A Dialogue System for Accessing Drug Reviews

Jingjing Liu and Stephanie Seneff

MIT Computer Science & Artificial Intelligence Laboratory
32 Vassar Street, Cambridge, MA 02139, U.S.A.
{jingl,seneff}@csail.mit.edu

Abstract—**In this paper, we present a framework which harvests grassroots-generated data from the Web (e.g., reviews, blogs), extracts latent information from these data, and provides a multimodal interface for review browsing and inquiring. A prescription-drug domain system is implemented under this framework. Patient-provided drug reviews were collected from various health-related forums, from which significant side effects correlated to each drug type were identified with association algorithms. A multimodal web-based spoken dialogue system was implemented to allow users to inquire about drugs and correlated side effects as well as browsing the reviews obtained from the Web. We report evaluation results on speech recognition, parse coverage and system response.**

I. INTRODUCTION

Increasingly, consumers are turning to the Web to seek information, and, increasingly, this information comes in the form of consumer-provided comments in discussion groups or chat rooms. User reviews of products and services have empowered consumers to obtain valuable data to guide their decision process. Recently, health care and prescription drugs represent a growing topic of discussion online. This is not surprising given that almost half of all Americans take prescription drugs each month, costing over $200 billion in a single year ([7]). Although drugs are subjected to clinical trials before reaching market, these trials are often too short, and may involve too few people to give conclusive results. Many of the side effects that occur in practice are not uncovered during these short trials. Also, drug interactions are highly controlled in such trials. In real life, however, people might take multiple drugs at the same time and the interactions among drugs may cause salient side effects that are unpredicted from drug trials. For example, the official Lipitor web site lists as potential side effects mainly muscle pain, weakness and digestive problems. However, many other side effects have been observed in more alarming areas ([2], [5], [12], [16]), such as heart failure, diabetes, memory problems, Parkinson's disease and ALS (Lou Gehrig's disease).

In contrast, there are various online forums such as *AskAPatient.com* and *Medications.com* which allow users to post comments and to share with other patients their experiences in using a certain drug. These drug reviews often provide evidence of unusual side effects, e.g., "*my legs started to feel heavy after using this drug for three months*" (from a drug review on AskAPatient.com). Thus, patient-provided drug reviews are an important resource for side effect discovery. By harvesting and analyzing these reviews, we may be able to corroborate suspected side effects or even uncover new knowledge. Through grassroots studies, we can augment our knowledge using social networks, which not only increases the sample size dramatically, but also allows patients to share their personal experiences.

In previous work, we have explored approaches to harvesting public reviews from the Web, summarizing them into an opinion summary database with linguistic parsing and sentiment scoring methods [8], and embedding the database into a dialogue system to support a speech-based interface for opinion inquiry [9] [10]. The approach has been applied mainly in the restaurant review domain, but can be applied to other domains as well, such as various product reviews. However, the task is more complicated when it comes to reviews with more complex semantic structure, as in the following movie review from IMDB.com: "*I'm very happy that the trailers didn't give the full story away. Lots of emotions are waiting for the viewers, laughter and tears too.*" Such reviews do not necessarily provide explicit sentiment. Thus, traditional topic extraction and sentiment estimation methods do not easily apply to these reviews. However, important messages are often hidden between the lines, and latent information extraction methods can be applied to reveal the underlying information.

This paper describes our efforts to extract latent information from implicit opinions and to develop a speech-based multimodal interface that allows users to inquire about and browse a large consolidated database of reviews obtained from the Web. A spoken dialogue system was implemented to increase patient awareness of drug-related side effects, enabling consumers of prescription drugs to discover possible side effects via speech and to access relevant reviews efficiently. First, we collected a large number of patient-provided reviews from health discussion sites. Next, we extracted potential drug side effects from these reviews and used log-likelihood ratios to quantify associations between drugs and symptoms. We then developed a speech-based multimodal system to allow spoken queries, which would be answered with summarized information about associations between drugs and side effects, as well as providing a set of succinctly summarized reviews that best match the query. An example conversation between a user and the drug side effect inquiry system is shown in Fig. 1.

In the remainder of this paper, we will first explain the methods for extracting latent information, such as drug side effects from grassroots-generated reviews, as well as measuring the association between drugs and side effects. Then we will describe our approach to implementing a speech-based interface for inquiring about the side effects of drugs and searching for relevant reviews. An evaluation of the dialogue system is also provided.

