

Trouble in Mind: Paralinguistic Indices of Effort and Uncertainty in Communication

Dale J. Barr

Beckman Institute, University of Illinois at Urbana-Champaign, U.S.A.

Abstract

When we speak, we often encounter problems that compromise our ability to produce fluent speech. Sometimes we can't recall a word, do not feel confident in a proposition we wish to assert, or need extra time to plan a complex utterance. In conversation, these problems manifest themselves paralinguistically; for instance, as hesitations, filled pauses (e.g., in English, "um" and "uh"), or in prosodic contour. I hypothesize that these phenomena constitute a form of "vocal gesture" (Okrent, 2000), that, like manual gestures, can provide listeners with insight into the speaker's mind. I review three psycholinguistic experiments that suggest that these signals contribute to linguistic and conceptual coordination.

How we interpret language depends on not just *what* is said, but on *how* it is said. The current work investigates how listeners use information that a speaker expresses "paralinguistically"—that is, not in the propositional "text" of discourse but as part of the acoustic packaging of utterances. Specifically, the work focuses on a class of paralinguistic signals that can convey information about a speaker's certainty and mental effort. This class includes a broad set of phenomena, such as hesitations, filled pauses (in English, "um" and "uh"), and prosodic contour. It is hypothesized that these phenomena constitute a form of "vocal gesture" (Okrent, 2000) that can provide a window into the speaker's mind. I report three experiments that addressed the following questions: 1) what is the potential information value of these signals? 2) can listeners exploit them in understanding referential descriptions? and 3) can they use them to learn about the speaker's underlying conceptual representation?

The first experiment tested whether the presence of filled pauses in referential descriptions correlates with the old/new status of referents. It was hypothesized that speakers would be more likely to encounter problems, and hence, produce filled pauses, when describing a new referent (a referent not yet described) than when retrieving a name for an old referent (a referent that had been described previously). Eighteen native speakers of American English contributed data to the study.

Each speaker described abstract shapes to a research assistant. Pairs of shapes (see Figure 1 for an example) appeared on a computer screen, one of which was designated the "target". Speakers were informed of the location of the target through a set of headphones. Their task was to describe the shape to the addressee so that she could identify it.

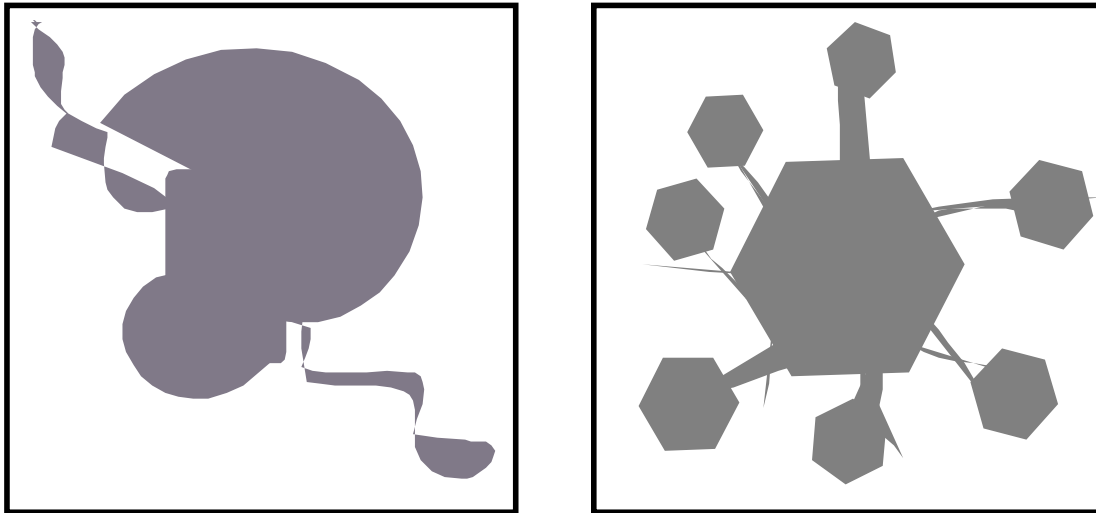


Figure 1. Sample stimuli used in filled pause production/comprehension experiments.

Statistical analysis revealed that speakers were more likely to produce a filled pause when describing a new than an old referent (mean proportions of .55 and .39, respectively; $t(17) = 2.838, p < .05$). More importantly, the form of an utterance-initial filled pause correlated with old/new status, $\chi^2(1) = 13.381, p < .001$. These data are presented in Table 1. An initial “um” was over twice as likely to precede a new than an old referent. Conversely, “uh” was twice as likely to be followed by an old than a new referent. A regression analysis revealed that hesitations following “uh” were shorter than those following “um”, $r = .46, F(1, 94) = 25.108, p < .001$, replicating the findings of Smith and Clark (1993). To sum up, speakers provide information both in the form and length of a disfluency that could potentially signal old/new status to an addressee.

