Non-Lexical Conversational Sounds in American English

Jack Brajcich

This short paper examines what some of the non-lexical conversational sounds in American English are and why these sounds are important. It also looks at some of the functions and meanings of non-lexical conversational sounds. Lastly, this paper takes a look at how the meanings of non-lexical conversational sounds are determined.

Background: Ward's and others' work

The inspiration for this paper came from Ward's work on the subject (Ward, 2006). To briefly explain that paper, Ward took a mid-sized corpus of casual conversations of 13 different speakers, both male and female, all American, aged 20 to people in their 50s, from a variety of geographical areas and found 316 non-lexical items, with one occurring about every 5 seconds on average in conversation. Some of the more common items were *yeah, uh, um,* and *oh* as well as what is called a *click* sound. Although, as Ward points out, most of the aspects of non-lexical items in conversation have been studied, less common items have not been given all that much notice. Within conversations in American English, there is a great variety of non-lexical items, and an infinite set of possible sounds in roles. However, Ward summarized the basic phonetic components (some 91%, $=286/316$ of grunts in his corpus) as follows;

- schwa (uh, uh-huh)
- /a/ (ah)
- /o/ (oh)
- /e/ (yeah)
- /n/ (uh-hn, uun, nyeah)
- /m/ (mm, um, hm, myeah)
- /j/ occurs initially in yeah
- /h/ (hm)
- clicks
- creaky voice

One example of these phonetic components and what they can mean is the /m/ sound. The /m/ sound is often used (as in *um*) as a filler in conversation (perhaps to give the speaker more time to think before they talk) and is considered to be what Ward calls "thought worthy" (as opposed to when a listener uses *uh* which is often considered to trivialize what someone else has said. *Uh*, in some cases, can even be considered rude!) The /m/ sound is also seen in many back-channels as well. Back-channels refer to short utterances produced by one conversational participant while another is talking (Ward & Tsukahara, 2000). Back-channels are used by speakers to respond directly to the content of utterances of another, are optional in conversation, and do not require acknowledgement by the other speaker. Examples include; *yeah*, *uh-huh*, and *hm*. Also, the use of such utterances shows that a person is a good listener or that a person is listening (Ward, et al, 2007). In dialog this includes; attention, interest, understanding, and/or willingness to let the other person continue. Without skills in the usage of back-channels, a learner could appear to be uninterested, ill-informed, thoughtless, discourteous, passive, indecisive, untrusting, dull, pushy, or even worse no matter their mastery of vocabulary and grammar.

In addition, as is in the case of the non-lexical utterance of *oh*, non-lexical items

can influence listeners' comprehension in three ways. They can have no effect, a beneficial effect, or a negative effect (Fox Tree & Schrock, 1999). Beneficially, *oh* can inform listeners they should process upcoming information. Negatively speaking, *oh* create havoc in the listener's attempt to understand the syntactic relationship among words in an utterance.

Identifying Meanings for Sounds

Every utterance someone makes in conversation can mean many things at many levels (Ward, 2006). [Fox Tree and Schrock (1999), for example, mentioned 17 different ways *oh* can be used in speech!] Take, for example again, the /m/ sound. The use of the /m/ sound can signal withdrawing, becoming serious, wanting to slow the interaction down, cuing the other to listen closely, holding the floor, hiding something, and so on (Ward, 2006). Other functions include the signaling of turn-taking, negotiation, recognition, control, emotion, attitude, and affect. Ward and Tsukahara (2003) note that conversational dialogs are rich in non-verbal signals, and many of these relate to attitudes and feelings.

In addition, Shriberg (2001) noted that disfluencies are significant because they affect up to ten percent of words and over one third of utterances in natural conversation. Moreover, and interestingly to note, men make more disfluencies per word than women do. Men tend to 'control the floor' to a greater extent than women and thus may cause their listener to wait for longer than women do. However, women tend to use back-channels more in conversation than men do (Mulac, et al., 1998). Mulac, et al. also found that men see back-channels as being more controlling and as a way to express uncertainty in conversation. Women see back-channels as more 'other-focused', that is, as being more interested in the other person's opinion. Both men and women think that back-channels spoken to female conversation partners are more controlling than those spoken to male partners.

