

Appurtenant: Enhancing Completeness and Efficiency of Bidirectional Patient-Physician Communication Using Automatic Speech Recognition

Guisi Ni
Department of Computer
Science
Wayne State University
Detroit, MI
ni.guisi@gmail.com

Weisong Shi
Department of Computer
Science
Wayne State University
Detroit, MI
weisong@wayne.edu

Prashant Mahajan
School of Medicine
Wayne State University
Detroit, MI
PMahajan@dmc.org

ABSTRACT

Optimal patient-physician communication is integral component in the rendering of medical care. Despite the importance of adequate communication, substantial barriers including time constraints, inadequate provider expertise and variation among patients in their ability to provide complete history of illness, exist which do not allow a comprehensive clinical evaluation at the patient's bedside. As governments and the public scrutinize healthcare costs and quality, it is imperative that we develop tools such as speech recognition based support systems that will assist the provider in obtaining a complete history and physical examination. Here we propose a mobile healthcare assistant system, Appurtenant, aiming at improving completeness and efficiency of bidirectional patient-physician communication for less experienced clinicians without interruption and reducing costs of care by ensuring a standardized approach to patient evaluation.

Categories and Subject Descriptors

J.3 [LIFE AND MEDICAL SCIENCES]: Medical information systems; C.3 [SPECIAL-PURPOSE AND APPLICATION BASED SYSTEMS]: Real-time and embedded systems

Keywords

Patient-Physician Communication, Automatic Speech Recognition, Mobile

1. INTRODUCTION

In the past years reams of researches have been invested in the e-health development from both technological and medical world. Emerged as subsegment of e-health, m-health (mobile health) which focuses on integrating health care systems to ubiquitous mobile platforms to improve the experi-

ence from physicians' point of view to patients' appreciate is considered a promising domain in academia[12, 16]. While a great deal of efforts have been involved, there exists a curse between technological innovations and large scale application due to many issues remain in the field to be addressed including adapting to realistic medical needs, cooperating with existing e-health systems and physicians' acceptance[9, 7].

Patient-physician communication is an integral component in the rendering of medical care. "Medicine is an art whose magic and creative ability have long been recognized as residing in the interpersonal aspects of patient-physician relationship." [11] Indeed, the initial communication between the patient and the physician i.e., elicitation of the history of current illness, past medical problems and ancillary history is integral to the provider for developing a list of provisional diagnoses or differential diagnoses, which then directs the provider to perform either a focused or comprehensive physical examination and obtain supportive diagnostic tests to arrive at the final diagnosis or narrow down the list of differential diagnoses. Effective patient-physician interaction in not just the start up move, the resultant communication is traditionally regarded as heart and art of medicine[10]. An optimal initial patient-physician encounter has the potential to create a harmonious interpersonal relationship between the physician and the patient, stabilize patients' attitude and status, assure adherence to the treatment procedure, enhance patient satisfaction thus achieving the goals of delivering high quality cost-effective health care.

Despite the importance of this patient-physician communication and the need to obtain a comprehensive history/physical examination, current research reveals that substantial barriers exist that preclude an adequate or complete encounter. A national survey reveals that 75% of the orthopedic surgeons believed that they communicated satisfactorily with their patients, but only 21% of patients reported satisfactory communication with their physicians[20]. The barriers lie in both the clinical competence of the physicians and their communication skills. For instance, trainees either as medical students/post-graduate interns or junior residents were often responsible for first evaluation, assignment of diagnosis and treatment for newly admitted patients in a teaching hospital and the general poor overall performance in the quality of history-taking and physical-examination skills was secondary to inexperience[17]. Furthermore, there are barriers such as language, cultural differences in the ter-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
MARS'14, June 16, 2014, Bretton Woods, New Hampshire, USA.
Copyright 2014 ACM 978-1-4503-2823-4/14/06 ...\$15.00.
<http://dx.doi.org/10.1145/2609829.2609830>.

minology used and importance given to medical complaints and finally lack of time spent in the actual patient-physician encounter. The United States is on the way to becoming a more diverse nation. According to projections by the US Census Bureau, minorities, consisting of all but the single-race, non-Hispanic white population now 37 percent of the U.S. population, are projected to comprise 57 percent of the population in 2060[3]. The natural cultural disparities make the bias even more difficult for physicians to conquer.

As governments and the public pay more and more attention to the quality of health care, research regarding the topic of observation and evaluation of the patient-physician consultations is being conducted in the field of medicine. Methods have been developed to evaluate physicians' clinical competence [6], assess patient satisfaction [14] and check the quality of physician-patient interaction[13]. However, few studies have focused on improving the completeness and efficiency of the information obtained during patient-physician encounter.