U: What side effects do Statin drugs cause?

S: *Some of the side effects that show up frequently on Statin drugs are memory problems, difficulty walking, liver damage and all body aches.*

U: Can you show me some reviews on memory problems?

S: *Some of the drugs in reviews that mention memory problems are Mevacor, Pravachol, Crestor, Lovastatin and Zocor. I've sampled a few reviews relevant to memory problems.* (Top-relevant reviews are displayed on the screen.)

U: What side effects may co-occur with memory problems?

S: *Some side effects that often co-occur with memory problems are kidney damage, muscle problems and Parkinson's disease.*

U: What antidepressant drugs can make me gain weight?

S: *Some of the antidepressant drugs that mention weight gain are Lexapro, Arnitriptryline hcl, Serzone, Anafranil and Tofrannil-pm. Other side effects in this category are weight loss, inability to lose weight and weight fluctuation.* (Reviews that are most relevant to weight gain are shown on the screen as well.)

Fig. 1. A real conversation between a user and our drug side effect inquiry system ("U" represents the user, and "S" represents the system).

II. Latent Information Extraction

For the review database, we have collected 107K patient-provided reviews on drugs to treat a broad range of problems such as depression, acid reflux disease, high blood pressure, diabetes, etc., from three drug discussion forums ("AskAPatient.com," "Medications.com" and "WebDB.com"). In these forums, patients fill in values for several fields, including the drug name, their age, side effects experienced, and a general comment field where they typically describe their personal story. An example review is shown in Fig. 2.

Drug: "Lipitor"
Dosage: "20mg 1X D"
Sex: "Female"
Age: "56"
Duration: "5 years"
Reason: "Heart disease"
Side effects: "Severe muscle pain in shoulders radiating through the chest, cramping in back muscles, calves and hamstrings. Severe muscle pain after working out with weights, all the while losing strength. Difficulty with memory at times..."
Comment: "My shoulder pain resulted in a visit to a specialist who said inflammation was present but no torn rotator cuff. Prescribed physical therapy which made it hurt even more. I first noticed the pain several months in to taking the drug. After an ER visit due to severe back spasm/cramp…"

Fig. 2. An example drug review from AskAPatient.com.

To obtain a clean set of side effect expressions, we first automatically extracted over 7,500 words and phrases that describe common side effects on various drugs from the side effect field. After an automatic process of noise and redundancy eliminating (e.g., stop-word filtering), the number of side effect phrases shrank to 2,314, which were manually clustered into 307 synonym groups (e.g., "mental slowness," "slow brain," and "fuzzy thinking" were clustered into the synonym group of "loss of mental clarity") and further grouped into 30 higher-level categories [11]. Table I shows a few examples of synonym groups of side effects and their categories.

TABLE I
Example Groupings of Side Effects into Categories

Category	Synonym groups of side effects
cognition problems	brain shocks, dementia, loss of mental clarity, memory problems, mental instability, problems concentrating, short attention span
mood issues	aggressive behaviour, anxiety, bipolar, bizarre thoughts, blunted emotions, crying easily, depression, despair, disoriented, euphoria

Given the extracted side effects, the next step is to discover from the full review in the comment field which side effects are strongly associated with a certain type of drug. As described in previous work [11], we use log likelihood ratio [3] for the association measurement and treat the drug side effect correlation problem as a coin toss model. The null hypothesis is that the review set R_1 (reviews on a certain type of drug D) and the set R_2 (randomly sampled reviews on drugs other than D) have the same probability of containing a side effect phrase t. The alternative model is that one review set (either R_1 or R_2) has a higher probability of containing the phrase t than the other. The measurement of the likelihood of the hypothesis that the side effect phrase t is more likely to occur in the set R_1 is calculated by:

$$L_1 = k_1 log \frac{p_1}{p} + (n_2 - k_2) log \frac{1-p_2}{1-p} \quad (1)$$

where k_1 and k_2 are the counts of reviews that contain the side effect phrase t in R_1 and R_2 respectively, p_1 and p_2 are the probability of the phrase t occuring in R_1 and R_2 respectively, p is the probability of the phrase t occurring in the whole review set ($R_1 \cup R_2$), and n_1 and n_2 are the sizes of R_1 and R_2.

