

# A multidimensional analysis method for think-aloud protocol data

Paul Hutchison, Isabel Monaghan, and Rachael Morgan

*Grinnell College, Department of Education*

*1115 8<sup>th</sup> Ave., Grinnell, IA, 50112*

**Abstract.** As part of a larger project we analyze think-aloud data to produce descriptions of thinking. This analysis requires inferring thinking from observable participant behaviors, primarily what participants say. To produce rich and reasonably accurate descriptions of the thinking we focus on several different features in the data. We analyze the participants' speech for both their description of their thinking and the insight provided into their context dependent expectations. We also attend to two non-verbal features in the data, gestures and pauses. In this paper we focus on each analytic feature, first describing the relevant research base and then explaining how we operationalize it in our analyses. We tentatively claim that coordinating the analyses of the four features produces more accurate descriptions of reasoning than traditional think-aloud analysis methods, which focus primarily on analyzing speech.

**PACS:** 01.40.Fk, 01.40.gf

## I. INTRODUCTION

Researchers use a variety of sources of data to gain insight into thought. We are using a think-aloud protocol [1] in our current study; others use videotaped classroom episodes, other types of interviews, or student generated products. In all of these, researchers must infer students' thinking based on the data, attending to both interpretive and contextual validity concerns. Interpretive validity refers to the match between the interpretation and the participants' actual thinking. We must also be concerned with the relationship between the data collection context and actual learning contexts, or contextual validity.

We illustrate interpretive validity first with an example from our data. In our study we collected think-aloud protocol (TAP) data from students in an introductory, calculus-based, college physics course. Participants were asked to think-aloud as they produced answers to a series of multiple-choice physical science questions. Charlie (all names are pseudonyms), is answering a Newton's Third Law question about the relative magnitude of the forces in a collision – in this question a large, moving truck collides with small, parked car. After reading the question aloud there is a brief pause before he says, “*I'm going to say A, um (pause) because it's (pause) moving and has the most weight*”. In this brief excerpt Charlie first announces his choice of option A - the truck exerts a larger force on the car than the car exerts back on the truck - without verbalizing anything about the cognitive process he used to produce the answer, although his justification includes information that suggests what his thinking might have been. He identifies movement and weight as relevant features, which suggest he may be drawing on physical intuitions. We feel somewhat confident about this inference, though we may be wrong. In other situations the data leaves us less certain about the participants' underlying

thought processes. Some participants provide very little verbal data to use to make inferences. Hence, when researchers infer thinking they must critically examine interpretive validity. We are grappling with this issue as we analyze our TAP data.

The other concern that must be addressed is contextual validity. This is particularly relevant when analyzing TAP data. Traditional analysis of this data assumes that a TAP verbal report provides direct access to a participant's thinking and that the process of verbalizing does not influence the thinking [1]. These assumptions seem implausible at first sight and are called into question by research evidence. For example, Smagorinsky [2] provides a theoretical critique based in cultural-historical activity theory. Deschambault [3] analyzes TAP data and shows features of the reports that indicate participants view it as talking-in-interaction rather than a direct report of thinking. Both of these papers call into question contextual validity.

Even with these concerns we see value in collecting and analyzing TAP data, primarily in the potential for stronger interpretive validity compared to other data sources. Consider the relative merits of TAP data and videotapes of classroom episodes. While a classroom episode provides greater contextual validity, the concurrent report of a TAP can give a researcher insight into thinking not available in the classroom data and hence provides the potential for greater interpretive validity. Some who analyze classroom data seek to get similar reports by inviting participants to view and comment on classroom videos after the fact. However, a participant's after-the-fact report provides their plausible guess at their own reasoning, which is less valuable than a concurrent report.

Our point here is that all of these data sources present a researcher with some combination of contextual validity and interpretive validity challenges. In analyzing our TAP data we sought to create a methodology that enhances

interpretive validity while also providing insight that will help us attend to concerns about contextual validity.

