Non-Lexical Conversational Sounds in American English

Nigel Ward
nigelward@acm.org
phone: 915-747-6827
fax: 915-747-5030
http://www.cs.utep.edu/nigel/
Computer Science, University of Texas at El Paso
El Paso, TX 79968-0518

[0]Acknowledgements: I thank Takeki Kamiyama for phonetic label checking, Gautam Keene and Andres Tellez for pragmatic function labeling and discussion, and all those who let me record their conversations. For general discussion I thank Daniel Jurafsky and Kazutaka Maruyama. I would also like to thank Keikichi Hirose, the Japanese Ministry of Education, the Sound Technology Promotion Foundation, the Nakayama Foundation, the Inamori Foundation, the International Communications Foundation and the Okawa Foundation for support. Most of this work was done at the University of Tokyo.

Non-Lexical Conversational Sounds in American English

Abstract

Sounds like *h-nmm*, *hh-aaaah*, *hn-hn*, *unkay*, *nyeah*, *ummum*, *uuh*, *um-hm-uh-hm*, *um* and *uh-huh* occur frequently in American English conversation but have thus far escaped systematic study. This article reports a study of both the forms and functions of such tokens in a corpus of American English conversations. These sounds appear not to be lexical, in that they are productively generated rather than finite in number, and in that the sound-meaning mapping is compositional rather than arbitrary. This implies that English bears within it a small specialized sub-language which follows different rules from the language as a whole. This functions supported by this sub-language complement those of main-channel English; they include low-overhead control of turn-taking, negotiation of agreement, signaling of recognition and comprehension, management of interpersonal relations such as control and affiliation, and the expression of emotion, attitude, and affect.

Biographical Note: Nigel Ward received a Ph.D. from the University of California at Berkeley in 1991. From 1991 to 2002 he was with the University of Tokyo. His primary research interest is human-computer interaction, especially sub-second responsiveness in spoken dialog systems.

[clear-throat]	2	hh-aaaah	1	nuuuuu	1	uam	1	uumm	1
[click]	22	hhh	1	nyaa-haao	1	uh	36	uun	1
[click]neeu	1	hhh-uuuh	1	nyeah	1	uh-hn	2	uuuh	1
[click]nuu	1	hhn	1	o-w	1	uh-hn-uh-hn	1	uuuuuuu	1
[click]ohh	1	hmm	2	oa	1	uh-huh	3	wow	1
[click]yeah	1	hmmmmm	1	oh	20	uh-mm	1	yah-yeah	1
[noisy-inhale]	1	hn	1	oh-eh	1	uh-uh	2	ye	1
achh	1	hn-hn	1	oh-kay	1	uh-uhmmm	1	yeah	70
ah	6	huh	2	oh-okay	2	uhh	4	yeah-okay	1
ahh	1	i	1	oh-yeah	1	uhhh	1	yeah-yeah	1
ai	1	iiyeah	1	okay	8	ukay	2	yeahh	2
am	1	m-hm	2	okay-hh	1	um	20	yeahuuh	1
ao	1	mm	2	ooa	1	um-hm-uh-hm	1	yegh	1
aoo	1	mm-hm	1	ookay	1	umm	5	yeh-yeah	1
aum	5	mm-mm	1	oooh	1	ummum	1	yei	1
eah	1	mmm	3	ooooh	1	unkay	1	yo	1
ehh	1	myeah	2	oop-ep-oop	1	unununu	1	yyeah	1
h-nmm	1	nn-hn	4	u-kay	1	uu	6		
haah	1	nn-nnn	1	u-uh	4	uuh	1		
hh	3	nu	1	u-uun	1	uum	6		

Table 1: All Conversational Non-Lexical Sounds in the Corpus, with numbers of occurrences

1 INTRODUCTION

American English conversations are sprinkled with large variety of non-lexical sounds, as suggested by Table 1. Along with such familiar items as *oh*, *um*, and *uh-huh*, there are a large number of less common sounds such as *h-nmm*, *hh-aaaah*, *hn-hn*, *unkay*, *nyeah*, *ummum*, *uuh* and *um-hm-uh-hm*. Similar variety is also seen in Swedish (Allwood & Ahlsen 1999), German (Batliner *et al.* 1995) and Japanese (Ward 1998).

While aspects of non-lexical items in conversation have been studied, these less common sounds have mostly escaped notice. In particular four basic questions have not been raised, much less addressed: first, the reason for such a large variety of sounds, second, what they all mean, third, their role in human communication, and fourth, their cognitive status.

The structure of this paper is as follows. The first three sections illustrate the phenomena, survey the current state of knowledge, explain the practical importance, and outline the overall approach. Section 4 presents a phonetic description and argues that most non-lexical conversational items, including both the rare and the common forms, are productive combinations of 10 component sounds. Sections 5, 6, and 8 present meanings for each of these component sounds and evaluate the power of a Compositional Model, in which the meaning of a non-lexical token is the sum of the meanings of the component sounds. The methods used to identify and check these meanings are presented as they arise, but mostly in Sections 2, 5, and 7. Sections 9 and 10 explore how the model helps clarify the role of non-lexical utterances in human communication and their relationship to phenomena such as interjection and laughter. Section 11 summarizes.

	total	back-channel	filler	dis-fluency	isolate	res-ponse	confirm-ation	final	other
[clear-throat]	2	.	.	1	.	.	.	.	1
[click]	22	.	12	2	1	.	.	.	7
ah	6	1	3	2	.	.	.	.	.
aum	5	.	4	1	.	.	.	.	.
hh	3	.	.	.	2	.	.	1	.
hmm	2	.	.	.	1	.	.	.	1
huh	2	.	1	.	1	.	.	.	.
m-hm	2	2	.	.	.	.	.	.	.
mm	2	2	.	.	.	.	.	.	.
mmm	3	2	1	.	.	.	.	.	.
myeah	2	2	.	.	.	.	.	.	.
nn-hn	4	4	.	.	.	.	.	.	.
oh	20	6	9	.	.	.	.	.	5
oh-okay	2	1	.	.	.	1	.	.	.
okay	8	2	2	.	.	1	2	.	1
u-uh	4	.	.	2	.	2	.	.	.
uh	36	.	13	20	1	.	.	1	1
uh-hn	2	2	.	.	.	.	.	.	.
uh-huh	3	3	.	.	.	.	.	.	.
uh-uh	2	.	1	1	.	.	.	.	.
uhh	4	.	3	1	.	.	.	.	.
ukay	2	1	1	.	.	.	.	.	.
um	20	.	10	8	.	.	.	1	1
umm	5	.	5	.	.	.	.	.	.
uu	6	3	2	.	.	.	.	.	1
uum	6	.	4	2	.	.	.	.	.
yeah	70	26	19	1	6	6	6	2	4
yeahh	2	2	.	.	.	.	.	.	.
(other)	69	32	18	3	8	3	.	1	4
Total	316	91	108	44	20	13	8	6	26

Table 2: Counts of Non-Lexical Occurrences in various positions and functional roles, for all items occurring 2 or more times in the corpus

2 THE NEED FOR A INTEGRATIVE ACCOUNT

For several reasons an integrative account of non-lexical items in conversation is needed. Although aspects of these phenomena have been addressed by a large number of studies, undertaken with a variety of aims, there has as yet been no attempt to integrate the findings. This section explains why it is worth doing so.

First, although there are many studies which have focused on one or a few of these items, the big picture has been missing. That is, there has been no attempt to explain how these items function as a system, meaning that, for example, there is no account of how speakers can chose among these items, especially the less common ones.

This lack hinders the construction of more useful spoken-dialog systems, in that non-lexical items have the potential to let spoken dialog systems give the user better, more motivating feedback, to deliver information more efficiently and smoothly, and in general to make human-

computer more pleasant (Schmandt 1994; Shinozaki & Abe 1998; Thorisson 1996; Rajan *et al.* 2001; Iwase & Ward 1998; Ward 2000a; Ward & Tsukahara 2003). This lack also hampers learners of English as a second language (Gardner 1998). Today there is no model or resource that describes even approximately, for example, the relation between *uh* and *uh-huh*, the ways in which the meaning of *uh-huh* resembles and differs from that of *uh-hn*, and when people use *myeah* instead of *yeah*. Thus, as a supplement to more detailed studies, a big-picture account would have great practical value.

Second, although there have been detailed studies of non-lexical utterances within certain roles, especially disfluencies and back-channels, there has been little work looking at the distribution of non-lexical items across such roles. This lack of category-spanning studies is unfortunate since, as McCarthy (2003) notes, many of these sounds are multi-functional. This is seen also in Table 2: for example, *oh* occurs both as a back-channel and turn-initially. An integrative account has the potential to reveal broader generalizations.

Third, although there have been on the one hand several phonetically sensitive studies of non-lexical utterances, and on the other hand many pragmatically sophisticated studies of their use in conversation and a few controlled experiments, there has been little connection between the two: the phonetically sensitive work has said little about those variations which are common in conversation or cognitively significant, and conversely the work based on conversation or dialog data has not paid much attention to phonetic variation. An integrative account, looking at variations in form and variations in meaning together, has the potential to improve our understanding of both aspects.

Ultimately, of course, the reason to seek an integrative account lies is the hope that it will be simpler overall.

3 APPROACH

To seek an integrative account it was necessary to approach the phenomena in a novel way.

3.1 Working with a Mid-Size Corpus

The basic strategy adopted was to take a mid-sized corpus of casual conversations and try to understand and explain everything about all of the non-lexical utterances. By looking at all occurrences it was easier to notice the relations between items and to examine items across a variety of functional and positional roles.

Conversations were used, rather than task-oriented dialogs or controlled dialog fragments, to allow the study of diverse dialogs and rich interactions, giving a broader view of when and how non-lexical utterances are used.

Analysis was limited to a mid-size corpus, rather than a large one, in order to allow a reasonably thorough examination of the phonetics and pragmatics of each occurrence. This also made it possible for all the analysis to be done by listening directly to the data, without having to rely on transcriptions.

A home-made corpus, rather than a standard one, was used because the author was familiar with it, as the sound engineer recording the conversations, as a friend or acquaintance of most of the conversants, and as a participant in a few of the conversations. (The author's own

non-lexical utterances were excluded from the analysis.) The extra information this gave was often helpful when interpreting ambiguous utterances.

The corpus used includes 13 different speakers, male and female, all American, aged from 20 to 50ish, from a variety of geographical areas. Most of the conversations were recorded for another purpose (Ward & Tsukahara 2000), and participants were not informed of the interest in non-lexical utterances. In some cases people were brought together to converse and be recorded, other times the conversations were already in progress. All recordings had only two speakers, and in most cases these two were doing nothing but conversing with each other, although some conversations included interactions with other people or pets, and one speaker was driving. Recording locations included the laboratory, living rooms, a conference room, a hotel lobby, a restaurant, and a car. The relationships between conversants ranged from relatives to close friends to acquaintances to strangers. Most conversations were recorded in stereo with head-mounted microphones; one was a telephone conversation.

3.2 Looking at a Wide Variety of Items

Given this corpus, the first thing to do was to identify all the non-lexical items. To avoid missing anything that might be relevant, the initial definition was made inclusive. Specifically, all sounds which were not laughter and not words were labeled as non-lexical items. A 'word' was considered to be a sound having 1. a clear meaning, 2. the ability to participate in syntactic constructions, and 3. a phonotactically normal pronunciation. For example, *uh-huh* is not a word since it has no referential meaning, has no syntactic affinities, and has salient breathiness. Although the distinction between words and non-lexical items is not clear-cut, as will be seen, this gave a reasonable way to pick out an initial set of sounds to examine.

To keep the scope manageable, attention was limited to sounds which seemed at least in part directed at the interlocutor, rather than being purely self-directed, even if the communicative significance was not clear. This ruled out stutters and inbreaths.

The corpus has 316 non-lexical items, with one occurring about every 5 seconds on average.

3.3 Listening to the Data

Rather than working from transcripts, all analysis was done by listening. This probably helped focus attention on the interpersonal aspects of the dialogs, rather than the information content. This research style was facilitated by the use of a special-purpose software tool for the analysis of conversational phenomena, didi (Ward 2003).

However, it being important to pay attention to the detailed sounds of non-lexical items, these were labeled phonetically. These labels were always visible while listening.

The phonetic labeling was done using normal English orthography, as discussed below. IPA was not used as it provides more detail than was needed, potentially obscuring generalizations. This is a common choice in studying dialog, for example Trager (1958) argued that the study of 'vocal segregates' such as *uh-uh*, *uh-huh*, and *uh*, requires 'less fine-grained' phonetic descriptions. The labels in the corpus included annotations regarding prosody and voice, although this information is not shown in this paper except where relevant. The labels in the corpus are as seen in Table 1.

Due to concern that native knowledge of English or theoretical predilections might bias phonetic judgments, about half of the items, including all difficult cases, were labeled independently or cross-checked by an advanced phonetics student with little experience of conversational English and no knowledge of the hypotheses presented below. However no biases were found, and the remaining items were labeled by the author alone.

3.4 Comparison to Alternative Approaches

Thus the method of analysis is unusual. Moreover, as will be seen in Section 5, it relies in part on subjective judgments. Although there are better established and more powerful methods and theoretical frameworks, none of these seemed quite appropriate for the task of attaining an integrative account of non-lexical items. Thus the approach taken here.

4 A MODEL OF THE PHONOLOGY

Revisiting Table 1, the variety of non-lexical items is striking. Phonological conditioning, a common cause of phonetic variety, can provide little explanatory power here, since these items mostly occur in isolation. This section shows how most of the variation can be accounted for by a relatively simple model.

