Analyzing Prosodic Features and Student Uncertainty using Visualization

Wenting Xiong and **Diane J. Litman** and **G. Elisabeta Marai**
University of Pittsburgh, Department of Computer Science
Pittsburgh, PA 15260
{wex12,litman,marai}@cs.pitt.edu

Abstract

It has been hypothesized that to maximize learning, intelligent tutoring systems should detect and respond to both cognitive student states, and affective and metacognitive states such as uncertainty. In intelligent tutoring research so far, student state detection is primarily based on information available from a single student-system exchange unit, or *turn*. However, the features used in the detection of such states may have a temporal component, spanning multiple turns, and may change throughout the tutoring process. To test this hypothesis, an interactive tool was implemented for the visual analysis of prosodic features across a corpus of student turns previously annotated for uncertainty. The tool consists of two complementary visualization modules. The first module allows researchers to visually mine the feature data for patterns per individual student dialogue, and form hypotheses about feature dependencies. The second module allows researchers to quickly test these hypotheses on groups of students through statistical visual analysis of feature dependencies. Results show that significant differences exist among feature patterns across different student groups. Further analysis suggests that feature patterns may vary with student domain knowledge.

Introduction

Current Intelligent Tutoring Systems (ITS) research aims to increase student learning by responding to students in a manner similar to human tutors. Previous experiments with several dialogue-based ITS have shown correlations between student learning gains and student knowledge states, as well as between learning and student affective and/or metacognitive states (Craig et al. 2004; Litman et al. 2009). Correlations have also been shown between student knowledge states and certain affective states (Graesser and D'Mello 2008). Researchers have thus hypothesized that responding to not only student knowledge states, but also to student states such as "uncertainty" about domain knowledge, may improve the student learning process (Forbes-Riley and Litman 2009; Pon-Barry et al. 2006). For this reason, significant work has been done on developing new tutorial dialogue system components that can automatically detect and adapt to both cognitive and other types of student states (Litman and Forbes-Riley 2006; Graesser and D'Mello 2008; D'Mello et al. 2008).

In most work so far, the detection of student states is primarily based on information from a single student turn; features such as lexical information and speech characteristics are first extracted from a student turn, then used to detect the student affective and/or knowledge states associated with the turn. While some projects also add contextual features such as the index of the student turn and the purpose of the dialogue move (Graesser and D'Mello 2008), these local-context features are still primarily used to predict the student state underlying the current turn. However, in a dialogue, turns are naturally connected together, and student states are likely to have a temporal component along the turns as well.

In order to better predict a student's current affective/metacognitive state, we wish to investigate potential temporal patterns across turns. However, computing such patterns directly is difficult, since it is not clear which features such patterns might involve and what the pattern might be. Given the complexity of the pattern-mining task, visualization becomes an interesting approach, as humans are remarkably good at perceiving patterns from visual signals.

In this paper we present an interactive tool for the visual analysis of prosodic features and student uncertainty, across student turns. The tool consists of two complementary visualization modules. The first module allows researchers to visually mine the feature data for patterns per individual student dialogue, and subsequently form hypotheses about feature dependencies. The second module allows researchers to quickly test these hypotheses on groups of students through statistical visual analysis of feature dependencies.

Methods

Data and Features

The dialogue dataset we investigate was collected in a previous experiment, where the ITSPOKE spoken dialogue tutor was enhanced to detect and respond to student uncertainty, over and above correctness (Forbes-Riley and Litman 2009). In this prior experiment, 81 students worked through college-level physics problems with a Wizard-of-Oz version of ITSPOKE, where a human "wizard" replaced ITSPOKE's speech recognition, natural language understand-

