

[ポスター講演] ナビゲーションシステムにおける音声認識を用いた直
観的操作平田 正保[†] 道満 恵介[†] 目加田 慶人[†] 森 健策[‡][†] 中京大学情報科学研究科 〒470-0393 愛知県豊田市貝津町床立 101[‡] 名古屋大学情報連携統括本部 〒464-8601 名古屋市千種区不老町E-mail: [†] h11220m@st.chukyo-u.ac.jp, {kdoman,y-mekada}@sist.chukyo-u.ac.jp, [‡] kensaku@is.nagoya-u.ac.jp

あらまし 近年はナビゲーションシステムが広く臨床医の間で利用されている。しかし従来のシステムでは手術器具の操作とナビゲーションシステムの操作を切り替えなければならなかったため、手術進行の妨げとなっていた。またジェスチャーを用いた操作も試みられてはいるが、利用できる機能は限定的であり、手術器具を持ったままナビゲーションシステムを操作することはできなかった。そこで本研究では音声操作によるナビゲーションシステムを提案する。拡大縮小や回転といった基本操作に加え、画像処理機能であるフィルタリングを利用することが可能なシステムを作成した。

キーワード ナビゲーションシステム, 音声認識

Controlling Medical Navigation System by Speech Recognition

Masayasu HIRATA[†] Keisuke DOMAN[†] Yoshito MEKADA[†] Kensaku Mori[‡][†] Graduate School of Sociology, Chukyo University 101 Tokodachi, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan[‡] Information and Communications Headquarters, Nagoya University Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 JapanE-mail: [†] h11220m@st.chukyo-u.ac.jp, {kdoman,y-mekada}@sist.chukyo-u.ac.jp, [‡] kensaku@is.nagoya-u.ac.jp

Abstract Recently, navigation systems play an important role in helping surgeons during surgery. This system provides a 3D representation of the patient and can be seen the patient from any point of view. However, a conventional navigational system is not very intuitive. Using a mouse and a keyboard for manipulating the navigation system is during surgery disruptive. This research proposes a navigation system interface using human voice as its input. This system can navigate and perform image processing by only human voice.

Keyword Navigation System, Speech Recognition

1. はじめに

現代の医療では高度な手術が行われる一方で、手術にかかる時間が増加しその内容も複雑化している。臨床医の負担軽減と手術の安全性の向上を目的とし、医工連携のもとナビゲーションシステムを利用した手術支援の研究が盛んに行なわれている。実際にナビゲーションシステムを利用している、または、利用したいと考える臨床医は多い [1]。これはナビゲーションシステムが、患者固有の臓器形状などの情報をボリュームレンダリング等によって表示でき、さらに事前にセグメンテーションした様々な情報を可視化でき、それらを術中に参照できることにある。一方で、ナビゲーションシステムで、医師に何をどのように提示すれば良いのか、どのような方法でナビゲーションシステム

を操作するのが良いのかについては、十分に検討されているとは言えない。

そこで本稿では、音声認識を利用したナビゲーションシステムの操作についての提案と、術中に行なわれる頻度が高い基本的な操作の実装についての報告を行なう。

2. 関連研究

画像誘導手術に関しては、対象部位や術式に応じて多岐に渡る [2]。情報提示に関しては、古くから様々な開発、改良が行なわれている。Tamakiらは乳がんの外科的手術において腫瘍の位置や大きさを把握するために、超音波画像から抽出した腫瘍領域と実際の患者の乳房を位置合わせし重畳表示した [3]。また、森らは腹

腔鏡下手術支援を目的に、自動抽出した腹部血管に自動的に解剖学的名称を付与し、実際の腹腔鏡の視点変化に追従してナビゲーションシステムで表示するシステムを開発している[4]。これは、ナビゲーションシステムの操作を腹腔鏡の視野と同じにすることにより、術者の操作を不要にするものである。

従来のナビゲーションシステムの操作は、マウスやキーボードを利用したものとなっている。医師は手術器具の操作とナビゲーションシステムの操作を切り替える、あるいはナビゲーションシステムの操作を他者に委託しなければならなかった。ジェスチャーを用いたナビゲーションシステムの操作も試みられているが[5]、利用できる機能は提示画像の切り替えなど限定的であり、手術器具を持ったまま自由に操作することはできない。