Clark and Fox Tree (2002) noted that *uh* and *um* are characteristically associated with planning problems in speakers. On the plus side, these fillers warn the listener about impending delays in speech (caused by finding words, formulating utterances, deciding what to say, and so on). *Uh* often indicates a minor delay in speaking, *um*, a major delay. *Uh* and *um* are also cliticized onto prior to words. For example, when cliticized onto *and, but, the, a,* and *that*, they form prosodic words with trochaic stress and syllable boundaries: "an.duh", "bu.tum", "thi.yum", "ai.yuh", and "tha.tuh".

Determining Meaning

The meaning of a component sound is largely determined by the context in which it is made (Ward, 2006). Meaning is also determined by the way that the utterance sounds. Although determining meaning is somewhat difficult, there are research methods which minimize or eliminate subjectivity, for example, controlled psychological experimentation, acoustical analysis, Conversation Analysis, statistical analysis, (with large corpora), etc. Another complication for determining meaning is that patterns of usage of non-lexical sounds vary across communities/groups (ethnic, region, and gender).

Sound-Meaning Correspondences

One meaning for the /m/ sound was described earlier. Below are some additional meanings of other common sound components (Ward, 2006).

Nasalization and the /n/ sound. Nasalized non-lexical sounds generally mean that the speaker has pre-knowledge ('old information', 'given information', or 'common ground'), as well as meaning something is already established, and known to the interlocutor too. This 'covering of old ground' through nasalizations accounted for 12 of 20 occurrences found in Ward's corpus. These nasalizations

included common sounds such as *nn-hn*, *nyeah*, and *nn-nnn*.

Breathiness and the /h/ sound. *Hmm*, unlike *um* and *mm*, occurs only as a back-channel (Ward, 2006). Moreover, *hmm*, compared to *mm*, seems to carry the meaning of respect and expresses a willingness to not only listen, but also give the other person's words some serious consideration. This conveyance of concern with breathiness was found in 23 of 43 tokens in Ward's corpus and are commonly known, for example, as *uhh* and *huh*.

Creaky voice. The "creaky" voice found in utterances often conveys the meaning of "claiming authority" (38/56 tokens) in common items like, *yeah* and *um*. By claiming authority, Ward means to say that people use creaky voice to indicate that they know what they are talking about. Creaky voice is found is authoritative statements, advice, opinions, decisions, recollections, and so on. (Speakers will indicate that these are indented as such.) These authoritative statements may come from expertise of some topic or from direct experience, and so on.

Click. The meaning of tongue clicks can involve personal dissatisfaction. Sometimes in conversation, personal dissatisfaction is temporary and speakers use clicks to show this dissatisfaction and then move on in the conversation. Such dissatisfaction can range from dissatisfaction about the conversation itself, to unhappiness about the topic, to about something the speaker finds disappointing. Of the 26 clicks found in Ward's corpus, 19 seemed to be expressing some form of dissatisfaction (self-remonstrance, dissatisfaction with the current topic, and dissatisfaction with the interlocutor, as a form of remonstrance).

Oh: the /o/ sound. The expression *oh* can mark receiving new information (Ward, 2006). Interestingly, *oh* often occurs not at the very moment new information is heard, but a fraction of a second later, after the information has been absorbed and the listener has decided how to regard it. A total of 44 tokens

of 46 in the corpus contain *oh*.

The /a/ sound. The /a/ sound in non-lexical utterances indicates some 'control'. People sometimes use this sound to show they are fully in control, knowing exactly what to do or say next, and people who feel this way express it. Also, pausing with the /a/ sound warns the listener something complex is coming in the conversation and the listener should pay close attention. The 'in control' /a/ sound accounts for 9 of the 18 tokens in Ward's corpus with such examples as *ah* and *ao*.

