design and discusses results so far of a study in progress which attempts to shed light on this question and to address some of the very difficult methodological issues in cross-language studies of vocal correlates of emotion. The study compares the cross-cultural decoding accuracy and phonetic correlates of emotion vocalisations encoded by native English and native Japanese speakers. Nonsense utterances and quasi-universally recognised facial expressions of emotions are used. These help deal with translation, ethical problems in data collection, the trade-off between artificiality of data and consistency and the masking of verbal utterances whilst allowing any influence exerted by specific vowel qualities to be highlighted.

1. INTRODUCTION

One means of addressing the question of the extent to which emotional expression is psycho-biologically or culturally determined is to conduct cross-cultural research into emotion encoding and decoding. This paper describes a study in progress comparing the cross-cultural decoding accuracy and phonetic correlates of emotion vocalisations encoded by native English and native Japanese speakers. Preliminary results of the decoding experiment are presented here: auditory and acoustic analysis of the data has not yet been conducted. For ease of comparison, the English terms *happy, sad, angry, fearful* and *calm* are used to represent both English and Japanese throughout this article.

There is very little research into cross-cultural recognition of vocally expressed emotion. This is possibly due to methodological problems of data collection and vocal parameters being more difficult to measure than facial expression which has been more extensively investigated. Frick [1] notes that of the very few cross-cultural studies made, some found that cross-cultural recognition was as accurate as monolingual recognition**,** whilst others found that cross-cultural recognition was adversely affected. For reviews of monolingual studies see [2,3,4].

The bimodal approach of the study described here uses as cues eight quasi-universally recognised photographs of *happy, sad, angry* and *fearful* facial expressions [5,6]. *Calm* in the sense of emotionally unperturbed is also included. Since there are no such photos for *calm*, this is represented by two pieces of music described by subjects as "calm" or "relaxed". The different stimulus for *calm* may perhaps influence its vocalisation. The study combines simulation and self-induction in a way which is not ethically compromising, but which encourages vocalisations which combine authenticity and consistency more than those generated by simulation, natural or induction methods alone.

This study uses nonsense utterances, phonotactically possible in both languages. These permit more cross-language consistency, avoid problems associated with translation and give no verbal cues. Conflicting suggestions regarding possible influence of vowel quality are hinted at in previous studies [7,8]: in the present study nonsense utterances were composed which should allow influence of vowel quality to be highlighted as well as to be compared cross-culturally.

The encoding and decoding procedures used in this investigation are biased towards highlighting phonetic correlates of emotion vocalisation which are recognised by both cultures, thereby indicating potential psycho-biological influence. Vocalisation of emotion in a normal social setting will tend to be controlled by the different social conventions of each group. The encoding procedure attempts to relax this control. However culturally influenced phonetic correlates, which do emerge, will also be analysed. The method does not aim to investigate possible different contexts in which the two cultures are more or less likely to use psycho-biologically determined phonetic correlates. See [9] for a consideration of methodological issues implicit in cross-language studies, suggestions as to how these may be addressed and a fuller explanation of the rationale for and procedure used in this current study.

2. HYPOTHESES BASED ON PREVIOUS RESEARCH

The following hypotheses were made in relation to the decoding experiment.

1. English subjects will decode English emotion vocalisations and Japanese subjects will decode Japanese emotion vocalisations with an accuracy of approximately 50% after accounting for chance, given the restricted number of response alternatives. This would be in line with accuracy rates found in previous monolingual studies reported in [4].

2. English subjects will decode Japanese emotion vocalisations and Japanese subjects will decode English emotion vocalisations with an accuracy of lower than 50%,

due to lack of awareness of cultural influence. However it is hypothesised that this figure will be well above chance, say at approximately 30% after accounting for chance due to the recognition of the effect of common, cross-language psycho-biological response mechanisms upon emotion vocalisations.

3. *Happy, angry* and *sad* will generally be decoded with greater accuracy than *fearful*. This hypothesis is based on results of previous studies, for example, [8,10]. It is also anticipated that *calm* will be decoded with a lower accuracy than *happy, sad* and *anger*, given that it is not classed as a basic emotion and may therefore not be as subject to the influence of psycho-biological response mechanisms.