4.1 Intuitions about Non-lexical Expressions

Not only is the variety great, the set of possible sounds in these roles appears not to be finite. For example, it would not be surprising at all to hear the sound *hm-ha-hn* in conversation, or *mm-ha-an*, or *hm-haun* and so on. However, there are limits: not every possible non-lexical sound seems likely to be used in conversation. For example *ziflug* would seem a surprising novelty, and would be downright weird in any of the functional positions typical for non-lexical items. The existence of this intuition — that only certain non-lexical sounds are plausible in conversation — is a puzzle that has not previously been addressed.

There have, of course, been attempts to describe the phonetics of such items by identifying all possible phonetic components (Trager 1958; Poyatos 1975). However the descriptive systems produced by these efforts cover wider ranges of sounds, including moans, cries and belches, and so they do not help with the task of circumscribing the set of conversational non-lexical items.

It is also possible to attempt to describe the set of possible items in terms of a list. Although it is possible, for purposes of linguistic theory, to postulate the existence of such a list, actually making one is problematic. The best attempts so far have been by researchers who are labeling corpora for training speech recognizers, who of course have an immediate practical need for some characterization of these sounds. For example, the best current labeling of the largest conversation corpus, Switchboard, uses a scheme (Hamaker *et al.* 1998) which specifies a small finite list, where hesitations are represented with one of *uh*, *ah*, *um*, *hm* and *huh*; 'yes/no sounds' are represented with one of *uh-huh*, *um-hum*, *huh-uh* or *hum-um* 'for anything remotely resembling these sounds'; and 'non-speech sounds during conversations' are represented with one of: 'laughter', 'noise' and 'vocalized-noise'. Comparison with Table 1

reveals how much information is lost by using such a list. Moreover, no mere list can account for intuitions about which sounds are plausible: a description in terms of a list of 10 or 100 items gives no explanation for why *hum-ha-hn*, but not *ziflug*, could be the 11th or 101st observed token. Of course a list-based model could be embellished with descriptions of the permitted phonetic variations or sub-forms — as in Bolinger's discussion which starts with the claim that '*Huh*, *hunh*, *hm* is [sic] our most versatile interjection', and then turns around and focuses on differences between these three forms. However such a hybrid approach seems unlikely to be concise or to have much explanatory power. Thus a satisfactory list-based account of conversational non-lexical items seems likely to be elusive.

4.2 The Phonetic Components

I propose that many non-lexical utterances in American English are formed compositionally from phonetic components (leaving open the vexed question of whether these components are phonemes or features (Marslen-Wilson & Warren 1994)). This claim is not without precedent: there are a number of works which have, more or less independently, attempted to characterize variation in non-lexical expressions in German, Japanese, and Swedish, and have done so using tables of non-lexical items or lists of rules relating or distinguishing different tokens (Ehlich 1986; Werner 1991; Takubo 1994; Takubo & Kinsui 1997; Kawamori *et al.* 1995; Shinozaki & Abe 1997; Ward 1998; Allwood & Ahlsen 1999; Kokenawa *et al.* 2004). These all imply the possibility of an analysis in terms of component sounds.

This subsection describes the main inventory of phonetic components in non-lexical conversational sounds in American English.

- Schwa is often present, as seen in *uh* and *uh-huh*. (In conversation this is a schwa, although when stressed, in tokens produced in citation form, it appears as ʌ.)
- An /a/ vowel can also be present, as seen in *ah*, which is distinct from schwa, at least for some speakers.
- An /o/ vowel occurs in some sounds, such as *oh*.
- An /e/ vowel occurs in *yeah* and occasionally elsewhere.
- /n/ and nasalization, of vowels or of the semivowel /j/, is a feature that can be present or absent, as seen in *uh-hn* (versus *uh-huh*), in *uun* (versus *uh*), in *nyeah* (versus *yeah*).
- /m/ can occur in isolation (*mm*) or as a component, as in *um* (versus *uh*), *hm* (versus *huh*) or *myeah* (versus *yeah*).
- /j/ occurs initially in *yeah* and variants thereof.
- /h/ occurs in isolation occasionally, as a noisy exhalation or a sigh. /h/ or breathiness is also present in items such as *hm* (versus *mm*), and in the back-channel *uh-huh*. Some such items involve breathiness throughout, others involve a consonantal /h/, while others are ambiguous between these two realizations.

- Tongue clicks occur often in isolation, and occasionally initially. (Specifically, there are cases where the click is followed by a voiced sound with no noticeable pause; the delay from the onset of the click to the onset of voicing ranged from 50 milliseconds to 170 milliseconds in the corpus for these cases.)
- Creaky voice (vocal fry) occurs often , including for example on *aummm*, *yeah*, *okay*, *um*, *hm*, *aa*. Creakiness sometimes spans the entire sound, but other times is present only towards the end.

Sound	Notes
/ə/	
/o/	
/a/	
/e/	limited distribution
nasalization	
/m/	
/j/	limited distribution
/h/ and breathiness	
click	limited distribution
creakiness	

Table 3: Phonetic Components of Common Non-Lexical Utterances.

The list above is summarized in Table 3. Although this summary may suggest that these phonetic qualities are binary, for example nasalization being either present or absent, it seems more likely that the phonetic components are in fact non-categorical, involving 'gradual, rather than binary, oppositional character' (Jakobson & Waugh 1979). This explains how the set of non-lexical items generated can be literally not finite.

It is also worth noting that the vowel identifications are approximate. Indeed, it is entirely likely that, as found for German hesitation particles, "the vocalic portions ... have their own quality", distinct from those used in lexical items (Patzold & Simpson 1995).

For expository convenience, this phonological analysis is presented here, before the semantic analysis, although in fact the set of relevant component sounds cannot be determined without reference to meaning. Actually a preliminary version of the semantic investigations described below was done before the list of sound components was drawn up. This is why, for example, the inventory of sounds groups together consonantal /h/ and breathiness, but not the nasals /m/ and /n/: the first grouping, but not the second, has a consistent meaning, as will be seen.

The fact that this inventory of sounds is fairly small makes it possible to concisely specify the phonetic values for all the labels seen in Table 1. Thus the non-obvious American English orthographic conventions for non-lexical items are (slightly regularized) as summarized in Table 4. Other Englishes apparently have other conventions, for example, British English uses *er* to represent a sound not unlike American English *uh* (Biber *et al.* 1999). Further discussion of spelling appears elsewhere (Ward 2000b).

notation	phonetic value	notes
h	a single syllable-final 'h' bears no phonetic value, elsewhere 'h' indicates /h/ or breathiness	
n	nasalization and /n/	
click	alveolar tongue click	
u	ə	
uu	as a syllable, indicates a short creaky or glottalized schwa	
repetition of a letter	duration and/or multiple weakly-separated syllables	
- (hyphen)	a fairly strong boundary between syllables or words	
yeah	/jeə/	
kay	/keɪ/, as in *okay* etc.	
gh	velar fricative	rare
chh	palatal fricative	rare
oop	/up/	rare

Table 4: Some Non-Obvious Facts about Conventional American English Orthography for Non-Lexical Sounds

4.3 Rules for Combining Phonetic Components

The full phonological model includes the above list of component sounds plus two rules for combining them.

The first way in which sounds are combined is by superposition. For example, a sound can be a schwa that is simultaneously also nasal and creaky.

The second way is concatenation. There are probably minor constraints on this, for example /j/ and /e/ have very limited distributions, and click seems to appear only initially. These remain to be worked out.

There seems to be a tendency for these sounds to have relatively few components, that is, the number of component sounds in a non-lexical token generally is less than the average number of phonemes in a word. There is also a tendency, rather stronger, for the number of *different* sounds to be few: most sounds have only one or two, and more than three is rare. This is also seen in the fact that these sounds often involve repetition.

4.4 The Power of the Phonological Model

The above components and rules constitute a simple, first-pass model of the phonology of these sounds. In effect, this describes the space of non-lexical utterances as based on "a phonological system which is different from those employed in lexical items" (Patzold & Simpson 1995), although the ultimate status of this phonological system remains to be determined.

However it is relatively easy to evaluate the model for descriptive adequacy. Ideally a model should generate all and only the non-lexical utterances of English.

As far as generating ONLY non-lexical items, the model does reasonably well. The key

explanatory factor is that the inventory of component sounds excludes most of the phonemes present in lexical items, including high vowels, plosives, and most fricatives. This provides a partial explanation for native speakers' intuitions that only certain sounds are plausible as non-lexical items in conversation. However this model does overgenerate somewhat; although Section 7.3 explains how it can be extended to reduce this.

As far as generating ALL the non-lexical items, this model does fairly well on this also. Evaluating it against the inventory of grunts in the corpus, the phonological model accounts for 91% (=286/316). It achieves this performance because, of course, it includes sound components not present in English lexical items. However it does not account for all the non-lexical items. The exceptions fall into 4 categories. First, there are 3 breath noises such as throat-clearings and noisy inhalations. Second there 2 exclamations including rare sounds, namely *achh* and *yegh*. Third, there are 5 items which only seem explicable as word fragments, extreme reductions or dialectal items, such as *i*, *nu* and *yei*. Finally, there are 20 tokens with phonemes missing from the model but normal for lexical English, including *okay* and *wow*. This last set includes items which are only marginally non-lexical, in the sense discussed in Section 10.2, so it is not entirely surprising that the model fails to handle them poorly.

Thus, although the model is not perfect, it accounts for rare non-lexical tokens and the common ones in the same way. It is also more parsimonious and explains intuitions better than the alternative, modeling these items with a finite list of fixed forms. In this sense, these sounds are truly non-lexical. Using this model as a base, subsequent sections extend the analysis to deal with meaning and dialog roles.

5 METHODS FOR FINDING SOUND-MEANING CORRESPONDENCES

Thus it seems that these sounds can be analyzed in terms of the composition of phonetic components. This leads inevitably to the question: what do they mean? This is the topic of this section.

Asking this question presupposes that sound components of this size can bear meaning. While most morphemes are syllable-sized or larger units, various studies have found a rich vein of sound-meaning mappings at a lower level, or "sound symbolism". That is, there exist phonesthemes, sounds which are smaller than normal morphemes but still bear meaning. The existence of such mappings is theoretically interesting in that they violate Saussure's principle of the "arbitrary nature of the sign", which postulates that the meaning of the whole cannnot be predicted from the meanings of the parts (de Saussure 1915 1959). However there is a wealth of evidence that sound symbolism is often productive in non-lexical items and also infuses large portions of the lexicon (Sapir 1929; Hinton *et al.* 1994; Abelin 1999; Magnus 2000). For example there appears to be a phonestheme common to words like *splash*, *crash*, *bash*, and *mash*. In such cases the meaning of the whole is predictable, at least in part, from the meanings of the component phonesthemes.

The specific mappings most commonly identified in studies of sound-symbolism relate mostly to percepts, including sounds, smells, tastes, feels, shapes, spatial configurations, and manners of motion. Thus few of the mappings previously identified seem relevant to non-lexical items with conversational functions. There are few exceptions: some work has discussed or examine the possibility of a sound-symbolic system operating in discourse particles and

related items. Jakobson and Waugh (1979), Ameka (1992), and Wharton (2003) have noted that sound symbolism may also be present in interjections. Bolinger (1989), in his discussion of exclamations and interjections, proposed specific meanings for vowel height, vowel rounding, and various prosodic features in a variety of non-lexical items, as detailed below. Finally, Nenova et al. (2001) examined various non-lexical items in a corpus of transcripts of task-oriented dialogs. Based on considerations of articulatory effort, they proposed a distinction between 'marked' items, those which involve nonsonorants, lengthening, multiple syllables or rounded, noncentral or tense vowels, and 'unmarked' items, those which are composed of only /m/ and /ə/. They showed that marked items are more common as indicators of 'dynamic participation', as opposed to the production of neutral back-channels during passive listening. The present paper goes beyond this level of analysis to ascribe specific meanings to specific sounds.

The analysis methods used in this paper combine and extend the methods used in these studies. Detailed discussion of the methodological issues appears after an example of the analysis.

5.1 A first example: /m/

In fillers, /m/ generally occurs while the speaker is trying to decide whether to speak or trying to decide what to say. This is illustrated in Example 1, where the *umm* occurs before a substantial pause preceding a restart of the explanation, in contrast to the *uh*, which occurs before minor formulation difficulties. There is a wealth of statistical and experimental evidence that *uh* indicates a minor delay and *um* a major delay (Fox Tree 2001; Barr 2001; Clark & Fox Tree 2002), although it may be that only speakers, not listeners, make this distinction (Brennan & Williams 1995; Barr 2001). Also Smith and Clark (1993) have observed, in the context of quizzes, that fillers *um* and *am*, compared to *uh* and *ah*, generally seem to indicate more thought. Also, the distributions of *uh*, *um* and *umm* in Table 2 show that the presence of /m/ correlates with the tendency to appear as a filler, utterance-initial, rather than as a simple disfluency.

(discussing effects of speaking rate on phonology probabilities)		
E:	going to be different than if they're, **uh**, talking much more slowly,	1
X:	*um-hm*	2
E:	so, **umm** [3 second pause] so, uh, the stuff that we did at ...	3

(1)

This meaning for /m/ is seen in back-channels also. The contemplation can be directed at various things, including trying to understand what the interlocutor is saying, trying to empathize with him, or trying to evaluate the truth or relevance of his statement. For example, in Example 2 M seems to be giving some thought to the situation X has related; specifically, he seems to be sympathizing and perhaps contemplating the complexity or inevitability of the situation. As a consequence, this *mmm* functions as a polite response, and in contrast to a neutral *uh-huh*, which would trivialize the matter, and be rude.

(after some talk about television, children, and violent play)		
X:	*and this video was about Ultraman . . . most of it's not too violent . . . but there is a little bit of stabbing and stuff*	1
M:	right	2
X:	*and so he came home and he was stabbing poor little Henry*	3
M:	nyaa-haao	4
X:	*yeah, I, I felt.*	5
M:	**mmm**	6
X:	*well, I mean, yeah. <click>I was pretty annoyed.*	7

(2)

A similar case is seen in Example 3, where T is telling a story, and has just introduced the people involved. O's *m-hm* seems to indicate that he's thinking, perhaps trying to visualize the complex situation described or perhaps speculating about what happened next.