A symmetric equation of (1) can be derived for L_2, the hypothesis that the side effect phrase t occurs more frequently in R_2. Whether the alternative model fits significantly better and should thus be preferred can be determined by deriving the probability or p-value of the obtained difference $L_1 - L_2$.

For each drug type D in our dataset, we take the set of reviews on this drug as R_1, and randomly select the same size of reviews from the reviews on other drugs as R_2. To avoid the age bias of review-providers in the data selection process, we follow the same distribution of reviewers' age on drug D for the random selection of R_2 (reviews on drugs other than D). Then, based on R_1 and R_2, we calculate the log likelihood ratio and p-value for each of the 307 side effect synonym groups (phrases within the same group are treated as identical); and we select the side effects which have p-value lower than a threshold (normally set as 0.05) as significantly associated with the drug type D. An example of top-relevant side effects to the drug type of *antidepressants* is shown in Table II, where k_1 and k_2 represent the numbers of reviews in R_1 and R_2 respectively that contain any phrase within the side effect synonym group, and $L_1 - L_2$ represents the log likelihood ratio. Lower p-value indicates higher significance of the association between the side effect and the drug type.

TABLE II
SIGNIFICANT SIDE EFFECTS ASSOCIATED WITH ANTIDEPRESSANT DRUGS

Side effect	k_1	k_2	$L_1 - L_2$	*p-value*
mouth issues	1601	469	2124.63	0.0000004
reduced sex drive	1654	517	2136.27	0.0000005
increased body temperature	793	180	1143.38	0.0000085
weight gain	2385	1279	2058.01	0.0000102
high energy	1009	322	1291.19	0.0000123
increased sweating	1012	334	1274.76	0.0000147
loss of appetite	1179	471	1331.48	0.0000244
dizziness	1904	1310	1064.56	0.0008830
nausea	2123	1524	1059.08	0.0011523

III. SPOKEN DIALOGUE SYSTEM

After discovering the significant side effects that are most relevant to each drug type, our next goal is to build a speech-based multimodal interface that allows users to inquire about these side effects and access the reviews in the database. In this section, we will explain our efforts on implementing such a spoken dialogue system, focusing on speech processing, language understanding, and dialogue management.

A. Speech Processing

For the speech recognition of users' utterances, we use the SUMMIT system [4], the acoustic models of which are trained with an English corpus unrelated to this domain. The class *n*-gram language model is trained by parsing a synthetic corpus [15]. To create this corpus, we first created a set of templates of English utterances covering plausible patterns of users' queries, based on the developers' experience and judgment. The templates use a recursive context-free grammar formalism to support a rich set of phrase and sentence patterns. 10,000 utterances were then automatically generated from these templates and were used to train the language model of the recognizer. The templates were later expanded based on real users' speech input from a pilot data collection effort.

For speech synthesis of the system response, we utilize an English text-to-speech system provided by Nokia. The open source WAMI[1] toolkit [6] is used to embed the recognizer, the synthesizer, the language understanding and generation components, and the dialogue manager into a web-based interface. Fig. 3 shows the interface of the spoken dialogue system. Users can talk to the system by clicking the microphone icon (on the top right of the interface). The conversation history is shown in text on the top, and users can browse previous dialogue turns. Below the history window, there is a type-in window where users could type their questions in text instead of speaking. On the left is an ontology tree, where users can click to expand each side effect category and click on individual side effects within each category. The click will also trigger a query and the system will retrieve reviews corresponding to the clicked category or side effect and display them in the middle of the interface. This tree is restricted to include only the side effects that are most strongly correlated with the on-focus drug class. The dialogue system currently supports two major drug classes, Statins (to treat high serum cholesterol) and antidepressants. Other drug types can be easily added. On the right of the review browsing window is a cartoonized figure of a human body. Users can click each body part in the figure, and the system will respond by providing the most significant side effects related to the clicked body part. Below the cartoonized figure is a list of a few suggestions of possible queries.

[1] WAMI is an open-source toolkit developed in our group to add speech recognition capabilities to a web page for developing, deploying and evaluating Web-accessible multimodal interfaces. http://wami.csail.mit.edu/

B. Language Understanding

Given the user's query utterance generated by the speech recognizer, the system uses a generic syntax-based grammar to parse the utterance [15]. The grammar captures syntactic structure through a set of heuristically constructed context-free grammar rules, and the parser employs a feature-passing mechanism to enforce long distance constraints. The grammar is lexicalized, and uses a statistical model to rank order competing hypotheses. The grammar probability model was trained automatically on the corpus of simulated sentences generated from our templates.

