Nasalization in Japanese Back-Channels bears Meaning

Nigel WARD[†] **and OKAMOTO, Masafumi**[‡]
†University of Texas at El Paso
‡University of Tokyo
nigel@cs.utep.edu, masafumi@isi.imi.i.u-tokyo.ac.jp

ABSTRACT

Many back-channels seem to be primarily sound-symbolic expressions, rather than fixed sequences of phonemes with arbitrarily associated meanings. The presence of a sound-symbolic meaning for one phonetic feature, nasalization, in back-channels in Japanese was investigated. Subjects were presented with tokens synthesized with and without nasalization and judged the differences in meaning for each pair. The results showed a fairly consistent meaning for nasalization, independent of the pragmatic context and of the phonetic context in which it occurred.

1. INTRODUCTION

Back-channels, such as *uh-huh, un-hn,* and *mm*, are ubiquitous in casual dialog in English and many other languages. Such utterances are known to play an important role on conveying attitude, information processing status, and conversation control intentions [1].

Back-channels display an unusual degree of fine phonetic variation. For example, in addition to the prototypical *uh-huh*, our small corpus of casual English conversations includes *m-hm, nn-hn, uh-hn, uh-uh, h-mmmm, hh-aaaah, hhh-uuuh, mm-hm, mm-mm, uh-mm*, and *ummum* (in coarse transcription*),* and also monosyllabic and further reduplicated forms. Such variation is difficult to explain as a result of assimilation or other phonological processes, since most back-channels occur in isolation.

The simplest way to account for such variation is probably to explain it in terms of meaning: subtle differences in pronunciation exist because of subtle differences in meaning. While it seems clear that an *uh-hn*, for example, does not mean the same thing as an *uh-huh*, pinning down the differences is difficult. Tentatively, however, for English, in many back-channels (and also in many fillers) the phonetic components appear to convey roughly the meanings shown in Table 1. Similar correspondences seem to be present in interjections in English [3], and in back-channels and fillers in Japanese [4].

In this account, such back-channels are not normal lexica items, with a fixed pronunciation and an arbitrarily associated meaning. Rather they are non-lexical creations, assembled on the fly from the sound components needed to express the speaker's precise intention and feeling at the time. That is, sound symbolism, well attested for semantic domains related to percepts, also seems to play a role in back-channels. If language involves two systems of mapping from sound to meaning mapping, the arbitrary and the motivated, then many back-channels seem to involve the latter type of mapping.

/m/	thought-worthy
nasalization	covering old ground
/h/ and breathiness	concern
creaky voice	claiming authority
clicks	dissatisfaction

Table 1: Some hypothesized sound-meaning correspondences in English back-channels, from [2].

These claims about the presence of sound-symbolism are based mostly on introspective analysis of the meanings for back-channels in corpus data [2]. This paper strengthens the case by presenting the first experimental evidence for a direct sound-meaning link in back-channels.

2. NASALIZATION IN JAPANESE BACK-CHANNELS

Nasal vowels are common in back-channels in Japanese. Indeed, the most frequent token is a nasalized schwa. (This sound is, interestingly, usually written *un* in (transliterated) standard Japanese orthography, although it in fact contains neither /u/ nor /n/. This sound violates the phonotactics for Japanese lexical items in two ways: first, the vowel is a central vowel, a schwa or slightly higher, distinct from the 5 vowels of Japanese lexical items, and second, the vowel is nasalized.) Nasal vowels are rare in Japanese overall, and nasalization is not phonemic in lexical items, as it occurs only in contexts where it is explicable by known phonological rules [5], and because there are no lexical minimal pairs with and without nasalization.

This prototypical back-channel occasionally alternates with a non-nasalized version, namely a simple schwa. Some other common back-channel forms, such as *e, o, a, ha:* and *he:,* also have two versions, sometimes appearing with a nasal vowel.

In Japanese back-channels nasalization is one of the sound components that seems to be meaning-bearing. Identifying the meaning is far from easy, but based on corpus data we have argued that the presence of nasalization indicates that the speaker is treating some contribution as old information and/or indicating agreement [4,6]. Since nasalization is uncontroversially without meaning in Japanese lexical items, a demonstration that it bears meaning in back-channels would be good evidence that back-channels are different from lexical items in the nature of the sound-meaning mapping, specifically in involving sound symbolism.

3. MODEL

The claim that back-channels involve sound symbolism entails that the meaning of a back-channel is (largely determined by) the sum of the meanings of each of its component sounds. For example, the corpus includes a token of [hã:], which exhibits 5 sound components: [h], [a], a long duration, nasalization, and a flat pitch, and whose meaning seems to include 5 components, each corresponding to a feature, namely deference, action orientation, the need for a moment to think, an intention to agree, and incomplete understanding.

For the current study we formalize this compositional model as follows:

1. The meanings of back-channels and of their sound components are expressed as continuous-valued real vectors ("meaning vectors"). Each dimension of a vector corresponds to one aspect of meaning, such as degree of understanding or degree of interest.