4. *Happy* will be least accurately decoded when it is encoded on the utterance containing [a] vowel quality compared to [i] or [u]. Bezooijen [8, page 30] suggests that *happy* may be easier to detect on [i] saying that "extra…lip spreading" due, for example, to smiling, "is easier to detect in unrounded vowels." Laver [7] suggests the reverse relationship between vowel quality and perception of lip-rounding, implying that *happy* would be easier to detect when encoded on [u], given the distortion of this vowel which would be created by smiling. This study attempts to investigate this further by controlling for vowel quality on nonsense utterances tested on native speakers of unrelated languages.

5. *Angry* will be most accurately decoded when encoded on [u] and least accurately decoded on [i]. This hypothesis is based on the suggestions made above and on research by Ohala [11] suggesting that the technique of vocal tract lengthening, thereby signalling a larger sound source, is used by certain animals when expressing anger or aggression and since [u] necessitates a lengthening of the vocal tract, it is more likely to be used sound symbolically to suggest aggression or anger.

3. ENCODING EXPERIMENT

Further information on the rationale for the subject profile, the material used and the encoding procedure may be found in [9].

8 Japanese and 8 English female university students aged between 18 and 35 encoded the data. The following is a summary of the procedure and rationale of the encoding experiment used in this study.

(i) A questionnaire on eight quasi-universally recognised facial expression photographs. This aims to:
- Focus attention upon emotions to be considered primarily via a visual stimulus which is common to both language groups, rather than primarily via verbal cues which would have to be translated between groups.
- Elicit suggestions from each subject as to emotions expressed in the photographs.
- Familiarise the subjects with the photos to be used in the recorded interview.

(ii) Practice of nonsense utterances aims to:
- Prepare subjects for the interview section.

(iii) Individual recorded interview - after "warm-ups", including listening to the *calm* music, there is a final game section from which the data is gathered in which the researcher tries to guess the emotion vocalised by the subject. This has the following aims:
- To elicit vocalisations of emotions with the aid of optional stimuli. Optional stimuli include:
  - ◊ Concentrating on the facial expression photos.
  - ◊ Imitating facial expressions on the photos.
  - ◊ Thinking of interjections appropriate to the emotions (suggested by the subject).
  - ◊ Remembering/visualising a situation or experience during which the subject felt or would feel the emotion.
  - ◊ Stimulus of the subject's own choice.
- To disinhibit subjects by playing a game and focusing attention on the researcher who tries to guess which emotions are vocalised by the subject. The researcher also turned her back so that the subjects' attempts at imitating the facial expressions could not be seen - the subject could thereby rely only on vocal cues to communicate the emotion on the nonsense utterances.

(iv) Self-report by subject aims to elicit:
- Which emotions subjects found easier or more difficult to vocalise.
- Which of the optional stimuli subjects found most useful. All subjects found concentrating on the photographs and imitating the facial expressions to be the most useful stimulus. This may have been influenced by facial feedback and/or interpersonal feedback mechanisms: see [12].
- Whether subjects felt any of the emotions encoded.

4. DECODING EXPERIMENT

4.1. Data used in decoding experiment

A reliability test was conducted in which eight raters, four for each language, decoded and rated emotion vocalisations produced by speakers of the same native language as themselves. Data from the three most reliable speakers for each language was used in a forced judgement decoding test. To qualify to be included subjects had to score an average of at least 3 out of 5 across all raters and emotions. The total number of items presented to decoders included 90 items (3 speakers x 2 languages x 5 emotions x 3 vowel qualities). These were preceded by 6 practice utterances. The 90 utterances were randomly ordered and edited out. Each utterance was preceded by a number, in order and repeated three times with a precise two-second interval between each and an eight second interval before the following number.

4.2. Decoding procedure

16 English subjects (12 female and 4 male) and 8 Japanese subjects (4 female and 4 male) performed a forced judgement decoding test on the edited data described above. Judges were offered 5 emotion words in their native language from which to choose a single response. They also gave a confidence rating on a scale of one to three which may be useful in the auditory and acoustic analysis: salience of relevant phonetic correlates may be analysed in relation to confidence ratings.