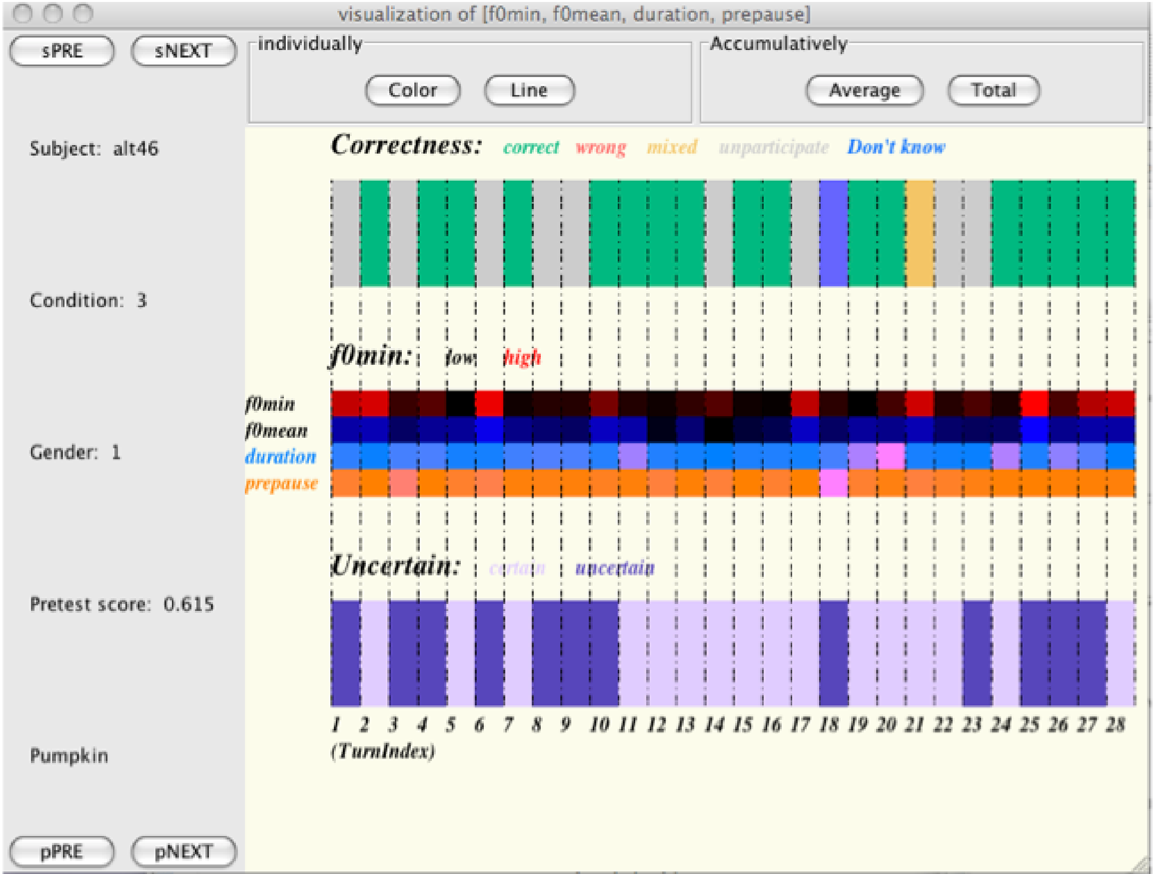

Figure 1: Visualization panel for mining metacognitive tutoring data at the level of each individual student dialogue. Dialogue-specific data is mapped to timelines as follows: the X-axis is mapped to the turn index, while separate rows correspond to separate features. Different features use different hues, and saturation is used to encode the numerical values of the speech features. The top and bottom timelines correspond to the wizard-annotated correctness and uncertainty student states. The middle rows correspond to speech features interactively selected by the user.

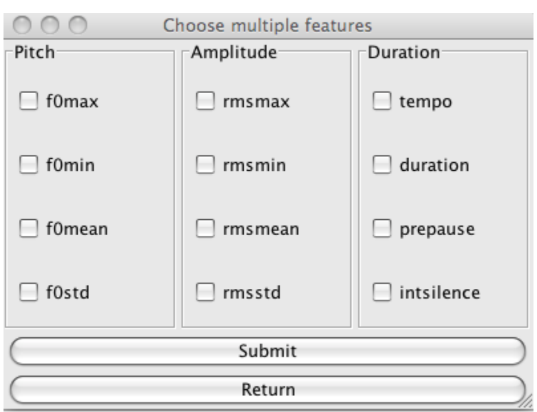

Figure 2: Feature selection interface.

ing, and uncertainty detection components. Students took a pretest, worked through 5 physics problems (i.e. engaged in 5 dialogues), then took a posttest. Since students interacted with ITSPOKE via speech, note that the collected data included not only what students said, but how they said it (i.e. their prosody).

