Real-time Multi-media Translation for Healthcare: a Usability Study

Mark Seligman, Ph.D.
Spoken Translation, Inc.
1100 West View Drive
Berkeley, CA 94705
mark.seligman@spokentranslation.com

Mike Dillinger, Ph.D.
Spoken Translation, Inc.
1100 West View Drive
Berkeley, CA 94705
mike.dillinger@spokentranslation.com

Abstract

This paper reports on a usability study applying Converser for Healthcare, Version 3.0, a real-time, multi-modal, broad-coverage, highly interactive translation system, in three departments (Pharmacy, Inpatient Nursing, and Eye Care) of a large hospital complex belonging to a major US healthcare organization. We survey issues concerning software; use cases and setups; equipment and technology; and logistical and processing matters.

1 Introduction

Worldwide institutions for healthcare and government services have in recent years experienced increasing demands for interpretation between English and other languages. San Francisco General Hospital, for example, receives more than 3,500 requests for interpretation per month, or 42,000 per year for 35 different languages. Requests for medical interpretation services are distributed among many wards and clinics (Paras, et al., 2002).

Responding to this demand, several groups have implemented and experimented with systems for automatic real-time translation, with special interest in spoken language translation (Bouillon, Ehsani et al., 2006, 2008).

This paper reports on a usability study applying our real-time, multi-modal, broad-coverage, highly interactive translation system, Converser for Healthcare, Version 3.0, in three departments (Pharmacy, Inpatient Nursing, and Eye Care) of a large hospital complex belonging to a major US healthcare organization. It is hoped that future work will test later versions of Converser, updated in view of the observations to be discussed below.

As we are reporting on work in progress, formal evaluation will remain for later papers. (Sixty-five interviews with caregivers and patients have been conducted by independent investigators. Compilation of this evaluation data is now under way.) Here, we restrict ourselves to informal assessments of several aspects of the study, based on our own observations and preliminary user feedback. Overall, our aim is to explore several dimensions of usability for real-time translation systems in healthcare settings. As even this early exploration will demonstrate, accuracy is only one such dimension, necessary but not sufficient for true usability.

Section 2 will briefly describe the Converser system, sketching its approach to highly interactive real-time translation. We will describe the system's facilities for instant translations via pre-translated phrases and its provisions for multi-modal input. For fuller description, see (Dillinger and Seligman, 2004); (Zong and Seligman, 2005); and (Seligman and Dillinger 2006, 2008).

Sections 3 to 6 will discuss the following aspects of the study: software; use cases and setups; equipment and technology; and logistical and processing issues. We conclude in Section 7.

2 The Converser System

We now briefly summarize our approach to real-time automatic interpretation.

In speech-enabled translation systems, the twin goals of accuracy and broad coverage have generally been in opposition: systems have gained tolerable accuracy only by sharply restricting both the range of topics that can be discussed and the sets of vocabulary and structures that can be used to discuss them. The essential problem is that both speech recognition and translation technologies are still quite error-prone. While the error rates may be tolerable when each technology is used separately, the errors combine and even compound when they are used together. The resulting translation output is generally below the threshold of usability – unless restriction to a very narrow domain supplies sufficient constraints to significantly lower the error rates of both components.

Converser's approach has instead been to concentrate on interactive monitoring and correction of both technologies.

First, users can monitor and correct the speech recognition system to ensure that the text which will be passed to the machine translation component is completely correct. Voice commands (e.g. **Scratch That** or **Correct <incorrect text>**) can be used to repair speech recognition errors.

Next, during the machine translation (MT) stage, users can monitor, and if necessary correct, one especially important aspect of the translation – lexical disambiguation.

The system's approach to lexical disambiguation is twofold: first, we supply a *Back-Translation*, or re-translation of the translation. Using this paraphrase of the initial input, even a monolingual user can make an initial judgment concerning the quality of the preliminary machine translation output. Other systems, e.g. IBM's MASTOR (Gao, Liang, et al., 2006), have also employed re-translation. Converser, however, exploits proprietary technologies to ensure that the lexical senses used during back translation accurately reflect those used in forward translation.