5. RESULTS

5.1. Comparing emotion vocalisation by English and Japanese subjects

After accounting for chance by Cohen's Kappa, the following percentages were obtained:

- English subjects decoded 60% of emotions English speakers attempted to encode.
- Japanese subjects decoded 42% of emotions English speakers attempted to encode.
- Japanese subjects decoded 36% of emotions Japanese speakers attempted to encode.
- English subjects decoded 35% of emotions Japanese speakers attempted to encode.

Tables 1 and 2 show confusion matrices for English (Table 1) and Japanese (Table 2) judges. For each Table, the emotions portrayed (English and Japanese) are indicated along the first column and the possible decoding responses are shown across the top row. Abbreviations are used representing both English and Japanese emotion words, the English translations of which are *happy* (H), *sad* (S), *angry* (A), *fearful* (F) and *calm* (C).

Where subjects decoded English vocalisations, of the five emotions, both Japanese (Table 1) and English (Table 2) subjects gave more responses indicating the speakers sounded *sad,* then *happy,* then *angry,* then *calm,* then *fearful.*

Where Japanese subjects decoded Japanese vocalisations, of the five emotions (Table 2), there were most responses indicating the speakers sounded *angry,* then *sad,* then *calm* and *happy* with a similar number of responses, then *fearful*. This is a different pattern to that described above for decoding of English emotions.

Where English subjects (Table 1) decoded Japanese vocalisations, of the five emotions, there were most responses indicating the speakers sounded *calm* or *fearful* with a similar number of responses, then *angry,* then *sad,* then *happy*. This is a markedly different pattern from that displayed by Japanese subjects (Table 2) decoding the same set of Japanese data. This difference may be partly due to chance responses if the Japanese data is unconvincing, which may be caused by the reputedly greater stigma attached to emotional expression in Japan.

An alternative explanation for this difference may be that it indicates more cultural influence upon phonetic correlates used in the vocalisation of these emotions in Japanese. However the former explanation seems more likely if we consider that both Japanese and English decoders score approximately the same overall accuracy score, 36% and 35% respectively, when decoding Japanese vocalisations. These are lower accuracy percentages than either group scored when decoding English vocalisations: Japanese decoders scored 42% and English decoders scored 60%.

	H	S	A	F	C	Totals
English H	**62**	2.5	2	1	4.5	72
English S	1.5	**57**	0	4.5	9	72
English A	3	4.5	**56**	1	7.5	72
English F	4	19	5	**35.5**	8.5	72
English C	8.5	27.5	1.5	0.5	**34**	72
Eng. Totals	79	110.5	64.5	42.5	63.5	360
Japanese H	**24.5**	1.5	21.5	8	16.5	72
Japanese S	0	**38.5**	0	18.5	15	72
Japanese A	11	0	**44**	3.5	13.5	72
Japanese F	9	2.5	11	**36.5**	13	72
Japanese C	3	24	0	16.5	**28.5**	72
Jap. Totals	47.5	66.5	76.5	83	86.5	360
Eng. And Jap. Totals	126.5	177	141	125.5	150	720

Table 1: Emotions decoded by English subjects

	H	S	A	F	C	Totals
English H	**50**	6	8	3	5	72
English S	2	**55**	2	6	7	72
English A	7	3	**50**	6	6	72
English F	4	35	11	**17**	5	72
English C	17	29	2	2	**22**	72
Eng. Totals	80	128	73	34	45	360
Japanese H	**39**	3	11	5	14	72
Japanese S	1	**40**	7	7	17	72
Japanese A	12	3	**45**	2	10	72
Japanese F	9	4	23	**29**	7	72
Japanese C	7	31	6	5	**23**	72
Jap. totals	68	81	92	48	71	360
Eng. and Jap. Totals	148	209	165	82	116	720

Table 2: Emotions decoded by Japanese subjects

It will be interesting to check if specific instances of highly accurately decoded emotions encoded by Japanese and English speakers, were encoded with similar acoustic correlates. If this is found to be the case, it would suggest that the vocal effects of possibly quasi-universal psycho-biological response mechanisms may be present.