Here we propose a technology driven mobile assistant - an "appurtenant" that will augment patient-physician interaction by ensuring completeness of history taking/physical examination of the initial encounter by using automatic speech recognition and reminder mechanism. Consequently, Appurtenant will improve the efficiency of patient-physician interaction to a great extent.

There exist long-term efforts to enable speech recognition technology in health care on the market. The benefits are apparent; including efficiency, flexibility, less manual effort, less costs to name a few. M-model is working on embedding speech recognition into electronic medical record system in healthcare institutions and allowing physicians to note records in real time on mobile devices during the clinical interaction[1]. Nuance, which is famous for the Dragon products, is even more advanced in its cooperation with Epic and Cerner[2]. But few efforts have focused on exploring the potential of technology in assuring the quality and completeness of the information obtained during the patient-physician encounter. The proposed approach is aimed at creating a paradigm in the patient-physician encounter that will eventually be used to train residents of all medical and surgical fields, reduce costs of care by ensuring a standardized approach to patient evaluation, enhance patient safety and improve quality and patient satisfaction by engaging more actively during the patient-physician encounter.

2. SYSTEM DESIGN

Our solution is an assistant system, which enables the technology of the automatic speech recognition synchronous with patient-physician interaction not only to facilitate the information gathering/documenting process but also to utilize the inherent nature of the technology to improve the quality of clinical encounter during and after the information has been captured.

The focus of our entire system is to ensure the completeness of information collection, processing and transformation during the initial communication between patient and physician. A comprehensive understanding of patient's illness condition is the crucial start move of the whole healthcare delivery process which affects physician's decision to perform focused physical examinations and narrows the region of final diagnoses to achieve the goal of high-quality and cost-effective health care providing. Traditional patient-

physician communication takes place in the form of questioning answering procedure, thus the completeness of the process can be measure in number of questions should be asked compared to number of questions actually asked.

$$Completeness = \frac{Number\ of\ questions\ should\ be\ asked}{Number\ of\ questions\ actually\ asked}$$

However, complete information collection in patient physician communication requires the physician involved to be both experienced in the specific situation and skilled in the dialogue with kinds of patients. Even though many medical education and training institutions have put emphasis on this and provide associated training program, it is still a challenge with pressure for medical students, post-graduate interns, junior residents or even in some cases experienced physicians. With mobile health development becoming cumulatively flourishing, we want to propose an approach to improve the completeness of patient-physician encounter utilizing the built-in nature of the increasingly powerful smartphones platform.

Though currently the healthcare organizations are undergoing the procedure of replacing the paper-based medical records with a more computerized way, known as Electronic Medical Records (EMR). The software currently used in hospitals and clinics is mostly like an electronic notebook that simply records every input physician provides without reaction, this is not helping with the complexity of the evaluation process and medical record. Part from the completeness, efficiency is not improved significantly comparing the transitional paper based medical record in the information obtaining process, which can be measured in time as

$$Tppe + Temr$$

where TPPE stands for time used in physician-patient encounter and TEMR stands for time used to format EMR.

The speech recognition technology allows our system to recognize the content and context of the conversation between the physician and the patient into text form as a real time function instead of the general way of completing the record in digital form by typing. This makes our system far more efficient and allows the system to work synchronously without further manual effort. Superabundant interference, which is considered as one of the appending drawbacks of frequent occurrence in many computer systems, is definitely inappropriate in health care system.

As we discussed before, to meet the requirement of completeness and achieve the decision-making assistant feature of computer-based system, after the information captured from speech recognition component, the function of keywords searching is implemented. And according to our specific requirement, a set of presupposed supplemental questions corresponding to the related keywords would be preimplemented in our system database. While the physician is communicating with the patient, the keyword searching function will be running at the background in comparison to the list of keywords in the database till the keyword or other form of the existing keyword is found. Then the reminder is called to assist the whole physician-patient interaction.

In order to improve the performance and flexibility of the system, a clinician framework is also implemented in the system to allow users to customize the tips and the keywords they are trying to get. Physician is one side of the whole

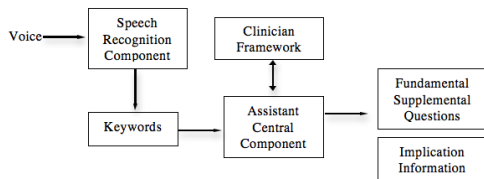


Figure 1: The functional blocks of the Appurtenant System

interaction procedure whose clinical competence and communication skills directly determine the outcome and quality of health care delivery. This brand new feature makes the whole assistant system transparent to physician in terms of total human language. Using clinician framework experienced physicians in clinic and hospital to provide and edit information will surely improve the performance and reliability of the system.

