# Linking Form to Meaning: The Expression and Recognition of Emotions Through Prosody

*Li-chiung Yang* *Nick Campbell*

CREST, Japan Science and Technology & Information Sciences Division, ATR
http://www.isd.atr.co.jp/esp
{yang,nick}@isd.atr.co.jp

## Abstract

Emotion is an integral component of human speech, and prosody uniquely represents the expressive meaning that is fundamental to communication. In this study, we demonstrate how the subtle and finely differentiated meanings permeating spontaneous speech are communicated by prosodic variations and show that it is the differences in shape that communicate the degree of uncertainty or certainty with respect to the speaker's knowledge state, specific emotional states, the intensity of emotion, and the effects of other co-occurring emotions.

## 1. Introduction

Recent research has pointed to the centrality of emotion in all aspects of human activity [1, 2]. In spoken language, emotion is indispensable in the ongoing communication of ideas, feelings, and judgements of importance towards topic and towards participants, and guides the exchange of information through mutual coordination of states of uncertainty and knowledge. Prosody is of paramount importance in this process in providing a forceful and flexible additional dimension that distinguishes and communicates often complex and finely differentiated layers of meanings, and is fundamental to progress in increasing intelligence and responsiveness in interactive systems and natural-sounding speech synthesis.

## 2. Speech corpus

### 2.1. Data and approach

In this study we investigate the expression and recognition of emotion through prosody by acoustically and perceptually analyzing natural interactive discourse data. Our approach differs from previous research in that we take an integrated approach of combining acoustic data from spontaneous conversation and experimental data from perceptual tests, with the goal of providing a more unified account of prosodic realizations of emotions in speech. The corpus consists of 6 hours of recorded conversation in Mandarin Chinese as well as 21 speech segments extracted from the corpus for the perceptual experiment.

### 2.2. Why use spontaneous speech?

We believe that it is crucial to study emotion using spontaneous speech because it is only in such speech that we will encounter the complex emotions occurring in real life [3]. This complexity arises because of the high degree of involvement in what is being communicated and the goal-directed motivations of the participants. In spontaneous speech, the high degree of interactive involvement is expressed in a rapidly varying stream of complex emotional states, in contrast to more constrained expressions of emotions present in controlled speech. These emotions can be highly varied and span a much wider range than typically recognized. In addition, the emotions expressed in spontaneous speech can be very subtle and finely differentiated, and can occur to varying degrees of intensity as well as combined together because of the large number of contextual variables simultaneously at work. To achieve human-like quality in intelligent interactive systems, it is crucial to understand and model how humans act in normal communicative situations.

## 3. Prosodic shapes and expressive meaning

Specific emotional and cognitive states such as disbelief, doubt, complaining, incomprehension, and puzzlement, can contribute greatly to intonation, and have systematic influences on the shapes of intonation. Our data show that in general, states such as continuation, expectation, hesitation, and uncertainty have a raising and lengthening effect, while states such as definiteness, finality, and negativity have a lowering effect on pitch [4]. In addition, the degree of tentativeness or definiteness of an expression is often correlated with the steepness of pitch slope. Another important consideration is that emotions occur to different degrees of intensity, intensity being some measure of physiological change, and the intensity may determine the magnitude of intonational influence in an utterance, and should be incorporated in the determination of realized prosodic forms.

### 3.1. Definiteness, tentativeness, and the level of intensity

Data from our corpus show that variations in pitch shape and the degree of intensity of an expression can give rise to the distinct quality and meaning of sounds. This is shown in the following extracted tokens of *dui* "right", a frequently used discourse marker of agreement in conversation, of one speaker from our corpus, shown in Figures 1 and 2.

Analysis of the data shows that the degree of emotional intensity is commonly signaled by pitch range and pitch height variation. In the consecutive responses w-dui8, w-dui9, and w-dui10, the speaker is getting progressively more involved and her *dui*s follow a corresponding progressively higher pitch pattern, from the gentle agreement of w-dui8 to the extreme level of emotional involvement and exaggerated emphasis of w-dui10. The intensity variation is also systematically indicated by the uniform stepwise increments in pitch level. Similarly, both w-dui22 and w-dui23 are intense expressions, but w-dui23 has a larger pitch range, exemplifying the speaker's higher degree of emotional intensity. In contrast, w-dui24, an immediate follow-up confirmation of w-dui23, has a convex shape and a moderate pitch level and pitch range because of the speaker's more normalized state.