For our current experiment, we formally represent each student turn in our previously collected dataset as a set of features. As shown down the lefthand side of Figure 1, some features encode information that remains constant either across all 5 dialogues (**subject**, **condition** in experiment, **gender**, and **pretest score**), or within each individual dialogue (**problem** being discussed, e.g. "Pumpkin").

As shown further right in the figure, other features represent information that potentially changes for each student turn: correctness and uncertainty states, and speech prosody. Student states were annotated by the wizard during the prior experiment, using the values "correct", "wrong", etc. for **Correctness**, and "certain" or "uncertain" for **Uncertainty**. The prosodic features shown in Fig. 2 were au-

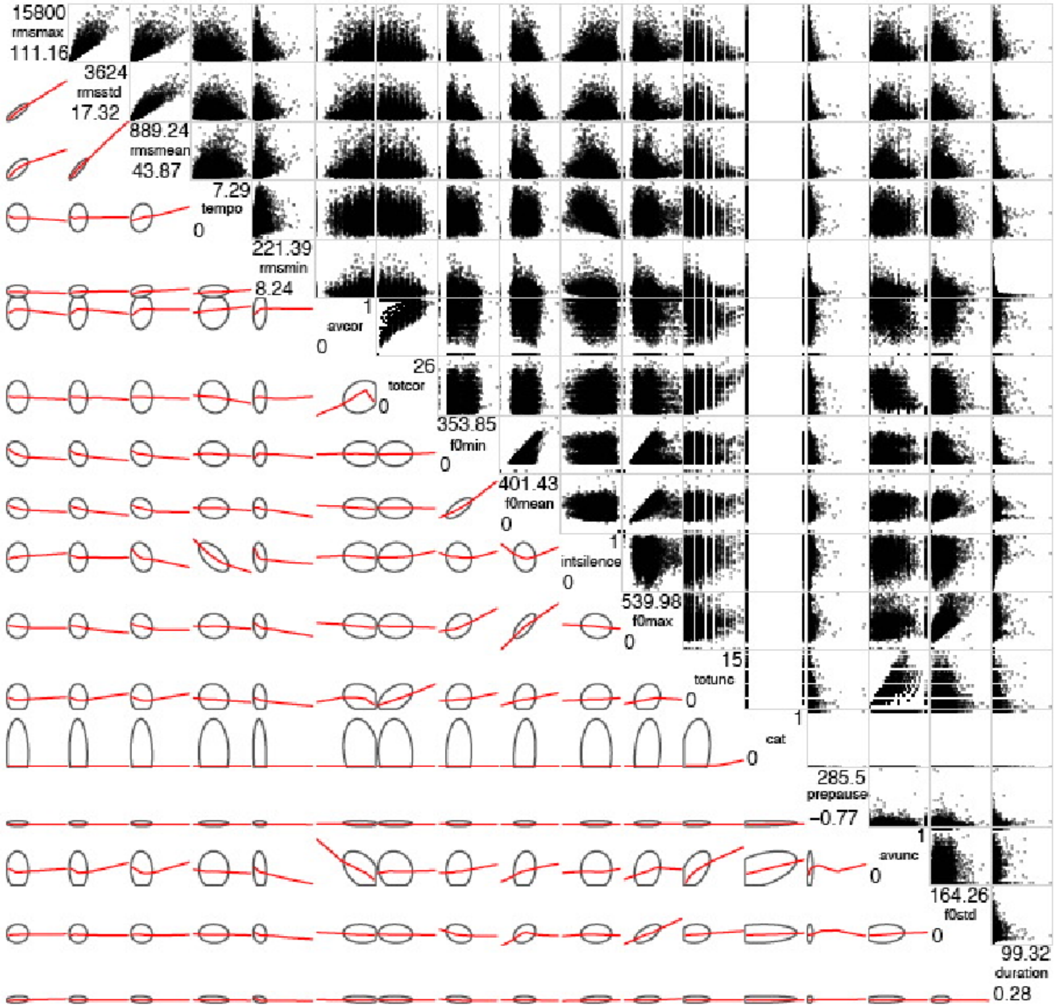

Figure 3: Scatterplots generated through the second, statistical analysis module, showing correlations between feature pairs. Correlation matrices are displayed using ellipse-shaped glyphs for each entry. The ellipse represents a level curve of the density of a bivariate normal with the matching correlation; correlations are computed across the entire dataset. Correlations (non-flat line segments) were observed between rmsmean and rmsstd; f0mean and f0min; rmsmean and tempo, etc.