In addition, if uncertainty remains about the correctness of a given word sense, the system supplies a proprietary set of Meaning Cues™ – synonyms, definitions, etc. – which have been drawn from various resources, collated in a database (called SELECT™), and aligned with the respective lexica of the relevant MT systems. With these cues as guides, the user can monitor the current, proposed meaning and when necessary select a different, preferred meaning from among those available. Automatic updates of translation and back translation then follow.

Such interactivity within a speech translation system can provide increased accuracy and confidence, even for wide-ranging conversations (Seligman, 2000).

Translation Shortcuts. The Converser system includes Translation Shortcuts™ – pre-packaged translations, designed to provide two main advantages:

First, re-verification of a given utterance is unnecessary, since it has been pre-translated by a professional (or, in future versions of the system, verified using the system's feedback and correction tools).

Second, access to stored Shortcuts is very quick, with little or no need for text entry. Two facilities contribute to quick access:

Shortcut Search can retrieve a set of relevant Shortcuts given only keywords or the first few characters or words of a string. The desired Shortcut can then be executed with a single gesture (mouse click or stylus tap) or voice command. If no Shortcut is found to match the input text, the system automatically and seamlessly gives access to broad-coverage, interactive speech translation.

A *Translation Shortcuts Browser* is provided, so that users can find needed Shortcuts by traversing a tree of Shortcut categories. Using this interface, users can execute Shortcuts by tapping or clicking.

Figure 1 shows the **Shortcuts Browser** facility.

- On the left, the **Translation Shortcuts Panel** contains the **Translation Shortcuts Browser**, split into two main areas, **Shortcuts Categories** (above) and **Shortcuts List** (below).
- The Categories section shows current selection of the **Conversation** category, containing everyday expressions. Categories for **Administrative topics** and **Patient's Current Condition** are also visible.
- The **Shortcuts List** contains a scrollable list of alphabetized Shortcuts in the selected Category.

The **Input Window** does double duty for **Shortcut Search** and entry of text for full translation. The search facility is shown in Figure 2.

- **Shortcuts Search** begins automatically as soon as text is entered – by voice, handwriting, touchscreen, or standard keyboard.
- The **Shortcuts Drop-down Menu** appears just below the Input Window. Here, the user has entered "Good" and a space, so the search program has received its first input word. The drop-down menu shows the results of a keyword-based search, with the first hit bolded for instant execution.
- Arrow keys or voice commands can be used to navigate the drop-down list.
- If the user goes on to enter the exact text of any Shortcut, e.g. "Good morning," the interface will confirm Shortcut recognition, indicating that verification of translation is unnecessary. However, final text not matching a Shortcut, e.g. "Good job," will undergo full translation with verification.

Multimodal input. In healthcare settings, speech input is not appropriate for every situation.

Current speech-recognition systems are unfamiliar for many users. Converser for Healthcare, Version 3.0, attempts to maximize familiarity by incorporating Dragon NaturallySpeaking, a standard commercial-grade dictation system for broad-coverage and ergonomic speech recognition – and a product which already has an established user base in the healthcare community. Even so, some training has been required. We will have more to say about this issue below.

To address such training and usability issues, we have provided a range of input modes. In addition to dictated speech, we enable handwritten input, the use of touchscreen keyboards for text input, and the use of standard keyboards. All of these input modes are completely bilingual, and language switching is arranged automatically when there is a change of active participant. Further, it is possible to change input modes seamlessly within a given utterance: for example, users can dictate the input if they wish, but then can make corrections using handwriting or one of the remaining two modes.

Despite this flexible range of input options, many issues remain. In particular, illiterate patients pose special problems. Naïve users tend to suppose that speech is the ideal input mode for illiterates. Unfortunately, however, the careful and relatively concise style of speech that is required for automatic recognition is often difficult to elicit, so that recognition accuracy remains low. Further, the ability to read and correct the speech recognition results is obviously absent. The remaining three text input modes will be equally ineffectual.

Our current approach to low literacy is to supply Translation Shortcuts for the minimally literate, and – in the future – to augment Shortcuts with text-to-speech and iconic pictures.

Staff members, while usually literate, present their own usability issues. Their typing skills may be low or absent. Handling the computer and/or microphone may be awkward in many situations, e.g. when examining a patient or taking notes.