Pitch shape characteristics such as pitch slope and concavity

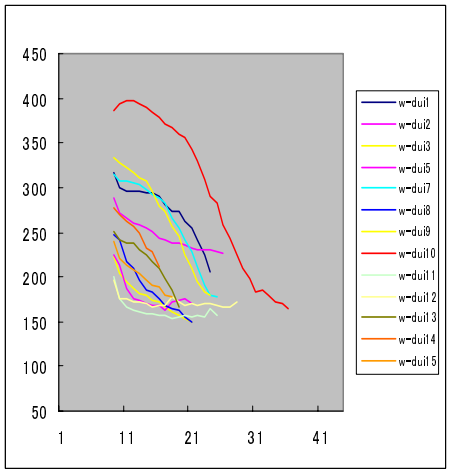



*Figure 1*: Variations of emotional intensity in expressions of agreement of *dui* .

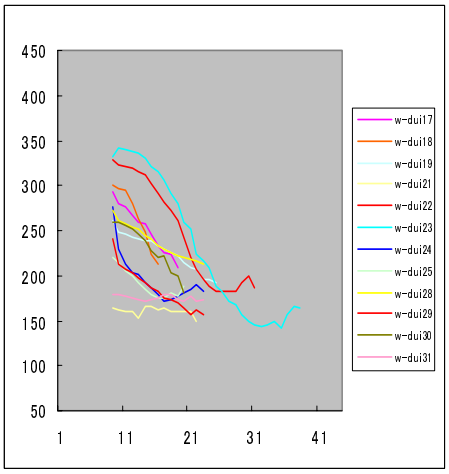



*Figure 2:* A comparable pattern of *dui*s in a subsequent series.

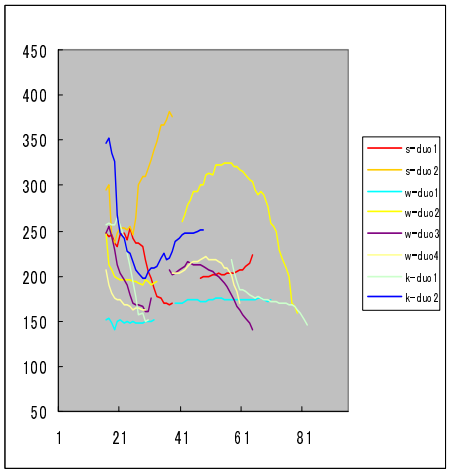



*Figure 3*: Different emotions cause *hen duo* to take on different pitch realizations.

and convexity are very important features in distinguishing intonational meaning, with more moderate slopes associated with more tentativeness, and sharper falling slopes with greater definiteness. The concavity or convexity of slope is also critically related to the perceived degree of harshness or softness of the utterance, and these shape characteristics reflect the underlying expressive states that often arise from the discourse process itself. For example, w-dui13, w-dui14, and w-dui15 are shorter and have sharper slopes than w-dui1 and w-dui2. In this sequence, the speaker first started to express her opinion in w-dui13, but was interrupted, and in w-dui14 she restarts, so her pitch level is higher. In w-dui15, the speaker was just providing further confirmation after an explanation, so the *dui* here is low-pitched with a convex shape, corresponding to the more gentle agreement. W-dui17, w-dui18, and w-dui19 show a similar pattern of intensity variation, with a higher-pitched sharper slope *dui* perceived as more definite.

While most of the high pitched instances of *dui* in these two figures are intense and have a concave shape, the remaining *dui*s in the mid and low pitch ranges exhibit mostly convex shapes. For example, w-dui2, w-dui3, w-dui5, and w-dui15 have a gradual pitch slope with a convex shape, giving an impression of gentle agreement. W-dui11, w-dui12 and w-dui31 are at the other extreme from harshness and definiteness and have the flattest slopes of all the instances. The neutral quality of these expressions is represented in the insignificant pitch range, low amplitude and pitch level, as well as the mild shape of these *duis.*

### 3.2. Emphasis co-occurring with different emotional states

The nature and intensity of the underlying emotional-cognitive state and the degree of emphasis all contribute to the realized pitch shapes. This is shown in Figure 3, where different tokens of the phrase *hen duo* "very many" of 3 speakers from different contexts were plotted. As seen, in w-duo1, both *hen* 'very' and *duo* `many' have a level shape, but they differ in pitch level and duration. The speaker's prominent focus on *duo* is signalled by both the lengthened duration and by the sustained sound quality.