3. 提案手法

本研究ではナビゲーションシステムの操作性向上を目的とし、音声認識技術の利用を提案する。これにより、術具操作への妨げが少なく、直感的なナビゲーションシステムの操作を提供することが出来ると考えられる。本システムの構成図を Fig.1 に示す。本システムは音声認識部とナビゲーションシステム部、制御部から成り立っている。音声認識部には名古屋工業大学と京都大学を中心に開発されている大語彙連続音声認識エンジンの Julius を用いた[6]。これは自作の文法で音声認識を行うことが可能なソフトウェアであり、高い汎用性をもつことで知られている。本来は連続音声認識に適したソフトウェアだが、今回は主にナビゲーション操作に利用される単語などを登録した辞書を作成し、単語認識に利用している。ナビゲーションシステムには名古屋大学で開発されている、高速ボリュームレンダリングが可能な NewVES を用いた[7]。NewVES は単に医用画像を可視化するだけでなく画像処理や表示したデータの操作が可能なナビゲーションシステムである。以下では音声認識部と、ナビゲーションシステム部、制御部について説明する。

3.1. 音声認識部

音声認識部では、バックグラウンドで Julius を動作させ、マイクから音声が入力されるのを待つ。音声入力があった場合 Julius は音声認識結果を Unicode 形式の文字列として出力する。今回 Julius で利用する文法ファイルの作成には Julius の学習用ツールの一部である文法コンパイラ mkfa に、次で説明する語彙辞書ファイルと構文規則ファイルを与えることで作成した。

語彙辞書ファイルとは単語とその音素列を定義したものであり、7クラス 35 語を登録した (Tab.2)。辞書の詳細は次の通りである。まず NewVES における視点

変更操作に対応した単語“前”や“右”など 9 語を、基本操作クラスとして登録した。次に画像処理を行う画像処理クラスとして“フィルタ”など 2 語を登録した。また間投詞クラスとして“あー”や“えっと”など 16 語を登録した。これは音声認識の精度向上のためである。修飾語クラスは基本操作クラスを修飾するクラスであり“少し”、“もっと”、“ずっと”の 3 語を作成した。停止操作クラスは基本操作クラスで実行された命令を停止するためのクラスである。今回は“ストップ”、“停止”などの 2 語とした。最後に Julius の辞書に必要なデフォルトクラス 2 語とノイズクラス 1 語を合わせ全 32 語とした。

構文規則ファイルは単語間の接続を定義したものであり、6 つの規則を作成した (Tab.2)。まず基本操作クラスと画像処理クラス、および停止操作クラス単体での単語認識の規則である。これらの規則はそのクラスの単語単体でのみの認識である。また、実際に人が発話する際には間投詞が多く発せられるため、間投詞を含む発話の認識精度を向上させるために、間投詞クラスと基本操作クラス、間投詞クラスと画像処理クラスを合わせた構文規則を作成した。最後に修飾語クラスと停止操作クラスを認識できるクラスを設計した。これは移動量を少なくして微調整を行う場合や、移動量を大きくして閲覧したい部分へすばやく視点移動するための構文規則である。

3.2. 制御部

制御部では Julius の文法ファイルを読み込んだ後、マイクから音声入力を受け付ける。入力された音声を音声認識部へと渡す。音声認識部からは認識結果として Unicode 形式の文字列が送信されてくるため、制御部では辞書に登録された文字列とのマッチングを行い、発話された文字列が辞書内の単語であるのかを求める。検出結果に対して何らかの操作が定められている場合はあらかじめ起動しておいた NewVES ウィンドウ、つまりナビゲーションシステム部に対して仮想キーコードを送信する。

3.3. ナビゲーションシステム部

ナビゲーションシステム部では制御部から受信したキーコードに合わせて対応する処理、つまりレンダリングパラメータの調整を行う。キーコードの情報を受信するため、NewVES 側ではあらかじめ対応する処理にキーコードを割り振っておく必要があるが、モディファイアキーを組み合わせるなどし、未使用かつ通常のナビゲーション操作に支障のないような割り当てをする。また基本操作クラスの単語を受信した場合には修飾語に応じてキーコードに対応した処理をリピートする事とした。例えば“ずっと右”という発話に対しては、視点の右回転移動を行い続ける。

