

# Meaning and Context: Prosodic Variation of Interjections in Conversational Speech

Li-chiung Yang

English Language Center, Department of Foreign Languages & Literature,  
Tunghai University, Taiwan

[yang\\_lc@thu.edu.tw](mailto:yang_lc@thu.edu.tw), [lyang@sprynet.com](mailto:lyang@sprynet.com)

## Abstract

In this paper we report on our research on the contextual meaning and prosody of three interjections *ey*, *wa*, and *oh*. A detailed qualitative-contextual analysis of our corpus shows that these interjections share important contextual and prosodic characteristics due to their similar functional status with respect to new or unexpected information. We show that there are also significant differences in contextual meaning arising from specific emotional or cognitive states, and that these differences are expressively communicated in the varied prosody of each interjection.

**Index Terms:** prosody, meaning, interjections, discourse

## 1. Introduction

A key aspect of natural discourse is the interactive exchange of information and the cooperative discourse process that functions to bring about a mutually satisfactory sharing of information. The encountering of new information gives rise to emotional responses that are based both on the specific content of the information, as well as on the uncertainty or uncertainty and degree of cognitive reorientation to the pre-existing knowledge state. Interjections often provide the first and most immediate expression of the emotional response and cognitive reorientation to new information, and are thus critical indicators of state that allow participants a window into the ongoing success of the speech process [1]. In this paper, we investigate the functions and contextual meaning of the interjections *ey*, *wa*, and *oh*, and how their contextual meanings are expressed systematically through prosodic form.

## 2. Speech corpus

In this study we present our results on prosody and contextual meaning of interjections with a special focus on the prosody and discourse functions of the interjections *ey*, *wa*, and *oh*. Our data consists of several sets of spontaneous conversations in Mandarin Chinese. Each conversation has two participants and the total speech data comprises about 10 hours. These conversational data were annotated by two experienced analysts for discourse relations, topic structure, discourse particles, interjections, and speaker turns. Instances of *ey*, *wa*, and *oh* were extracted from the corpus and analyzed both acoustically and contextually. Identification of contextual meaning including subcategories of emotional content, cognitive state, and interactive speech acts was performed by the analysts. The judgements of the specific contextual meanings were based on perceptual evaluation as well as acoustic-prosodic information.

## 3. Prosodic shapes and contextual meaning

Interjections have special importance in measuring conversational participant responses to information exchange

because of their high frequency of occurrence and their ability to compress valuable contextual meaning in an immediate and compact prosodic form. *Ey*, *wa*, and *oh* are similar in that each intrinsically expresses the essential newness or unexpectedness of the received information, but each also expresses a different underlying emotional response to the new information. Because of the non-lexical nature of these interjections, as well as to the more direct vocal path from emotion to expression that they follow, each interjection also expresses a range of varied emotional and cognitive states through variations in acoustic parameters of pitch, amplitude, and voice quality. By examining how the contextual meaning of these interjections is expressively revealed through prosody, we hope to contribute to a deepened understanding of the close relationship of emotion and cognitive states to their acoustic representation in expressive speech [2] [3].

### 3.1. Surprise, alert and disagreement: the interjection *ey*

The importance of emotions to prosodic form can be seen in stark contrast when we compare discourse particles or feedback utterances which are pervasively used in spontaneous dialogue to express specific emotional or cognitive states. *Ey* 'eh' in Mandarin is an interjection frequently used as an attention marker as well as a frequent expression of doubt or surprise. Our analysis shows that both pitch shape and duration provide important clues in distinguishing and disambiguating these cases.

*Ey* as an attention marker often occurs when one speaker suddenly remembers a salient fact. In these cases, the resulting pitch level of *ey* is usually high, and the duration short, expressing iconically both the intensity of speaker state and urgency of calling attention to the idea, as shown in s15, s16, s17, s1, s2, and s3 in Figure 1. As seen, these comprise a mix of short rises, together with short, high-pitched, but flatter shapes, suggesting that the attention marking stems from the high pitch level of the expression. This result is also in agreement with other researchers' findings in the field [4].