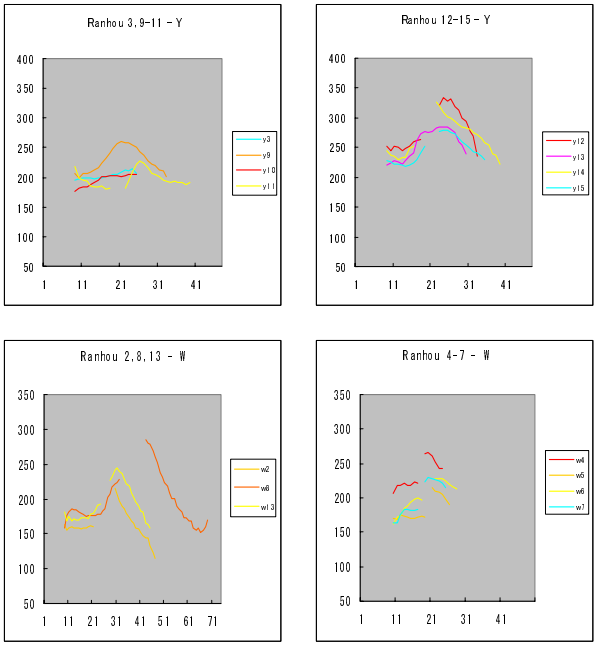
The prosodic realization of a dramatic exaggerated expression can differ from a more informational type of emphasis. Instead of the level shape seen in w-duo1, the *duo* in w-duo2 is greatly modified and has a dramatic rise-fall arch shape, due to the exaggerated and persuading emphasis of the speaker. Comparison of w-duo2 and w-duo3 further shows the effect of different emotional intensities and the progression of focus on prosody. In w-duo3, the speaker's emotion is more settled, both the pitch level and magnitude of pitch movement have correspondingly decreased, and the rising slope is smaller, representing the diminishing force. This reinforces the view that the steepness of the pitch slope is very critical, and is correlated with the strength of emotion. W-duo4 echoes the shape of w-duo2 but with a lesser magnitude.

Pitch shapes can take a drastically different pitch direction in the presence of a strong intonational force as shown in k-duo1. In this utterance, the speaker is emphasizing *hen duo* with a negative emotion, expressing her disapproval of `too many' and the *duo* here has a distinct and complex falling shape. The perceptible pause of .19 sec between *hen* and *duo* also enhances considerably the expression of disapproval present. As compared to k-duo1, the *hen* in k-duo2 drops sharply, with a much higher pitch level and a large pitch range, signalling the greater focus. The light accusing tone here gives *duo* a rising shape. An even more striking rising *duo* is seen in s-duo2, where strong emphasis on *duo* and high involvement combine to give the sharp rising pitch slope within a very short time frame, indicating the speaker's urgent and intense state. In contrast, s-duo1, with light prompting intention, is at a mid pitch level with a slight rising shape.

The examples presented show that focus and emphasis can perturb the pitch relationship and play an important role in the overall prosodic system; however, the specific realization or manifestation of the pitch relationship will depend upon the particular emotional relationship as well. It is the specific underlying expressive nature of a particular focus which ultimately determines the actual realized pitch shape.

### 3.3. Certainty and uncertainty

Prosodic signals of definiteness and tentativeness are pervasive in discourse and critical to the development of a conversation. The degree of a speaker's certainty or uncertainty on the content of the conversation is an important element in the participants' success in communicating their ideas. Uncertainty and tentativeness also occur with hesitation, where uncertainty and implied continuation cause a rising and lengthening effect on




*Figures 4-7*: (in top left-right, and lower left-right order) Shape variations of the marker *ranhou* 'then' of 2 speakers showing uncertainty, certainty, and emphasis.

pitch shape. Figures 4-7 show how remarkably the expression of finely differentiated intentions and meanings present in spontaneous conversation are realized in the prosodic form.