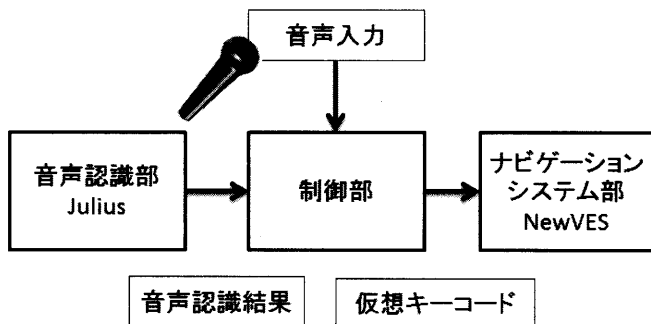


Fig.1 本システムの構成図。制御部がナビゲーションシステム部と音声認識部の橋渡しを行う。

Table.1 Julius で使用した辞書。クラス内の括弧は実際のクラス名。ノイズクラスの sp は無音が挿入されたときに出力される文字列。単語の読みは省略。

クラス	辞書
基本操作クラス (SOUSA)	右, 左, 上, 下, 前, etc
画像処理クラス (GAZOU)	フィルタ, メディアン
間投詞クラス (KANTOU)	あー, えっと, etc
修飾語クラス (SHUSYOKU)	少し, もっと, ずっと
停止操作クラス (TEISHI)	ストップ, 停止
Julius デフォルトクラス (NS_B, NS_E)	<s>, </s>
ノイズクラス (NOISE)	Sp

Table.2 Julius で使用した構文規則。

構文規則	内容
NS_B+SOUSA+NS_E	基本操作クラスに登録された語句の単語認識
NS_B+GAZOU+NS_E	画像処理クラスに登録された語句の単語認識
NS_B+TEISHI+NS_E	英紙操作クラスに登録された語句の単語認識
NS_B+KANTOU+NOISE+SOUSA+NS_E	間投詞が入力された場合の基本操作クラスの認識
NS_B+KANTOU+NOISE+GAZOU+NS_E	間投詞が入力された場合の画像処理クラスの認識
NS_B+SHUSYOKU+SOUSA+NS_E	修飾語クラスと基本操作クラスの認識

4. 実装機能

今回実装した3つのナビゲーションシステムの基本的な操作機能について紹介する。前述したように、回転、拡大縮小に加え画像処理機能を実装した。

4.1. 回転機能

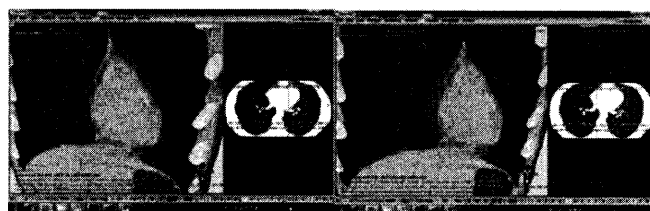
“右”, “左”, “上”, “下”の音声入力で、視点を現在の状態から指定した方向に移動する機能である。“もっと”や“少し”の修飾語で移動量を調整することができる。移動量は“少し” < “右” < “もっと”の順に移動量が増加する。なお“ずっと”の修飾語が発話された場合は視点変更を行い続けるため、“ストップ”等の音声認識で処理を停止する必要がある。実際に動作している状態を Fig.2 に示す。これは音声認識処理によって動作中のものを画像として保存したものである。

4.2. 拡大縮小機能

“前”, “拡大”, “後”, “縮小”の音声で現在閲覧しているデータを拡大または縮小する。これは視点位置を前後方向に移動させる機能である (Fig.3)。回転機能と同様“もっと”や“少し”で移動量の調節が可能である。

4.3. 画像処理機能

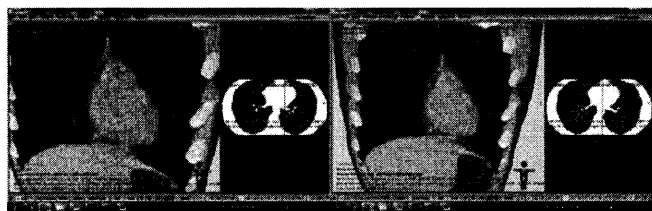
今回は音声入力に対してフィルタ処理を行う機能を実装した。“フィルタ”の音声入力でも操作することが可能であり、閲覧している画像の原画像 (CT 画像など) に対して3次元のメディアンフィルタ (フィルタサイズ 3×3×3) をかける。実行例を Fig.4 に示す。この図では、フィルタが適用されている事がわかるために、原画像に上記のフィルタを5回適用している。



(a)

(b)

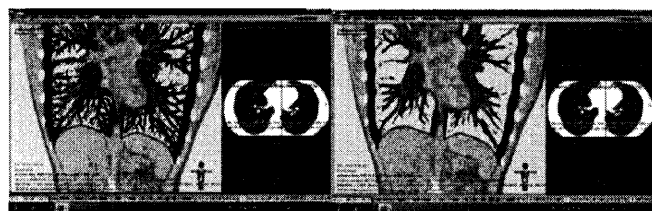
Fig.2 回転機能の例。右回転中のものを撮影した画像。(a) 停止している状態。(b) “右”の音声入力が与えられた状態。



(a) (b)

Fig.3 拡大縮小の例. 縮小中のものを撮影した画像.