tomatically computed from the speech file for each turn, and have been used in our prior work to detect positive and negatively-valenced student affect (Litman and Forbes-Riley 2006). **F0** and **RMS** values, representing pitch (fundamental frequency) and amplitude (i.e. loudness), respectively, are computed via Entropic Research Laboratory's pitch tracker. We compute the maximum and minimum F0/RMS value in each turn, as well as the mean and standard deviation of all F0/RMS values in a turn. **Internal silence** is approximated as the proportion of zero f0 frames for the turn. **Turn duration** and **previous pause** are computed via the start and end student and tutor turn boundaries. **Tempo** is computed based on the number of syllables in the turn transcription, divided by the turn duration. We encode not only a feature's value at the current turn, but also two cumulative values (the average and the total value so far within the current dialogue; when moving to the next problem/dialogue, the cumulative values are reset to zero).

Visual Analysis Tool

The first module of the visual tool allows researchers to mine the tutoring data at the level of each individual student dialogue (Fig. 1). As a visual anchor, the visual interface displays on top, for a given student, the correctness values along the student turns. In this timeline-visualization (Plaisant et al. 1996), the turn index is mapped to the X-axis, while vertical lines of varying color encode the correctness value. A similar timeline encoding the uncertainty values along the student turns is displayed on the interface bottom. The middle rows correspond to features interactively selected by the user. Different features use different hues, while color saturation is used to encode numer-

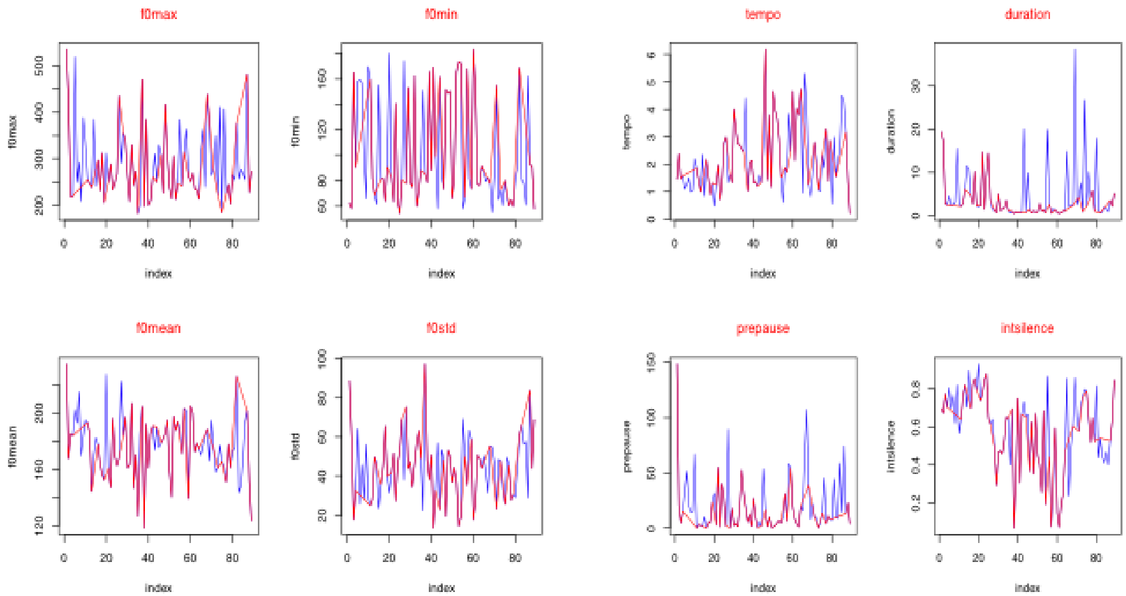

Figure 4: Eight exploratory plots of various prosodic features against human-annotated uncertainty, visually selected by the researcher. Shown is data across turns (the turn index across problems is mapped to the X-axis) for a single student, with the feature shown in blue for turns labeled as "uncertain" and in red for "certain" turns. Note how uncertain turns correlate with higher values of the features than certain turns, most notably for the duration and prepause features.