We compare the different forms of the discourse marker-connective *ranhou* 'then' below. In y-ranhou3 in Figure 4, the speaker is hesitating, and the leveling and lengthening effect of hesitation is expressed in the flattened slight rising pitch contour, reflecting the uncertainty and continuation. Unlike the narrow pitch range of y-3, y-9 has a well-defined rise-fall pitch pattern with a large pitch range, making it perceptually more definite, and this prominence acts to signal the break from the immediately preceding topic and carries more emphasis. Conversely, the following y-10 introduces a phrase that is a natural continuation of the preceding phrase as the speaker tries to recall information. By contrast to the strong rise-fall contour of y-9, the continuation and uncertainty in this case cause *ranhou* to take on a rise-level contour, reflecting the temporary holding state. By the following phrase, the speaker has successfully retrieved the relevant piece of information, and is more confident of what to say, and this is reflected in the downward pitch slopes in y-11, although the level ending suggests that some uncertainty is still present.

Uncertainty is inherent in discourse interactions, and gives rise to many discourse strategies for resolving uncertainty. Figure 5 presents paired instances of *ranhou* in interactive floor negotiation reactions, with salient attention markers and subsequent lower pitched repetition in each pair. In each case, the main speaker immediately reacts to an interruption by repeating *ranhou* two times with a high pitch level, loud amplitude, longer duration and expanded pitch range, with a systematic pitch lowering of about 50Hz between the first and second instances, representing the normalization from the initial immediate reaction.

Figure 6 presents an interesting mix of co-occurring emphasis and different degrees of certainty and uncertainty. The

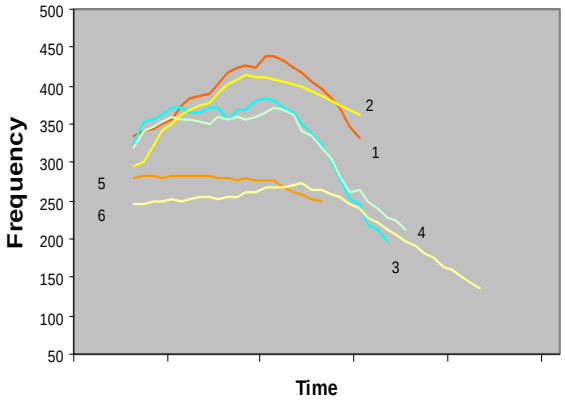



*Figure 8*: Different underlying meanings correlate with shape.

3 *ranhous* of w2, w8, and w13 are all very extended with striking shapes highlighting the contrastive emphasis that the speaker is making, and the pitch variations are correspondingly much greater. The duration and pitch changes in these cases exhibit very prominent shapes that emphasize the contrastive nature of the phrases: in w13, the prominent shaped contrastive emphasis is used to signal a turn in topic, and in both w2 and w8 the prominent form emphasizes a point of contrast, while in w8, the speaker is simultaneously taking time to reflect, and the strong emphasis is combined with hesitation, resulting in the exceptionally long duration and ending rise in pitch. By contrast, the *ranhou*s in w4-7 are all short in duration, moderate in pitch height, and generally have a well-defined rise-fall shape, reflecting the very orderly topic succession in this section.

### 3.4. Surprise to matter-of-fact

Surprise is closely related to tentativeness, uncertainty and doubt, and is typically characterized by a rise-fall shape. The cognitive conflict between the pre-existing belief and the newly encountered awareness that occurs in surprise reflects a strong element of doubt and uncertainty. But surprise also contains an acceptance and belief that the new knowledge is true. Prosodically, the doubt and uncertainty is manifested in an initial rising pitch shape, and the acceptance gives the contour a declining pitch, reflecting the certainty of realization. The doubt and uncertainty together with the ultimate acceptance may give surprise its characteristic rise-fall pitch contour.

A change in degree of emotional intensity may also change the nature of the emotion itself, such as from the emotion of surprise to matter-of-fact acknowledgement. Figure 8 plots several pitch manifestations of the expression *zhende* "really", which is often used to express surprise. Both the high patterns of extreme surprise in k-zhende1 and s-zhende2 and the more moderate surprise in s-zhende3-4 exhibit the rise-fall slopes that exemplify the pattern of uncertainty and acceptance that is characteristic of surprise. By contrast, s-zhende5 and k-zhende6 are similar in having lower pitch level and flatter slopes, and both express light or matter-of-fact acceptance or acknowledgment, the result of an already completed normalization process subsequent to strong surprise in each case. The paired similarities seen in Figure 8 correlate with the different levels of emotion expressed, and this may suggest that differences in shape and level provide a systematic categorization of intonational meanings either locally or globally based.