Broadly, our analytic method is to seek multiple dimensions of analysis in an effort to triangulate our inferences with as much data as possible. Traditional TAP analysis focuses entirely on analyzing the transcript. We also analyze three additional dimensions of the data. We assess how participant's expectations about the interview context impact what they say by analyzing the transcripts for linguistic markers of expectations [4]. Then, because we videotape our TAP interviews, we are able to identify and interpret gestures, which provide data that can enrich our understanding of the thinking process [5]. Finally, in selective moments of interest we attend to the participant's speech pauses, which indicate moments of cognitive planning [6] and can be useful to interpret how speech relates to thinking. For each of these analytic dimensions we draw on an existing body of literature that provides guidance on how to reliably interpret the various features.

In what follows, we describe our analytic methodology in more detail and use our data to illustrate it. For each analytic feature we describe relevant pieces of the research base and then explain how we code our data.

## II. ANALYTIC METHODOLOGY

### A. Transcribing, segmenting, and coding

The traditional approach to analyzing TAP data [1] is to create a transcript of the participant's speech, segment the transcript into sensible chunks, and code each segment. Those are at is a reasonable description of our initial steps. We transcribe each TAP and then segment the transcript into "chunks of thought". Ericsson and Simon's [1] analogy of segments as something like phrases in a sentence was useful to us. We separated the text in the Charlie example from the introduction is two segments, the first is the announcing the answer ("*I'm going to say A*") and the second providing a justification ("*because it's (pause) moving and has the most weight*"). Our coding categories largely emerged and were refined through discussion. In our coding scheme we separate inferences about participant reasoning from the codes that describe the content of segments. We do this to keep in mind that these are tentative inferences.

Many verbalizations announce information, either produced through some cognitive process or necessary for a cognitive process. Required and produced information are our two most frequent segment codes. There are also some verbalizations that explicitly describe the cognitive processes, though they are less common. The Charlie example is typical in not providing a direct verbal report of the process by which he produced his answer. Beyond categories for information and processes, we created codes to flag features of the transcript of interest for our study. For example we code for participant justifications and any

explicit comments about the questions or the problem solving processes.

Returning to the example of Charlie in the introduction, the first segment, "*I'm going to say A*", is a piece of information he has produced. The second segment, "*because it's (pause) moving and has the most weight*", is Charlie noting information from the problem that he used in the cognitive process that produced his answer. That information also serves as a justification for him in this instance, so we code the second segment as both required information and justification.

The combination of the two segments allows us to make an inference about Charlie's reasoning, which we treat as a separate coding category. It appears he draws on the information about the relative sizes and speeds of the two vehicles to reach the conclusion that the force the truck exerts on the car is larger than the force the car exerts. This suggests he has used his intuition that the greater speed and weight of the truck will result in a larger force.

Again, that is an inference about his cognitive process. In our analysis these codes and inferences serve as a starting point. We then move on to analyze the evidence of expectations, gestures, and pauses to support, refute, or refine our initial interpretations.

### B. Expectations

The notion of expectations we use comes out of the literature on framing. Framing refers to an individual's interpretation of a situation; this is often characterized as an individual's tacit answer to the question "What is it that's going on here?" [7] This kind of interpretation activates a set of expectations about what is appropriate in the situation. For example, if you interpret a situation as an academic talk you would likely expect it is inappropriate to interrupt the speaker with a question. But in another type of situation, say talking with a friend, it might seem quite appropriate to interrupt the speaker with a question.

A participant's expectations related to how they frame or interpret the TAP can impact their reasoning and/or the way they report the reasoning in their concurrent verbalization. So we attempt to identify their expectations.

To do this we use Tannen's [4] linguistic markers of expectations. Tannen is a sociolinguist whose work focuses on identifying features of speech that provide insight into a speaker's expectations. After we code the transcript as described in the previous section, we review the transcripts to identify linguistic markers of expectations. Here too our coding scheme was refined and expanded as we went, as some types of Tannen's markers are not present in our data and we also see markers that are not ones Tannen identifies.

An example of this latter category are words or phrases that only make sense in the context of a dialogue between people. For example, some participants say "sorry" during their think-aloud. To us that indicates the participant perceives their report as directed at someone else, discourse

rather than a report of thinking. When a participant says “sorry” after a pause of several seconds we interpret it as their acknowledging they have violated an expectation of the interview context, namely we ask them to speak as continuously as possible during their think-aloud interview. That is an example of how linguistic markers provide insight into a speakers’ expectations.