The parser provides a hierarchical representation (called a linguistic frame), which encodes the syntactic and semantic information of a sentence. A set of generation rules [1] is heuristically constructed to paraphrase the linguistic frame into a set of [key:value] pairs, which represents the semantic meaning of the sentence. Table III shows some examples of input sentences and the corresponding meaning representation (i.e., [key:value] pairs). For example, Sentence I inquires about drugs that might cause a specific side effect, which is encoded by "*drug name: *what*; side effect*: *weight gain.*" Sentence II mentions two side effects, which are numbered and identified with corresponding categories (e.g., "*side effect #1*: *headache*; *category #1*: *cognition problems*").

TABLE III
EXAMPLES OF [KEY:VALUE] PAIRS GENERATED FROM USERS' UTTERANCES

Sentence I	What antidepressant drugs can make me gain weight?
[Key:value] pairs	*Drug class: antidepressant; drug name: *what*; side effect: weight gain; category: weight problems*
Sentence II	Does Lipitor cause headache or general weakness?
[Key:value] pairs	*Drug class: Statins; drug name: Lipitor; side effect #1: headache; category #1: cognition problems; side effect #2: general weakness; category #2: muscle problems*
Sentence III	What side effects often co-occur with heart failure when using Statin drugs?
[Key:value] pairs	*Side effect: heart failure; category: heart problems; command: list_co_occur_side_effects; drug class: Statins*
Sentence IV	Can you show me the reviews on SSRI related to memory loss?
[Key:value] pairs	*Drug class: antidepressant; drug group: SSRI; side effect: memory loss; category: cognition problems; command: list_reviews*

