A voice command interface for real-time interventional MR imaging

J. von Spiczak[1,2], E. Samset[1,3], D. F. Kacher[1], F. A. Jolesz[1,4], S. P. DiMaio[1]

[1]Radiology, Brigham & Women's Hospital, Boston, MA, United States, [2]University of Karlsruhe, Karlsruhe, Germany, [3]University of Oslo, Oslo, Norway, [4]Harvard Medical School, Boston, MA, United States

Problem (Background and Purpose)

As Magnetic Resonance Imaging (MRI) is used to guide complex interventions such as catheterization, biopsy, brachytherapy and robot assisted procedures, there is an increased need for the clinician performing such tasks to have direct control of the imaging from inside the scanner room. Both hands may be occupied when performing complex tasks inside the scanner, thus preventing the use of physical user interfaces. Even for the lone researcher, the ability to start and stop imaging while testing new techniques inside the scanner would be welcome. At present, primitive approaches are used to overcome the difficulty of controlling the scanner from within the scanner room. Two-way intercom systems facilitate verbal communication between the user located inside the room and the equipment operator outside. With the "yell and click" paradigm, the equipment operator adjusts the scan plane position and other parameters based on verbal commands from inside the scanner room.

The purpose of this work was to develop a voice command interface for use inside the MRI scanner room, to overcome such difficulties. Voice command recognition has been successfully used in the medical field in the past; for example, it was used for intraoperative control of the AESOP laparoscope positioning robot developed by Computer Motion Inc. In addition to commercial success, the concept was the subject of several studies to evaluate usability and advantages over other interfaces [1].

Methods

A voice command interface for real-time interventional MR imaging was developed for interactively controlling scan plane position and orientation from inside the scanner room. The system incorporates several established software components, as shown in **Figure 1**. The *Microsoft Speech Software Development Kit (MS Speech SDK)* is a freely distributable speech recognition (SR) and a text-to-speech (TTS) engine for American English. The *OpenTracker (OT)* library, which was designed for the interface and integration of multiple devices for applications requiring spatial localization and tracking, offers standardized data exchange, network support, and includes wrapping for the MS Speech SDK [2]. The *General Imager Hardware Interface (GIHI)* library provides a framework for interfacing with medical image equipment, with respect to retrieving real time 2D images and embedded tracking information, and controlling the scanner. Based upon these components, a new voice command interface and a scan controller component were developed. These two components were implemented as OpenTracker modules to allow for easy integration into the OpenTracker framework. Making use of OpenTracker's SR wrapping, a new software module called *SpeechControlModule* was developed. It implements a set of relevant commands, which can be recognized and translated into the corresponding modification of the MRI scan plane position, or an adjustment of the program's internal state. The command-set includes different kinds of plane translations and rotations, as well as three predefined scan planes. Moreover, there are commands to start and stop continuous movements, to save and restore imaging states, commands for undo/redo functionality, scaling, state logging and additional control commands. The command-set was implemented using the command and factory design pattern, and can therefore easily be constrained, modified or extended. The TTS functionality is used to provide feedback for the user. Every recognized voice command that results in a modified scan plane position causes an event to be dispatched with the updated position information. A new software module called *MedScanModule* was developed to interpret the information that is received by the OpenTracker configuration running on the MRI host. This component allows control of the MRI scanner via data provided in the standard OT data format, relying on the MRI interface of the underlying GIHI library. The system was tested with closed-bore GE Excite 12.x systems and an open-bore GE Signa SP/i scanner; however, it can be used with any system supported by the GIHI library.

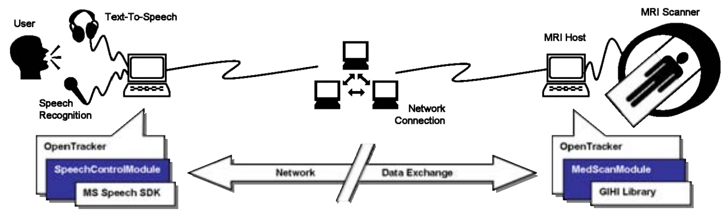

Figure 1: *Flow chart of the new voice command interface.*

Results

The new voice command interface was developed and tested. Interactive control of scan plane positioning was demonstrated using closed-bore GE Excite 12.x systems and an open-bore GE Signa SP/i scanner. The workflow parameters presented in **Table 1** illustrate the potential of this system to save time and improve efficacy by moving away from the "yell and click" paradigm to the voice control paradigm in order to perform repeated imaging tasks. The approximate number of executed commands and the time that can be saved, are provided. While the fifth application is for research experiments, all others are interventional applications. The first column (p) counts the number of positions at which images need to be taken, e.g., the number of targets during a biopsy or the number of needles that need to be positioned during a brachytherapy. The second column (m) describes the percentage of such target points that are missed and require re-positioning of the scan plane. For each target, images may be taken in a number of different orientations (e.g., axial, sagittal, coronal) as listed in the third column (o). The time that can be saved by using the voice command approach when setting the scan plane position to a certain target point is described by $\Delta t_{targetpoint}$, while $\Delta t_{orientation}$ is the time saved when changing the orientation of the scan plane. The time saved in total is estimated using the formula $t_{saved} = ((p+p\cdot m)\cdot\Delta t_{targetpoint})+(p\cdot o\cdot\Delta t_{orientation})$. Applications – such as prostate brachytherapy – that use a stereotactic grid for needle placement, offer the highest time saving potential. The software can translate voice commands for encoding a position (e.g., grid row and column, "A7") instantly, while a human operator currently needs to look up target points in a table in order to determine imaging position. All values are rough estimations based upon clinical practice at the Brigham and Women's Hospital and need to be investigated further.

Table 1: *Overview of potential applications and estimated time savings.*

	Application	Scanner	#target points (p)	misses (m)	#orient./point (o)	$\Delta t_{targetpoint}$ [seconds]	$\Delta t_{orientation}$ [seconds]	template grid used	t_{saved} **[minutes]**
1	*Robot assisted procedures*	open	6-30	75%	2-3	1-3	1-3		**0.4-10.9**
2	*Catheterization*	closed	20	25%	2-3	3	1-3		**1.9-3.3**
3	*Prostate Biopsy*	open/closed	6-20	125%	3	15	1-3	√	**3.7-13.3**
4	*Brachytherapy*	open	20-25	75%	2	20	1-3	√	**12.3-17.1**
5	*MRI Physicist scans self*	open/closed	1-100	n/a	1-3	n/a			

Conclusions

Voice controlled scan plane positioning was shown to be feasible and promising. Automated, voice controlled MRI scan control provides significant advantages over common practice for a number of applications. Several points of failure occurring in the traditional workflow can be eliminated. Moreover, the operator of the MRI scanner can concentrate on more complex tasks during the surgery, or perhaps oversee multiple procedures in adjacent rooms. In summary, the new interface gives complete control of MRI scanning back to the responsible user and allows for faster, direct adjustments of the scan plane. The reliability of command recognition in the noisy environment of the MRI scanner room will depend upon imaging protocols and applications, and should be investigated further.

References

1. M.E. Allaf et al., Laparoscopic visual field. Voice vs foot pedal interfaces for control of the AESOP robot, *Surg Endosc*, 1998, 12, 1415-1418
2. G. Reitmayr and D. Schmalstieg, An open software architecture for virtual reality interaction, *Proceedings of the ACM symposium on Virtual reality software and technology*, 2001, 47-54