(T is halfway into a story involving himself, his son, and his daughter)		
T:	my son was working in Bo-, uh, Boston, actually Cambridge, at the time	1
O:	**m-hm**	2

(3)

This can be contrasted with the /m/-less version, *uh-huh*, in Example 4, where O responds to a simple utterance whose point and relevance is immediately obvious.

(T and O have just donned head-mounted microphones for recording)		
T:	*once had to wear one of these riding in the back seat of an airplane, because the airplane was so noisy*	1
O:	**uh-huh**	2
T:	*that the only way the four people in it could talk, was with earmuff earphones*	3

(4)

In general, the meaning of /m/ in non-lexical conversational sounds can be described as follows.

(Thought-worthy. People in conversation sometimes interact relatively superficially and sometimes at a deeper level. Deeper places in conversation sometimes involve the sharing of some emotion, but more often just the communication of something that requires thought. The speaker may mark something said by the other as meriting thought, or he may mark something that he himself has just said, or is trying to say, as involving or meriting thought. This may correlate with the intention or need to slow down the pace of the conversation in order to give time for this thought or contemplation. Note that deepness in this sense does not usually involve intellectually deep thinking, just that the conversation turns relatively deeper for a moment or two.)

Of the 57 tokens in the corpus containing /m/, 49 appeared to be indicating this sort of meaning, and 8 seemed not to.

5.2 Identifying Meanings for Sounds

The first methodological issue to discuss is that of how to discover and demonstrate that some component sound S has some specific meaning M. While there exist good methods for testing a hypothesized S (Magnus 2000), here the primary task is to discover the S in the first place.

One basic strategy is to seek an M is shared by all (or most) tokens which include S. This is the first basic strategy employed in this paper. This can be done using direct evidence for the presence of meaning M in each case, or indirect evidence, such as the prevalence of S in tokens serving functional or positional roles which correlate with M.

However this is not easy, because every utterance means many things at many levels (Schiffrin 1987; Traum 2000; Louwerse & Mitchell 2003). For example, the *umm* in Example 1, which was presented as indicating that the speaker was thinking, might also be interpreted as meaning that he was withdrawing, or becoming serious, or wanting to slow the pace of the interaction, or foreshadowing the imminent discussion of something significant, or showing a polite reluctance to dominate the conversation, or cuing the other to listen closely, or holding the floor, or hiding something, and so on. In past, sophisticated studies of some such functions at various levels have been done, and there are a number of useful frameworks for analysis. These, however, are mostly limited in that they focus on one level or one type of function. There are, for example, studies which consider some non-lexical items as discourse particles, connectors, acknowledgements, continuers, assessments, turn-taking cues, and so on. However, as Fischer (2000) notes, these items 'actually form a more homogeneous group than suggested by the number of different descriptive labels'. For this reason the analysis here was not done within any specific theory or framework; rather the shared meanings were sought bottom-up, by observing similarities across the corpus.

The task of the analyst is to examine the entire set of tokens containing S, and pick out the 'best' meaning, that is, the meaning component M which is (most) common across the set. While examining the data various possible Ms were kept constantly in mind, namely those identified as important in previous studies of conversation, non-verbal communication, and inter-personal interaction. These include various functions involving discourse structure marking, signaling of turn-taking intentions, negotiating agreement, signaling recognition and comprehension, managing interpersonal relations such as control and affiliation, and expressing emotion, attitude, and affect.

For lack of a formal procedure for finding the best M, the method used was to simply consider various possible Ms and see how well each matched the set of tokens which include S. This time-intensive process was simplified somewhat by homemade tools to help find and quickly listen to all tokens sharing some phonetic property or semantic annotation. The Ms presented in this article are the result of iterative refinement to minimize the number of exceptions and simultaneously avoid unnecessary vagueness. However there is no guarantee that these Ms are in any sense optimal.

The second basic strategy for determining that S means M is to find a minimal pair of non-lexical tokens, one with S present and one with S absent, and show that the difference in meaning is M. Sometimes minimal pairs or near minimal pairs were found in the corpus; if not, sometimes it is possible to appeal to intuition, considering what it would mean if some non-lexical utterance in the corpus had occurred instead with some component S added or subtracted.

Fortunately, for each of the component sounds studied, except schwa, evidence of both

kinds (shared meaning across the set and difference in a minimal pair) was found, and both types of evidence pointed to the same meaning M in each case.

5.3 Determining the Meanings of Tokens

Identification of the meaning of a component sound using the methods above relies heavily on the ability to identify the meaning of a non-lexical utterance as a whole. This also is not trivial. Two basic sorts of information are available. The first is the context, primarily the nearby utterances of the speaker and the nearby utterances of the interlocutor, both before and after the token: from this it is generally possible to infer how the speaker meant it and/or how the listener interpreted it. While these are not invariably aligned, as misunderstandings and willful misinterpretations do occur, such cases are rare, and in the corpus all non-lexical sounds appeared to be interpreted compatibly by both speaker and listener. (Although ultimately a full understanding will require consideration of non-obvious differences in the information content of such items to speakers versus listeners (Nicholson *et al.* 2003; Brennan & Schober 2001; Corley & Hartsuiker 2003).) The second sort of information is the way that the utterance sounds in itself, based on native speaker intuitions. For this study, meanings are ascribed to non-lexical sounds only if both types of information are available and consistent.

This means that tokens for which only one sort of information is available, or where the two sorts of information are in conflict, are not ascribed meanings; they are characterized below as UNCLEAR IN MEANING. For example, in three of the tokens including /m/, the token itself, considered in isolation, does appear to be contemplative, but the context does not suggest any need for the speaker to be thinking, as in *myeah* in Example 24. (Perhaps the speaker in these cases had a private thought, not related to the conversation, or perhaps he was momentarily distracted, producing an utterance that was not strictly appropriate for the context. As it happens these 3 cases were all back-channels, where lapses of attention can often pass unnoticed.) For lack of a technique for further investigating such examples, such non-lexical utterances are simply considered to be unclear in meaning and thus providing no evidence for or against any sound-meaning correspondence.

It is worth stressing that both sorts of information are subjective, especially the second. There are alternative research methods which minimize or eliminate subjectivity, for example, controlled psychological experimentation, acoustical analysis, Conversation Analysis, statistical analysis over large corpora, and validation with labels by analysts unfamiliar with the hypotheses. All of these methods are superior in various ways to the current methods, and ultimately the claims made here will stand or fall as they are supported or rejected by more powerful methods. However for the present purpose, identifying meanings in the first place, the sorts of information given by simple approaches are adequate.

Another complication for this approach is that patterns of usage of non-lexical sounds vary across communities. It is well known that the timing and frequency of non-lexical usage varies with ethnicity, region, and gender (Erickson 1979; Tannen 1990; Mulac *et al.* 1998), and the meanings ascribed to non-lexical tokens almost certainly do also. While such differences are interesting sociolinguistically, for present purposes they raise a difficulty: there will be examples where the interpretation presented here will not be shared by all readers. As a partial back-up, all of the claims in the next section are multiply supported, so that none is dependent on the interpretation of a single example.

Since subjective interpretations are unavoidably involved, the main purpose of the dialog excerpts is to allow the reader engage his or her own intuitions, rather than, say, to support tight demonstrations that each token must mean what is claimed. Thus the dialog excerpts are presented concisely and in standard orthography and punctuation, although of course there exist alternative conventions which are more descriptive in terms of phonetics, prosody, timing, etc. (Edwards & Lampert 1993; Hutchby & Wooffitt 1999; Jefferson 2002). Concise presentation is necessary for another reason also: given the goal of an integrated account and the concomitant need to examine a large number of tokens, space does not permit an exhaustive presentation of any single example.

It is not uncommon for people looking at a non-lexical item to have different interpretations. In my experience most such differences arise not from dialect differences or fundamentally different judgments, but rather from noticing different aspects of the dialog; this is the problem of multiple levels mentioned in the previous subsection. Such different interpretations generally turn out to be compatible. Differing interpretations are easier to resolve if the audio itself is available. To give more readers access to this data, sound waves for the non-lexical items discussed, with timing, pitch and energy information for the utterances in the contexts, are available at the website for this paper, `http://nigelward.com/egrunts/`, mirrored at `http://www.cs.utep.edu/nigel/egrunts/`.

5.4 Using the Compositional Hypotheses

These analysis methods presume that the meaning of each component sound is evident in the meaning of the whole, or, more strongly, that the meaning of each non-lexical utterance is compositional. This is the Compositional Hypothesis. Its validity will be discussed later, but for now it is an working hypothesis, and an essential one, since it makes the investigation possible.

This hypothesis makes analysis possible but not easy. In particular, the hypothesis implies that the meaning contributions of all sounds in the token are also active, in addition to the meaning for the sound under study. Thus contributions of some sounds may be more salient. For this reason careful listening is required, to detect not only the obvious meanings but also the more subtle ones.

This is especially true for prosodic features, which often seem to be trump cards, dominating other contributions to the perceived meaning (although Bolinger (1989) probably overstates the case with the suggestion, with reference to *huh*, *hunh*, and *hm*, that 'prosody is fairly decisive, in fact this interjection might almost be regarded as a mere intonation carrier'.) Fortunately, highly expressive non-lexical utterances, with complex contours carrying complex meanings (Luthy 1983), were rare in this corpus. Indeed, almost all tokens had an almost flat pitch, so this was not a big problem in practice. Prosody is discussed further in Section 8.

The Compositional Hypothesis also implies that the sound-meaning mappings are context-independent: that each sound bears the same meaning regardless of the context. This may not be completely true, for both phonetic context and discourse context. First, it is possible that the contribution of one sound could be masked or shifted by the meaning contributions of neighboring sounds. Second it is clear that the discourse context affects interpretations; this will be discussed in Section 7.4.

6 SOUND-MEANING CORRESPONDENCES

Having already a meaning for /m/, this section looks at the other common sound components.

6.1 Nasalization and /n/

(C has applied for a summer-abroad program)		
H:	*I bet you'll hear something soon.*	1
C:	I hope so. I just turned that in, though, like. A couple weeks ago, so.	2
H:	*yeah (slightly creaky)*	3
C:	you know what I mean, so	4
H:	*yeah, it might take a little longer*	5
C:	**nn-hn**	6

(5)

In Example 5 C's *nn-hn* seems to indicate that C had held this opinion all along; it effectively closes out this topic. Had the sound been *uh-huh*, without nasalization, it would instead imply that somehow H had offered new information, and leave open the possibility of more talk on this topic. Other nasalized versions, such as *uh-hnn* would, however, share the same meaning component seen in *nn-hn*.

(A is illustrating the difficulty of working with the International Phonetic Alphabet)		
A:	*she had to count them by hand from the print-out, because she didn't have any way of searching for these weird control characters*	1
J:	**nyeah-nyeah (low flat pitch, overlapping as A keeps talking)**	2
A:	*now I mean she could have gotten something that might have been able to do it, but*	3
J:	(interrupting) It's a pain, yeah	4

(6)

Similarly in Example 6, which occurs moments after J had mentioned a problem of using the IPA for corpus work, the *nyeah-nyeah*[1] seems to be serving to remind A of this, that she is already well aware of such difficulties, and by implication encouraging closure of this topic. An unnasalized *yeah-yeah* (in the same flat pitch) here would sound merely bored, without laying claim to prior knowledge.

The *nyaa-haao* back in Example 2 line 4 is slightly different. In this case the fact which it reacts to has not been previously mentioned explicitly, but it is nevertheless obvious — from the previous context it is clear where the story is leading, and when X finally gets to the point, it seems that M has already seen it coming, as indicated by this nasalized token.

Nasalized non-lexical sounds generally mean not just that that the speaker has pre-knowledge of something, but that the something is already established, and known to the interlocutor too. (This 'pre-knowledge' is related to the notions of 'old information', 'given information' and 'common ground', but is often based on extra-linguistic knowledge.)

[1] not to be confused with the *nyah-nyah* of playground taunts, which is creaky, has a low vowel, and has a pitch downstep

In Example 7, V's *nn-nnn* conveys not only a negative answer[2] but also that V is surprised by the question, probably because he considers that M should have already known the answer, to the extent that his statement that he slept for most of the train ride implies that he experienced no problems. A similar usage is probably also present in the examples of Jefferson's (1978) study, in which she characterizes 3 nasalized tokens, *ne:uh*, *nyem*, and *mnuh*, occurring in response to questions, as indicating that the person who asked the question already 'knows the answer' or should be able to infer it easily.

(at the start of a recording session M throws out a first topic)		
M:	*So, V, tell me, tell me what you saw on the train, that, because I slept for an hour in be-, sort of in the middle*	1
V:	well, I slept, I slept for most of the train ride, actually. The one up here?	2
M:	*yeah*	3
V:	the Tokyo, the Tokyo, train ride, the Shinkansen	4
M:	*so, did you have a problem with your ears popping?*	5
V:	**nn-nnn**. You did?	6
M:	*Yeah, I did, actually . . .*	7

(7)

Nasalization and /n/ often seem to signal the following function:

(Covering Old Ground. Conversations often re-cover old ground: things that came up earlier get repeated or referred back to. While this may involve literal repetition, it may just be the expression of things that are obvious or redundant since inferable from what has gone before. People sometimes indicate it when they are expressing something that is somehow covering old ground, or to indicate that they think the other person is doing this, whether deliberately or inadvertently.)