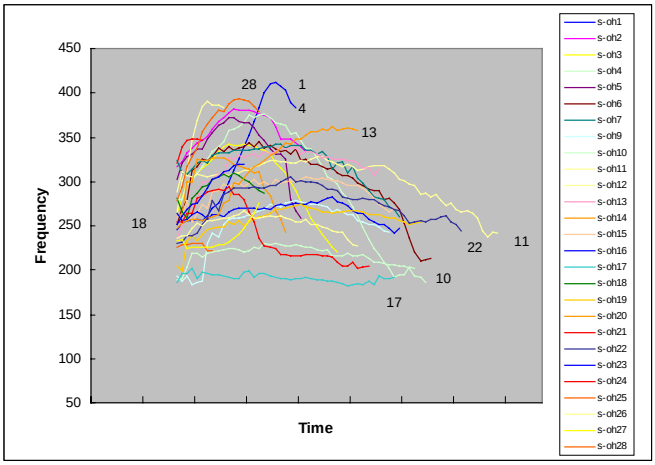



*Figure 9*: Expressive meaning of 28 instances of the interjection *oh* of one speaker (speaker S).

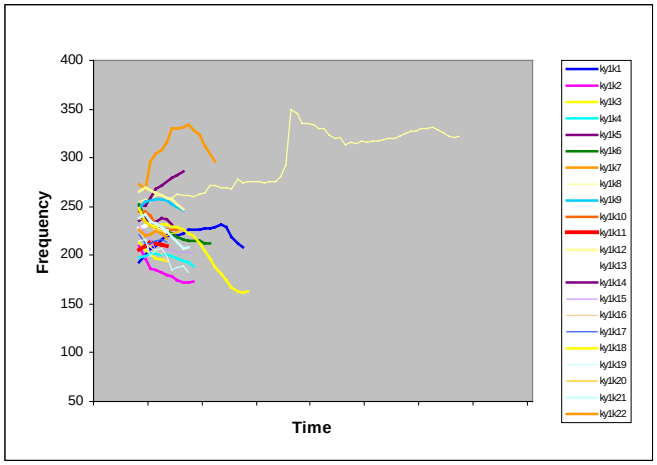



*Figure 10:* 22 instances of the expressive use of the interjection *oh* of another speaker (speaker K).

### 3.5. Dimensions of meaning: Surprise to dawning realization to acknowledgement

Intonation expresses fine gradations in meaning even when lexical information is largely absent, as in the case of the particle *oh* *'o'* and its variant *ah* *'a'*, which, like *zhende*, communicate a range of uncertainty-based states, including doubt, surprise, acceptance, acknowledgement, and registering of information. Three basic patterns for *oh* are evident in the plot containing 28 instances of *oh* of one speaker in Figure 9. Like *zhende*, *oh* often expresses surprise in a rise-fall shape, with an arched and extended concave pattern communicating different intensities of dawning realization. It is the differences in shape, height, and duration that communicate the degree of uncertainty or certainty with respect to the speaker's knowledge state, the intensity of emotion, and the effects of other co-occurring emotions.

Our data show that intense surprise causes a high rise in pitch. In contrast to the mild gradual arch shapes of dawning realization seen in s-oh13, s-oh15, and s-oh19, a sharper and narrower arch shape indicates the presence of surprise with co-occurring emotions, as in the high amazement of s-oh4 and the horror expressed in s-oh3. A lower pitch range often reflects acceptance and registering of information, with a lesser degree of surprise, as in s-oh11, s-oh13, and s-oh22, and a matter-of-fact acceptance of information that offers little challenge to the speaker's knowledge state causes the pattern of nearly flat pitch slopes in s-oh10 and s-oh17. Emotions that are closely related to acceptance, such as sympathy and approval also tend to be expressed in a low pitch level.

S-oh1 and s-oh18 are at the other extreme of uncertainty, with rapid rises in pitch within a short time-frame exemplifying incomprehension, alertness, and a need for further information, in contrast to the completely realized acceptance of information accompanying more extended duration pitch shapes. The uncertainty in s-oh1 in particular stands out because of the convex steep rise to at a high pitch level, with nearly no subsequent fall, reinforcing the final incomprehension, while the moderate pitch and gradual rise of s-oh27 expresses the speaker's doubt and heightened interest. By contrast, concavity of pitch shape is associated with greater comprehension, as in s-12, s-24 and s-28 which express surprise, interest and quick recognition upon encountering unexpected new information.