By contrast, the association of puzzlement, surprise and questioning with the use of *ey* typically occurred in cases where *ey* extends over a medium or longer duration. This is seen in s16 and k1, where medium length and clear arching rises occur as the speaker expresses her surprise and puzzlement on receiving new information. In s19 and s20, the speaker's surprise and puzzlement are reflected in the long, rising pitch shapes. In particular, in s14, the speaker expresses a very strong doubt and questioning emotion with a very dramatic rising pitch over an extended duration, and this prosodic questioning is followed with an explicit lexical questioning exclamation: *zhende a!* "Really!"

Analysis of our data further shows that *duration* plays a significant role in distinguishing between a more straightforward attention marker and an expression of puzzlement and questioning as well. When the speaker's emotional focus is on the upcoming fact itself, the length of *ey* is often very short, and

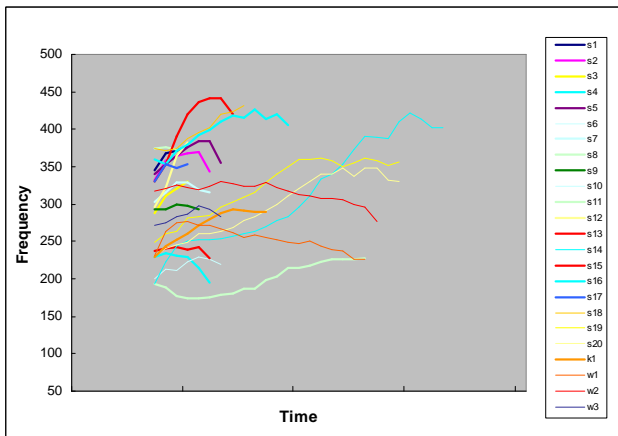


Figure 1: the interjection *ey* as marker of attention and expression of surprise

acts simply as an attention-getter and reflection of cognitive reorientation. On the other hand, when the speaker's impulse is to express confusion, puzzlement, and questioning, a greater amount of time is spent just to express or to expel that emotion, and this places the emotion itself as the focus, allowing a full release of the cognitive tension caused by the uncertainty, as well as necessitating cooperative efforts to resolve any discrepancies in participants' information status [5]. By comparison, pitch variation is more directly representative of the degree of cognitive uncertainty. In our data, *w1* and *w2* each occurs as the speaker directs strong attention towards the subsequent phrase, and *ey* in these cases acts as a command for emphasis, rather than uncertainty, and therefore have extended duration with falling pitch, reflecting the cognitive certainty of the speaker.

### 3.2. Surprise and Amazement: the interjection *wa*

Like *ey*, the Mandarin interjection *wa* characteristically functions as a prominent and immediate response to new information, but commonly adds a strong positive or negative emotional component in expressing a reaction to new information. The open sound quality of *wa* contributes to its use as a natural expression of surprise and amazement. Because *wa* communicates such a wide range of emotional responses, the prosody is especially valuable in disambiguating the emotional meaning in conversational speech.

#### 3.2.1. Surprise and amazement: symmetry of rise-fall shape

In contrast to the prototypical use of *ey* to express questioning and doubt, the archetypal use of the discourse marker *wa* 'wow' is to express amazement. *Wa* does not generally function as an attention marker like *ey*, therefore the focus of *wa* is usually on the expression of the emotion itself, and this causes the length of most cases of *wa* to have medium to long duration. In addition, *wa* is frequently used as a reaction to something said by the other speaker or as a speaker narrates their own reaction to the current topic idea. Surprise and amazement are directly related to the degree of uncertainty and certainty caused by encountering information that conflicts with a previously held belief and is simultaneously recognized as certain and true. It is this combination of uncertainty and certainty that gives *wa* its questioning as well. When the speaker's emotional focus is on the upcoming fact itself, the length of *ey* is often very short, and but differ in pitch height, sharpness of slope, and duration. In