The fact that Japanese subjects more accurately decoded English than Japanese vocalisations may suggest that despite possibly being exposed to less emotional expression, Japanese subjects are capable of decoding emotions where

acoustic correlates are due to psycho-biological response mechanisms.

Japanese decoders discriminate English emotion vocalisations and English decoders discriminate Japanese vocalisations with 42% and 35% accuracy respectively, after accounting for chance, thus supporting the second hypothesis. This suggests that common acoustic correlates are used by native speakers of English and native speakers of Japanese in the expression of at least some of the emotions some of the time. English vocalisations of *sad, angry* and *happy* are much more accurately decoded than *fearful* or *calm* by both groups. Japanese vocalisations *of sad, angry* and *happy* are also more accurately decoded than *fearful* or *calm* by Japanese decoders. However, although English subjects decoded Japanese vocalisations of *sad* and *angry* with a higher percentage accuracy than Japanese *fearful* or *calm*, of the five emotions vocalised by Japanese speakers, English subjects were least accurate in their decoding of *happy*: see section **5.2.** for details of possible vowel influence. The third hypothesis is therefore only partially supported.

It is anticipated that for each emotion, forthcoming auditory and acoustic analysis may highlight common phonetic correlates used by Japanese and English subjects in the vocalisation of the most accurately decoded items. Any relation between confidence ratings given by decoders for each item and salience of acoustic correlates will also be considered.

This is a study in progress and further statistical analysis is to be conducted on the significance of differences in decoding of emotions according to encoder language, decoder language, vowel quality, specific emotion and decoder gender.

5.2. Vowel Quality

Tables 3, 4, 5 and 6 show confusion matrices with a similar format to Tables 1 and 2, except that these tables also indicate, along the first column, the vowel quality within the vocalisation. Cohen's Kappa has been calculated where indicated in the tables to allow for chance.

English subjects most accurately decoded emotions encoded by English subjects on [i], then [a] then [u] (Table 3). Here, *fearful* appears especially difficult to decode if encoded on [u], when it is much more often decoded as *sad*. There is also less accurate decoding of *sad* on [u] where it is more often confused with *calm* than with other emotions. *Calm* is decoded as *sad* on [a] and [u]. However on [i] when *calm* is confused with another emotion it is most often confused with *happy*. Subjects slightly more accurately decode *happy* on [i] supporting Bezooijen's suggestion: see the fifth hypothesis. Interestingly, subjects are also most likely to categorise vocalisations overall as *happy* when they are encoded on [i]. They are least likely to categorise vocalisations overall as *sad* on [i] than on [a] or [u]. *Angry* is least accurately decoded on [i] and vocalisations are least often categorised as *angry* on [i]: this perhaps lends support to the fifth hypothesis.

E. of E.	H	S	A	F	C	Total	Kappa
Happy a	**20**	1	1	0.5	1.5	24	
Sad a	0	**20.5**	0	1	2.5	24	
Angry a	0.5	1.5	**19.5**	0.5	2	24	
Fearful a	0	4.5	0.5	**14.5**	4.5	24	
Calm a	0.5	12.5	0.5	0	**10.5**	24	
Total a	21	40	21.5	16.5	21	**120**	0.6354
Happy i	**22.5**	0	0	0	1.5	24	
Sad i	0.5	**20.5**	0	2.5	0.5	24	
Angry i	0	2	**16.5**	0.5	5	24	
Fearful i	2	2	2	**16.5**	1.5	24	
Calm i	6.5	4.5	0.5	0.5	**12**	24	
Total i	31.5	29	19	20	20.5	120	0.6667
Happy u	**19.5**	1.5	1	0.5	1.5	24	
Sad u	1	**16**	0	1	6	24	
Angry u	2.5	1	**20**	0	0.5	24	
Fearful u	2	12.5	2.5	**4.5**	2.5	24	
Calm u	1.5	10.5	0.5	0	**11.5**	24	
Total u	26.5	41.5	24	6	22	120	0.4948
Total aiu	79	110.5	64.5	42.5	63.5	360	0.5990

Table 3: English decoders' discrimination of emotions encoded by English speakers showing vowel quality