The functions of an interjection are closely related to the nature of discourse and individual speaking style. In contrast to the highly varying expressive *oh*s of the previous speaker, the other speaker, speaker K's *oh*s in Figure 10 seem much more subdued and are much shorter on average as they function mostly as quick acknowledgement or quick recollection responses. Due to the particular nature of the discourse, there are many instances of reorientation, recalling, acknowledging, and sudden occurrence of ideas by the speaker, and this is why these particular shapes dominate, in contrast to the varied reactions to new information experienced by speaker S.

It is worth noting that even within all these short expressions of *oh*s, there still exist finer variations in shape, height, range, direction, duration and intensity, and these variations are systematic and are related to interpolation, status of information, and emotional state. For example, there is the short falling type as in 3 and 16, slightly longer falling type as in 2, 6, 10 and 20, the curvy twist type as in 14, 17 and 22, the arch type as in 4, 9, and 11, and the more intense emotional type, characterized by wider pitch movement and longer duration, as seen in 1, 5, 7, and 18, for expressing the speaker's disgust, sudden remembrance of an important or exciting event, and appreciative acknowledgment, respectively. How expressive a speaker is in a particular conversation also depends on the degree of topic relevance. This is exemplified in the striking extended *oh* of 8 expressing the speaker's mock terror and protesting emotion on encountering an unexpected event.

## 4. The recognition of emotions through prosody

### 4.1. Experimental evidence

*Materials and Procedure.* To help us understand more fully the complexity of emotions inherent in spontaneous speech, and how well these subtle and highly varied emotional expressions could be identified by ordinary people in a more constrained context, without prior knowledge of the discourse or participant speaking style, we carried out a small perceptual experiment in which 21 excerpts containing a variety of different emotions were taken from the original corpus, randomized and played back 3 times with a 2-second pause to 13 phonetically untrained native speakers of Chinese (5 females and 8 males) and 5 Americans (1 female and 4 males), with no knowledge of Chinese, for identification of the speaker's emotion, following a multiple forced-choice.

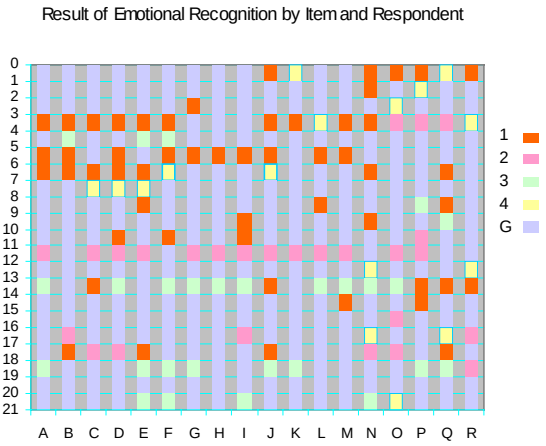



*Figure 11*: Individual responses in identifying emotions, with colors indicating the specific choice (Chinese: A-M, Americans: N-R).

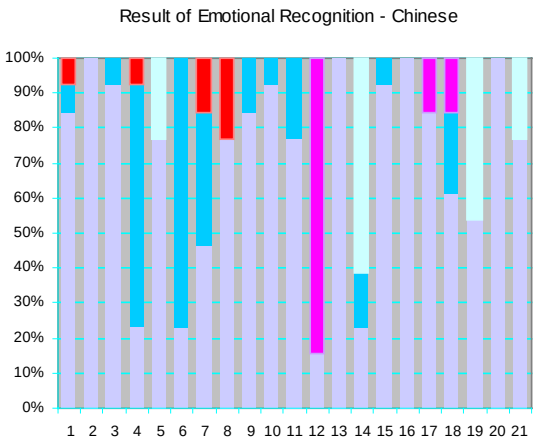



*Figure 12*: Percentage of correctly identified states and alternative answers by color and item (Chinese only), G= correct choice.