		Form			Total
		none	“uh”	“um”	
Referent	Old	63	23	18	104
	New	45	11	44	100
	Total	108	34	62	204

Table 1. Distribution of filled pauses in old/new referent conditions.

The next experiment examined whether listeners are sensitive to filled pauses in interpreting referential descriptions. Participants viewed the pairs of abstract figures while they listened to a speaker’s pre-recorded descriptions. Some of the shapes were referred to multiple times. The test trials included an “old” shape that had been seen before but never described, and a “new” shape that the speaker had described twice. In half of the test trials, the speaker’s description was preceded by a filled pause. In one condition, a description of the new referent was preceded by an “um” and a pause of 1.815 sec. In the other, a description of the old referent was preceded by an “uh” and a pause of .618 sec. As a baseline, in the other half of test trials the pauses were replaced with incidental noise (a cough, a sniffle, clearing of the throat).

Three response variables were considered: eye movements, mouse movements, and response time. If listeners are sensitive to differences in the distribution of forms in old/new contexts, then when the “um” is utterance-initial they should be faster to look at, move the mouse toward, and respond to new referents than in the corresponding baseline condition. Conversely, in the “uh” case they should be faster in these measures for old referents.

Listeners were faster when descriptions of new referents were preceded by an “um” compared to incidental vocal noise. This was evident from a comparative advantage in response time of 308 ms ($t(15) = 2.458, p < .05$). Interestingly, listeners in the “um” condition were also comparatively more likely to begin moving the mouse toward the new referent during the pause, even before they had any supporting linguistic information ($t(15) = 2.380, p < .05$). It was surprising to find neither of these results corroborated in the patterns of eye movements. Nor was there any meaningful pattern in any of the variables for the “uh”-initial descriptions of old referents. In fact, the pattern was the opposite of what was predicted: after an “uh”, listeners were more likely to look at *new* instead of old referents, but this preference for new referents was virtually identical to baseline. The finding that listeners were sensitive to “um” but not “uh” suggests that listeners might use, albeit in a limited way, information about the form of a filled pause to constrain referential interpretation.

The last experiment used a concept learning task to investigate how speakers’ expressions of uncertainty in classification affect how listeners learn concepts. It was assumed that a speaker’s certainty would depend on the typicality of the instance to the category; typical instances would elicit confident assertions containing few hesitations, while less typical instances will produce more hesitations and an unsure intonation. In this manner, hesitations and intonation can provide information about category structure.

Participants in this experiment learned six sets of three novel color categories from a person who they believed was a “trained expert”. Each category had five instances, varying in typicality and proximity to the category boundary. The categories were associated with made-up names, such as “riallo”, “donlon”, “murray”, etc. The sets were presented in blocks divided into “study” and “test” phases. In the study phase, participants heard recordings of the speaker naming the color patches. In the test phase, participants attempted to correctly classify the patches. The phases were repeated until the participant met a learning criterion. Half of all of the study blocks were in the “consistent” condition, where the speaker’s certainty or uncertainty was consistent with the underlying structure of the category. The others were in the “inconsistent” condition, where the speaker’s certainty was inconsistent with the category structure.

Preliminary data from twelve participants indicate that they learned categories faster when supplied with confidence information that was consistent versus inconsistent with the underlying structure (means = 1.4 and 1.9 blocks, $t(11) = 2.171, p < .06$). Clearly, then, people attempt to use paralinguistic cues about certainty to infer the structure of a speaker’s representation of a category.

Conclusion

The vocal coordination signals that I have examined share some important properties with manual gestures (McNeill: 1992): They are ubiquitous, have an intimate relation with the timing of speech, and can provide a window onto the speaker’s mind. While it is unlikely that these signals

are “m-intended” in the sense that speakers “intend” to communicate referential or conceptual content with a filled pause, nonetheless I have shown that listeners readily exploit these signals to enhance linguistic and conceptual coordination. A point of departure for future research is the observation that the interpretation of these signals depends on the kind of trouble the speaker is experiencing. Speakers can encounter problems at different levels of language production, e.g., lexical access, speech planning, or articulation. In other words, the “meaning” of a disfluency depends on context, which would seem to require a sophisticated sensitivity on the part of the listener to the circumstances of language production.

References

McNeill, D., 1992, *Hand and mind: What gestures reveal about thought*, Chicago, University of Chicago Press.

Okrent, A. 2000, A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics (or at least give us some criteria to work with), in R. Meier, K. Cormier, & D. Quinto (eds.), *Modality and structure in signed and spoken languages*. CUP.

Smith, V. L., Clark, H. H., 1993, On the course of answering questions, *Journal of memory and language*, 32, 1, 25-38.