E. of J.	H	S	A	F	C	Total	Kappa
Happy a	**7**	0	11.5	1.5	4	24	
Sad a	0	**12**	0	7.5	4.5	24	
Angry a	2.5	0	**15**	1	5.5	24	
Fearful a	1.5	1	7	**12**	2.5	24	
Calm a	2	6.5	0	5.5	**10**	24	
Total a	13	19.5	33.5	27.5	26.5	120	0.3333
Happy i	**15**	0	5	1	3	24	
Sad i	0	**11.5**	0	6	6.5	24	
Angry i	7	0	**12.5**	1.5	3	24	
Fearful i	4	1	3	**11.5**	4.5	24	
Calm i	1	4	0	7.5	**11.5**	24	
Total i	27	16.5	20.5	27.5	28.5	120	0.3958
Happy u	**2.5**	1.5	5	5.5	9.5	24	
Sad u	0	**15**	0	5	4	24	
Angry u	1.5	0	**16.5**	1	5	24	
Fearful u	3.5	0.5	1	**13**	6	24	
Calm u	0	13.5	0	3.5	**7**	24	
Total u	7.5	30.5	22.5	28	31.5	120	0.3125
Total aiu	47.5	66.5	76.5	83	86.5	360	0.3472

Table 4: English decoders' discrimination of emotions encoded by Japanese speakers, showing vowel quality.

As in Table 3, English subjects (Table 4) also scored highest on decoding Japanese vocalisations of emotions on [i], then [a], then [u]. *Calm* is more often confused with *sad* on [u] than on [i] or [a]. On [a] *calm* is more often confused with *fearful*. Subjects much more accurately decode *happy* on [i] than on [u] or [a], again supporting Bezooijen's suggestion. In addition, where subjects categorise a vocalisation as *happy*, this is most likely to be on [i]. They are more likely to decode an emotion as *sad* on [u] than on [i]. *Happy* is most often confused with *angry* on [a] and [i] but with *calm* on [u], which may be relevant to the fifth hypothesis. However they were most likely to decode vocalisations overall as

angry on [a], not on [u]. *Calm* was most often confused with *sad*, especially on [u], or *fearful*, especially on [i] being more often decoded as *sad* than *calm* on [u].

J. of E.	H	S	A	F	C	Totals	Kappa
Happy a	**16**	1	3	2	2	24	
Sad a	1	**14**	2	4	3	24	
Angry a	0	1	**17**	3	3	24	
Fearful a	1	15	2	**5**	1	24	
Calm a	1	11	1	0	**11**	24	
Total a	19	42	25	14	20	120	0.4063
Happy i	**19**	0	3	0	2	24	
Sad i	0	**23**	0	1	0	24	
Angry i	2	2	**17**	2	1	24	
Fearful i	1	10	4	**8**	1	24	
Calm i	8	8	0	1	**7**	24	
Total i	30	43	24	12	11	120	0.5208
Happy u	**15**	5	2	1	1	24	
Sad u	1	**18**	0	1	4	24	
Angry u	5	0	**16**	1	2	24	
Fearful u	2	10	5	**4**	3	24	
Calm u	8	10	1	1	**4**	24	
Total u	31	43	24	8	14	120	0.3438
Total aiu	80	128	73	34	45	360	0.4236

Table 5: Japanese decoders' discrimination of emotions encoded by English speakers, showing vowel quality

As with English subjects' decoding of English vocalisations (Table 3), Japanese subjects scored lower on decoding English vocalisations using [u] than either [i] or [a] (Table 5). *Fearful* is more often decoded as *sad*, as was the case for English decoding subjects. However, in the case of Japanese decoders, this confusion occurs for all vowels, not just for [u]. Despite the lower overall decoding score of English vocalisations by Japanese subjects - they scored 42% accuracy compared to English decoders' score of 60% - Japanese subjects scored slightly higher than English subjects on their decoding of English *sad* on [i], scoring almost 100%. English subjects were less likely to categorise a vocalisation as *sad* where it was expressed on [i] than on [a] or [u] whilst vowel quality did not appear to influence the Japanese decoders in this way. Where Japanese decoders confused *calm* with another emotion, it was generally confused with *sad* on [a] and *sad* or *happy* on [i] or [u]. Like English subjects they more accurately decoded English *happy* on [i] than on [a] or [u]. However unlike English subjects, accuracy in decoding angry seems uninfluenced by vowel quality.

