NeuroDialog: An EEG-Enabled Spoken Dialog Interface

Seshadri Sridharan, Yun-Nung Chen, Kai-Min Chang, Alexander I. Rudnicky
Language Technologies Institute, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213-3891, USA
{seshadrs,yvchen,kkchang,air}@cs.cmu.edu

ABSTRACT

Understanding user intent is a difficult problem in Dialog Systems, as they often need to make decisions under uncertainty. Using an inexpensive, consumer grade EEG sensor and a Wizard -of-Oz dialog system, we show that it is possible to detect system misunderstanding even before the user reacts vocally. We also present the design and implementation details of *NeuroDialog*, a proof-of-concept dialog system that uses an EEG based predictive model to detect system misrecognitions during live interaction.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: User interface management systems; Voice I/O; Natural Language; I.2.7 [**Natural Language Processing]**: Speech recognition and synthesis.

General Terms

Experimentation, Languages.

Keywords

Brain-computer interaction (BCI), electroencephalogram (EEG).

1. INTRODUCTION

Traditionally, dialog systems base their decisions solely on the acoustic signal. Although vision based multi-modal systems exist, detecting dialog critical metrics such as user frustration, attention, fatigue etc. visually is nonviable. Awareness of the user's state of mind can help interactive systems act in accordance with user goals. Given the advances in brain computer interaction (BCI) research [1] and the recent availability of low cost sensors for electroencephalography (EEG), we can now interface neural signals to dialog systems as a secondary sensory stream that helps augment system perception.

Misrecognition of speech, incorrect parsing and mistaking user intent are common problems in dialog systems. Bulyoka et al showed that detecting errors in advance and adjusting the response strategy could improve the performance of spoken dialog systems. By monitoring the user's brain activity, it is possible to detect system mistakes as they happen, which could trigger error-rectifying mechanisms. In this paper, we describe such a spoken dialog system that tailors responses to user's EEG signal.

ICMI'12, October 22–26, 2012, Santa Monica, California, USA.
ACM 978-1-4503-1467-1/12/10.

2. DIALOG INTERACTION

A Wizard-of-Oz spoken dialog system (Figure 1) is used to collect EEG data. A Wizard controls the dialog interaction remotely, while the participant interacts with a dialog interface, designed based on the process monitor in the Olympus framework [3]. The interactions are scripted, where in each interaction a query is made for a particular place in the local neighborhood. Figure 2 shows examples of a correctly recognized and a misunderstood interaction. The System's speech response and the text-output (TTY) are the auditory and visual cues to the user about possible misunderstanding. The EEG signal over the period between these cues and the user-confirmation is considered.

Six adults from our laboratory participated in this pilot study. The participants interacted with the Wizard system while wearing a wireless single-channel BrainBand [4] EEG headset that measured their frontal lobe activity (Figure 3). No participants reported discomfort while wearing the EEG sensor. Besides the timestamps from the dialog interaction, we used the headset to collect the following signal streams: a) the raw EEG signal sampled at 512 Hz, b) proprietary "attention" and "meditation" measures reported at 1 Hz, c) power spectrum reported at 8 Hz, consisting of Delta (1-3Hz), Theta (4-7Hz), Alpha (8-11Hz), Beta (12-29Hz), and Gamma (30-100Hz) frequency bands, and d) an indicator of signal quality reported at 1 Hz. To remove potential EMG artifacts, Neurosky [4] applies a 3-100Hz band-pass filter to the raw EEG signal.

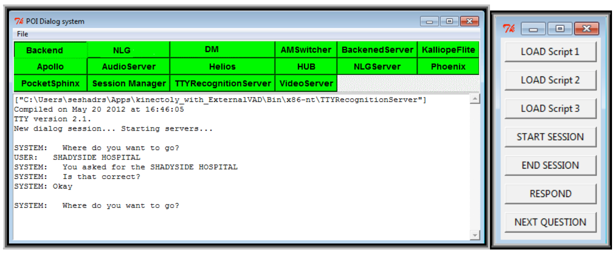

Figure 1: The spoken dialog system interface (left) and the Wizard-of-Oz's command interface (right)

SYSTEM:	Where do you want to go?	*SYSTEM*:	Where do you want to go?
USER:	I want to go to HOMEWOOD.	*USER*:	I want to go to BEACON.
TTY:	I want to go to HOMEWOOD.	*TTY*:	I want to go to BAYARD.
SYSTEM:	You asked for HOMEWOOD.	*SYSTEM*:	You asked for BAYARD.
SYSTEM:	Is that correct?	*SYSTEM*:	Is that correct?
USER:	Yes.	*USER*:	No.
SYSTEM:	Okay.	*SYSTEM*:	Okay.

Figure 2: Example interactions with correct recognition (left) and incorrect recognition (right)

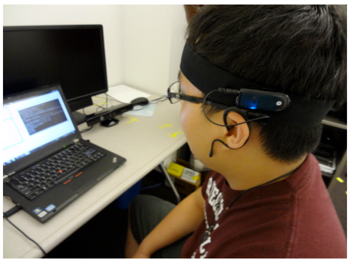

Figure 3: Participant wearing the BrainBand EEG sensor.