The reminder mechanism is the guarantee of the system completeness and intelligence. The reminder for supplemental questions should have two options – (a) simultaneous reminder so that supplemental questions can be asked in real-time and (b) delayed reminder to ensure completeness. To make sure the supplemental questions can be asked in real time, the questions should turn up in the fixed zone on the screen right after the keyword has been detected without any operation. In this way, we plan to reduce both the distraction between the physician and the patient and the information-missing rate to the lowest level. And to assist decision making and increase the interaction completeness, the delayed reminder contained the knowledge of implications of the given symptoms is implemented before saving and uploading the information. Also, the directly saving and uploading is allowed for simple case and experienced physicians.

2.1 Speech Recognition Component

The first adoption of practically functional automatic speech recognition systems appears in late 1980's and the interest to apply this technology in the area of healthcare has been revealed since then[5]. Even though the quality and accuracy at that stage is far from mature for medicine system, the effort has never stopped. But until recent days, as the promotion of powerful mobile devices, the incorporation of speech recognition technology and medicine systems is just emerging.

As shown in Fig 2, the performance of our system is applied on the iOS platform. Server-based automatic speech recognition is conducted due to the hardware power limit of mobile devices. The speech recognition engine converts the voice captured and recorded into time-based digital signal and views it as the input for the acoustic model and grammar model. A list of words corresponding to the input signal will be given in the order of calculation results after computed in the acoustic model. And the grammar model is responsible for language consistency of the given input. A best match of the recorded voice content will be transmitted to the mobile devices from the cloud as the final text converted.



Figure 2: Speech Recognition in Appurtenant

This speech recognition approach shares the common benefits of most technology solutions as improving efficiency and reducing costs, comes with the advantages of reusable data, flexibility and productivity. Collaboration with the following mechanism of keywords searching, clinical framework and supplemental questions and implication information providing will allow quality and completeness improving without interference. To fit our specific purpose, the adoption of our system on mobile devices is another guarantee of efficacy due to the nature of the devices. The smaller, touchscreen environment may also contribute to reduce the distraction between the physician and the patient during the communication.

Thus the time used to input medical records into EMR system with manual effort have been eliminated to improve the efficiency of bidirectional patient-physician communication.

2.2 Assistant Central Component

Although the final diagnosis may differ in each patient, it is usually the presence of or absence of a constellation of signs and symptom that lead to the eventual diagnosis. The arrival at the correct diagnosis depends on the completeness of this evaluation. For instance, vomiting can be a presenting symptom for more than 30 different illness, however, presence of projectile non-bilious vomiting in an infant 3 months of age and younger is sine quo non i.e. diagnostic for hypertrophic pyloric stenosis which is a medical emergency. Thus if a trainee were to obtain a history of vomiting but overlook the important question of projectile nature, then the differential diagnosis is now possibly reflux disease, hypertrophic pyloric stenosis, gastritis, gastroenteritis, meningitis and overfeeding to name a few. This would have led the provider to obtain unnecessary investigations such as radiographs or blood tests. No single human being can remember all ancillary questions that need to be asked for every symptom and this is what Appurtenant is trying to achieve - use of certain key word triggers will prompt specific questions that will ensure completeness of this encounter.

Other than that, physicians will need reminders in case they may skip important information that is necessary to acquire from the patients. To meet the requirement of completeness and achieve the decision-making assistant feature

of computer-based system, the function of keywords searching is implemented.

After the detection of keywords the related supplemental questions should turn up on the screen in the format of simultaneous reminding to the questions can be asked by the physician in real time. And before saving and uploading the text file, delayed reminder will be implemented to ensure completeness. This function is depending on the physician's willing and the direct saving and uploading is also allowed. Uploading new question and it corresponding key word will contribute to the completeness of the system reminder service.

With Appurtenant, the completeness of the information obtaining procedure is improved on account of fundamental supplemental questions and delayed implication information corresponding to triggered key word thus strengthening the bidirectional patient-physician communication and saving money and time at the same time.