When decoding Japanese emotion vocalisations (Table 6), Japanese subjects did not follow some of the patterns which seemed to be forming from analysis of Tables 3, 4 and 5. They decoded Japanese vocalisations with greatest accuracy overall when they were encoded on [a], then [u], then [i]. In particular, they were much more accurate at decoding *happy* on [a]. However they were less likely to decode Japanese vocalisations overall as *happy* when they were expressed on [u] and were more likely to categorise any Japanese emotion as *sad* when they were expressed on [u]: this is a similar pattern as was found for English subjects' decoding of Japanese *happy* and *sad*. Like the English decoders of Japanese vocalisations, calm was most often confused with *sad* on [u].

J. of J.	H	S	A	F	C	Totals	Kappa
Happy a	**20**	0	1	0	3	24	
Sad a	1	**10**	2	3	8	24	
Angry a	4	1	**15**	1	3	24	
Fearful a	3	2	5	**12**	2	24	
Calm a	4	8	3	1	**8**	24	0.4271
Total a	32	21	26	17	24	120	
Happy i	**11**	2	4	3	4	24	
Sad i	0	**15**	3	2	4	24	
Angry i	6	2	**12**	1	3	24	
Fearful i	3	1	10	**8**	2	24	
Calm i	3	7	3	3	**8**	24	0.3125
Total i	23	27	32	17	21	120	
Happy u	**8**	1	6	2	7	24	
Sad u	0	**15**	2	2	5	24	
Angry u	2	0	**18**	0	4	24	
Fearful u	3	1	8	**9**	3	24	
Calm u	0	16	0	1	**7**	24	0.3438
Total u	13	33	34	14	26	120	
Total aiu	68	81	92	48	71	360	0.3611

Table 6: Japanese decoders' discrimination of emotions encoded by Japanese speakers, showing vowel quality.

5.3. Sex of Encoder/Decoder

Encoder and decoder sex may also have influenced results. Emotions were encoded by female speakers only and decoded by male and female subjects for each language - the reasons for this are explained in [9]. There is no reliable evidence in previous research to suggest that females are generally better at decoding emotion than males but initial bald figures suggest this may be the case in this study. Further analysis of the data will also reveal whether sex is an influence on ability to decode any of the emotions under consideration here.

There is also the possibility that emotions are more accurately decoded where they have been encoded by a speaker of the same sex as the decoder. If this is the case, male decoders would have been at a disadvantage in this experiment and therefore the Japanese decoder sample as a whole would have been disadvantaged given the balance of males and females compared to the English decoder sample. This possibility cannot be tested for in this experiment given the female only encoded data available. There is no space to expand upon this factor here, but further details will be provided in a future article.

6. SUMMARY

English subjects decoded English emotion vocalisations with a 10% higher accuracy than that anticipated by the first hypothesis. This is after accounting for chance, given the number of emotions to choose from. This higher percentage may not be statistically significant but otherwise may be due to the effectiveness of the encoding experiment. The greater

accuracy may also be due to the fact that of the emotions studied, three of them tend to be highly recognisable.

Japanese subjects decoded Japanese vocalisations with a lower accuracy percentage (36%) than that suggested by the first hypothesis (50%). If social stigma attached to emotion vocalisation results in less experience of emotion encoding and decoding in Japanese culture, this may influence native Japanese subjects' ability to encode and decode data. As mentioned, this lower percentage may also be partly due to the influence of encoder and decoder sex.

Despite the overall greater accuracy of English subjects in decoding emotion, they were particularly inaccurate at recognising Japanese *happy* encoded on [a] and [u] (9.5/48) compared to Japanese decoders (28/48). This suggests that there may be more culturally conditioned influence on the expression of Japanese *happy* than on *sad* and *angry* which are more generally decoded with greater accuracy. This question will be considered further in the forthcoming acoustic analysis.