To help deal with such awkwardness issues, our system provides voice commands, which enable hands-free operation. Both full interactive translation and the Translation Shortcut facility (using either the Browser or Search elements) can be run hands-free. To some degree, the system can be used *eyes*-free as well: text-to-speech can be used to pronounce the back-translation so that preliminary judgments of translation quality can be made without looking at the computer screen.

Having surveyed the Converser system, we now go on to discuss the current usability study. We begin with observations concerning software.

3 Software

Several software improvements in Converser and supporting programs have been suggested for future use:

- *Speech recognition:* During Phase One, the Dragon NaturallySpeaking (DNS) voice training (enrollment) process worked reasonably well for individual staff personnel, thus enabling spoken translation input from the English side. However, since voice training remained impractical for Spanish speakers, Spanish voice input was postponed. In the future, however, speech input requiring no voice training or enrollment should be enabled for both sides.

Also, Converser should add an on-screen push-to-talk button, thus eliminating the need to set up tablet buttons for each new Dragon user.

- *Interface and training requirements:* Staff trainees had no difficulty learning to use the Converser interface during the study. However, they also had to learn two other interfaces: the pen interface of Windows XP, Tablet Edition, and the voice interface of Dragon NaturallySpeaking, Version 10.x. Learning of all three elements should be eased going forward. With respect to the Converser GUI itself, among other improvements: (1) switching between English and Spanish users should be made less error-prone; (2) a no-check mode ("I'm feeling lucky mode") should be added so that translation verification can be bypassed when desired; and (3) large fonts should be enabled for the Converser Transcript Window and Translation Shortcuts Panel. With respect to handwriting and speech input, the respective upgrades, Windows 7 and DNS 11.x, are greatly improved and simplified.
- *Translation Shortcuts:* The *How to Use Converser* Translation Shortcut category should be further refined and merged into a new *Introductions* category in order to provide a smoother introduction for first-time users. In addition, facilities are needed for saving verified translations as Personal Translation Shortcuts™, and for sharing new Shortcuts with qualified colleagues.
- *Text-to-speech:* It should be possible to control the speed of text-to-speech, especially to slow it when desired. Alternation between male and female voices as appropriate should also be enabled.
- *Handwriting:* Handwritten entry of text to be translated has been popular with both staff and members. However, the correction process has required some learning, and checking for errors has been difficult for users with limited vision. Upgrade to Windows 7 will eliminate both issues, since the built-in handwriting facility is greatly improved.
- *Typing:* During the study, the Windows onscreen keyboard has been inadequate: the characters are too small and a stylus must be used. Thus typed input has been practical only at fixed locations and only for users familiar with keyboards. Windows 7 will feature a much-improved onscreen keyboard with large characters. Future Converser computers should also feature touchscreens, enabling onscreen text entry by finger.
- *Translation verification:* To enable tuning for specific use cases, Converser should include facilities for saving word meaning preferences within personal profiles – indicating e.g. that "tablet" should by default mean "pill" for individual or Pharmacy-wide use, rather than "tablet computer", etc. Tools should also be provided for quick tuning of group-wide vocabulary.
- *Transcripts:* Facilities are requested for saving Converser transcripts directly to the healthcare organization's electronic medical records system.
- *Upgrade to Windows 7:* Converser should be compatible with Windows 7 as well as Windows XP and Windows 2000.
- *Centralized installation and maintenance:* Facilities are also need for installing and maintaining Converser from a centralized location, to minimize the need to service separate machines in person.

4 Use Cases and Setups

During the study, Converser use and evaluation has been enabled in three departments. There were four use cases in Pharmacy, and one each in Inpatient Nursing and Eye Care. Each use case had its own work-flow and equipment setup.

- In the Pharmacy, the master computer could be stationary (in the **Consulting** or **Drop-off** use case); handheld (in the **Pick-up** use case); or on a cart (in the **Greeter** use case).
- In Inpatient Nursing, a handheld tablet PC was used throughout.
- In Eye Care, stationary use of the tablet was preferred in order to facilitate typing.