2.3 The Clinical Framework

The role of physician played in clinical interaction is well established in literature. Though there are doubts raised about physicians' ability in clinical encounter, human beings as the central elements of the entire physician-patient encounter are still considered more flexible than machines since a frequent update is needed for any software in order to keep it advanced and capable enough to meet the requirement of users.

In order to improve the performance and flexibility of the system, we allow experienced users to customize both keywords and corresponding supplemental information. Though iOS application is mainly programmed in high level programming, it is still hard for common users who do not have computer science background. However, we also notice that in database the content showing on the UI Interface are and always will be Human Language. Parsing value in human-language form to each key in the database for experienced user to input or change will customize the predefined database in general purpose or in one particular medical domain aiming at.

The schema is every time users make some changes, data will be stored in HTML form as plist file and then connected to local database or cloud database. And for the point of users' experience, this serves as the enchantment to modify the code of application.

3. CASE STUDY

Let's take the case scenario of baby vomiting as a case study. Our system will prompt clinicians in real-time to ask supplemental questions based on patient's history to embellish the encounter and ensure completeness. For instance, a two-month-old infant with vomiting can potentially have multiple etiologies. But if this history is embellished with sub-questions including frequency, color, temporal relation to feeding and history of weight loss a diagnosis of hypertrophic pyloric stenosis can be made clinically thus obviating the need for investigations that are not only costly and unnecessary but are actually harmful to the patient. There are routine questions that work in specific situation while consulting physicians, the system will save these questions and finish these parts together with physicians, which will avoid this both time-consuming and repeating part, in ad-

dition, far less omission could occur. The patient-physician interaction may appear as in Fig 3.

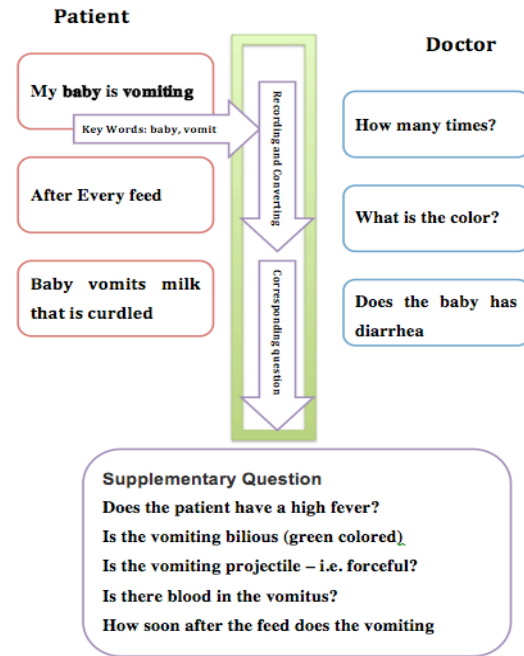


Figure 3: A Simple Scenario of Baby Vomiting

From this scenario, five questions are pushed forward. For example, every time the keyword of "vomiting" has been detected; a reminder is required to pump up with the option of switching to further questions. Human-centered interface will be designed in this part to reach the goal of avoiding physician distraction. And finally the received electronic-version medical records will be managed by the database center for searching, reviewing, sharing and storing. Based on the speech recognition we implemented in our system, function for keywords and supplement information become feasible. Speech recognition converts audio data into verbal data so searching the content of the discussion after convert if any word matches the key word in database will remind physicians as the reminder in the user interface.

With Appurtenant, the questions asked by physician increase from three to eight in our specific situation with vomiting due to our case scenario thus improving the completeness during the information obtaining process and saving money and time for further investigations which might be potentially harmful to the patient.

4. PERFORMANCE EVALUATION

This section evaluates our Appurtenant system based on the case study illustrated in the previous section. Table 1 shows the experimental results using our baby vomiting scenario where APA is for Audio Processing Activity and GA stands for Graphics Activity. The accuracy of speech recognition plays as a vital role in the performance of the system. In the experiment, we use the google speech-to-text API as the speech recognition server. Martin Cutts offers the standard average sentence length to be 15-20 words for effective communication in Oxford Guide To Plain English[8]. And the average word length in syllables is 1.74[15]. Plus