Supporting the second hypothesis, English subjects decoded Japanese vocalisations and Japanese subjects decoded English vocalisations with an accuracy of between 30% and 50%. This suggests that the vocal effects of possibly quasi-universal psycho-biological response mechanisms may be signalling distinctions between the emotions under consideration here. This is further supported by the fact that Japanese subjects decoded English vocalisations with greater accuracy than they decoded Japanese vocalisations. Even though they may be less used to encoding and decoding vocal expressions of emotion, Japanese subjects still recognise emotions expressed by English subjects with an accuracy close to that found for same language encoders and decoders in previous studies (50%).

The third hypothesis that happiness, *sad* and *angry* would be decoded with greater accuracy than *fear* or *calm* is supported except for English decoding of Japanese vocalisations, where of the five emotions expressed, English subjects are least accurate at decoding *happy*. This suggests possible cultural influence upon Japanese encoding of *happy*. However, whilst there may be cultural influence upon the acoustic correlates signalling *happy* in Japanese, Japanese subjects can still decode English *happy* with a high degree of accuracy, scoring 50 out of a possible 72. This suggests the possible influence of psycho-biological response mechanisms upon English *happy* vocalisations which Japanese subjects are able to decode despite possibly not using these acoustic signals themselves. Acoustic analysis of this data may shed more light upon this suggestion.

The fourth hypothesis that *happy* would be least accurately decoded on [a] is not supported. In this data, *happy* is generally least accurately decoded on [u]. Overall recognition figures are [u] 44/96, [a] 63/96 and [i] 67/96. With the exception of Japanese subjects decoding Japanese *happy* (where *happy* is more easily recognised on [a]), *happy* vocalisations are most accurately decoded on [i]. This tends to lend support to Bezooijen's suggestion that "extra…lip spreading is easier to detect in unrounded vowels" [8, page 30]. Possibly there was often not enough perceived distortion of the [u] vowel for it to be heard as signalling a smile.

English subjects decoded both Japanese and English *angry* and Japanese subjects decoded Japanese *angry* most accurately when it was encoded on [u] and least accurately on [i]. This tends to support the fifth hypothesis. However Japanese decoding of English *angry* does not follow this pattern since discrimination appears to be more independent of vowel quality.

Further statistical analysis and auditory and acoustic analysis currently being conducted should lead to a more in-depth understanding of what this data may reveal.

REFERENCES

[1]Frick, R.W. 1985. Communicating Emotion: The Role of Prosodic Features. *Psychological Bulletin*, 97, 3, 412-429.

[2]Scherer, K.R. 1986. "Vocal Affect Expression: A Review and a Model for Future Research", *Psychological Bulletin,* 99, 2, 143-165.

[3]Murray, I. and Arnott, J. 1993. Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion, *J.Acoustic.Soc.Am.,* 93, 2, 1097-1108.

[4]Banse, R. & Scherer, K.R. 1996. Acoustic Profiles in Vocal Emotion Expression, *Journal of Personality and Social Psychology*, 70, 3, 614-636.

[5]Ekman, P., Friesen, W.V., and Ellsworth, P. 1972. *Emotion in the human face.* Pergamom Press. New York.

[6]Matsumoto, D. 1996 *Unmasking Japan.* Stanford University Press, Stanford, California.

[7]Laver, J. 1981. *Users' manual for vocal profile analysis protocol: A perceptual guide.* Unpublished paper, University of Edinburgh.

[8]Bezooijen, R. 1984. *Characteristics and Recognizability of Vocal Expressions of Emotion.* Foris Publications, Dordrecht Holland/Cinnaminson USA.

[9]Tickle, A.A. 1999 Cross-language vocalisation of emotion: methodological issues, *ICPhS 99 Proceedings*, 305-308, ICPhS 99 San Francisco.

[10]Scherer, K.R., Banse, R., Walbott, H.G., & Goldbeck, T. Vocal Cues in emotion encoding and decoding, *Motivation and Emotion*, 15, 123-148.

[11]Ohala, J.J. 1984. An Ethological Perspective on Common Cross-Language Utilization of Fo of Voice. *Phonetica* 41: 1-16.

[12]Capella, J.N. 1993. The facial feedback hypothesis in human interaction: Review and speculation. *Journal of Language and Social Psychology*, 12, 13-29.