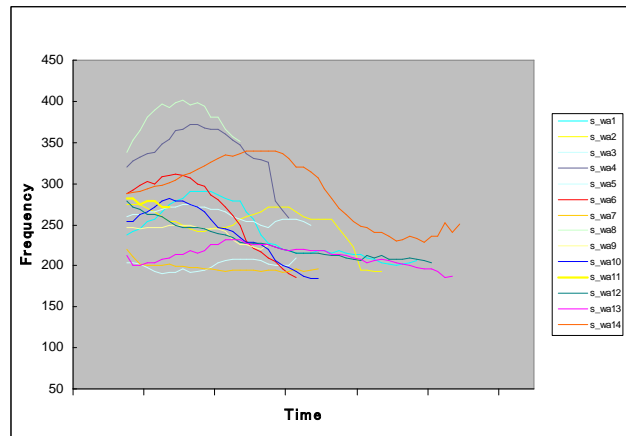


Figure 2: the interjection *wa* expressing emotional variation and amazement

addition, they have significant differences in extent of the downward section after the pitch peak.

#### 3.2.2. The narrow vs. wide arch shape

The complicating factor is that although *wa* has the basic meaning of amazement or being impressed, it can be used to express both negative and positive amazement, and simultaneously expresses other emotions that represent the specific reaction to the object of amazement. In our analysis, we have found that *wa* with a narrow arched shape generally indicates surprise together with caution and negativity, whereas a wide arch shape often expresses surprise and a high degree of awe or being impressed. For example, the narrow high-pitch rise followed by a steep fall of *S4* and *S8* both indicate a high degree of surprise concurrently with a strong negative emotion: in *s4*, the speaker is both impressed and signaling caution, while *s8* has a similar arch, but with a higher pitch level and a narrower arch, reflecting the high degree of terror and disgust expressed by the speaker. *s14*, on the other hand, has a wider rise-fall arch shape and is said with a nasal twang voice quality, expressing the speaker's surprise with strong disapproval and disagreement.

#### 3.2.3. Asymmetry in rise-fall shape

The relative proportion of rise and fall in an arch shape makes a significant difference in the meaning communicated. *Wa* with a longer duration and a mid-range pitch level with a proportionally greater falling extent is often associated with a lesser degree of surprise and with a more protesting emotion than the more pronounced rise-fall shape. In the examples of *s2*, *s6*, and *s10*, the initial rise is much smaller and there is a generally falling pitch contour; the shorter rise indicates the mild surprise, and the prolonged downward slopes signal the speaker's disapproval and disappointment in these cases.

#### 3.2.4. Degrees of emotional intensity

High-pitch *wa* is associated with greater degree of emotional intensity and a higher level of excitement as in *k1* and *k2*. By contrast, a longer and lower pitched *wa* is frequently used to show speaker's support, sympathy, or more matter-of-fact reaction to new unexpected information, as in *s7* and *s12*. A similar low-pitched expression with a different shape as in *s13*, however, can be associated with a mixture of surprise, caution and admiration. The convex shapes of *s7* and *s12* are more supportive, while the low pitch level and very sustained

concave shape of s13 reflects the warning implicit in what the speaker is saying. S12 is extended in duration, but the convex shape combines with the length to express the sympathetic attitude of the speaker in this case.

### 3.2.5. Combination of features

Pitch and voice quality are often used by speakers to mimic the physical state that accompanies a particular emotion. The pitch values of s3 and s9 both occur at a mid range level, and both have a similar slight rise, followed by a deeper fall. In both of these cases, the speaker was expressing an overwhelmed or exhausted response to information that was “too much”. The mild shape changes are accompanied in both cases with very audible expelling of air, which is crucial in providing this effect. S11, on the other hand, is short, quick, and with a breathy quality, iconically representing the speaker’s urgent and dismay reaction in missing an important event.