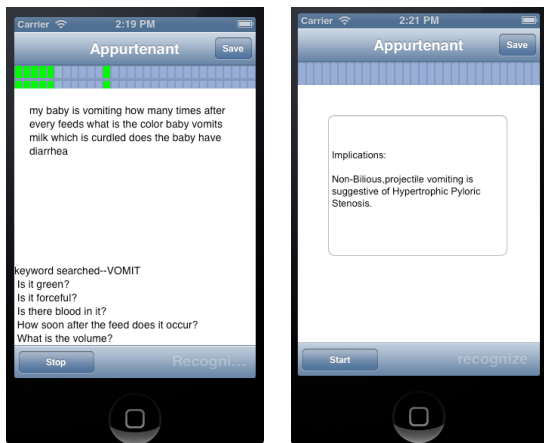


Figure 4: Performance of Supplemental Questions and Implications in Appurtenant

the English rates vary between 3.3 and 5.9 syls/sec during speaking according to Arnfield and Roach[4]. We can expect the spoken time for each sentence length to be between 4.4 and 10.5. In the experiment, we calculate the performance accuracy on a 10 sec base. As revealed from the table, the overall accuracy will be 82.17%.

Table 1: Evaluation of Appurtenant System

Timeline	Accuracy	Data In	Data Out	APA	GA
0-10 s	100%	3 kb	78 kb	2.2%	9.1%
10-20 s	90%	6 kb	155 kb	1.4%	7.45%
20-30 s	42%	6 kb	149 kb	1.4%	7.8%
30-40 s	85%	6 kb	163 kb	1.4%	6.8%
40-50 s	76%	8 kb	174 kb	1.4%	7.9%
50-60 s	100%	7 kb	178 kb	1.5%	7.9%

The deployment of server-based speech recognition technology makes the networking condition a significant factor affecting the overall accuracy of the system. As an important step of server-based speech recognition, the transmission of audio data to server makes dominating contribution to the outgoing network traffic and requires sustainable audio processing activity from CPU. In the mean while, the data incoming which includes the speech to text result will be much less compared to the outgoing data. And the overall energy consumption based on a 20-level measure in as shown in Fig 5.

As a general view, from the baby vomiting scenario the questions asked during the patient-physician encounter increase from three to eight with efficiency enhancement using server-based speech recognition at the same time and assuring the quality of health care delivered.

5. CURRENT STATUS AND NEXT STEP

5.1 Current Status

As shown in Fig 2 and Fig 4 , we have implemented the speech recognition component, key word detection, and two kinds of assisting mechanism, both simultaneous reminder so that supplemental questions can be asked in real-time and delayed reminder giving implications to ensure com-

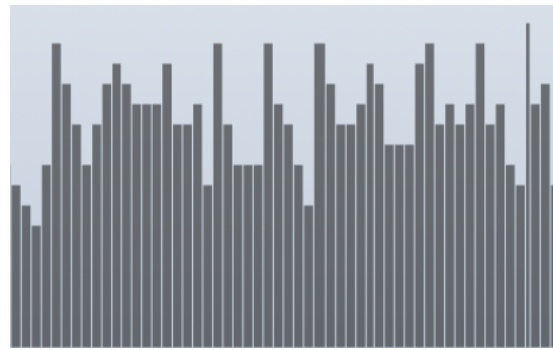


Figure 5: Levels of Energy Consumption

pleteness. Currently our appurtenant system is carried out on the platform of iOS and can be put aside during the physician-patient interaction to improve efficiency and completeness without additional operation.

5.2 Speaker Distinguishing

The first challenge we encounter in the practical implementation is to distinguish different speakers when they are interacting in front of the device. However, at the present stage, there is no effective approach to solve this problem regarding our system environment and situation. The voiceprint recognition technology only allows speaker recognition with voice information collecting in advance. And other approach with extra hardware implementation will add unexpected burden to the existing system and create inappropriate interference in the physician-patient communication procedure.

However, distinguishing multi-users' voice is of importance in assuring the system readable and editable, improving accuracy in converting audio data into verbal information and promoting the use of the speech recognition technology in the standard EMR system during clinical encounter process.

5.3 Complete Scenario

Due to time and resource limitations, the complete scenario for the central database has not been implemented at the present stage. Other than that, data source can be close to infinity since the medical knowledge is broad and profound. Practical diagnose process will always contribute to the database setting up which will eventually benefit user satisfaction. Separate the big data into detailed category is also necessary and of value since medical filed has various department and each physician has what one is skilled in or expert in aiming at treating various patients more professionally.

5.4 Collaborating with Existing EMR System

Ever since the current technology of speech recognition accuracy and strict format requirement in medical record in medicine system of America, the system output still cannot meet the requirement the EMR system with no modification. A better mechanism can be explored to make sure passing rigorous assessments in medical field.