Of the 20 occurrences of nasalized non-lexical items in the corpus, 12 seem to mark the covering of old ground or expression of information already known. Of these, 11 were in reference to something said by the other person (7 after a restatement of something that had already surfaced in the conversation or was otherwise obvious, and 4 after the other person has said something that the speaker could have predicted or seems to consider well-known). 1 occurs after the speaker himself has said something that he appears to consider well known, as part of an apparent bid to close out the topic. There are 4 cases which seem to lack any meaning of pre-knowledge[3]. 4 cases are unclear in meaning[4].

[2] probably due to the sharp pitch downstep and a glottal stop

[3] 2 of these occur where the speaker is somewhat taken aback.

[4] Of these, in 2 cases the sound itself, considered in isolation, does appear to connote a claim of pre-knowledge, but it occurs in a context where it seems unlikely the speaker could really have already known the information the interlocutor had just conveyed. Listening to the conversation after-the-fact, these items seem slightly rude, making the speaker sound like a know-it-all. Given the context, however (all were back-channels, all overlapped long continued speech by the interlocutor, and all occurred at times when the speaker seemed uninterested in the topic), perhaps meaning was merely 'I already know as much about that topic as I want to, so we can move on to another topic'. Under this interpretation there is a similarity to the already-known meaning.

6.2 Breathiness and /h/

hmm, unlike *um* and *mm*, occurs only as a back-channel. Moreover, *hmm*, compared to *mm*, seems to be bearing some extra respect and expressing a willingness to not only listen, but to give the other person's words some weight. A correlation with deference is also seen by the fact that *hmm*, *hm* and *mm-hm* tend to be produced by lower-status speakers: in the 3 conversations in the corpus where the interlocutors were significantly unequal in age and social status, 12 out of the 13 occurrences of these items were produced by the younger speaker.

Breathiness is also a factor distinguishing the agreeable *uh-huh* from the *uh-uh* of denial[5].

/h/ being related to relative social status and functional role (back-channel versus filler), it is hard to find clear minimal pairs, with and without breathiness, for the same speaker and the same functional role. Example 8 is a rare example: here the *uhh*, unlike O's other fillers, is breathy. This may be marking some trepidation, in that this occurs at the point where O, for the first time in a long conversation with a senior person in his research field, ventures to make a joke.

(after some talk on the merits of goats versus llamas as pack animals)		
O:	uu (creaky) they, they carry quite a bit, compared to their body weight, um, ... And **uhh (breathy)**, if you bring a female goat you can (pause) drink her milk and make yogurt and (pause) (laughs)	1
T:	*(laughs)*	2
O:	(pause) you don't need to turn back, head back to town ever, you know (laughing)	3

(8)

In Example 9, the *huh* (in falling pitch and of moderate duration) is a challenge, but it is a polite one, an attempt to engage the other person in discussion, in contrast to the flat contradiction which an *uh* would convey.

(X thinks that a reporter is biased)		
X:	*and he always hypes everything up*	1
M:	wow	2
X:	*is what I've heard*	3
M:	**huh**, that isn't the impression I've gotten	4

(9)

(Concern. Sometimes people in conversation are lacking in confidence or somehow dependent on the other person, and they sometimes signal this. While speaking, they may be solicitous or tentative, as if fearing that the other person will find their words stupid or inappropriate; while listening, they may listen with extra concern and attention. This often occurs when the other person is older or in a position of power, but arises more generally at points in a conversation when one person for a moment treats the other person's words or thoughts with extra respect or consideration.)

[5] and is probably stronger than the other two factors: absence of glottal stop and final pitch rise

Of the 43 tokens with /h/ or breathiness, 23 appear to bear a meaning of concern, deference or engagement. Of the remainder, 9 were two-syllable sounds which would have seemed rude had breathiness not been present. There were also 3 cases which were borderline laughter, 2 sighs, and 1 where the breathiness seemed to soften a contradiction.

6.3 Creaky Voice

(discussing who is likely to be at the party)		
H:	*and, um, and that other guy, K, majoring in Psychology.*	1
C:	**yeah(creaky)**. (two second pause) **yeah**, they're so fun.	2
H:	*(pause) That's cool*	3
C:	**yeah**	4

(10)

In Example 10 C's first *yeah* offers confirmation of a factual matter, in response to H's uncertain-sounding statement of what she thinks K's major is. The subsequent *yeah*s relate to subjective impressions. Perceptually the first *yeah* sounds authoritative and the others do not. The most salient phonetic difference is creakiness. This can be considered as indicating a sort of detachment in the sense that C's creaky response is not merely a polite acknowledgment of H's statement for the sake of continuing the conversation, but reflects that C is stepping back and providing an evaluation of H's statement based on C's independent knowledge.

(talking at a conference, resuming after an interruption)		
R:	let's see, so we were talking about what my favorite	1
X:	*yeah*	2
R:	talks were. **Um**, actually right now I'm sort of interested in what this U-tree algorithm is, because	3
X:	*yeah*	4
R:	I've done a, **um (creaky)**, a search, or, a literature search a while back on, on reinforcement learning and ...	5

(11)

Similarly, in Example 11 R's second *um* is creaky; at that point R is about to reveal that he is somewhat of an authority on the topic of learning algorithms, not merely chatting about them to politely pass the time. His first *um* was not creaky, and, as a statement of personal taste, would have sounded strange if it were.

(T is driving, O is navigating)		
T:	*shall I just go in here and turn around, n (and)*	1
O:	**yyyeah**. **Yeah (creaky)**, that might be best	2

(12)

Example 12 has a pair where the first *yeah* is uncertain, and the second, creaky one, produced after due deliberation, sounds authoritative.

(discussing whether it would be fun to go to the beach)		
H:	*I want it to be sunny*	1
C:	I know, this weather is no good	2
H:	*No, it makes me like groggy, kind of, you know what I mean, like*	3
C:	like you want to stay in bed and	4
H:	*yeah (slightly creaky)*	5
C:	just like watch a movie or something	6
H:	*and like not really do anything*	7
C:	**yeah (very creaky)** I know. I'm trying to fight it (laughs)	8

(13)

In Example 13 H complains about the weather and how it affects her, but C then reveals that she feels exactly the same way. Her *yeah*, being creaky, seems to indicate that C has personal experience, indeed she seems to be taking a moment here to actually indulge in that feeling. A non-creaky *yeah* would be less appropriate here, although it would be fine in an expression of merely perfunctory sympathy.

Creak also has a possibly related function in which it occurs with items which indicate detachment in the form of a momentary withdrawal to take stock of the situation. In Example 14 J infers that A's mother was from North Germany, but A then corrects her. After A clarifies the location of Wiesbaden, J talks to herself for a moment while he continues, then she produces a creaky *okay* and then a normal *yeah*. It seems as if J is withdrawing from the conversation to consult and correct her mental map of German dialects, and indicating this with the creakiness of the *yeah*, before returning to full attention and participation with the *okay*.

(regarding trills in German)		
A:	*but I think my mother does the, ah, the uvular one all the time . . .*	1
J:	. . . your Mom's from North?	2
A:	*no, she's from ah Wiesbaden, which is uh*	3
J:	don't know	4
A:	*in the central west;*	5
J:	mm	6
A:	*it's right near Frankfurt, west of Frankfurt;*	7
J:	center, north, west, **yeah (slightly creaky)** okay	8
A:	*and ah, or she's from that area, she's actually from a small town . . .*	9

(14)

The slightly creaky *yeah* back in Example 5 is another example of creakiness in response to correction; H produces it after realizing that she had misunderstood the situation.

(Claiming Authority. Although people sometimes say things lightly, other times they really know what they are talking about. Thus some things people say in conversation are intended as authoritative statements, advice, opinions, decisions, recollections, etc., and often speakers will indicate that these are intended as such. Authoritative statements may be based on expert knowledge of some topic, on direct experience, and so on.)

Of the 56 tokens in the corpus which were creaky or partially creaky, 38 seemed to indicate authority[6].

6.4 Click

The meaning of tongue clicks can be subsumed under the term *personal dissatisfaction*.

(C has suggested going to the beach; H responds by describing her homework assignments)		
H:	*like I haven't like corrected my paper, and re-printed it*	1
C:	<**click**>**-oh** (slightly breathy, low fairly flat pitch)	2

(15)

In Example 15, C's click seems to be indicating dissatisfaction with the situation, namely the fact that H can't come, and perhaps dissatisfaction with H herself, in the form of a mild remonstrance.

Some clicks seem to indicate dissatisfaction with the current topic, or the lack of one; these uses often occur near topic change points. In Example 16, M produces clicks while searching for a topic, before introducing a new topic, and when closing out a topic.

(M is trying to find a new topic at the start of the recording session)		
M:	aoo (creaky), let's see what other exciting things have been, worth chatting about. <**click**> uuuuuuu (creaky). (3 second pause) <**click**> Really good low budget movie you might want to rent . . . (M describes movie for 25 seconds, X seems uninterested) . . . <**click**> was quite well done	1
X:	*(3 second pause) I'm probably not going to rent that any time soon, because (changes topic)*	2

(16)

The click in Example 17 seems to express E's dissatisfaction with his own performance as a conversationalist, and marks the point where he gives up on one formulation and re-starts his explanation on a new tack.

(E is trying to describe simply a highly technical line of research)		
X:	*so, what are you doing, actually?*	1
E:	well, hhh-uuuh, at the moment I'm doing phonological modeling, and essentially trying to get, umm (pause) <**click**> Trying to develop models of	2

(17)

[6]Of the remainder, 5 back-channels seemed to indicate boredom, lack of interest, or impatience, 1 annoyance, 3 taking stock after being corrected by the other person, as in examples Example 5 and Example 14, and 1 occurred as the speaker (while driving) was executing a turn and apparently signaled concentration on that to the exclusion of attention to the conversation.

(Dissatisfaction. People in conversation are sometimes momentarily unhappy but then move on, and they often indicate when they do this. This momentary unhappiness can be about the conversation itself, as when the conversation hits a rough spot, one runs out of things to talk about, or when one has a problem expressing oneself fluently; or the unhappiness can be about the topic, as discussing something of which one disapproves or finds disappointing.)

Of the 26 clicks in the corpus, 19 seemed be expressing some form of dissatisfaction. Of these 9 seemed to express self-remonstrance, either at forgetting something, at getting off track, or at explaining something poorly (these sometimes co-occurring with the close of a digression or a re-start of an explanation on another tack), 4 seemed to indicate dissatisfaction with the current topic, co-occurring with a bid to close it off. 3 seemed to express dissatisfaction with the situation under discussion, and 3 seemed to be dissatisfaction directed to the interlocutor, as a form of remonstrance[7].

6.5 /o/

It is well known that the expression *oh* can mark the receipt of new information, among other functions (Heritage 1984; Schiffrin 1987; Fox Tree & Shrock 1999; Fischer 2000), and this is seen in the corpus too, as in Example 18. Other times it performs related functions, such as indicating the successful identification of a referent introduced by the other speaker, and the uptake of self-produced new information, as a result of figuring something out or noticing it.

(after X has explained that he is collecting conversation data)		
E:	is there any particular topic that we should, uh	1
X:	*no*	2
E:	**oh**.	3
X:	*so*	4
E:	So it's just	5
X:	*so, yeah*	6
E:	ookay	7

(18)

It is worth noting that the *oh* often occurs, not at the moment where the new information is heard, but a fraction of a second later, after the information has been assimilated somewhat and the listener has decided what stance to take regarding it, as seen in Example 19.

(regarding who buys Sailor Moon comic books in Japan)		
X:	*there's two audiences for that, one is the junior high school girls, and the other is the pervert, the uh, the, the perverts*	1
M:	yeah, **oh** absolutely, yeah, yeah	2

(19)

[7]Of the remainder, 5 seemed to simply mark the introduction of a new topic, and 1 marked a shift in conversation style from serious to facetious.

okay seems to share with *oh* some element of meaning, as Beach (1993) has observed, and this is likely due to the shared /o/. This is seen by the fact that the newness is downgraded in cases where the /o/ is reduced to a schwa (*ukay* as in Example 23), elided completely (*kay*), or replaced by a nasal (*m-kay*, *n-kay*, and *unkay*). On the other hand, where the newness of the information is significant, the /o/ is lengthened or repeated, forming *ookay* (as in Example 18) or *oh-okay*.

(New Information. People in conversation sometimes encounter information which is new to them, and may signal that they are aware of, or want to draw attention to, that newness. This may be done in reference to one's own utterances or in reference to the other's utterances. The new information may have been introduced by the other speaker, or may be self-produced, as a result of figuring something out or noticing it. This new 'information' may also include a new topic or referent, or a surprising turn of the conversation, etc.)

Of the 46 tokens containing /o/, 44 bear a new-information meaning.

6.6 /a/

Sometimes people in conversation are passive or at a loss, and other times they are fully in control and know exactly what they're doing. /a/ seems to signal the latter: that the speaker is fully on top of the situation and *ready to act* [8].

(X is winding up a roundabout explanation of why he's recording conversations)		
X:	*. . . when does it happen in English? is the question*	1
E:	right	2
X:	*and I have no data*	3
E:	**ah (creaky), okay (slightly creaky)**	4

(20)

In Example 20 the /a/ seems to indicate this. Indeed, it could be glossed as 'I've got it, I understand the whole picture, I'm very familiar with that kind of situation, I could finish your story for you'. This is in contrast to an /o/, which would stress the novelty of the information that X lacked data, and in contrast with schwa, which would imply that E was not sure what to say, perhaps having failed to understand the statement or its significance.

(over dinner after a conference)		
E:	did you go to the talk?	1
X:	*which one?*	2
E:	did you go to my talk? I should say	3
X:	*I missed it, I'm sorry*	4
E:	**ao (creaky)** *that's fine, that's fine.*	5

(21)

[8] This claim appears to conflict with Bolinger's (1989) remark that 'since the vowel is neutral, it fluctuates nonsignificantly, easily verging on [o] or [a]', but Bolinger was focusing on exclamations of surprise, which may not follow the same rules as more conversational non-lexical utterances (Section 9.4).