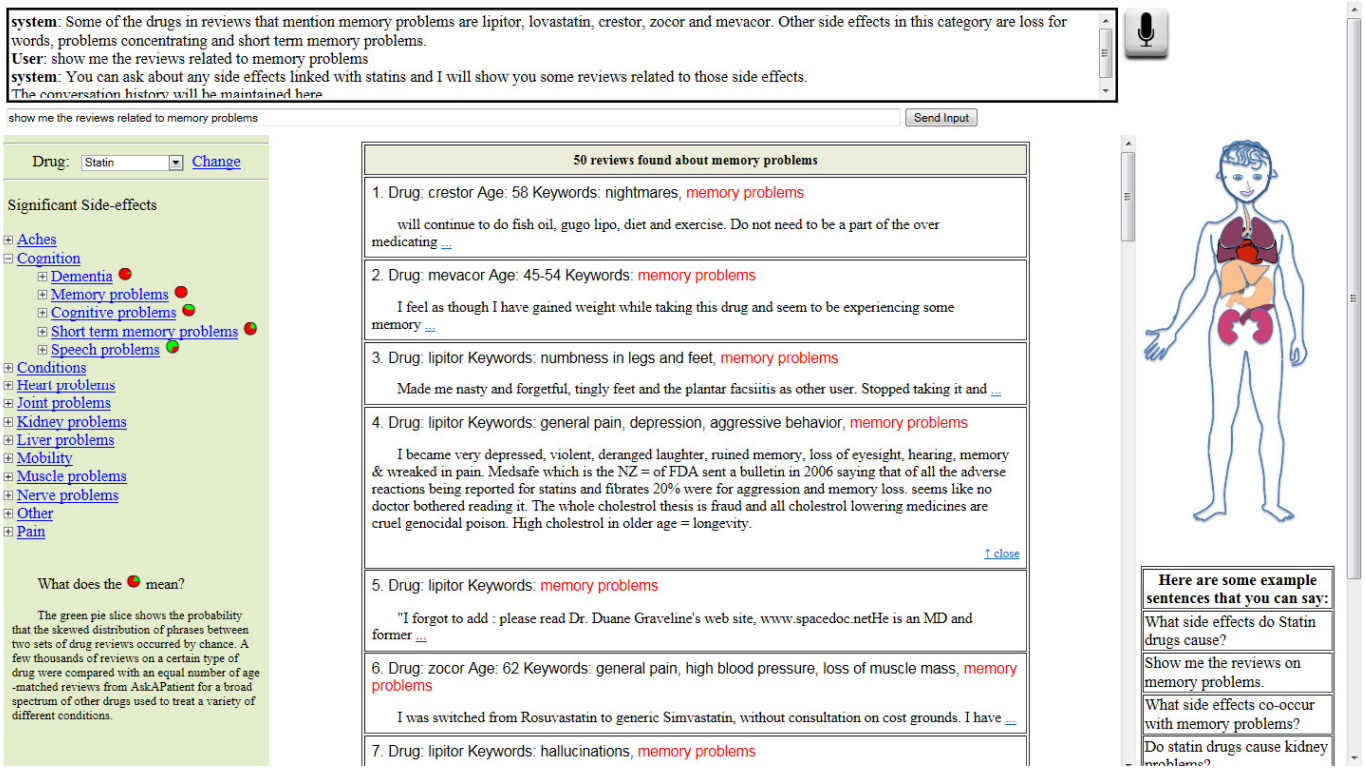

Fig. 3. Screenshot of the speech-based multimodal interface of the drug side effect inquiring and review browsing system.

C. Dialogue Management

For dialogue management (DM), we utilize a goal-based dialogue planning framework developed in our group [17]. It treats the user as a knowledge source, so that the entire framework is DM-centered. A declarative entity-based specification encodes the domain logic, and customized task actions handle any domain-dependent computations.

Under this DM framework, the system searches the database following the corresponding command in the [key:value] meaning representation,. Take Sentence I in Table III as an example. "*Drug name: *what**" will trigger a database search for drug names ("*what*" indicates a search on this key) which have significant correlation with the side effect "weight gain". As aforementioned in Section II, side effects with a *p*-value lower than a threshold 0.05 are considered as strongly associated with the drug. Thus, the system will retrieve as candidates those drug names for which this specific side effect has a *p*-value lower than 0.05.

When users ask the system to show some reviews about a side effect related to a specific drug or drug group (e.g., Sentence IV in Table III), a review searching event will be triggered by the command "*list_reviews*". Reviews are ranked by their relevance to the specific side effect using standard BM25 ranking algorithms [14]:

$$score(D, Q) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{f(q_i, D) \cdot (k_1+1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{avgdl})} \quad (2)$$

where D represents a review, $|D|$ represents the length of the review D, $avgdl$ is the average review length in the review collection, Q represents a side effect, q_i represents a term in Q, and k_1 and b are free parameters. $IDF(q_i)$ is the IDF (inverse document frequency) weight of the query term q_i. It is usually computed as:

$$IDF(q_i) = log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5} \quad (3)$$

where N is the total number of reviews in the collection, and $n(q_i)$ is the number of reviews containing q_i.

The top-ranked reviews are then listed in an abbreviated form on the interface, each ranked by its relevance score $score(D, Q)$. For each review, the side effect phrases that were extracted from the review are listed, to serve as a succinct summary (e.g., "*Keywords: general pain, depression, aggressive behavior, memory problems*" for Review #4 in Fig. 3), and the specific side effect mentioned by the user is highlighted in red. Users can browse through the displayed set of review summaries and open up any summary to see the expanded text. The type of drug that each review commented on is also listed, as well as the age of the reviewer.

To enrich the response, the system not only gives a direct answer to users' questions regarding to a specific side effect or a specific drug, but also provides additional relevant information, such as mentioning those side effects that are in the same category of the queried side effect, or side effects that often co-occur with the queried side effect, or specific

drugs that may trigger the side effect. In this way, the system can give users feedback that will suggest possible follow-on queries and help the users to stay within the knowledge domain of the system.

For example, a potentially useful operation is to assess whether two side effects tend to co-occur. Sentence III in Table III inquires about side effects that may co-occur with a specific side effect, in which case a co-occurrence retrieval event will be triggered by the command "*list_co_occur_side_effects*". The system then extracts the most frequent side effect topics from the set of reviews relevant to the queried side effect. These topics are considered as the side effects that most probably co-occur with the queried side effect.

After retrieving the candidate results from the review database and encoding the information into a response frame, the response generation component converts the response frame into a well-formed English string, via GENESIS generation rules [1], which is further synthesized as the speech response from the system.

IV. System Evaluation

For the system implementation and evaluation, we conducted two runs of data collection, where the first run was used as a development set to help us improve the recognizer language model and the NL (natural language) parse coverage, and the second run served as our test set. In each user session, we first show the user a demo video to demonstrate the capabilities of the system. Then the user is asked to conduct a short practice interaction, in order to become comfortable with our speech interface and the overall system configuration. When the user is ready, he/she starts a new session as the formal recording session. Users' utterances were captured through a headset microphone and later transcribed.