### 4.2. Results and discussion

*General Discussion.* The results of the experiment are presented in Figures 11-12. The results demonstrate that there are clearly identifiable patterns among observers in the recognition of cognitive-emotional states, suggesting that cognitive-emotional states are systematically organized and expressed. This conclusion is supported by both the cases on which there is complete unanimity and by those cases where there are some seeming discrepancies, but in which there are clear internal consistencies. Overall, 70.62% (193 of 273) of the responses were identified consistent with the analysis. In 4 cases, there was 100% agreement among all observers. In two-thirds of the cases, 14 of 21, over 70% of the respondents agreed in their identification of speaker states. In only 28.6% of the cases, 6 of 21, were agreement rates less than 50%. There are also 4 problem cases, where agreement with our analysis is very low, although there is some internal consistency among the respondents. If we exclude those 4 cases, the overall agreement for the rest of the 17 cases correlates with our analysis 82.3% of the time, which is very much in agreement with previous research on emotion recognition [5].

Our results from both the Chinese and the American data also indicate that there is a basic sound-meaning correspondence. For example, strong disgust and dismissive emphasis, as well as the degree of intensity, were clearly differentiated by both the Chinese and American respondents, and both had a single unambiguous interpretation with unanimous agreement on the supportive and sympathetic state. The fact that the American respondents do not have any knowledge of Chinese and were unaware of the semantic content, yet were still able to differentiate the speaker state purely based on the sound pattern further supports the idea of a sound-meaning link.

It is revealing in this respect even when we look at the cases in which there are variations in the responses. For example, the emphatic and definite state was correctly interpreted by 11 Chinese and 2 Americans, with 2 Chinese choosing surprised and 2 Americans choosing harsh. However, none of the respondents chose gentle over emphatic. A similar pattern occurs in the responses to the gentle agreement example, which was correctly identified by 8 of 13 Chinese and by 2 of 5 Americans, and although some respondents chose "reluctant or uncertain agreement", none mistook it as "strong and emphatic agreement". The emphasis state was identified by all but 1 in both the Chinese and the American groups. In the case of "eager and encouraging agreement", the proposed interpretation was largely confirmed, though some chose "matter-of-fact agreement" instead. Interestingly, the emotion was unambiguously identified by all 5 Americans. For the negative emphasis case, 3 of 13 Chinese chose "fear". However, none of the respondents chose "anger" or "happy". Remarkably, the Americans unanimously identified the state as negative emphasis, even though they have no knowledge of the language and the semantic content.

These results suggest that there exist some basic distinctions between clearly different cognitive-affective states and that there are close similarities between finely differentiated states which have some *shared* element. For example, the distinction between supportive vs. angry or gentle vs. harsh are clearly recognized in our data. Conversely, some expressions of cognitive-affective states may be more complex and more difficult to differentiate in constrained situations. Strong emphatic agreement was also interpreted as surprise or harshness by 2 respondents in each group. This may be reasonable considering that both surprise and emphasis are characterized by a large pitch range. For gentle agreement, some respondents chose reluctant or uncertain instead, perhaps because all of these answers share an element of non-intensity. For complaining, the three answers chosen, uncertainty, complaining, and surprise, may be difficult to differentiate in some cases because of the shared feature of an unexpected and unsettling state.

*Factors affecting the respondent judgement.* There are a number of problems which can affect the judgement and interpretation of respondents. The first is a problem of *context*. In the perceptual experiment, respondents were exposed to only a very brief and narrow context, in contrast to the familiarity with topic and with a speaker's normal pitch and amplitude variation present in normal conversation. Such a contextual problem may be at work in the choice of "matter-of-fact" chosen over the proposed "exaggerated emphasis". Lacking an appropriate frame of reference, some listeners may find it

difficult to form a judgement of a contrastive pattern. A larger environment may be necessary to interpret more reliably.

On the other hand, respondents may be distracted and misled by a context which is not itself the intended focus. In spontaneous discourse, there is a progression of cognitive-affective states throughout, and different emotions can co-occur within an utterance. In such cases the perception of a later state may override the perception of an earlier one. For example, hesitation was identified by all but 3 Chinese respondents, who chose definiteness instead, perhaps because of the presence of the elaboration and summarizing discourse marker *jiushi* "that is", which is often used in more definite situations.

When there are conflicting cues, the stronger cue may dominate. In the light and prompting example, responses varied, including "insistent and demanding prompting", and "astonished". In this utterance, elements of both insistent prompting and light prompting are present, which may be the reason that there is a near split between those two answers.