A similar meaning is seen in Example 21, where *ao*, instead of *oh*, seems to connote that E had half-expected X to have missed the talk, and is already prepared and willing to give him the gist of it, as he then goes on to do.

A similar distinction between /a/ and schwa may be seen in fillers and disfluency markers. *ah* seems to be used (for those speakers who use both *uh* and *ah*) in cases where the filled pause is being produced mostly for the benefit of the listener. That is, /a/ occurs when the speaker knows exactly what he wants to say, and the purpose of the filled pause is only to give the listener time to re-orient or catch up. This is seen in the last line of Example 14, where *ah* introduces a parenthetical remark, and in the third line of Example 14, where the *ah* precedes code-switching from English pronunciation to German pronunciation.

$$\left(\begin{array}{l}\text{In Control. Although sometimes people in conversation are momentarily passive}\\ \text{and drifting or at a loss, there are times when they are fully in control, knowing}\\ \text{exactly what to say or do next, and people in this state often indicate it. As a special}\\ \text{case, this is seen when a speaker is pausing, not because he's stuck for how to say}\\ \text{something, but semi-deliberately to warn the listener that something complex, like}\\ \text{a borrowing from a foreign language is coming up.}\end{array}\right)$$

This seems compatible with Fischer's (2000) observation that *ah*, in comparison to *oh*, 'does not display emotional content' and indicates that 'I want to say some more'.

Quantifying the strength of the association between /a/ and readiness to act is complicated by the fact that some speakers use *ah* but not *uh* as a filler, and others always use *aum* but not *um*. Thus for some speakers /a/ is perhaps a mere allophone, the variant of schwa used in fillers and disfluency markers.

Of the 18 tokens in the corpus containing /a/, 9 seem to manifest some such meaning of being in control, including 4 which preceded foreign language words.

6.7 Schwa

Schwa is the most common sound in non-lexical items in the corpus. It seems to be *neutral*, bearing almost no information. This is seen in the stereotypical back-channel *uh-huh*, which sometimes conveys essentially nothing but 'I'm still here', and in the stereotypical filler *uh*, which often conveys almost nothing but 'I'm starting to talk'.

This neutrality can be seen by contrasting schwa with /o/. Consider Example 22, where the long *oh* indicates that B now understands why A is upset about having missed the meeting. In contrast, a schwa-based sound, such as *uh* or *uh-huh* here would not indicate this at all.

(after some talk about a meeting that H feels bad about having missed)		
H:	*and then, like, at the end you're supposed to, like, split up into, like, program groups, but.*	1
C:	**ooooh**	2

(22)

Similarly in Example 23, A acknowledges receipt of new information with *okay*, but produces *ukay*s (with schwa) when he is acknowledging only the receipt of confirmation of what he already knew.

sound	meaning	strength of support in corpus	other
/m/	thought-worthy	strong	strong
nasalization	covering old ground	weak	weak
/h/ and breathiness	concern	moderate	–
creaky voice	claiming authority	moderate	–
clicks	dissatisfaction	moderate	–
/o/	new information	strong	weak
/a/	in control	weak	weak
schwa	neutral	weak	–

Table 5: Summary of the Meanings Conveyed by some Common Sound Components. Descriptions in the 'meaning' column are highly abbreviated.

(23)

(J is starting to describe an interesting conference talk)		
J:	*she works at Bell Labs, and what they do is, they do diphone concatenation. okay*	1
A:	**okay**, yeah, now, that's like the TrueTalk system, right? that's the AT&T one	2
J:	*iiyyeahh,*	3
A:	**ukay**	4
J:	*that's AT&T,*	5
A:	**ukay**	6
J:	*yeah. So, um, basically they're doing diphone concatenation and ...*	7

While there are minimal pairs with and without schwa, such as *um* and *mm*, there seems to be no difference in the basic meaning; rather the versions with vowels just seem to express the basic meaning more confidently, as one would expect from the greater loudness (Section 8).

6.8 Summary of Correspondences

Thus there is a candidate for the meaning associated with most of the sound components common in non-lexical utterances. These sound-meaning correspondences are summarized in Table 5. Of the sound-meaning mappings identified, some are strong, obvious, and almost invariant, some are fairly limited, weak, or tentative, and the others are in between, as seen in the table.

It is worth noting that these sound-meaning correspondences show up even when working under the assumptions of compositionality and context-independence, which are probably not entirely correct.

Interestingly, some of the sound-meaning correspondences identified above for English

also appear to be present in Japanese (Ward 1998; Okamoto & Ward 2002; Ward & Okamoto 2003), raising the question of whether universal tendencies are at work.

At this point is is worth mentioning some other phonetic features that various researchers have implicated in various meanings. Lip rounding may indicate surprise, and it has even been suggested, for tokens indicating astonishment, that it is not tongue position but 'the rounding of *oh*, which distinguishes it from *ah*' (Bolinger 1989). Glottal stops are often implicated with a meaning of negation or denial. Vowel height may also be significant: 'the 'importance' of *ah*, for example, is consonant with the 'size' implication of the low vowels' (Bolinger 1989). Throat clearing has been observed to function as an indicator of upcoming speech (Poyatos 1993).

The sound-meaning correspondences provide answers for the second question posed in the introduction: what all the variants mean. The existence of these correspondences also answers the first question: the reason for the existence of so many variants is just that people in conversation have a large variety of (combinations) of meanings that they need to express.

7 THE STRENGTH OF THE SOUND-MEANING CORRESPONDENCES

This section discusses the strength of the proposed sound-meaning correspondences and the power and limits of the compositional hypothesis.

7.1 Evaluation of the Compositional Hypothesis and the Compositional Model

According to the compositional hypothesis, the meaning of a non-lexical utterance is predictable from the meaning of its component sounds. Although this was a useful working hypothesis, it is clear that it is not invariably true, as witnessed by the exceptions noted above. In some cases it is clear where compositionality fails. For example, examining the properties of the four tokens which are exceptions to the correlation between /n/ and preknowledge (Section 6.1), it turns out that two of these were the shortest of the all sounds involving /n/, which suggests that somehow the lack of duration is canceling or overriding the contribution of /n/. Also, very quiet sounds appear to convey little or no meaning, regardless of their phonetic content.

It therefore is necessary to reject the Compositional Hypothesis as a full account.

However, since most non-lexical tokens in conversation seem to be largely compositional in meaning, the idea is worth salvaging. I therefore propose a COMPOSITIONAL MODEL for non-lexical conversational sounds, specifying that the meaning of a whole is the sum of the meanings of the component sounds. Compared to the alternative, a list-based model which associates meanings arbitrarily with fixed sequences of sounds, the compositional model explains the meanings of rare items as well as common ones, and does so parsimoniously. Thus, in this sense also, these items are truly non-lexical.

Having seen that the compositional model is better than the alternative, it is of interest to consider how well it does absolutely, so the rest of the section examines how much of the corpus data is accounted for by the compositional model.

The power of the sound-meaning correspondences can be quantified by counting failures, that is, cases where one or more predictions is not borne out. These were listed in Sections 5.1 and 6 for each sound, and summarized again in Table 6. Summing across all 316 tokens, for 77 (24%) the model predicts some meaning element that is not found. In other words, in 86% of the tokens all of the meanings associated with the component sounds were found. It is important not to ascribe too much significance to this number, since it depends on the specific corpus and on the subjective judgments of one person, nevertheless it does suggest that the model has substantial explanatory power.

Thus the model can predict meanings given sounds. The model also gives predictions in the reverse direction: starting from the meanings to be expressed and predicting which sound combination a person will use. The strength of these reverse predictions can be quantified by counting failures, namely cases where some expected sound component S is absent, that is, where a non-lexical sound bears some meaning which is not associated with any of the components of that token. This is unfortunately difficult to measure. One big problem is that almost every non-lexical item has a meaning which is richer or more specific than that predicted by the model. This 'failure' is pervasive, because the current model does not say anything about the way the context contributes to the full interpretation (Section 7.4). The second big problem is that there are alternative ways to express any given meaning, and meaning M may be conveyed not with S but with prosody or timing, etc. Just as a reference to a pet as *a stupid feline* does not constitute a counterexample to the mapping between the word *cat* and the meaning 'cat', the expression of an M without the use of the corresponding S does not count as a counterexample to the S-M mapping.

Nevertheless it is easy to make a rough count of cases where the unpredicted elements of the meaning of a token include one or more of the meanings in Table 5: that is, one of the meanings on the list is present somewhere when it is not expected. There are 101 such tokens, 32% of the total. This means that in the corpus, when a speaker used a non-lexical utterance to express some of the meanings on the list, 68% of the time he used all the sound components associated with those meanings.

This 32% is small enough to lay rest one concern: that the meanings identified for the sounds might be so vague that they can be found just about anywhere. Rather, the meanings identified are found in only a fraction of the tokens, and they occur mostly with the sounds they map to.

In a future study it would be interesting to attempt a quantitative formulation of the compositional model and the sound-meaning correspondences. This could allow testing the extent to which the meaning of a non-lexical utterance is actually equal to the SUM of the meanings of the component sounds, as has been attempted for Japanese (Okamoto & Ward 2002). It might also allow direct evaluation of the explanatory power of compositionality, without the need to identify any specific sound-meaning correspondences. It might also allow the quantitative statement of sound-meaning correspondences, for example, relating the degree of nasalization to the degree of fore-knowledge expressed.

7.2 A Complex Token

The primary value of the compositional model is the explanations it provides and the number of observations it organizes, as seen in the previous section. But it also provides simple explanations for some complex cases.

sound	meaning	total	predictions correct	incorrect	unclear
/m/	thought-worthy	57	54 (95%)	3	0
nasalization	covering old ground	21	11 (52%)	2	8
/h/ and breathiness	concern	43	29 (67%)	2	12
creaky voice	claiming authority	56	46 (82%)	3	7
clicks	dissatisfaction	26	20 (77%)	5	1
/o/	new information	47	44 (94%)	2	1
/a/	in control	11	5 (45%)	5	1
schwa	neutral	109	-	-	-
all tokens	composite of all predictions	316 (100%)	273 (86%)	18 (6%)	25 (8%)

Table 6: Summary of the Evidence for each Sound-Meaning Mapping. 'Total' is the total number of tokens containing the given sound component; this is also the number of tokens for which the model predicts the presence of the given meaning. 'Correct predictions' is the number of tokens with that sound which do in fact bear the predicted meaning. 'Incorrect predictions' is the number of tokens with that sound which do not bear the predicted meaning. 'Unclear' is the number of tokens for which it is impossible to tell whether the meaning includes the predicted meaning, generally because the token is 'unclear in meaning' in the sense of Section 5.3. The last row is not the sum of the others due to tokens including multiple sounds, sound components not covered by the model, and schwa, whose meaning cannot be observed directly.

Consider the example *<click>-naa(creaky)* in Example 24, a token with four sound components: a click, /n/, creakiness, and a neutral vowel (for this speaker /a/ and schwa do not appear to be contrasted in non-lexical utterances). The meaning of this utterance includes the speaker's chagrin or annoyance (the click) at finding that he had just described a video for 40 seconds to someone who has no possibility of seeing it, plus an indication that he has recalled that he knew (the /n/) that X lacked a T.V., plus a momentary withdrawal (the creakiness) to take stock of the newly recalled information. Thus the meaning is as predicted by the model.

(M has recommended a movie for X to rent)

X:	*I'm probably not going to rent that anytime soon because ...*	1
M:	myeah	2
X:	*because I don't have a video. (punchline intonation)*	3
M:	**<click>-naa(creaky)**	4
X:	*I don't have a T.V. <click>*	5
M:	<click>-neeu, that's right, you're one of those	6

(24)

7.3 The Compatible Meaning Constraint

The compositional model is useful in another way also. As mentioned in Section 4.4, the first-pass, purely phonological model of which non-lexical sounds can occur in conversation is inaccurate. For example, it generates such items as *mo* and *yeom*, which are implausible as English non-lexical expressions in conversation. Given the compositional model, there is an obvious way to explain why some of these items are implausible: a Compatible Meaning Constraint, stating that a non-lexical utterance can only contain sounds whose meanings are compatible.

Thus *uh-huh* is a plausible sound: deference and a non-committal, neutral attitude go well together, so the combination of /h/ and /ə/ is allowed. *mo*, however, is less plausible. It could only be appropriate in a situation where the speaker is both contemplating something and being in a state of having just assimilated some new information. Although not unimaginable, this would require a rather unusual state of mind or a rather unusual context. The addition of this constraint thus gives an improved model of intuitions about which non-lexical utterances can be used in conversation.

It can also provide a test for the model, at least in principle. It predicts that no non-lexical utterance will be observed which combine sounds whose meanings are incompatible. Applying this in practice, however, is not straightforward. Consider creakiness and /h/. These two sounds can co-occur, the model says, if there exist contexts in which a speaker is speaking with detached authority and yet concerned about the other's reaction. These two properties seem, at first blush, to be incompatible, and from this assumption one might predict that no token will both contain /h/ and be creaky. However such a sound exists in the corpus, as seen in Example 25.

(25)

(at the start of the conversation, after X has told E the purpose of recording it)		
E:	so maybe you should have told me at the end,	1
X:	*that's true*	2
E:	but then I'd have, I would have been nervous all the way through, so,	3
X:	*right, yeah*	4
E:	you know, it's like, 'hmm, why is he recording my voice, **hmmmmm (creaky)**', (laughs)	5

Here E is evoking a situation in which he is withdrawn and thinking suspiciously about X's intentions, but at the same time he is describing this imagined scene as part of making a joke. Thus in this context both detachment and engagement are present and expressed in a single non-lexical item. *nyaa-haao* in Example 2 is also a case where two superficially incompatible meanings, of /o/ and of nasalization, co-occur: here the speaker is indicating that something is simultaneously new (the stabbing) and yet predicted (the occurrence of some kind of violence). The moral of these examples is that it is not always trivial to predict compatibility of meaning from first principles; in order to reliably apply the Compatible Meaning Constraint will probably require an inventory of all the communicative needs which arise during conversation, in all their multi-faceted complexity, clearly a long-term goal.