5 Equipment and Technology

During the study, following extensive pre-testing of computers and auxiliary equipment, the following equipment has been field tested:

HP EliteBook 2740P (folding convertible tablet PC) *(Use cases: Drop-off, Consultation)*
Wacom Cintiq Pen Display
TableMike microphone, with optional pedal
Keyboard for patient

Motion Computing F5v (tablet PC with handle) *(Use cases: Greeter, Pickup; Inpatient Nursing; Eye Care)*
Docking station
(optional) Logitek wired speaker
(optional) Motorola wireless (BlueTooth) speaker

The setups were evaluated as follows:

HP EliteBook 2740P setup

- *Good points:* The EliteBook setup with auxiliary display can be useful for relatively roomy over-the-counter situations, e.g. at the Pharmacy drop-off station. The EliteBook is fast, includes a touchscreen, and can run the healthcare organization's standard image. The Wacom Pen Display is high quality, and allows patients to see the Converser screen and use handwriting or typed input without handing the master computer back and forth. The TableMike provides excellent noise cancellation, has helpful hands-free operation and on-signal features, and can easily be traded back and forth between staff and member.
- *Drawbacks:* There is too much equipment for crowded work areas (such as the Pharmacy's consultation window). The EliteBook can be moved to make space, but staff personnel prefer to avoid this effort.
- *Conclusions:* This setup can be retained for over-the-counter situations with adequate space and no need for frequent movement of the equipment. Note that the HP can be used for portable as well as stationary situations: while it lacks a handle and is not liquid-tight, it can be folded into slate configuration and has its own keyboard attached.

Motion Computing F5v setup

- *Good points:* All Converser functionality is contained within the F5v unit for one-handed portability, and staff personnel seem comfortable with it. The unit is liquid-tight for leak-proof sanitation.
- *Drawbacks:* Sound volume is too low for noisy settings. The auxiliary speakers worked, but must be carried and handled as extra items. The docking station is very heavy, so it is usually left in one place. Peripherals (keyboard, etc.) can be connected through a clip-on EasyConnect bar to avoid moving the docking station, but fully integrated solutions are preferable. This machine did not yet have the healthcare organization's standard image during the study period.
- *Conclusions:* We can consider upgrading to e.g. the Motion Computing J3500, which has twin speakers, giving ample volume, and a portable clip-on keyboard, avoiding the need for a docking station and improving portability. The J3500 also has a touchscreen, minimizing the need for stylus use.

Additional points concerning both machines:

- The noise level and lighting in the interpretation settings can influence the technology solution and the effectiveness of interpretation.
- Securing the computers remained a challenge throughout the study. Additional research is needed to find the most practical solutions.

NOTE: During the study period, the popularity of mobile computing platforms, especially smartphones and new tablets in the iPad class, grew explosively. Converser releases for these platforms are under development, and can be considered for future use.

6 Logistical and Practical Issues

The following logistic and practical challenges have been addressed:

- *IT issues:* Initially, it was thought necessary to restrict even experimental Converser use to computers officially recognized by the healthcare organization and running its authorized software builds. These restrictions were challenging, since (1) Spoken Translation, Inc., as an outside vendor, was not authorized to have official accounts on those builds and (2) Converser makes use of facilities for handwriting recognition which were unavailable on many existing official machines. It was eventually decided that both official and non-official computers could be used temporarily for proof-of-concept purposes. *Conclusions:* (1) Plans should be made at project start to arrange limited working accounts for the use of software vendors. (2) The handwriting software issue is now becoming obsolete: with Windows 7, all machines will have the necessary handwriting programs.
- *Scheduling issues:* During the study, it was necessary to coordinate training, equipment preparation and installation, and experimental Converser use among several departments at several locations. This scheduling proved to be time-consuming, and the usability study activities added to the staff's current work load. There were delays when important participants took personal time off or holidays. In the Pharmacy, schedules had to be aligned among Converser users, members, and evaluators. *Conclusions:* In future projects, explicit individual scheduling can be planned once key personnel are fixed, and delays can be anticipated and built into timelines. However, most scheduling issues are restricted to usability studies or pilot projects: they will be resolved once software and equipment use of Converser is standardized and training is minimized.
- *Training issues:* During the study, a one-hour Converser training was necessary for participating staff. A half-hour training in the use of Dragon NaturallySpeaking software was also required for English speech input. As mentioned, both requirements will be reduced or eliminated during future projects. It was initially hoped that, once a core staff group was trained as Converser users within Pharmacy, they could spread the training to others, thus yielding a critical mass. However, the initial users proved to be too busy; so all users (eleven in three departments) were trained by the first author. To aid independent and user-led training, Converser contains four animated tutorials in both English and Spanish, plus full online User Manuals in both languages. For quick instruction of members/patients, an on-the-spot Spanish-language guide and special set of Translation Shortcuts were provided. *Conclusions:* In general, busy users cannot be counted on as trainers. Scheduling training sessions proved to be time-consuming, so ample time must be budgeted to the extent that training remains necessary. Shortcuts should be refined as experience dictates.
- *Coverage and guidance issues:* In the Pharmacy setting, only six out of 32 pharmacists and 18 clerks were trained in Converser use; so Spanish speakers too often wound up in lines leading to untrained pharmacists. The **Greeter** role was intended to guide members to appropriate lines, but the pharmacist designated for this role was often busy elsewhere, and in any case the coverage problem would remain if many lines were without Converser. *Conclusions:* In similar situations in future projects, procedures to guide members to Converser personnel can be rethought.