ical value: higher saturations correspond to higher numerical values. User-selected features are stacked-up in the order in which they are interactively selected from a selection panel (Fig. 2); to avoid overwhelming the researcher with too much information, we limit to six the number of features to be analyzed simultaneously (Halford et al. 2005). Finally, the researcher can quickly scroll through students and physics problems to look for further patterns and correlations in the data. This first visualization module was designed through multiple prototyping-and-feedback iterations, in tight collaboration with an affective dialogue tutoring researcher. The current module version was implemented in Java2D.

While the first, timeline-based visualization module can be used to mine for patterns at the individual student dialogue level, it is most useful for comparing data across two dialogues from at most two students. On the other hand, feature patterns may be specific to different subgroups of users. This means we may need to divide the dataset and look at different types of users separately. The second, complementary visualization module addresses this need, by allowing researchers to test hypotheses related to feature dependencies on groups of subjects.

Our second module enables statistical visual analysis and is developed in R (R 2009), a flexible free software environment for statistical computing and graphics. The module allows the user to display plots showing dependencies between uncertainty and user-selected speech features, as well as scatterplot graphs emphasizing correlations between various features. The scatterplot correlation matrices (Fig. 3) are displayed using ellipse-shaped glyphs for each entry. The ellipse represents a level curve of the density of a bivariate normal with the matching correlation (Murdoch and Chow 1996), with red line segments emphasizing the correlation trend. In the non-scatter plots, prosodic feature values and uncertainty are also color-coded, allowing the user to quickly identify relevant regions in the plot (Fig. 4 and Fig. 5). For both scatter and non-scatter plots the user can select features interactively.

Validation and Results

Validation

We validate our analysis tool by verifying that researchers can use the tool to replicate results previously obtained through different methods. For example, a previous ITSPOKE corpus study using standard statistical methods (Litman and Forbes-Riley 2006) had found positive correlations between negatively-valenced affective student states (including uncertainty) and longer student turn durations; also between such states and longer pauses between the end of the tutor's turn and the onset of the student's speech (prepause). A senior researcher with extensive experience in speech-based tutoring systems and a novice researcher were both asked to mine the current ITSPOKE data for as many such patterns as possible, using the visual analysis tool.

Using the timeline and the statistical tool, both researchers observed that duration and prepause correlated with the uncertainty states across all students, confirming earlier, non-visual findings. Additionally, the researchers

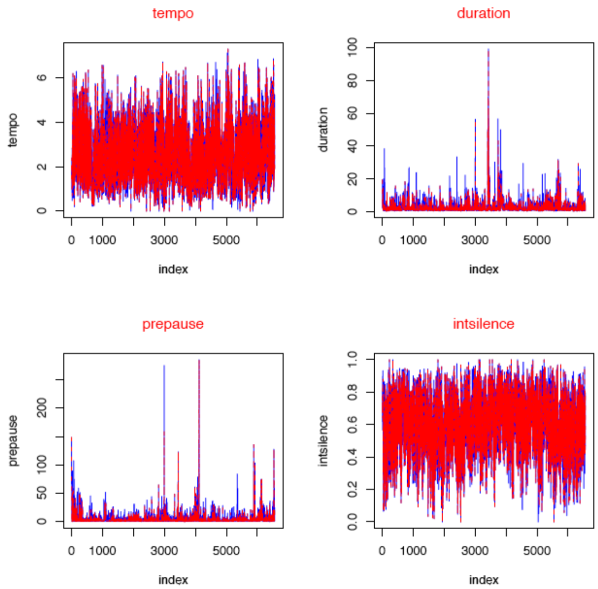

Figure 5: Plots of various speech features (tempo, duration, previous pause, internal silence) across tutoring turns, this time for the entire student population. As in Fig 4, “uncertain” turns are shown in blue, and “certain” turns in red. Turns are concatenated for the 5 physics problems, resulting in very dense plots; nevertheless, the researcher quickly identified correlations between low uncertainty and both short durations and short prepauses.

found subtle relationships between uncertainty and mean and also minimum frequencies across the turns (Fig. 4).