Application of the Compatible Meaning Constraint to longer utterances may be further complicated by the fact that speakers' mental state can change rapidly. If, as seems likely,

the sound at each instant within a non-lexical utterance reflects the speaker's mental state at that instant, incompatible meanings may be seen at different points within the utterance. *nyaa-haao* may be an example of this also.

7.4 Compositional Meaning and Pragmatic Force

Considering the small list of meanings identified for the sound components, it is obvious that the compositional model does not provide a complete account of how non-lexical utterances are used. What it can specify is how a sound maps to a basic meaning. What it can not address is the meaning and pragmatic force at other levels. (The difficulty of making a clean semantics-pragmatics distinction here, as more generally, does not alter the fact that it appears wise not to require a single model to cover all levels of meaning.)

In particular, the compositional meaning does not fully describe the role a sound in a specific conversational context. As Fischer (2000) observes, dialog tokens may have 'under-specified meanings ... that are specified by means of reference to particular aspects of the communicative situation'. For example, for clicks the model ascribes a meaning of dissatisfaction, but it does not specify when a speaker can use this meaning to reproach someone else or to reproach himself or to echo someone else's dissatisfaction or to mark dissatisfaction with the current topic. More subtly, the actual pragmatic force borne by any specific occurrence may depend on the exact timing relative to the discourse.

This problem is, of course, not unique to non-lexical utterances. When someone uses the word *reliable* the semantics is clear, but it is not until a context is given that you know whether the pragmatic implication is that the car is stodgy, or able to skip an oil change, or deserving of affection, or worth more than your offer. But the role of context is even more significant for these items, since their basic meanings are so vague. Certainly some non-lexical items seem to have special roles in specific activities, such as joint projects (Bangerter & Clark 2003), joking, making plans, telling stories, explaining, complaining, and so on. (In this regard it is worth mentioning the existence of items which are generated by the phonological model of Section 4.2, but which are only marginally non-lexical, in that they have fairly fixed forms, fairly specific roles, and appear only in fairly specific, non-conversational contexts, such as *yo*, *oomm*, *hohoho*, *ahem*, and the gustatory *mmm* (Wiggins 2002).) In unstructured conversational interaction also, non-lexical items may assume specific meanings or roles in specific contexts.

Fortunately there is a large body of research focusing on just this problem, namely that in the 'Conversational Analysis' tradition. That body of work pays attention to observing and characterizing in detail the diverse situations in which people in conversation find themselves, and the various ways they employ language resources to meet their goals in those situations. In the words of Hutchby and Wooffitt (1999), Conversation Analysis is 'only marginally interested in language as such, its actual object of study is the interactional organization of social activities', and in Conversation Analysis the items 'used in talk are not studied as semantic units, but as products or objects which are ... used in terms of the activities' in the talk.

Thus the approach of the present study is complementary to the work in the Conversation Analysis tradition. This is one reason why the analyses here make little direct contact with the central concerns in that body of work. However the connections can be made. For example, Gardner (1997) characterizes *mm* as a 'weak and variable acknowledging token', in comparison to items such as *yeah* (Jefferson 1984), whereas the present analysis characterizes /m/, and by

implication *mm*, as indicating contemplation. The two descriptions are compatible: Gardner's work describes the various kinds of pragmatic force that /m/ can bear in the diverse contexts where it appears, while the current model specifies the basic, context-invariant meaning of /m/. Moreover a connection can easily be made — it seems reasonable that a speaker being contemplative is likely to acknowledge only weakly. Working out such connections is left as a topic for future work.

8 PROSODY-MEANING CORRESPONDENCES

While a proper analysis of the prosody of non-lexical utterances is beyond the scope of this paper, it is worth considering at least the most accessible such property, namely syllabification. This is so salient that it is reflected in the conventional spellings, as in *mm-mm* vs. *mm*, *uh-huh* vs. *uh* and *yeah-yeah* vs. *yeah*.

Two-syllable items seem to signal the intention to take a listening role, to indicate that the person who produces them intends to say no more. Evidence for this includes the fact that *yeah-yeah* only functions as a back-channel, in contrast to *yeah* which appears in many roles (Table 2). Similarly *uh-huh* and *um-hm* are overwhelmingly back-channels, versus single-syllable *uh* and *um* which are overwhelmingly fillers and disfluency markers.

One speaker produced four-syllable items, *uhn-hm-uh-hm* and *um-hm-uh-hm*, and these appeared to contrast with *um-hm*: the four-syllable forms signaled a posture of continued listening, but the two-syllable *um-hm* was less passive, sometimes produced only shortly before he interrupted and took a turn.

Gardner (1997) has also noted that *mm-hm*, in comparison to *mm*, is typically 'passing up an opportunity to speak, handing the floor straight back to the prior speaker'.

By implication, the fact that you have nothing to add can serve to be encouraging the interlocutor to continue. Often, as with *uh-huh*, this is a purely passive posture. Other times, as with *yeah-yeah*, this can encourage the interlocutor to stop repeating himself and get to the point, as in Example 26. (Also, *yeah-yeah* in a creaky voice, and with a sharp downstep in pitch to add brusqueness, is a stereotypical way to say 'enough already, let's drop this topic'.)

(discussing a party they might go to)		
H:	Is it like a party, like, 'rave' type party? or like	1
C:	*well, it's someone's house*	2
H:	**yeah**	3
C:	*there's going to be, I mean there's like, they're going to be spinning. So, in that sense, maybe, but it's just at someone's house, like*	4
H:	**yeh-yeah**	5
C:	*it's in the middle of the night, that too, but.*	6

(26)

Although multiple syllables are most common in back-channels, some syllabification also occurs in other positions, and with the same meaning. In Example 27 the *uuuh* has three energy peaks, and sounds frustrated: this can be ascribed to the fact that O wanted to say what to do next (for the sound appears where it can only be interpreted as a filler), but is

simultaneously realizing that he doesn't know and so can say no more, as conveyed by the syllabification.

(T is driving, O is navigating)		
O:	can we turn here? can, can we make a right turn here?	1
T:	*If you say so*	2
O:	um, oh, I guess we can't (embarrassed laugh). No. (laugh)	3
T:	*what? no.*	4
O:	**uuuh.** hmm	5
T:	*should we turn around and go back?*	6
O:	uh-mm . . . (waits until the next intersection comes up before deciding)	7

(27)

Thus the meaning conveyed by syllabification seems to be as follows:

$$\left(\begin{array}{l}\text{Lack of Anything to Say. Sometimes people in conversation have nothing to say;}\\ \text{for a moment or two they are just content to listen and/or remain silent, and they}\\ \text{sometimes indicate this.}\end{array}\right)$$

Examining the tokens with syllabification (excluding *okay* and variants, which are intrinsically double-syllabled), 38 of the 60 seem to be indicating such a lack of anything to say. Interestingly, reduplication in Japanese back-channels also seems to be meaningful, although the meaning may not be the same (Katagiri *et al.* 1999).

It is worth noting that the multiple-syllable cases are generally not simply repetitions of a single syllable. Rather they generally include one or more additional phonetic features which mark the syllable boundaries, most commonly energy dip, pitch dip, breathiness, or creakiness. The choice of how to realize syllabification is perhaps independent of the choice of syllabification itself; thus, for example, when a syllable boundary marked with breathiness is present, it may convey both the meaning of breathiness and the meaning of syllabification. Incidentally, the term 'syllabification' is more appropriate than 'reduplication', or 'repetition' because the syllable boundaries appear in various realizations, with various strengths, and in various numbers.

sound	meaning
syllabification	lack of desire to talk
duration	amount of thought
loudness	confidence, importance
pitch downslope/upslope	degree of understanding / lack thereof
pitch height	degree of interest

Table 7: Meanings Speculatively Attributed to Some Prosodic Features

Other prosodic features clearly also contribute to the meanings of non-lexical utterances. Table 7 summarizes other likely correspondences, based on analysis reported elsewhere (Ward 2004). Interestingly, these prosodic features generally seem to bear much the same meanings

here as they do in sentences (Tench 1996). There are probably also other meaningful prosodic features; for example, abruptness of energy drop, giving a clipped sound, may be a 'gesture of finality' (Bolinger 1946). Multi-syllabic sounds occasionally have more complex prosodic contours (Hockey 1992), and the meanings are probably also more complex. There is also evidence that longer durations may strengthen the perceived meaning (Fox Tree 2002).

9 ROLES OF NON-LEXICAL UTTERANCES

The English language, it is generally acknowledged, includes 40-some phonemes, which are concatenated to form words, and the meanings of those words are arbitrary (in Saussure's sense). Non-lexical conversational sounds, however, appear to form a subsystem based on 10 sounds, which are concatenated or superposed to form items whose meanings are largely predictable from the sounds.

Is this plausible? Why on earth should the English language include a subsystem like this? The ultimate explanation may lay outside linguistics, perhaps referring to neural structures (Lamendella 1977; Jaffe 1978), evolution (McCune 2000), ethology (Ohala 1984), or articulatory effort (Nenova *et al.* 2001). This paper, however, examines only acoustic considerations and some aspects of human cognitive processing in the functional and positional roles where non-lexical utterances are common, and this is the topic of this section.

9.1 As Back-Channels

As language users, humans suffer from two fundamental cognitive limitations. First a person generally cannot produce coherent utterances while listening to someone else (Jaffe 1978). Second, symmetrically, in general a person cannot really listen (process speech input) while talking.

Back-channels somehow escape both these limitations. (For present purposes back-channels are optional responses to something said by the other which do not require acknowledgement.) First, back-channels are produced while the other person has the turn, and often while the other is talking (Ward & Tsukahara 2000). Second back-channels can be heard and understood — at least well enough to get a sense of whether the other person is confused, bored, excited, knowledgeable, supportive, and so on — by a person who is himself talking. Thus, people, both as speakers and as listeners, can process back-channels simultaneously with processing the 'content' of the conversation on the 'main channel' (Yngve 1970).

Given these characteristics of the back-channel role, it is significant that non-lexical back-channels use sounds which are distinguishable from those in the main channel. This can explain why a limited inventory of sounds is common (Section 4.2): these sounds are relatively non-interfering, acoustically, with the sounds of English lexical items, since some are not found in words, and the others are few in number and lack sharp transitions. Cross-linguistically also, back-channels are drawn largely from a small set of phonetic components (Allwood 1993).

These characteristics of the back-channel role also have implications for the sound-meaning mappings. If dealing with the main message keeps the main language-processing resources of the participants' brains occupied, one might expect the sound-meaning mappings for back-channel items to be computationally simple so that they can be dealt with elsewhere. For the

person hearing the back-channels, this would allow decodability by simple neural pathways (Jaffe 1978), distinct from those used to handle arbitrary (lexical) sound-meaning mappings. Similarly, a simple encoding would be desirable for the person producing the back-channels, whose brain is not only busy, but is also operating under a time constraint, the need to produce a back-channel within a narrow time window before the opportunity to be relevant is lost. This can explain why the sound-meaning mappings in non-lexical back-channels are simple.

This may also be why non-lexical back-channels have long been an area of interest within the study of conversation, human communication, and interpersonal interaction (Yngve 1970; Duncan & Fiske 1985; Schegloff 1982). In terms of the mental processes involved they may have as much in common with gestures (with which they often co-occur) as with the rest of verbal language.

These characteristics of the back-channel role also have implications for the sorts of information that can be conveyed. Because the participants' primary attention will generally be occupied with the main message (the content), one might expect the information in the back-channel to be of semantically different kinds. This explains why non-lexical back-channels convey the limited meanings they do.

9.2 As Fillers and Disfluency Markers

Many utterances are framed or interrupted by hesitations and formulation problems. In this paper the terms filler and disfluency have been used for turn-initial or utterance-initial items, and turn-initial items, respectively (although this two-way taxonomy of hesitations etc. is perhaps not the best possible taxonomy (Hieke 1981).)

In these roles similar constraints are present. For the speaker's sake, the items in these roles also should have sound-meaning correspondences which are simple, so he can generate them while he is busy working out what to say and how to say it. For the listener's sake, they should be phonetically distinguished from the sounds used in the main channel, so that they can be easily filtered out (below the level of conscious processing) and processed separately from the main message. It is well known that fillers are phonetically and prosodically distinguishable from words (O'Shaughnessy 1992; Shriberg 1999; Goto *et al.* 1999; Shriberg 2001; Wu & Yan 2001); indeed they tend to be even less complex and varied than the back-channels. Cross-linguistically also, non-lexical fillers exhibit only limited variation (Clark & Fox Tree 2002).

Thus, considering both the speaker's and listener's needs, and for both phonetic and cognitive reasons, the phonetic and sound-symbolic properties identified are well suited to the roles of fillers and disfluencies.

9.3 As Confirmations and Clause-Final Tokens

There are also less common roles for non-lexical utterances. These include 'confirmations', that is, responses to back-channels, for example the *nn-hn* in Example 5 and the first *ukay* in Example 23. There are also clause-final tokens, which typically express attitude, such as the *yeah* at the end of Example 6.

While the cognitive and acoustic considerations above apply to these roles weakly, if at all, they have a clear family resemblance to back-channels, fillers, and disfluency markers. These roles can all be characterized as closely relating to, but not themselves part of, the main channel.

9.4 As Isolates

Rather different are the constraints on 'isolates', which in this paper means utterances produced when neither person has the turn; these typically more self-directed than other-directed.