### 3.3. Surprise to dawning realization to acknowledgement – the interjection *oh*

As pointed out in our previous work, 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 communicate a range of uncertainty-based states, including doubt, surprise, acceptance, acknowledgement, and registering of information [1] [3] [6]. Three basic patterns for *oh* are evident in the plots containing instances of *oh* of two speakers in Figures 3-6. Like *ey* and *wa*, *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 analysis shows 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 in Figure 3, 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 in Figure 6. 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 in Figure 3. Emotions that are closely related to acceptance, such as sympathy and approval also tend to be expressed in a low pitch level [cf. 2]

S-oh1 and s-oh18 (see Figure 4) are at the other extreme of uncertainty, with rapid rises in pitch within a short time-frame conveying 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.

Our analysis shows clearly that the specific nature of discourse and individual speaking style are inseparable from the functions of an interjection.. For example, In contrast to the

highly varying expressive *ohs* of the previous speaker, speaker K’s *ohs* in Figure 5 seem much more subdued and are much shorter on average as they occur mostly as quick acknowledgement or quick recollection responses. Because of the particular characteristic 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.

Even within all these short expressions of *ohs*, there still exist finer variations in shape, height, range, direction, duration and intensity, and our analysis of the data shows that these variations are not random, but 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.

Dealing with authentic spontaneous conversation has led us to the understanding that dialogue responses to newly perceived information frequently contain both a *cognitive* reaction of the new information as well as an *emotional* component that expresses the speaker’s evaluation of the specific information. Figure 6 shows instances of surprise co-occurring with different emotions, each expressed within the duration of a single *oh*.

Our extensive analysis of the discourse contexts indicates that the high degree of uncertainty inherent in intense surprise causes a high rise in pitch, as in s-oh2, and s-oh5. The sharper and narrower arch shape of s-oh4 and s-oh3 reflect the presence of surprise with co-occurring emotions. The relatively shorter period of rise and extended low pitch level in the second half of *oh* in s-oh3 iconically expresses the fear and caution at work here. The narrow high arch of s-oh4 reflects the speaker’s amazement at the new information, and the narrow arch that has a lengthy falling pitch to a low level in s-oh3 expresses both surprise and horror at a description of mummies [cf. 5].

### 4. Universal & specific features of *oh*, *wa*, & *ey*

Comparing our data on *oh*, *wa*, and *ey*, we can see that they have both areas of differences and areas of similarities in shape, meaning, and function. As a response to updated knowledge throughout a conversation, *oh* has a greater diversity of functions, and occurs with the greatest frequency in our data, by far. *Oh* exhibits a wide variety of shapes, depending on the specific nature of the emotion and degree of uncertainty or certainty, from arch-shaped surprise to rising doubt or falling acknowledgement. By contrast, the intrinsic meaning of *ey* is more specialized to situations where there is surprise and doubt, but also when attention needs to be directed to something new. *Ey* thus occurs fairly often but much less frequently than *oh*. As an attention marker *ey* is usually short, often with a sharp rise in pitch for sudden recollection. In these situations, *ey* is usually an exclamation within the continued speech of the current main speaker. Longer duration for *ey* is used for the full expression of emotions such as surprise, puzzlement, and disagreement: the longer time iconically expresses the cognitive difficulty the new information gives rise to. Both speech forms have their counterparts in the contours for *oh*, which can also express recollection and emotions of cognitive difficulty, but