(a) もとの状態. (b) “後”の音声入力を与えられたときの状態.



(a) (b)

Fig.4 画像処理機能の例. メディアンフィルタを閲

覧している画像に対して施した. (a) 音声入力前の状態. (b) “メディアン”の音声入力後の状態.

5. 考察

音声認識を用いたナビゲーションシステムの基本的な操作の実現と、共同研究医師の術中利用頻度が高い画像処理機能を実装した。このような音声を用いた直観的な操作システムにより、臨床医の手術進行を妨げないナビゲーションシステムが実現できる。不織布マスクを着用して発話した研究室内での実験において、音声認識の精度は良好であった。しかしながら、手術室は様々な機器から発せられる音が多い空間であるため、実際に利用できるシステムにするためには、音声認識の誤動作などに対する工夫が必要である。今回は Julius を単語認識に用いているが、本来は連続音声認識を得意とするソフトウェアであるため、さらに複雑な音声操作システムの構築も期待できる。また、事前に登録してある視点への自動移動や、回転角度を音声で直接入力など、閲覧したい部分へ素早く移動するための様々なインタフェースを用意することで、音声認識を利用したナビゲーションシステムが、操作者の煩わしさを軽減できると期待する。

6. まとめ

従来のナビゲーションシステムを術中に使用する際に問題点であった操作性を改善するため、音声認識によるナビゲーションシステムの操作を実現した。ナビゲーションシステムの利用時に特化した構文規則や

登録単語により、ナビゲーションシステムの基本操作や頻りに利用される命令が音声により実行できることを確認した。

今後は機能面だけでなく、利用できる構文規則や単語も利用する医師の要望を調査し、拡張していくことが必要である。また手術室は様々な機器から音が発生する非常に雑音の多い空間であり、音声認識による操作の場合、誤動作してしまう危険がある。医師が意図せず口にした語句で誤動作することがないように、音声認識の ON/OFF を簡単にできるようなインタフェースの構築に取り組む予定である。

謝 辞

本研究の一部は日本学術振興会科学研究費補助金の援助による。本研究では画像処理ソフトウェアライブラリ MIST [8]を使用している。

文 献

- [1] 鎮西清行, “手術ナビゲーションの臨床での使用実態に関するアンケート結果報告 (1),” JSCAS, vol.9, no.2, pp.116-135, 2007.
- [2] 佐藤嘉伸: 外科手術支援システム研究の現状と将来展望, 電子情報通信学会誌, 89(2): 144-150, 2006.
- [3] Tamaki Y., Sato Y., Nakamoto M., Sasama T., Sakita I., Sekimoto M., Ohue M., Tomita N., Tamura S., and Monden M: Intraoperative navigation for breast cancer surgery using 3D-ultrasound images, Computer Aided Surgery, vol. 4, no. 1, pp. 37-44, 1999.
- [4] Kensaku Mori, Huy Hoang Bui, Tetsuro Matsuzaki, Masahiro Oda, Takayuki Kitasaka, Michitaka Fujiwara, and Kazunari Misawa, “Automated anatomical labeling of abdominal arteries extracted from CT images based on machine learning,” Int'l Journal of Computer Assisted Radiology and Surgery, Vol.7, Sup.1, pp.S46-S47, 2012
- [5] 千葉慎二, 吉光喜太郎, 丸山隆志, 遠山仁啓, 伊関洋, 村垣善浩, “KINECT センサを用いた非接触画像閲覧システム: Opect,” JSCAS, vol14, no.3, pp150-151, 2012
- [6] 李晃伸, 河原達也, 堂下修司, “単語トレリスインデックスを用いた大語彙連続音声認識エンジン JULIUS,” 信学技報, SP98-3, pp.17-24, 1998-04.
- [7] 森健策, “コンピュータ画像診断支援とコンピュータ支援外科の連携,” CADM 特集, no.48, pp.2-5, 2006.
- [8] MIST (Media Integration Standard Toolkit), <http://mist.murase.m.is.nagoya-u.ac.jp>