Here again there is a need for items whose sound-meaning mappings are simple enough to process while doing something else. For back-channels, the 'something else' was just listening, for fillers, formulating, but here it includes extra-linguistic activities such as thinking one's own thoughts, looking around, working, and so on. Isolates also need to be easily distinguishable from normal language, but for a different reason than that seen for back-channels, namely, so that the interlocutor or bystanders know that you're not talking to them, nor to voices in your head (Goffman 1981). Acoustic non-interference with the main channel is not, however, a requirement for these items, and indeed these items often involve sounds outside the inventory of Section 4.2.

While some of the isolates in the corpus are explicable within the model, others are not; for example *wow*, *oop-ep-oop* (produced while trying to catch a falling lamp) and *achh* in the corpus are flagrant violations. Although these tokens did occur during conversations, they are different from the others in the corpus in that they stand outside the dialog more than they belong to it, or, in Ameka's (1992) phrase, they 'do not have addressees'. These items bear relatively little relation to other utterances in the conversation, and in that respect are not as conversational as back-channels and fillers.

9.5 As Responses

The final role where non-lexical utterances are common is as responses to direct questions and to high-rise statements.

Here again there is a family resemblance to back-channels, but many responses seem to be more in the main channel than not.

This may be a good place to address the question of *no*, which seems at first glance to be a clear exception to the sound-meaning correspondences (old ground for /n/ and new information for /o/). However *no* is fairly clearly a word. Since the model only attempts to account for non-lexical utterances, there is no reason to expect the same sound-meaning correspondences to apply. (Interestingly, however, in British English *no* is well attested as a back-channel, and in that role it is often not an expression of disagreement, but rather an acknowledging or even affiliating token (Jefferson 2002).)

It is worth noting that the use of non-lexical tokens as responses to direct questions can indicate social status (Andersen *et al.* 1999) and is often stigmatized. In a court case I recently participated in, one witness, a child, often used *uh-huh* and the like during cross-examination. To us on the jury this was informative: her exact choice of token told not only whether she agreed with the attorney's characterizations of events, but also whether she fully understood the question, whether she accepted the presuppositions, whether she thought the

answer obvious, or difficult to recall, or beside the point, and so on. But this was deemed unacceptable: the attorney considered it imprecise, and indicative of a flippant attitude, and admonished her with a little lecture to the effect that 'the people now in the courtroom may understand what you mean, but the court recorder is going to transcribe it as *uh-huh*, and later people who look at the transcript will have no idea'. The child, suitably chastened, thereafter restricted her responses to *yes* and *no*.

For the sake of completeness, other positions in which non-lexical items appeared in the corpus (the 'other' category in Table 2) include within quotations and in a few other positions difficult to characterize.

9.6 Summary

Thus the phonetic and sound-symbolic properties which the model ascribes to conversational non-lexical sounds are well suited to the characteristics of the contexts where they occur. Specifically, there are several roles which are outside the main channel but which are nevertheless conversational, or, in Clark's (1996) terminology, are 'collateral' signals rather than part of the 'official business' of the dialog. In these roles there are three types of considerations — involving the acoustic properties, involving the properties of the sound-meaning mapping, and involving the types of meanings conveyed — all of which make these roles hospitable to a class of sounds which have special properties — being phonetically distinctive, compositional in meaning, and related to conversation control and a few other functions.

Crystal's notion of a 'scale of linguisticness' is also useful here (Crystal 1974). 'At the 'most linguistic' polarity would be classified those features of utterance most readily describable in terms of closed systems of contrasts, which have a relatively clear phonetic definition, which display evidence of a hierarchical structure, and which are relatively easily integrated with other aspects of linguistic structure ... At the other 'least linguistic' end would be placed those features of utterances which seem to have little potential for entering into systemic relationships, which are relatively indiscrete, and which have a relatively isolated function and little integrability with other aspects of language structure ...' Although this scale was initially proposed for characterizing paralinguistic vocal effects, it also can be used to describe positional and functional roles in terms of the sorts of items which tend to inhabit them: the main-channel is very linguistic; short responses somewhat less so; filler, disfluency and back-channel positions even less linguistic, and interjections barely linguistic at all. The compositional model best accounts for items in roles towards the middle of this continuum.

As Goffman (1981) suggested, there seems to be a 'division of linguistic labor' between 'nonword vocalizations' and 'full-fledged words', and 'the character of the word bears the mark of the use that is destined for it'. Although Goffman had a different focus in mind, this is also apt as a description of the complementary roles of lexical and non-lexical utterances in conversation.

Thus the idea that within English there exists a separate subsystem with these properties is not so implausible after all; rather there is a natural match for a certain communicative niche.

It may be possible to extend the idea of "natural match" further, to argue that specific sounds naturally represent specific meanings, as has been argued for other forms of sound symbolism since at least Darwin. Although there may be similarities across languages, as

noted earlier, it is also clear that different languages and cultures use non-lexical items in different ways; so to a large extent they clearly form a social code, interpretable only with reference to that code.

10 CONVERSATIONAL GRUNTS AND RELATED PHENOMENA

10.1 Conversational Grunts

The research strategy taken here was to explain everything about all non-lexical utterances in a corpus of conversations. The phenomena to examine were chosen using a negative criterion, as the set of sounds which are conversational but not laughter or words. This set did not, however, turn out to be coherent: there were items which did not pattern with the others, notably breath noises, interjections, and word fragments.

However the vast majority of the non-lexical tokens are well covered by the model. Since this set of items has several distinctive and co-occurring properties, it is worth inventing a term. Thus I will use CONVERSATIONAL GRUNTS to refer to those items which are generated by the phonological model, exhibit the sound-meaning correspondences, have compositional meanings, and occur in conversational roles other than the main-channel.

This appears to be a graded category. At the core are pure grunts, such as *uh* and *uh-huh*. *oh* is also mostly grunt-like, although it sometimes appears in the main channel (James 1972). More marginal as a grunt is *yeah*, which includes component phonemes which have rather constrained distributions and whose meaning seems less compositional. Even less grunt-like is *okay* which has a phoneme which is not productive in grunts and which has no identifiable intrinsic meaning (the /k/). The marginal status of these items as grunts was in fact seen earlier: they were major sources of exceptions to the phonological model and to the sound-meaning correspondences.

10.2 Near Grunts

Since the properties of conversational grunts are well suited to certain functional positions, one might expect these same properties to also infuse non-grunts occurring in these positions. And indeed, the pronunciation of at least one word, *right* appears to vary in accordance with the sound-meaning correspondences identified above.

Right was the third most frequent non-grunt item occurring in the corpus in any of the functional categories seen in Table 2. Right was also the most frequent lexical response token in McCarthy's (2003) larger corpus.

(28)

(V had asked M to go shopping with him, but M had then changed the topic)		
V:	So I have to find a pair of running shoes, still; to get back to that topic, because, I need a pair.	1
M:	*Well, we'll see what we can do.*	2
V:	**Alright**. Because I want to go running ...	3

In Example 28, the *alright* (pronounced roughly as /araɪt/) is clearly similar in sound and meaning to *right*, but the presence of an additional /a/ seems to indicate only provisional acceptance, coupled with the intention to get his way; compatible with the meaning for /a/ identified in Section 6.6.

(M is discussing a disturbing Japanese web site found while surfing)		
M:	It was very much an emphasis on youth and . . . fascination with rape	1
X:	*uh-hm*	2
M:	it seemed,	3
X:	*uh-hm*	4
M:	which was kind of, odd.	5
X:	*It is true, that Japanese comics, right, there's a lot of them are really violent*	6
M:	**mright**, right(creaky)	7
X:	*like you can, just, <click>, but those, yeah, yeah*	8
M:	but I mean it, I, you know, better to exist on the funny pages than to exist on the street, I suppose, assuming that's the choice ...	9

(29)

In Example 29, the *mright* is clearly similar to *right*, but also bears a meaning of contemplation, as identified for /m/ in Section 5.1.

Thus *right* can be considered to be a NEAR GRUNT. Although a word, it is near the borderline: not only does it allow sound-symbolic modification, but it also lacks a clear referential meaning and clear participation in syntactic constructions.

More speculatively, items like *you know* and *and*, although clearly lexical, also bear interesting resemblances to conversational grunts in function and in phonetic inventory, at least when reduced.

To summarize with a metaphor, the speaker of English is a cook with many options. He can use a pure grunt — a sauce made fresh for the occasion, from scratch from basic ingredients. Or he can use near-grunts or frozen grunts, pre-prepared and usable as-is, but still amenable to freshening up with a sound-symbolic ingredient.

10.3 Laughter

Laughter, although prosodically unique, phonetically mostly falls with in the coverage of the phonological model of non-lexical utterances (Section 4.2). There have been suggestions that laughter involves sound-meaning correspondences, for Japanese, French, and English (Kori 1987; Léon 1991; Mowrer *et al.* 1987), and some of these correspondences seem to relate to those seen in grunts. Investigating this would go beyond the scope of the present paper, but it is interesting to note that the general acoustic properties of laughter — breathiness, large pitch range, lack of a clear pitch contour, multiple syllables — predict, using the present model, that laughter will generally be polite and engaged, display interest, be obscure in intent, and display a willingness to continue to listen; and in fact these properties are generally present in laughter as it occurs during conversation. This is true both for laughter as such and for words 'said with a laugh'.

11 SUMMARY

This paper has discussed the non-lexical conversational sounds of American English in an integrative way: considering sound, meaning, and function together, examining these items across variety of roles and functions, and treating unusual items such as *myeah*, *uh-nh* and *un-kay* together with better-known items such as *uh-huh* and *mm*.

This has led to a model in which non-lexical conversational utterances are productive combinations of 10 component sounds. This phonological model differs from that for the phonology of words, in that it includes nasalization, clicks, breathiness and creakiness, and in that it excludes all but a few of the phonemes of lexical English. This model provides a good match to the 316 conversational tokens observed in a corpus, and does even better when augmented with the Compatible Meaning Constraint.

The model also associates with each component sound a meaning or function, implying that these non-lexical utterances exhibit sound symbolism and that their meanings are largely compositional. These mappings apply to conversational non-lexical items across roles and across contexts, explaining most of the meaning of core members of the category of conversational grunts and also explaining part of the meanings of some other tokens. Although this claim, that sound symbolism is present in conversational grunts, is largely an extrapolation and synthesis of some suggestions in the literature, this paper is the first to identify many specific sound-meaning mappings, the first to present a systematic list of mappings, and the first to present detailed evidence for mappings.

This paper has further shown that this model of conversational grunt sounds and meanings is plausible in view of the roles that non-lexical utterances play in human conversation and the constraints of human cognitive processing.

This paper was based on an integrative (or broad-and-shallow), bottom-up approach to the phenomena. As such, there are several drawbacks with the results so far: clearly the analysis is incomplete, is lacking solid theoretical foundations, and is suggestive rather than definitive. Many questions about non-lexical items of course remain open, for example: regarding the phonetic details, regarding the details of how meanings function in specific contexts, regarding mental representation and processing (Bergen 2004), regarding the nature of variation in the use and interpretation of these tokens among speakers and among hearers, and regarding their status as a semiotic system. However this approach was successful to the extent of leading to a simple model, to a number of far-reaching generalizations, and to a host of interesting further questions.

References

Abelin, Asa (1999). Phonesthemes in Swedish. In *International Congress of the Phonetic Sciences*, pp. 1333–1336.

Allwood, Jens (1993). Feedback in Second Language Acquisition. In Clive Perdue, editor, *Adult Language Acquisition: Cross Linguistic Perspectives, II: The Results*, pp. 196–235. Cambridge University Press.

Allwood, Jens & Elisabeth Ahlsen (1999). Learning how to manage communication with special reference to the acquisition of linguistic feedback. *Journal of Pragmatics*, 31:1353–1389.

Ameka, Felix (1992). Interjections: The Universal yet Neglected Part of Speech. *Journal of Pragmatics*, 18:101–118.

Andersen, Elaine S., Maquela Brizuela, Beatrize DuPuy, & Laura Gonnerman (1999). Cross-Linguistic Evidence for the Early Acquisition of Discourse Markers and Register Variables. *Journal of Pragmatics*, 31:1339–1351.

Bangerter, Adrian & Herbert H. Clark (2003). Navigating Joint Projects with Dialog. *Cognitive Science*, 27:195–225.

Barr, Dale J. (2001). Paralinguistic correlates of discourse structure. In *Poster presented at the 42nd annual meeting of the Psychonomic Society*.

Batliner, Anton, A. Kiessling, S. Burger, & E. Noeth (1995). Filled Pauses in Spontaneous Speech. Technical Report 88, Verbmobil Project. also in ICPhS'95.

Beach, Wayne A. (1993). Transitional Regularities for 'Casual' "Okay" Usages. *Journal of Pragmatics*, 19:325–352.

Bergen, Benjamin K. (2004). The Psychological Reality of Phonaesthemes. *Language*, 80:290–311.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, & Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. Pearson Education Limited.

Bolinger, Dwight (1946). Thoughts on 'Yep' and 'Nope'. *American Speech*, 21:90–95.

Bolinger, Dwight (1989). *Intonation and Its Uses*. Stanford University Press.

Brennan, Susan E. & Michael F. Schober (2001). How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, 44:274–296.

Brennan, Susan E. & Maurice Williams (1995). The Feeling of Another's Knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34:383–398.

Clark, Herbert H. (1996). *Using Language*. Cambridge University Press.

Clark, Herbert H. & Jean E. Fox Tree (2002). Using *uh* and *um* in Spontaneous Dialog. *Cognition*, 84:73–111.

Corley, Martin & Robert J. Hartsuiker (2003). Hesitation in Speech can ... um ... Help a Listener Understand. In *Proceedings of the 25th Meeting of the Cognitive Science Society*.