Social factors may also play a role in the judgements of respondents. The negative disapproval state presents the most interesting case in our experimental data. The data show that there is clear internal consistency within each group. While within the Chinese group, the predominant choice was "positive approving", all of the Americans unanimously chose "negative disapproval", which is the proposed interpretation. This drastic split in responses between the groups may arise because of social convention and expectations of the Chinese group towards Americans, which in general is one of approval, and that may have led the Chinese respondents to ignore the sound cues, and to follow the social expectation instead.

The most puzzling case for us is the surprise state, as we thought the appropriate interpretation is clearly signalled by both the discourse context and semantic content. To our surprise, only 3 Chinese respondents chose "surprise", whereas 9 chose "approval" and 1 chose "matter-of-fact acknowledgment". None of the Americans chose the proposed interpretation, either. Since evidence for the surprised expression is fairly clear, the failure of identification does not invalidate the analysis. We need to do further investigation to find out why this mismatch occurs in this case.

Another unexpected test result is that, contrary to what we anticipated, the expression of definiteness and hesitation seem to be the hardest to recognize in our test. In one case, 11 of 13 chose the reverse interpretation from the proposed one. Similarly in the second case, 10 of 13 chose definiteness for hesitation. Interestingly, 7 of the 8 respondents who favored certainty in the second case also chose definiteness for hesitation in the first case. It is possible that there are some cues which those respondents tended to focus on more. This suggests that there was a mismatch in the concepts of definiteness and hesitation which was not clearly distinguished in the construction of the test or that more context is needed in making an appropriate interpretation. Further testing would be helpful in resolving this ambiguity.

### 4.3. Summary of experimental results

On the whole, the results of the perceptual experiment are in agreement with the analysis presented. The consistency of responses of both Chinese and American groups provides evidence of the systematic nature of cognitive and affective expression, and the responses from the Americans lend support to the idea of a sound-meaning link. The experiment also brought out the complex nature of cognitive and affective states, and points to the need for further research to explore the similarities between finely differentiated cognitive and emotional categories.

This complexity notwithstanding, consistent patterns do emerge when we compare our results cross-linguistically with both descriptive and experimental studies [6]. For example, surprise is often associated with a high pitch level and a large pitch range; emphasis generally correlates with a large pitch range, longer duration and louder amplitude; definiteness, finality or certainty have been associated with a low or falling pitch while uncertainty or tentativeness is often high in pitch; intensity, involvement or degree of arousal are associated with a large pitch range, as evidenced by many researchers.

The fact that our analysis is generally consistent with the results of a considerable number of other studies suggests that the expression of emotional and cognitive states through prosody may have underlying basis of evolutionary adaptation, speech mechanism structure, and neurological abilities and constraints. It also suggests that a detailed analysis of data is of crucial importance in forming a more adequate understanding of language, and that the use of various alternative approaches can help lead to a more complete understanding of prosody in speech.

## 5. Conclusion

In this paper we have shown that prosody and emotion interact in systematic ways so that participants successfully communicate the many levels of finely differentiated meaning present in conversational speech. Emotion is an integral component of human speech, and prosody is the principle conveyer of the speaker's state and hence is significant in recovering information. The significance of prosodic meaning to communicating judgements, attitudes, and the cognitive state of the speaker thus makes it essential to speech understanding projects such as emotion and intention tracking and to the development of natural-sounding spoken language systems.

## 6. Acknowledgements

We would like to thank Japan Science and Technology Agency for support of this research under the Expressive Speech Processing Project in the research area of "Information Processing for an Advanced Media Society".

## 7. References

1. Damasio, A., *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G. P. Putnam's Sons. 1994.
2. Scherer, K., Emotion effects on voice and speech: Paradigms and approaches to evaluation. *ISCA Workshop on Speech and Emotion*, Belfast, Northern Ireland. 2000.
3. Campbell, W.N., "Synthesizing spontaneous speech, Sagisaka, Y., Campbell, W.N., and Norio, H., editors, *Computing prosody*. Springer-Verlag, 165-186, 1996.
4. Yang, L.C., *Intonational Structures of Mandarin Discourse*. Ph.D. dissertation, Georgetown University. 1995.
5. van Bezooijen, R. A. M. G., *Characteristics and Recognizability of Vocal Expressions of Emotion*. Foris, 1984.
6. Murray, I. R. and Arnott, J. L., 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 (2). 1097-1108, 1993.