From the initial data collection episode, we collected 198 utterances from real users. All the utterance transcripts were subjected to parse analysis. The grammar was then expanded to accommodate any utterances that failed to parse but were well formed and within the domain of the system. We also expanded our templates for generating training sentences to include any new patterns observed from these utterances. A new set of training utterances was created from the expanded templates, and used to rebuild the recognizer language model, prior to the second data collection run. We also improved the semantic meaning extraction and response generation by modifying grammars and generation rules, when necessary.

Table IV (the column labeled "Development set") gives the parsing and speech recognition performance on the development set before and after expanding the grammar and the template platform with the collected users' utterances. The parsing rate is evaluated based-on the transcripts, which represents the NLP (natural language processing) performance achieved if recognition were perfect. Thus, the parsing rate also demonstrates the performance of the system on text queries.

The performance on the development set improved significantly after expanding the grammar and the templates with the collected user data. The original parsing rate was 88.4% and the word error rate on speech recognition was 32.3%. After retraining the grammar probability model and the recognizer language model, the NL parsing coverage increased to 98.0% (improved by 10.9%) and the speech recognition word error rate dropped to 11.8% (improved by 63.5%). This shows that the data collection from real users contributes significantly to the grammar construction and the language model training.

TABLE IV
PARSING AND RECOGNITION PERFORMANCE ON DEVELOPMENT AND TEST SETS

Dataset	Development set		Test set
	Before retraining	After retraining	
Parsing rate on transcripts	88.4%	98.0%	94.0%
Word error rate in speech recognition	32.3%	11.8%	27.3%

The second run of data collection was used for system evaluation. The procedure of the user experiment was identical to the first run: a video demo, a short practice session with the drug system, and a formal recording session. A total of 184 utterances were collected from this data collection episode, which were taken as the test set. The performances of parsing as well as speech recognition on the test set are shown in Table IV (the column labeled "Test set").

On the test set, the parsing rate was 94.0%, outperforming the development set (88.4%) by 6.3%. The word error rate in recognition (27.3%) was lower than that of the development set (32.3%) by 15.5%, which indicates an improved recognition capability on unseen data. The parsing rate (94.0%) was intermediate between the rates on the development set before and after grammar enhancements. The recognition word error rate (27.3%) was significantly above that of the development set (11.8%) after tuning. The most likely contributing factor is that the language model was retrained with the utterances in the development set, while the test set is unseen. Most of the recognition errors are due to uncovered sentence patterns and out-of-vocabulary words. With more iterative data collection and language model retraining, the recognition on unseen data can be expected to converge towards that on the development set. Meanwhile, the dialogue system can be used in the typing mode as well even as we improve its performance in the speaking mode, as the good parsing performance suggests that the system can be used very robustly for text input.

A second factor of recognition errors is the poor recording quality of some utterances in the second data collection episode. We examined each session and found that 35 utterances (19% of the test set) had very poor recording quality due to low voice volume of the user and high background noise. We also examined the system response in each dialogue turn, and observed that 72.3% of users' queries were answered correctly. Among the failed cases, 80.4% were due to recognition errors and others were due to parsing or discourse failure.

A questionnaire was also collected for an intuitive evaluation, where each user was asked to give a numerical rating from 1 to 5 on each aspect of the system. The aggregated results are given in Table V, where the columns of "I" and "II" represent the average rating from users in the first and the second data collection respectively. Compared to the first run, the rating in the second data collection on "speech" increased by 43.3% (from 3.0 to 4.3); the rating on "responses" increased by 32.1% (from 2.8 to 3.7); and the rating on "interface" increased by 25% (from 3.2 to 4.0). There was little difference on "understanding" and "helpfulness" (3.2 vs. 3.0 and 3.6 vs. 3.3). The increased ratings are likely due to improvements in the recognizer and the NL components.