Crystal, David (1974). Paralinguistics. In Thomas A. Sebeok, editor, *Linguistics and Adjacent Arts and Sciences*, pp. 265–295. Mouton.

de Saussure, Ferdinand (1915; 1959). *Course in General Linguistics*. McGraw-Hill.

Duncan, Jr., Starkey & Donald W. Fiske (1985). The Turn System. In Starkey Duncan, Jr. & Donald W. Fiske, editors, *Interaction Structure and Strategy*, pp. 43–64. Cambridge University Press.

Edwards, Jane & Martin Lampert (1993). *Talking Data*. Lawrence Erbaum Associates.

Ehlich, Konrad (1986). *Interjektionen*. Max Niemeyer Verlag, Tuebingen.

Erickson, Frederick (1979). Talking Down: Some Cultural Sources of Miscommunication in Interracial Interviews. In Aaron Wolfgang, editor, *Nonverbal Behavior: Applications and Cultural Implications*, pp. 99–126. Academic Press.

Fischer, Kerstin (2000). *From Cognitive Semantics to Lexical Pragmatics: The functional polysemy of discourse particles*. Mouton de Gruyter.

Fox Tree, Jean E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition*, 29:320–326.

Fox Tree, Jean E. (2002). Interpreting Pauses and Ums at Turn Exchanges. *Discourse Processes*, 24:37–55.

Fox Tree, Jean E. & Josef C. Shrock (1999). Discourse Markers in Spontaneous Speech: Oh What a Difference an Oh Makes. *Journal of Memory and Language*, 40:280–295.

Gardner, Rod (1997). The Conversation Object *Mm*: A weak and variable acknowledging token. *Research in Language and Social Interaction*, 30:131–156.

Gardner, Rod (1998). Between Speaking and Listening: the Vocalisation of Understandings. *Applied Linguistics*, 19:204–224.

Goffman, Erving (1981). Response Cries. In Erving Goffman, editor, *Forms of Talk*, pp. 78–122. Blackwell. originally in *Language* 54 (1978), pp. 787–815.

Goto, Masataka, Katunobu Itou, & Satoru Hayamizu (1999). A Real-time Filled Pause Detection System for Spontaneous Speech Recognition. In *Eurospeech '99*, pp. 227–230.

Hamaker, J., Y. Zeng, & J. Picone (1998). Rules and Guidelines for Transcription and Segmentation of the Switchboard Large Vocabulary Conversational Speech Recognition Corpus, Version 7.1. Technical report, Institute for Signal and Information Processing, Mississippi State University.

Heritage, John (1984). A Change-of-State Token and Aspects of its Sequential Placement. In J. Maxwell Atkinson & John Heritage, editors, *Structure of Social Actions: Studies in Conversation Analysis*, pp. 299–345. Cambridge University Press.

Hieke, Adolf E. (1981). A Content-Processing View of Hesitation Phenomena. *Language and Speech*, 24:147–160.

Hinton, Leanne, Joanna Nichols, & John J. Ohala, editors (1994). *Sound Symbolism*. Cambridge University Press.

Hockey, Beth Ann (1992). Prosody and the role of *okay* and *uh-huh* in discourse. In M. Bernstein, editor, *Eastern States Conference on Linguistics*, pp. 128–136.

Hutchby, Ian & Robin Wooffitt (1999). *Conversation Analysis*. Blackwell.

Iwase, Tatsuya & Nigel Ward (1998). Pacing Spoken Directions to Suit the Listener. In *International Conference on Spoken Language Processing*, pp. 1203–1206.

Jaffe, Joseph (1978). Parliamentary Procedure and the Brain. In Aron W. Siegman & Stanley Feldstein, editors, *Nonverbal Behavior and Communication*, pp. 55–66. Lawrence Erlbaum Associates.

Jakobson, Roman & Linda Waugh (1979). *The Sound Shape of Language*. Indiana University Press.

James, Deborah (1972). Some aspects of the syntax and semantics of interjections. In *CLS 8*, pp. 162–172.

Jefferson, Gail (1978). What's in a 'Nyem'? *Sociology*, 12:135–139.

Jefferson, Gail (1984). Notes on a Systematic Deployment of the Acknowledgement Tokens "Yeah" and "Mm hm". *Papers in Linguistics*, 17:197–216.

Jefferson, Gail (2002). Is "no" an Acknowledgement Token? Comparing American and British uses of (+)/(-) tokens. *Journal of Pragmatics*, pp. 1345–1383.

Katagiri, Yasuhiro, Miyoko Sugito, & Yasuko Nagano-Madsen (1999). Forms and Prosodic Characteristics of Backchannels in Tokyo and Osaka Japanese. *International Congress of the Phonetic Sciences*, pp. 2411–2414.

Kawamori, Masahito, Takeshi Kawabata, & Akira Shimazu (1995). A Phonological Study on Japanese Discourse Markers. In *9th Spoken Language Processing Workshop Notes (SIG-SLP-9)*, pp. 13–20. Information Processing Society of Japan.

Kokenawa, Yoko, Minoru Tsuzaki, Hiroaki Kato, & Yoshinori Sagisaka (2004). An analysis of speaking attitude manifested as fundamental frequency characteristics (in Japanese). In *52th Spoken Language Information Processing Workshop Notes (SIG-SLP-52)*, pp. 87–92. Information Processing Society of Japan.

Kori, Shiro (1987). Perceptual Dimensions of Laughter and their Acoustic Correlates. In *XIth International Congress of the Phonetic Sciences*, pp. vol. 4, 67.4.1–67.4.4.

Lamendella, John T. (1977). The Limbic System in Human Communication. In *Studies in Neurolinguistics, Volume 3*, pp. 157–222. Academic Press.

Léon, Pierre R. A. (1991). Riez-vous en Hi! Hi! Hi! ou en Ah! Ah! Ah! Oh! Oh! In *XIIth International Congress of the Phonetic Sciences*, pp. 310–314.

Louwerse, Max M. & Heather Hite Mitchell (2003). Toward a Taxonomy of a Set of Discourse Markers in Dialog: A Theoretical and Comptuational Linguistic Account. *Discourse Processes*, 35:199–239.

Luthy, Melvin J. (1983). Nonnative Speakers' Perceptions of English "Nonlexical" Intonation Signals. *Language Learning*, 33:19–36.

Magnus, Magaret (2000). *What's in a Word? Evidence for Phonosemantics*. PhD thesis, University of Trondheim.

Marslen-Wilson, William & Paul Warren (1994). Levels of Perceptual Representation and Process in Lexical Access: Words, Phonemes and Features. *Psychological Review*, 101:655–675.

McCarthy, Michael (2003). Talking Back: "Small" Interactional Response Tokens in Everyday Conversation. *Research on Language and Social Interaction*, 36:33–63.

McCune, Lorraine (2000). Grunts: A gateway to vocal communcation and language? In *3rd Conference on The Evolution of Language*. Paris.

Mowrer, Donald E., Leonard L. LaPointe, & James Case (1987). Analysis of Five Acoustic Correlates of Laughter. *Journal of Nonverbal Behavior*, 11:191–199.

Mulac, Anthony, Karen T. Erlandson, W. Jeffrey Farrar, Jennifer S. Hallett, Jennifer L. Molloy, & 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*, pp. 641–668.

Nenova, Nikolinka, Gina Joue, Ronan Reilly, & Julie Carson-Berndsen (2001). Sound and Function Regularities in Interjections. In *Disfluency in Spontaneous Speech*, pp. 49–52. ICSA.

Nicholson, H. B. M., E. G. Bard, A. H. Anderson, M. L. Flecah-Garcia, D. Kenicer, L. Smallwood, J. Mujllin, R. J. Lickley, & Y. Chen (2003). Disfluency under Feedback and Time Pressure. In *Eurospeech*, pp. 205–208.

Ohala, John J. (1984). An Ethological Perspective on Common Cross-language Utilization of F_0. *Phonetica*, 41:1–16.

Okamoto, Masafumi & Nigel Ward (2002). Aizuchi no Onkyoteki Yoso no Imi no Teiryoteki Suitei (Quantitative Estimation of the Meanings of the Phonetic Components of Backchannels). In *Special Interest Group in Spoken Language Understanding and Dialog (SIG-SLUD-A201-08)*, pp. 47–42. Japan Society for Artificial Intelligence.

O'Shaughnessy, Douglas (1992). Recognition of Hesitations in Spontaneous Speech. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. I–521–524.

Patzold, M. & A. Simpson (1995). An Acoustic Analysis of Hesitation Particles in German. In *International Congress of the Phonetic Sciences*, pp. 512–515.

Poyatos, Fernando (1975). Cross-Cultural Study of Paralinguistic "Alternants" in Face-to-Face Interaction. In Adam Kendon, Richard M. Harris, & Mary R. Key, editors, *Organization of Behavior in Face-to-Face Interaction*, pp. 285–314. Mouton.

Poyatos, Fernando (1993). *Paralanguage*. John Benjamins.

Rajan, Sonya, Scotty D. Craig, Barry Gholson, Natalie K. Person, & Arthur C. Graesser (2001). AutoTutor: Incorporating Back-Channel Feedback and Other Human-Like Conversational Behaviors into an Intelligent Tutoring System. *International Journal of Speech Technology*, 4:117–126.

Sapir, Edward (1929). A Study in Phonetic Symbolism. *Journal of Experimental Psychology*, 12:225–239.

Schegloff, Emanuel A. (1982). Discourse as an Interactional Achievement: Some Uses of "Uh huh" and Other Things that Come Between Sentences. In D. Tannen, editor, *Analyzing Discourse: Text and Talk*, pp. 71–93. Georgetown University Press.

Schiffrin, Deborah (1987). *Discourse Markers*. Cambridge University Press.

Schmandt, Chris (1994). *Computers and Communication*. Van Nostrand Reinhold.

Shinozaki, Tubasa & Masanobu Abe (1997). Kisoku Gosei Onsei de Yakudokan o Jitsugen suru Horyaku ni tsuite (A Strategy for Realizing Live Interaction with Synthesized Speech). In *17th Spoken Language Processing Workshop Notes (SIG-SLP-17)*, pp. 81–88. Information Processing Society of Japan.

Shinozaki, Tubasa & Masanobu Abe (1998). Development of CAI system employing synthesized speech responses. In *International Conference on Spoken Language Processing*, pp. 2855–2858.

Shriberg, Elizabeth (2001). To 'errr' is Human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31:153–169.

Shriberg, Elizabeth E. (1999). Phonetic Consequences of Speech Disfluency. In *Proceedings of the International Congress of the Phonetic Sciences, Volume 1*, pp. 619–622.

Smith, Vicki L. & Herbert H. Clark (1993). On the Course of Answering Questions. *Journal of Memory and Language*, 32:25–38.

Takubo, Yukinori (1994). Towards a Performance Model of Language. In *1st Spoken Language Information Processing Workshop Notes (SIG-SLP-1)*, pp. 15–22. Information Processing Society of Japan.

Takubo, Yukinori & Satoshi Kinsui (1997). Otoshi, Kandoshi no Danwateki Kino (The Conversation Functions of Responses and Exclamations). In *Bunpo to Onsei (Speech and Grammar)*, pp. 257–279. Kuroshio, Tokyo.

Tannen, Deborah (1990). *You Just Don't Understand: Men and women in conversation.* William Morrow.

Tench, Paul (1996). *The Intonation Systems of English.* Cassell.

Thorisson, Kristinn R. (1996). *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills.* PhD thesis, Massachusetts Institute of Technology, Media Laboratory.

Trager, George L. (1958). Paralanguage: A First Approximation. *Studies in Linguistics*, pp. 1–12.

Traum, David R. (2000). 20 Questions on Dialogue Act Taxonomies. *Journal of Semantics*, 17(1):7–30.

Tsukahara, Wataru & Nigel Ward (2001). Responding to Subtle, Fleeting Changes in the User's Internal State. In *CHI '01*, pp. 77–84. ACM.

Ward, Nigel (1998). The Relationship between Sound and Meaning in Japanese Back-channel Grunts. In *Proceedings of the 4th Annual Meeting of the (Japanese) Association for Natural Language Processing*, pp. 464–467.

Ward, Nigel (2000a). The Challenge of Non-lexical Speech Sounds. In *International Conference on Spoken Language Processing*, pp. II: 571–574.

Ward, Nigel (2000b). Issues in the Transcription of English Conversational Grunts. In *First (ACL) SIGdial Workshop on Discourse and Dialog*, pp. 29–35.

Ward, Nigel (2003). Didi, a Dialog Display and Labeling Tool. http://www.cs.utep.edu/nigel/didi/.

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

Ward, Nigel & Masafumi Okamoto (2003). Nasalization in Japanese Back-Channels bears Meaning. In *International Congress of the Phonetic Sciences*, pp. 635–638.

Ward, Nigel & Wataru Tsukahara (2000). Prosodic Features which Cue Back-Channel Feedback in English and Japanese. *Journal of Pragmatics*, 32:1177–1207.

Ward, Nigel & Wataru Tsukahara (2003). A Study in Responsiveness in Spoken Dialog. *International Journal of Human-Computer Studies*, 59:603–630.

Werner, Stefan (1991). Understanding "Hm", "Mhm", "Mmh". In *XIIth International Congress of the Phonetic Sciences*, pp. 446–448.

Wharton, Tim (2003). Interjections, Language, and the 'Showing/Saying' Continuum. *Pragmatics and Cognition*, 11:39–91.

Wiggins, Sally (2002). Talking with your Mouth Full: Gustatory *Mms* and the Embodiment of Pleasure. *Research on Language and Social Interaction*, 35:311–336.

Wu, Chung-Hsien & Gwo-Lang Yan (2001). Discriminative Disfluency Modeling for Spontaneous Speech Recognition. In *Proceedings of Eurospeech2001.*

Yngve, Victor (1970). On Getting a Word in Edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567–577.