7 Conclusions

This paper has reported on a usability study applying Converser for Healthcare, a real-time, multi-modal, broad-coverage, highly interactive translation system, in three departments of a large hospital complex within a national US healthcare organization. We have discussed a range of usability issues concerning software; use cases and setups; equipment and technology; and logistical and processing matters. Each section of the discussion has included preliminary conclusions, with implications for future versions of this and other real-time translation systems. We believe that the study has demonstrated the wide range of factors affecting usability in such systems.

Acknowledgments

Contractual commitments prevent disclosure of the identity of the healthcare organization where the current usability study is in progress until its completion. The authors look forward to expressing their gratitude to the participants in future publications.

References

Pierrette Bouillon, Farzad Ehsani, et al. 2006. *Proceedings of the First International Workshop on Medical Speech Translation*, Association for Computational Linguistics. New York, NY, June, 2006.

Pierrette Bouillon, Farzad Ehsani, et al. 2008. *Proceedings of the Workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*, COLING 2008. Manchester, UK, August, 2008.

Mike Dillinger and Mark Seligman. 2004. "A highly interactive speech-to-speech translation system." In *Proceedings of the VI Conference of the Association for Machine Translation in the Americas*. Washington, D.C., September-October, 2004.

Yuqing Gao, Gu Liang, Bowen Zhou, Ruhi Sarikaya, Mohamed Afify, Hong-Kwang Kuo, Wei-zhong Zhu, Yonggang Deng, Charles Prosser, Wei Zhang, and Laurent Besacier. 2006. "IBM MASTOR system: multilingual automatic speech-to-speech translator." In *HLT-NAACL 2006: Proceedings of the Workshop on Medical Speech Translation*. New York, NY, June, 2006.

Melinda Paras, O. Leyva, T. Berthold, and R. Otake. 2002. *Videoconferencing Medical Interpretation: The Results of Clinical Trials*, Health Access Foundation. Oakland, CA, April, 2002.

Mark Seligman and Mike Dillinger. 2006. "Usability issues in an interactive speech-to-speech translation system for healthcare." In *Proceedings of the First International Workshop on Medical Speech Translation,* Association for Computational Linguistics. New York, NY, June, 2006.

Mark Seligman and Mike Dillinger. 2008. "Rapid portability among domains in an interactive spoken language translation system." In *COLING 2008: Proceedings of the Workshop on Speech Processing for Safety Critical Translation and Pervasive Applications*. Manchester, UK, August, 2008.

Chengqing Zong and Mark Seligman. 2005. "Toward Practical Spoken Language Translation." *Machine Translation,* 19(2): 113-137, June, 2005.