Figure 3: Varying heights correlated with varying degrees of dawning realization

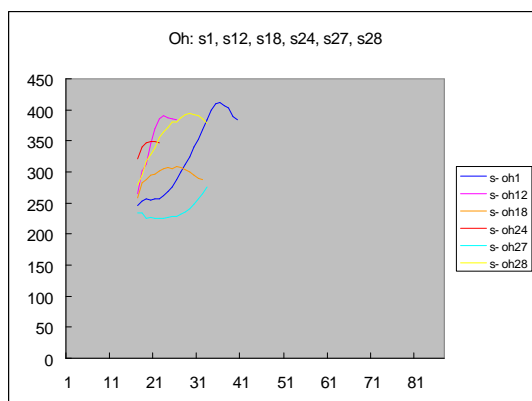


Figure 4: Instances of oh for Speaker S showing varying degrees of rise

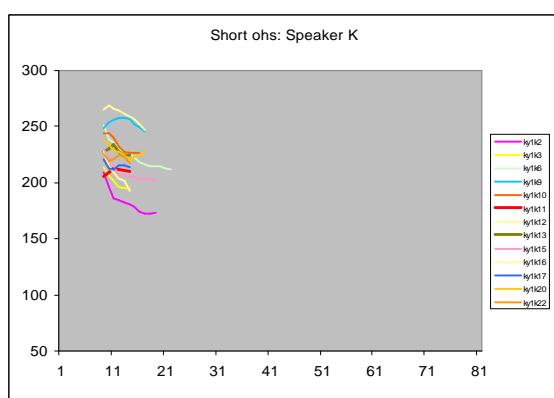


Figure 5: 13 Instances of Speaker K's short ohs

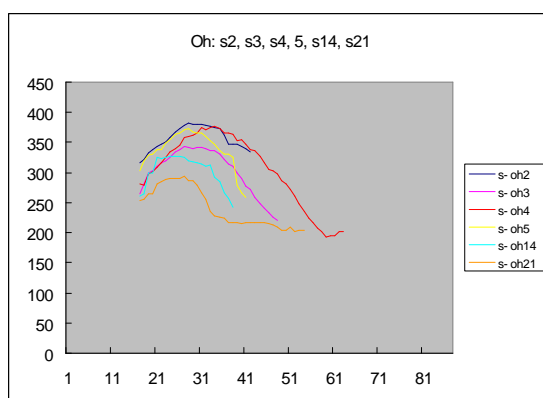


Figure 6: Dawning realization in combination with different emotions

the attention marking function of *ey* adds a dimension of meaning of more resistant receipt of the new information to each of these. Whereas a rising pitch is related to urgency, uncertainty, surprise, or doubt, with falling pitch, *ey* expresses less surprise and a more requesting and demanding attention signal.

*Wa*'s intrinsic meaning is to express amazement, and as amazing situations tend to be rare, *wa* occurs with the least frequency of these three in our data. *Wa* shares with *oh* the rise-fall arch shape that is typical of surprise and realization: with *wa* the surprise is accompanied by the expression of an immediate judgment on the impressiveness of the new information. The specific nature of this judgment or reaction determines the particular variants within the overall prosodic shape.

## 5. Conclusion

In this paper, we have shown that that prosody plays an indispensable role in communicating the multi-dimensional contextual meanings carried by the 3 interjections *ey*, *wa*, and *oh*. We have shown that variations in overall pitch direction, duration, and specific shape correlate *systematically* with changes in the level of uncertainty and the nature of emotions that develop throughout natural discourse. The characteristic functions and prosody of *ey*, *wa*, and *oh* presented in this study suggest that the interjections present a system of discourse contextual meaning in a crystalline and compact form, containing the essential elements of topic coherence, cognitive state, and emotional expressiveness that prosody brings to an

entire conversational exchange. We conclude that the ability of prosody to simultaneously signal both the cognitive degree of uncertainty and a specific emotional reaction to new information within the short time-scale of interjections are convincing evidence of the powerful expressive role of prosody in interactive discourse and should be utilized greatly by spoken language and robotic systems.

## 7. References

- Schiffrin, Deborah. 1987. *Discourse Markers* [Studies in Interactional Sociolinguistics 5], Cambridge: CUP.
- Ishi, C.T., Ishiguro, H., and Hagita, N., "Evaluation of prosodic and voice quality features on automatic extraction of paralinguistic information", Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 374-379, 2006.
- Yang, L-C, Integrating Prosodic and Contextual Cues in the Interpretation of Discourse Markers. In Kerstin Fischer (ed.), *Approaches to Discourse Particles. Studies in Pragmatics*, Amsterdam: Elsevier, pp.265-298, 2006.
- Ohala, J. J. The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (eds.), *Sound symbolism*. Cambridge, 1994.
- Fonagy, Ivan. 1971. Double coding in speech. *Semiotica*.
- Heritage, John. "Oh-prefaced Responses to Inquiry". *Language in Society* 27:291-334, 1998.
- 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.