### C. Gestures

Gesture analysis contributes to a more exhaustive understanding of an individual’s reasoning. Unlike language, gesticulation is not bound by “a codified, recognizable system,” [8]. Therefore, gestures can sometimes provide a more accurate representation of an individual’s “idiosyncratic imagination” than their verbal report [9]. As an individual is thinking aloud, gestures work in conjunction with language to help the speaker negotiate a personal mental process [10]. Alibali and her colleagues suggest that speech and gesture analysis provide an opportunity to more completely comprehend mental problem solving strategies and mental representation [10]. Goldin-Meadow contends that heightened difficulty or mental stress can be apparent in an interviewee’s increased gesticulation [8]. McNeill notes the particular relevance of gesticulation in areas that are traditionally less dependent on linguistic and verbal units, such as physics [9].

Our coding for gestures involves noting the placement within the transcript and descriptors of the gestures. Initially it is helpful to characterize the gestures without accompanying audio, so that they can be described without the influence of the participant’s verbal report. After the audio is reintroduced, this process allows for a broader understanding of possible meanings associated with a gesture. Some gestures hold meaning independently of speech, and others are simply an aspect of speech performance and have no independent meaning. We refer to the latter as “beats”. These codes allow for a more thorough account of the relationship between the gestures, and cognitive processes.

### D. Speech Pauses

Analyzing pauses and hesitations during speech in think-aloud protocols can also contribute to our understanding of the individual’s reasoning and speech planning. Goldman-Eisler [11] and others found a clear relationship between the difficulty of a cognitive task and the duration and frequency of pauses in spontaneous speech. Hesitations in the middle of phrases, or “ungrammatical pauses”, often indicate an individual is searching for a word that is unpredictable in the situation [12]. Grammatical pauses, which occur between phrases, are indicative of planning the upcoming speech [6], and these pauses become less frequent as an individual’s certainty about upcoming speech increases [13]. To illustrate this difference we return to the short excerpt from

Charlie in the introduction: “*I’m going to say A, um (pause) because it’s (pause) moving and has the most weight*”. The first of Charlie’s pauses is grammatical; it falls between two phrases. The second one occurs in the middle of a phrase and hence is ungrammatical. The pattern of changes in the frequency of pauses and the individual’s speech rate creates a “cognitive rhythm” which aids in the organization of information during speech production [14]. Analyzing the pauses in the think-aloud protocols provides evidence of when individuals are actively in cognitive processes rather than planning their explanations of their thinking, when they are experiencing periods of increased cognitive strain, and the framing that influences their ability to retrieve certain contextual words while speaking.

For us, analyzing pauses is tedious work. There is software available to automate this analysis, but as it is relatively new to us we do not presently have access to such a package. Because of the laborious nature, we only conduct pause analysis on portions of the transcript that are of interest for some reason. For example, we might identify an ambiguous point in the data where it is unclear whether an individual has already arrived at an answer and is explaining her reasoning after the fact, or is still actively engaged in producing her answer and explaining reasoning as she does so. We are interested in moments like these because we expect in-the-moment descriptions of reasoning are more likely to provide accurate insight into the reasoning process than after the fact reports, which can be composed in light of the speaker’s expectations of the interview context. In analyzing pauses we find it useful to slow the video playback speed. We identify pauses of one second or more and then plot the pauses on a timeline to illustrate the individual’s general speech pattern, noting on the timeline whether pauses are grammatical or ungrammatical.

## III. AN ILLUSTRATION

By analyzing only the text of the TAP it can be unclear when a participant is actively engaged in reasoning and when they are explaining a previously decided, but unannounced, conclusion. Through analysis of gestures and pauses in speech, however, it is possible to make reasonable inferences about the relationship between the participant’s cognitive and verbal processes.

An example of this is a participant, Misha, who provides verbal explanations for the first two questions in his think-aloud that have a similar structure. He reads the question, provides and evaluates pieces of evidence to support or negate each of the multiple-choice options, and then states his answer. From simply evaluating the transcript, it would be sensible to assume that his decision and explanation process is similar in the two questions. However, after considering Misha’s speech pauses and gestures, we argue that he produced his answer before describing his reasoning

in the first question, and was actively engaged in reasoning while evaluating the evidence in the second.