2. Each sound component has its own meaning vector, which is constant, not dependent on context.

3. The meaning vector of a back-channel token is the sum of those of its sound components.

According to this model, the meaning vector of a feature, such as nasalization, can be obtained by computing the difference between the meaning of two tokens, such as a nasalized back-channel and one which is non-nasalized but identical in every other respect.

4. HYPOTHESIS

Ultimately we would like to test the model and evaluate the explanatory power of sound symbolism. The aim of the current study, however, was only to test whether, in Japanese back-channels, nasalization bears meaning, as a demonstration that at least one phonetic component does bear meaning in at least one language.

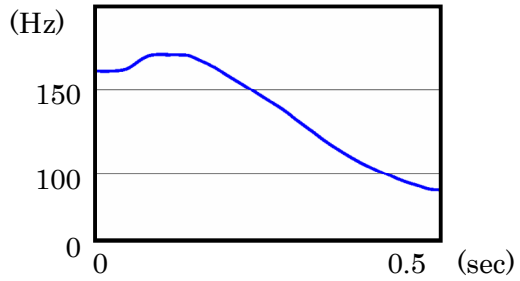

Figure 1: Pitch contour of the synthesized back-channels.

5. METHOD

Although ideally we would like to investigate nasalization across a widely representative set of back-channel forms, for practical reasons we limited attention to three pairs, namely [a] with and without nasalization, [o] with and without and [e] with and without. These were chosen because they seemed quite different in meaning, were relatively easy to synthesize, and were observed in corpus data.

As standard synthesizers of Japanese do not support the generation of nasal vowels, a Klatt synthesizer was used, with the height and strength of the formants adjusted following Maeda [8]. During the experiments subjects were asked which of the pair sounded more nasalized; the results showed that the tokens intended to sound nasalized were indeed so perceived. All tokens were synthesized with the same duration and intonation, as seen in Figure 1.

The contexts in which the back-channels tokens were presented were 3 utterances which in the corpus preceded back-channels and which were representative of contexts in which back-channels occur:

we are students from Mechanical Engineering,
: a statement made apparently with the intention of getting a perfunctory acknowledgment and then continuing

tomorrow I'm thinking of interviewing for a job,
: a statement introducing an interesting fact apparently made with the intention of eliciting a display of interest before continuing

if we split the cost three ways it's still 9000 yen,
: a statement introducing a problem apparently with the intention of getting an expression of sympathy before complaining in detail

These lead-in utterances were presented together with the synthesized back-channels, with the delay between lead-in and back-channel set to be the same as in the corpus.

Since it is impossible to quantitatively investigate all possible aspects of meanings, we limited attention to 8 dimensions of meanings which are commonly conveyed in

	[a]	[e]	[o]	[m]	[h]
1. I understand/agree	-0.51	-0.31	0.41		
2. I know about that topic			-0.86		-0.66
3. Please go on					
4. I've just recalled/thought of something		-0.27	0.31		
5. I am thinking	-0.51		-0.31	0.81	
6. I am interested	-0.46		1.24	-0.50	
7. I am surprised			1.26	-0.60	
8. I have a question/doubt					

Table 2:Significant portions of the meaning vectors for 5 component sounds, inferred from the pilot study.

back-channels, as seen in Tables 2 and 3. Nasalized and non-nasalized tokens were presented in pairs. For each of these 8 dimensions of meaning, subjects rated the meaning difference between the two versions: whether the second token, compared to the first token, was expressing more or less understanding, pre-knowledge, and so on, for each of the 8 dimensions of meaning. Rating was done on a 7-point scale, from –3 to +3.

The subjects were 6 naive native speakers of Japanese. Each subject judged 9 pairs of back-channels in all: each of 3 vowels in each of 3 contexts. Subjects were able to replay the samples as often as they wished before making the rating.

6. PILOT STUDY

As a pre-check, a study was run using this method but with a different token set. These tokens varied in vowel --- [a], [e], or [o] --- and in presence or absence of [h] and [m]. Table 2 shows those results that met a 1% confidence level by the t-test. Strong correlations were found between [o] and surprise and interest, and between the presence of [m] and the state of being in thought. Insofar as these correlations match introspection and common sense, the pilot study increases confidence in the experimental method.

7. RESULTS

Using the model of Section 3 and the method presented in Section 5, we computed the average meaning vector of nasalization across subjects, as seen in Table 3.

To investigate whether the values of this vector are reliable, we performed a one-sided t-test for each dimension of meaning. The null hypothesis is that the value of each dimension is 0: that is, that the presence or absence of nasalization has no clear effect on the meaning of the back-channel. At a 1% confidence level there is statistical significance on all dimensions except dimension 2 ("I know about that topic"); that is, nasalization conveys some meaning on at least these 7 dimensions.