TABLE V
AVERAGE RATINGS FROM REAL USERS IN BOTH DATA COLLECTION EFFORTS

Questions	I	II
Did the system ***understand*** your questions well?	3.2	3.0
Is the information from the system ***helpful***?	3.6	3.3
Were the ***responses*** from the system natural?	2.8	3.7
Is the ***interface*** to the system intuitive and easy to use?	3.2	4.0
Does the ***speech*** interface enhance the system?	3.0	4.3

V. CONCLUSION

In this paper, we have described a new Web-based multimodal system providing users with a rich facility for exploring the association of prescription drugs with possible side effects. Through standard log likelihood ratio estimation, we have revealed a statistically significant association between drugs and various side effects by examining drug reviews collected from health-related web sites. We implemented a multimodal interface that allows users to inquire about these drug side effects and browse through patient-provided drug reviews via speech, text and gestures.

For future work, we plan to maintain an online version of the system and make it available to the general public. We will continue to improve its performance through a larger-scale data collection from general users. The number of drug classes represented will be expanded as well. A more ambitious future goal is to develop a speech interface for *harvesting* spoken review data, which can allow users to add their own experience on drug side effects through natural speech and text. We will also explore methods to use crowd-sourcing to aid in the transcription of these recordings [13]. Another broader direction is to explore a framework that supports multimodal access to user-generated content in other domains, with a universal speech interface and a generalized platform for unstructured data processing.

ACKNOWLEDGMENT

This research is supported by Quanta Computers, Inc. through the T-Party project.

REFERENCES

[1] L. Baptist and S. Seneff, "Genesis-II: A versatile system for language generation in conversational system applications," *Proc. ICSLP*, 2000.
[2] J. Cable, "Adverse Events of Statins - An Informal Internet-based Study," *JOIMR*, 7(1), 2009.
[3] T. Dunning, "Accurate methods for the statistics of surprise and coincidence," *Computational linguistics*, 19(1):61–74, 1993.
[4] J. Glass, "A Probabilistic Framework for Segment-Based Speech Recognition," *Computer Speech and Language 17*, 137-152, 2003.
[5] B.A. Golomb, M.H. Criqui, H. White, and J.E. Dimsdale, "Conceptual foundations of the UCSD Statin Study: a randomized controlled trial assessing the impact of statins on cognition, behavior, and biochemistry," *Archives of internal medicine*, 164(2):153, 2004.
[6] A. Gruenstein, I. McGraw, and I. Badr, "The WAMI Toolkit for Developing, Deploying, and Evaluating Web-Accessible Multimodal Interfaces," *Proc. ICMI*, 2008.
[7] Q. Gu, C.F. Dillon, and V.L. Burt, "Prescription drug use continues to increase: U.S. Prescription drug data for 2007-2008," *NCHS Data Brief*, (42):1, 2010.
[8] J. Liu and S. Seneff, "Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm, " *Proc. EMNLP*, 2009.
[9] J. Liu, S. Seneff and V. Zue, "Utilizing Review Summarization in a Spoken Recommendation System, " *Proc. SIGDIAL*, 2010.
[10] J. Liu, S. Seneff, and V. Zue, "Dialogue-Oriented Review Summary Generation for Spoken Dialogue Recommendation Systems," *Proc. NAACL-HLT*, 2010.
[11] J. Liu, A. Li and S. Seneff, "Automatic Drug Side Effect Discovery from Online Patient-Submitted Reviews: Focus on Statin Drugs," *Submitted to the First International Conference on Advances in Information Mining and Management*, 2011.
[12] M.G. Mohaupt, R.H. Karas, E.B. Babiychuk, V. Sanchez- Freire, K. Monastyrskaya, L. Iyer, H. Hoppeler, F. Breil, and A. Draeger, "Association between statin-associated myopathy and skeletal muscle damage," *Canadian Medical Association Journal*, 181(1-2):E11, 2009.
[13] I. McGraw, C. Lee, L. Hetherington, S. Seneff, and J. Glass, "Collecting Voices from the Cloud," *Proc. LREC*, 2010.
[14] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford, "Okapi at TREC-3," *Proc. the third Text REtrieval Conference (TREC)*, 1994.
[15] S. Seneff, "TINA: A natural language system for spoken language applications," *Computational Linguistics*, 18(1):61--86, 1992.
[16] G. Wainwright, L. Mascitelli, and M.R. Goldstein, "Cholesterol-lowering therapy and cell membranes: stable plaque at the expense of unstable membranes? " *Arch Med Sci*, 5:3, 2009.
[17] Y. Xu and S. Seneff, "Dialogue Management Based on Entities and Constraints, " *Proc. SIGDIAL*, 2010.