During the first question the pauses in Misha's speech are primarily grammatical during his answers to the first question, indicating more cognitive activity dedicated to speech planning. This suggests he has already arrived at his answer prior to "reasoning" to the answer he eventually states. In the 30 seconds between the time he finishes reading the question and the time he announces his answer there are five pauses, all grammatical.

His verbal report in the second question has a similar structure; Misha reads the question, provides evidence to support or negate answers, and then states his choice. In this case there are 45 seconds between his reading the question aloud and then stating his answer. Those 45 seconds contain 8 pauses, and this time only 3 of the 8 are grammatical, indicating speech planning. The other 5 are ungrammatical pauses, which indicate active information processing and/or retrieval [6]. The fact Misha is actively processing or retrieving information during these ungrammatical pauses suggests he is reporting an active reasoning process, not reporting a past reasoning process. The second question also contains pauses of greater length, consistent with Goldman-Eisler's conclusion that the more cognitively challenging the task, the longer the speech pauses [11].

Misha's gestures in question two also suggest increased cognitive difficulty and a change in his problem-solving pattern. His answer to question two includes the greatest number of gestures of any of his questions. The increased gesturing we see in question two suggests on-going problem solving or reasoning, as well as heightened cognitive difficulty [8], which is consistent with the interpretation of the pause-type patterns in the questions. Considering this evidence together we infer that despite the fact the structure of the talk is the same, in question one Misha has decided on his answer before he begins speaking and in question two his verbalizations correspond with an active figuring-out process.

We argue that based solely on the text of the transcript researchers would not distinguish between these two possible interpretations. Misha's pauses and gestures, however, indicate he has reasoned to his answer before beginning his verbal report in the first question and is actively constructing his answer during his verbal report in the second question.

#### IV. CONCLUSION

Our analysis is of TAP as a methodology for studying cognitive processes. We tentatively claim that the interpretive validity of TAP analysis is improved by attending to multiple elements of speech performance. These features include the investigation of linguistic markers of expectations, gestures, and pauses within the think-alouds. Through the example of Misha we have illustrated that attending to multiple aspects of speech performance in analyzing TAP data can provide a more accurate depiction of the cognitive process than simply attending to the transcript for the participant's description of thinking.

In Misha's case we believe that an analysis of only the transcript, as traditional TAP analysis would do, could provide an inaccurate interpretation of his reasoning. The structure of his verbal report is similar in questions one and two. If a researcher were looking only at the transcript they would have no reason to think there was a difference between the reasoning in the two problems.

Why does this matter? We treat Misha's response to the second question as a more accurate description of reasoning because it may actually be a concurrent report. We give less value to his response to question one (and three and four) because it appears to be an after-the-fact construction. A post-facto description he has constructed is more likely influenced by his expectations about the context. So our attention here is ultimately to the intersection of interpretive and contextual validity concerns.

---

[1] Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis*. Cambridge, MA: The MIT Press

[2] Smagorinsky, P. (1998). Thinking and...protocol analysis. *Mind, Culture, and Activity*, 5(3) 157-177.

[3] Deschambault, R. (2012). Thinking-aloud as talking-in-interaction. *Language learning*, 62(1), 266-301.

[4] Tannen, D. (1993). *Framing in discourse*. New York: Oxford University Press.

[5] Scherr, R. (2008). Gesture analysis. *PRST-PER*, 4.

[6] Kircher, T. T. et al. (2004). Pausing for thought. *NeuroImage*, 21(1), 84-90.

[7] Goffman, E. (1974). *Frame analysis*. New York: Harper & Row.

[8] Goldin-Meadow, S. (2003). *How our Hands Help Us Think*. Cambridge, MA: Harvard College.

[9] McNeill, D. (1992). *Hand and Mind*. Chicago, IL: University of Chicago Press.

[10] Alibali, M.W. (1999). Illuminating Mental Representations. *Psychological Science*, 10, 327-333.

[11] Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. Academic Press.

[12] Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.

[13] Brubaker, R. S. (1972). Rate and pause characteristics of oral reading. *Journal of Psycholinguistic Research*, 1(2), 141-147.

[14] Oliveira, M. (2002). Pausing strategies as means of information processing in spontaneous narratives. In *Speech Prosody 2002, International Conference*.