Interpreting this result, nasalization in back-channels bears, albeit weakly, the following meanings:

- I don't understand/agree
- Please don't go on
- That doesn't bring anything to mind
- I am thinking
- I am not particularly interested
- I am not particularly surprised
- I have a question/doubt

To investigate the effect of the other factors on the meaning of nasalization, we also performed an ANOVA, with the explanatory variables being "subject", "base vowel", and "context of presentation". Considering both the main effects and the interactions, and using a 1% confidence level, a main effect of "subject" was seen in dimensions 6, 7, and 8 ("I am interested", "I am surprised", and "I have a question/doubt"). No other main effects or interaction effects were found. Thus, although individual differences did affect the interpretation of nasalization, base vowels and pragmatic context did not. This is somewhat counter-intuitive but is as predicted by the model of Section 3.

meaning dimension	value
1. I understand/agree	-0.63*
2. I know about that topic	-0.02
3. Please go on	-0.43*
4. I've just recalled/thought of something	-0.37*
5. I am thinking	+0.30*
6. I am interested	-0.78*
7. I am surprised	-0.48*
8. I have a question/doubt	+0.44*

Table 3: Inferred Meaning Vector for Nasalization.
* indicates significance.

8. DISCUSSION

This result is the first experimental demonstration that sound-symbolism can explain part of the meaning of back-channels.

However this does not show that a sound-symbolic model is better than the alternative, a traditional model in which the lexicon contains explicit lexical entries for both nasalized and non-nasalized versions of each back-channel. One way to help settle the issue would be to repeat the experiment with a larger variety of back-channels; another would be to investigate correlations between perceived meaning and continuous-valued phonetic features, such as the *degree* of nasalization.

The question of whether back-channels are indeed best modeled as sound-symbolic expressions is not just of theoretical importance. Back-channels, and non-lexical utterances in general, are a hallmark of spontaneous and casual speech. Current algorithms for speech recognition, speech synthesis, and dialog management are designed for lexical items, that is, they assume the tokens come with relatively fixed meanings and that they are arbitrarily associated with relatively fixed phoneme sequences. Until models suitable for handling tokens involving sound symbolism are developed, spoken dialog systems may be doomed to handle back-channels poorly, resulting in stilted dialog from the user's perspective.

Parenthetically, it is interesting that we did not find the expected positive correlations between the presence of nasalization and meaning dimensions 1 and 2. This might be due to differences in the amount of context considered: on the one hand the entire context of the conversation and on the other just a single lead-in utterance. Certainly there is a need for careful study of how the pragmatic force of a back-channel depends on its context.

Future work should also examine the meanings of the other components of back-channels, including not only phonetic features but also prosodic ones, including those identified in previous studies [9,10].

9. SUMMARY

An experiment with synthesized tokens shows that nasalization in Japanese back-channels bears meaning. This suggests that a model of the sound-meaning relation in these items may have to include a sound-symbolic component.

ACKNOWLEDGEMENTS

We thank the Inamori Foundation, the Sound Technology Promotion Foundation, the Japanese Ministry of Education, Professor Keikichi Hirose, and our subjects.

REFERENCES

[1] Nigel Ward and Wataru Tsukahara. Prosodic Features which Cue Back-Channel Feedback in English and Japanese. *Journal of Pragmatics,* 32, pp 1177—1207, 2000.

[2] Nigel Ward. A Model of Conversational Grunts in American English. manuscript, 2003.

[3] Nikolinka Nenova, Gina Joue, Ronan Reilly and Julie Carson-Berndsen. Sound and Function Regularities in Interjections. In Disfluency in Spontaneous Speech, pp. 49-52. International Speech Communication Association, 2001.

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

[5] Tsutomu Akamatsu. *Japanese Phonetics: Theory and Practice.* Munich*:* Lincom Europa, 1997.

[6] Nigel Ward. Ndakko (in Japanese). *Nihongogaku*, pp 4-5, April 2002.

[7] Masafumi Okamoto and Nigel Ward. Quantitative Estimation of the Meanings of the Phonetic Components of Back-channels (in Japanese). 35th Spoken Language Understanding and Discourse Workshop, pp. 47-52. Japan Society for Artificial Intelligence, 2002.

[8] Shinji Maeda. Acoustics of Vowel Nasalization and Articulatory Shifts in French Nasal Vowels. *Phonetics and Phonology 5: Nasals, Nasalization and the Velum*, pp. 147-167. Academic Press, 1993.

[9] Tubasa Shinozaki and Masanobu Abe. A Strategy for Realizing Live Interaction with Synthesized Speech (in Japanese). 17th Spoken Language Processing Workshop Notes (SIG-SLP-17), pp. 81-88, Information Processing Society of Japan, 1997.

[10] Stefan Werner. Understanding "Hm", "Hmm", "Mmh". International Congress of Phonetic Sciences XII, pp 446-448, 1991.