Exploratory Visual Analysis

In a second experiment, the researchers were instructed to look for feature patterns across individual problems. However, this time we found that the user hypotheses based on the timeline visualizations were not consistent across problems and students. This observation indicated that certain patterns are specific to a problem, and other patterns are student-specific. Next, the researchers used these findings to subdivide the dataset into separate subcategories.

Following this exploratory phase and based on visual observation, the original dataset was divided into groups according to a joint condition based on both the student pretest scores and the wizard-annotated assessment of the correctness of the student's turn. Visual analysis of the results shows a difference in the distribution pattern between datasets corresponding to students with lower versus higher pretest scores (Fig. 6). The multiple scatterplots in Fig. 6 show for each subgroup the distribution of the average duration feature against the average uncertainty. In the leftmost scatterplots — corresponding to low pretest scores —, note the clusters of high average duration points. No such high-duration points appear in the scatterplots corresponding to higher pretest scores. Since the student-computer interaction was carefully controlled in the original ITSPOKE experiment, we know that students were not allowed to stay in the same turn for unreasonably long times. Therefore, these long duration points can be interpreted as signals that the student is taking a long time thinking over how to solve the problem. This interpretation is supported by the lower pretest scores of this student subgroup.

Conclusion

We have presented an interactive tool for the visual analysis of metacognitive data in the ITSPOKE system; the tool employs multiple visualization techniques to assist researchers in mining the data. To evaluate the tool, we have conducted a series of experiments in which researchers used the tool to mine the data for feature patterns across student turns. Results show the tool can correctly visualize relationships between features, and also steer the analysis process. In particular, researchers using the tool were able to confirm that the duration of student turns, and previous pause duration before student turns, were correlated with the uncertainty states across all students. Additionally, the researchers found subtle relationships between uncertainty and voice frequencies across turns. Further study of the data indicated that certain patterns are specific to a problem, and other patterns are student-specific. Finally, a visual analysis of the data suggested that feature patterns vary with students' previous knowledge about the target domain.

In order to design stronger, self-aware tutoring systems, researchers in intelligent tutoring systems need tools for visualizing the domain knowledge and for correctly profiling the students. With respect to the ITSPOKE system, our interactive tool is a first step in this direction. In the long run, the tool and the findings derived from using the tool may be used to dynamically monitor and diagnose the student activities, and ultimately to propose user-tailored strategies to help students maximize their learning.

Acknowledgments

Thanks to K. Forbes-Riley for the feature representation. Thanks to NSF (0631930) for support of the creation of the corpus used in this work.

References

Craig, S. D.; Graesser, A.; Sullins, J.; and Gholson, B. 2004. Affect and learning; an exploratory look into the role of affect in learning. *Journal of Educational Media* 29:241–250.

D'Mello, S.; Craig, S.; Witherspoon, A.; McDaniel, B.; and Graesser, A. 2008. Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction* 18(1-2):45–80.

Forbes-Riley, K., and Litman, D. 2009. Adapting to student uncertainty improves tutoring dialogues. In *Proc. Intl. Conf. on Artificial Intelligence in Education*.

Graesser, A. C., and D'Mello, S. K. 2008. The relationship between affective states and dialog patterns during interactions with auto tutor. *Journal of Interactive Learning Research* 19:293–312.