6. CONCLUSIONS

The Appurtenant system exploits the innovative technology of the Automatic Speech Recognition in the procedure of

physician patient interaction to facilitate the information obtaining and documenting process efficiently and effectively. The central assistant system utilizes the inherent nature of the technology to improve the quality and completeness of clinical encounter after the information has been captured. The whole system is conducted under iOS platform to guarantee efficiency and flexibility from another perspective.

Our goal is to create a brand new bridge to experience and qualified clinical interaction for interns/residents and less experienced physicians during physician-patient interaction without interference. Screenshots of the implementation performance has been displayed in the paper. Future work need to be done to distinguish different speakers, complete practical scenario and corporate with existing EMR system in medicine institutions.

7. REFERENCES

- [1] M-modal. <http://mmodal.com>.
- [2] Nuance for healthcare. <http://www.nuance.com/for-healthcare/index.htm>.
- [3] U.s. census bureau projections show a slower growing, older, more diverse nation a half century from now, 2013. [CIC Report 146 January 2013].
- [4] S. Arnfield, P. Roach, J. Setter, P. Greasley, and D. Horton. Emotional stress and speech tempo variation. In *Speech under Stress*, 1995.
- [5] S. L. B Bergeron. Speech recognition as a user interface. *M.D. computing: computers in medical practice (M Comput)*, 7(5):329-34. DOI:10.1121/1.2355999, 2006.
- [6] F.-G. Boucher, W. H. Palmer, G. Page, R. Barriault, and J. Seely. The evaluation of clinical competence. *Canadian Family Physician*, 26:151, 1980.
- [7] V. W. Consulting. mhealth for development: the opportunity of mobile technology for healthcare in the developing world. *Washington DC and Berkshire, UK*, 2009.
- [8] M. Cutts. *Oxford guide to plain English*. Oxford University Press, 2013.
- [9] P. Germanakos, C. Mourlas, and G. Samaras. A mobile agent approach for ubiquitous and personalized ehealth information systems. In *Proceedings of the Workshop on Personalization for e-Health Of the 10th International Conference on User Modeling (UM 2005)*, pages 67–70, 2005.
- [10] J. F. Ha and N. Longnecker. Doctor-patient communication: a review. *The Ochsner Journal*, 10(1):38–43, 2010.
- [11] J. A. Hall, D. L. Roter, and C. S. Rand. Communication of affect between patient and physician. *Journal of Health and Social Behavior*, 1981.
- [12] M. Kay, J. Santos, and M. Takane. mhealth: New horizons for health through mobile technologies. *World Health Organization*, 2011.
- [13] F. Lehmann, L. Côté, A. Bourque, and D. Fontaine. Physician-patient interaction: a reliable and valid check-list of quality. *Canadian Family Physician*, 36:1711, 1990.
- [14] F. Lehmann, D. Fontaine, A. Bourque, and L. Côté. Measurement of patient satisfaction: The smith-falvo patient-doctor interaction scale. *Canadian Family Physician*, 34:2641, 1988.
- [15] Nirmaldasan. The average sentence length. <http://strainindex.wordpress.com/2008/07/28/the-average-sentence-length/>.
- [16] G. Phillips, L. Felix, L. Galli, V. Patel, and P. Edwards. The effectiveness of m-health technologies for improving health and health services: a systematic review protocol. *BMC research notes*, 3(1):250, 2010.
- [17] S. Sharma. A single-blinded, direct observational study of pgy-1 interns and pgy-2 residents in evaluating their history-taking and physical-examination skills. *The Permanente Journal*, 15(4):23, 2011.
- [18] T. Tamrat and S. Kachnowski. Special delivery: an analysis of mhealth in maternal and newborn health programs and their outcomes around the world. *Maternal and child health journal*, 16(5):1092–1101, 2012.
- [19] Z.-H. Tan and B. Lindberg. *Automatic Speech Recognition on Mobile Devices and over Communication Networks (Advances in Pattern Recognition)*. Springer Publishing Company, Incorporated, 2008.
- [20] J. R. Tongue, H. R. Epps, and L. L. Forese. Communication skills for patient-centered carereasearch-based, easily learned techniques for medical interviews that benefit orthopaedic surgeons and their patients. *The Journal of Bone & Joint Surgery*, 87(3):652–658, 2005.
- [21] J. A. Tulskey, R. M. Arnold, S. C. Alexander, M. K. Olsen, A. S. Jeffreys, K. L. Rodriguez, C. S. Skinner, D. Farrell, A. P. Abernethy, and K. I. Pollak. Enhancing communication between oncologists and patients with a computer-based training programa randomized trial. *Annals of internal medicine*, 155(9):593–601, 2011.