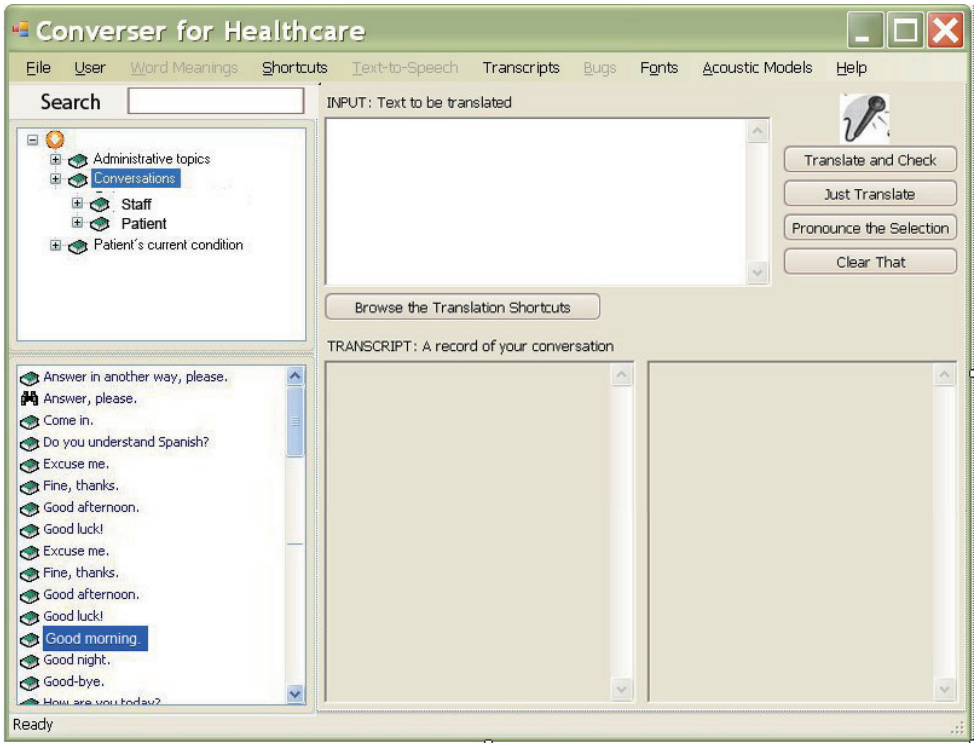

Figure 1: The Input Screen, showing the Translation Shortcuts Browser.

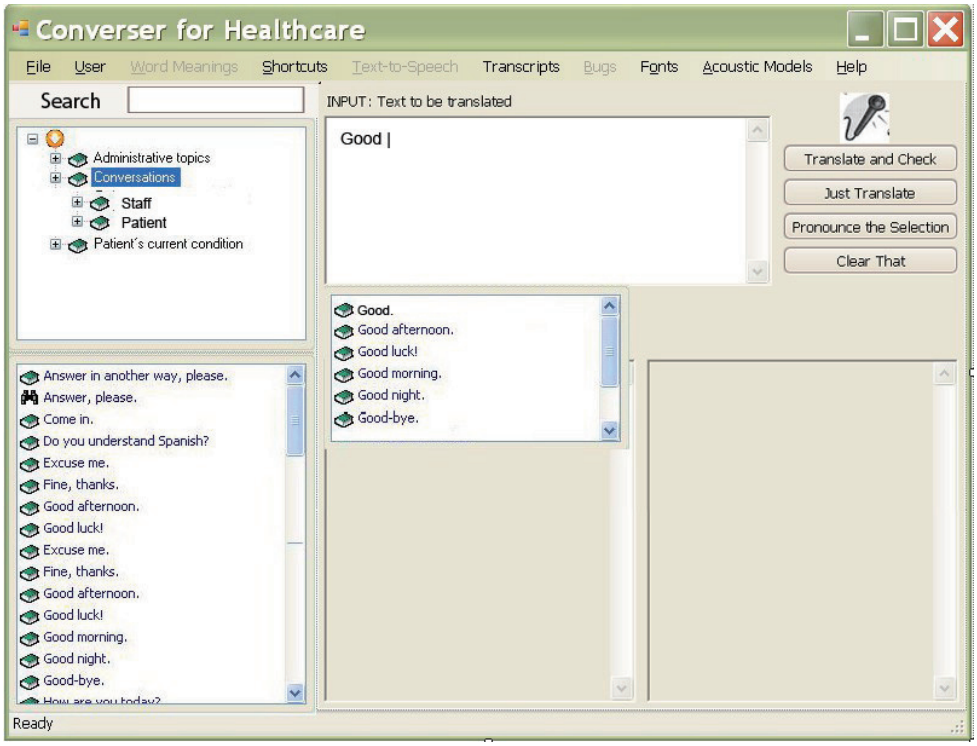

Figure 2: The Input Screen, showing automatic keyword search of the Translation Shortcuts.