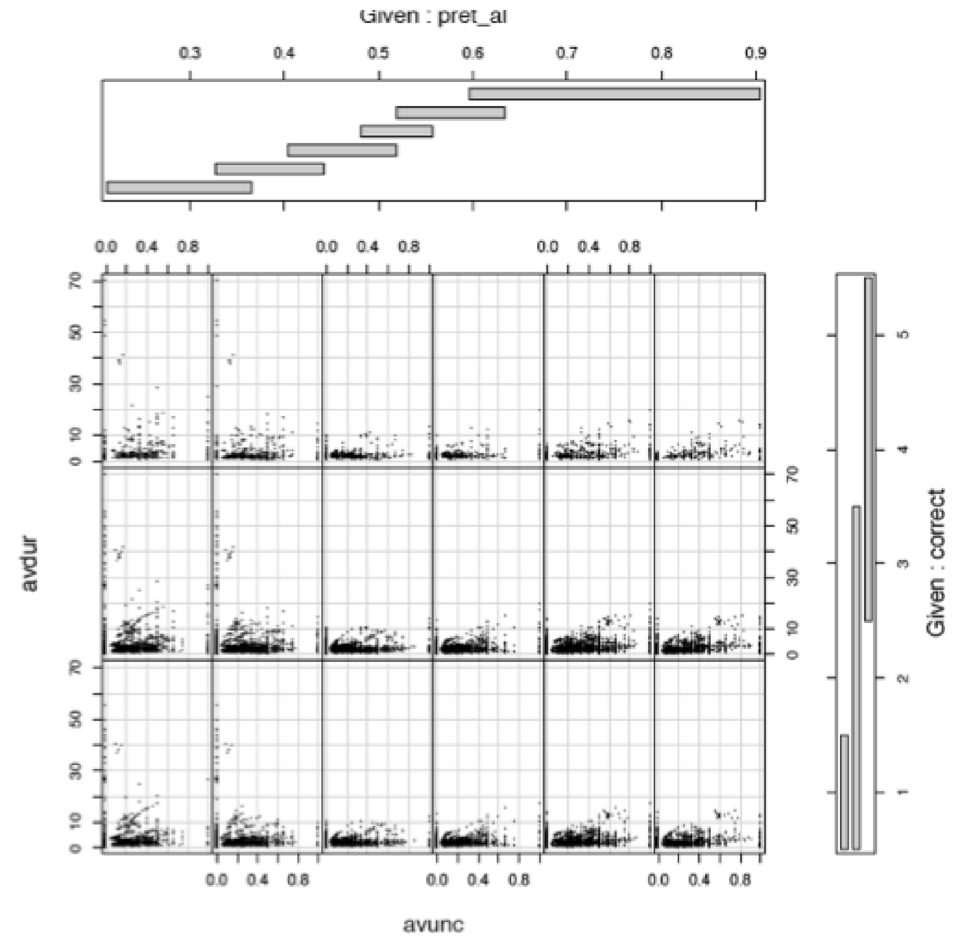

Figure 6: Duration and uncertainty distribution given pretest scores and the current correctness assessment of the turn. The top box shows the pretest score distribution along the X-axis, while the right-side box shows the correctness assessment distribution along the Y-axis. The multiple scatterplots show for each subgroup the distribution of the average duration feature against the average uncertainty. In the leftmost scatterplots — corresponding to low pretest scores —, note the clusters of high average duration points. No such high-duration points appear in the scatterplots corresponding to higher pretest scores.

Halford, G. S.; Baker, R.; McCredden, J. E.; and Bain, J. D. 2005. How many variables can humans process? *Psychological Science* 16(1):70–76.

Litman, D., and Forbes-Riley, K. 2006. Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors. *Speech Communication* 48(5):559–590.

Litman, D.; Moore, J.; Dzikovska, M.; and Farrow, E. 2009. Using natural language processing to analyze tutorial dialogue corpora across domains and modalities. In *Proc. Intl. Conf. on Artificial Intelligence in Education*.

Murdoch, D., and Chow, E. 1996. A graphical display of large correlation matrices. *The American Statistician* 50:178–180.

Plaisant, C.; Milash, B.; Rose, A.; Widoff, S.; and Shneiderman, B. 1996. Lifelines: Visualizing personal histories. 221–227. ACM Press.

Pon-Barry, H.; Schultz, K.; Bratt, E. O.; Clark, B.; and Peters, S. . 2006. Responding to student uncertainty in spoken tutorial dialogue systems. *International Journal of Artificial Intelligence in Education* 16.

R. 2009. The R project for statistical computing.