Schwa. The schwa is the most common sound in the non-lexical items in the corpus (Ward, 2006). It seems to be neutral, bearing almost no information. *Uh-huh* is one such example as it serves mostly as a filler.

Prosody-Meaning Correspondences

It is worth noting that several prosody-meaning correspondences can help make sense of non-lexical sounds in American English. Prosody is important to note because it often conveys more meaning than does the phonetic content (Ward, 2004). Back-channels, fillers, disfluency markers, and the like rely heavily on prosody to perform their functions, which include such things as turn-taking control, negotiating agreement, and signaling recognition and comprehension. Below is what Ward (2006) summarized as being essential to understanding non-lexical utterances and their prosodic meanings.

sound	meaning
syllabification	desire to talk
duration	amount of thought
loudness	confidence, importance
pitch	degree of understanding
pitch height	degree of interest

Generally, the more syllabification in an utterance, the more the listener is showing their willingness to engage in conversation (Ward, 2006). For example, two-syllable items are often back-channels whereas one-syllable items often act as fillers or disfluency markers (Ward, 2004). As for the duration of utterances, the longer the duration of the sound, the more thought is being put into the conversation (Ward, 2006). That is, duration seems to correlate with thought (Ward, 2004). When a person is speaking with confidence or of something of importance, they tend to speak more loudly (Ward, 2006). Therefore, loudness often correlates with assertiveness, self-confidence, and the importance of the utterance (Ward, 2004). In addition, the pitch of an utterance can show their degree of understanding. For example, a falling pitch indicates more understanding and a falling pitch slows some lack of understanding (Ward, 2006). Finally, the pitch height can indicate how interested a listener is in the conversation (Ward, 2004, 2006). Higher pitch height show a greater degree of interest (Ward, 2006).

Summary

This paper discussed some non-lexical conversational sounds and their sounds, meanings, and functions. Ward (2006) found that there are 10 component sounds among several non-lexical conversational utterances all with sound symbolism and compositional meanings. Ward's work has also given rise to the roles non-lexical utterances play in human conversation and the limits of human cognitive processing. The brief descriptions of non-lexical utterances in this paper emphasize that these conversational utterances are important in meaning and function. Therefore, I believe that there is something to be gained if learners can be made aware of these sounds and their meanings. Perhaps more explicated instruction of non-lexical items would be beneficial. Learners then may be able to understand and comprehend spoken English more efficiently and effectively.

References

Cark, HH & Jean E. Fox Tree (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84:73-111.

Fox Tree, J. E. & Josef C. Schrock (1999). Discourse markers in spontaneous speech: Oh what a difference oh makes. *Journal of Memory and Language*, 40:280-295.

Mulac, A., Karen T. Erlandson, W. Jeffery Farrar, Jennifer S. Hallett, Jennifer L. Mollory, & Margaret E. Prescott (1998). "Uh − huh". What's that all about?". Differing interpretations of conversational backchannels and questions as sources of miscommunication across gender boundaries. *Communication Research*, 25 (6) 641-668.

Shriberg, E. (2001). To 'errrr' is human: ecology and acoustics of speech disfluences. *Journal of the International Phonetic Association*, 31 (1) : 153-169.

Ward, Nigel & Rafael Escalante, Yaffa Al Bayyari, Thanmar Solorio (2007). Learning to show you are listening. Submitted to *Computer Assisted Language Learning*.

Ward, Nigel (2006). Non-Lexical Conversational Sounds in American English. *Pragmatics & Cognition*, 14 (1) : 129-182.

Ward, Nigel (2004). Pragmatic Functions of Prosodic Features in Non-Lexical Utterances. *Speech Prosody 04*, pp. 325-328.

Ward, Nigel & Wataru Tsukahara (2003). A study in responsiveness in spoken dialog. *Human-Computer Studies 59*, pp. 603-630.

Ward, Nigel & Wataru Tsukahara (2000). Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32:1177-1207.