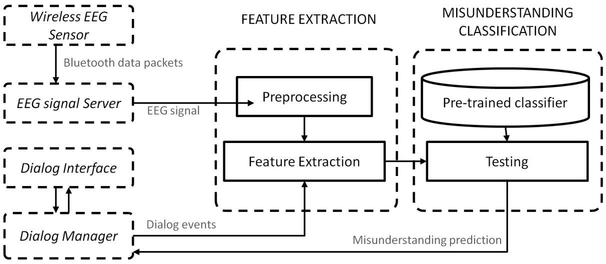

Figure 4: *NeuroDialog* system architecture

We trained a Gaussian Naïve Bayes classifier to detect system misunderstanding based on the brain activity. Initial results of our predictive models on the 360 aggregate dialog interactions we collected were encouraging. We observed a statistically significant average accuracy of 61% for the user-dependent classification task. In the *NeuroDialog* system, we use a similarly trained classifier.

3. NEURODIALOG SYSTEM

The *NeuroDialog* system has the same interface and dialog-task as the Wizard-of-Oz system, except that 1) there is no wizard involved, instead, the system misunderstands the user with a binomial probability of 50% over a predetermined query order that the user follows; and 2) there is no explicit system confirmation, instead, the system directly tailors its response to the EEG based misunderstanding prediction. The architecture schema of the system is shown in Figure 4. The wireless EEG sensor sends Bluetooth data packets to an EEG-signal server that runs locally. The signal stream from this server is preprocessed and stored in a buffer. Timestamps from the asynchronous dialog events from the dialog manager are used to extract the EEG signal over a particular period following the system's initial answer to the user query. Features extracted from this data are sent to the pre-trained classifier model for testing, and the signal buffer is emptied. The classifier's prediction on system misunderstanding and its confidence measure are sent back to the dialog manager.

The system adapts its response to user query depending on the classifier's prediction and the confidence interval. For high confidence of correct system recognition, the system proceeds with the dialog. When the classifier is adequately confident about the misunderstanding, the response ranges from "Could you please repeat? I misheard that." to "I think I may have got you wrong." At the end of the dialog session, we also generate a visualization of the trend in the attention, meditation and normalized signals for the right and wrong system understandings (Figure 5) along with the annotations of the dialog events.

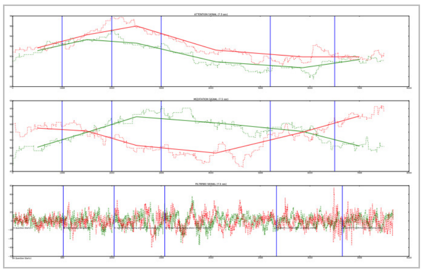

Figure 5: Trend in attention, meditation and raw signal for right (green) and wrong (red) answers for a user.

4. APPLICATIONS IN DIALOG

In a dialog system, the misunderstanding model can act as a sensory mechanism that aids it to act in accordance with the user's goal, without additional effort from the user. The classifier's prediction could be used as a confidence metric while the system advances to the next level of the dialog. A live classifier could publish its predictions to the dialog manager, which in turn could use the metric to make informed decisions. Collecting misunderstanding labels for the EEG data based on the dialog outcomes can be used to learn better predictive models with time.

From the user perspective, the ability to detect misunderstanding can potentially avoid many barge-in situations, where the system could initiate error-rectifying mechanisms before the barge-in event. The system may even skip multiple confirmation steps, common in dialog systems, saving time for the user. For example, in time sensitive applications where explicit user validation is expensive, such as gaming systems or command interfaces, the system can base its immediate decisions on the classifier confidence. As the next phase of our work, we will evaluate the impact of our system on dialog using objective metrics such as reduction in dialog length, barge-in events, as well as subjective metrics such as user satisfaction surveys, etc.

5. CONCLUSION

We present a Wizard-of-Oz dialog system capable of EEG data collection for training a system misunderstanding classifier. We also detail the design of our proof-of-concept multimodal system *NeuroDialog* that responds based on the predictions of a pre-trained system misunderstanding classifier. This predictive ability can be used to develop effective strategies to build perceptive dialog systems.

6. REFERENCES

[1] A. Nijholt and et al., *Brain-Computer Interaction: Can Multimodality Help?*. In *ICMI,* 2011.

[2] I. Bulykoa and et al., *Error-correction detection and response generation in a spoken dialog system*. In *Speech Communication*, 2005.

[3] B. Dan and et al., *Olympus: An Open sourse framework for converational spoken language research*. In *Dialog Technology workshop at HLT/NAACL*, 2007.

[4] NeuroSky, Brain Wave Signal (EEG): NeuroSky. Inc Neurosky, 2009.

[5] D. L. Donoho, *De-noising by soft-thresholding.* IEEE Trans. on Infomation Theory, 1995.