



HAL
open science

Interaction vocale de commande de visualisation médicale

Mohamed Ikbel Boulabiar

► **To cite this version:**

Mohamed Ikbel Boulabiar. Interaction vocale de commande de visualisation médicale. 30eme conférence francophone sur l'interaction homme-machine, Oct 2018, Brest, France. 2p. hal-01900063

HAL Id: hal-01900063

<https://hal.science/hal-01900063v1>

Submitted on 20 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interaction vocale de commande de visualisation médicale

Off-line voice interaction for medical visualization control

Mohamed Ikbel Boulabiar
 Aptum
 12 rue de Solférino
 92100, Boulogne-Billancourt, France
 boulabiar@concepta.me



Figure 1: Salle de chirurgie (Photo par Guillaume Piron sur Unsplash)

ABSTRACT

In this demonstration, we show a practical use case of voice recognition in a sterilized surgery room. Due to medical constraints of hygiene, all alternative means of interaction are forbidden, or are limited in terms of their possible bandwidth. We use the last state-of-the-art tools for off-line voice recognition for privacy reasons. By employing this system we can control a web-based visualization of the patient medical 3D images using a set of around 20 voice commands or more.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**;

KEYWORDS

Voice recognition, medical application, command, visualization.

RÉSUMÉ

Dans cette démonstration, nous montrons un cas pratique d'utilisation de la reconnaissance vocale dans une salle de chirurgie stérilisée. En raison des contraintes médicales

d'hygiène, tous les moyens d'interaction alternatifs sont interdits, ou sont limités en termes d'expressions possibles. Nous utilisons les dernières avancées de l'état de l'art pour la reconnaissance vocale hors-ligne pour des raisons de confidentialité. Grâce à ce système, nous pouvons contrôler la visualisation des images médicales du patient à l'aide d'un ensemble d'environ 20 commandes vocales ou plus.

MOTS-CLEFS

Reconnaissance vocale, application médicale, commande, visualisation.

1 INTRODUCTION

Les chirurgiens opèrent dans une salle d'opérations chirurgicales stérile. Ils ne doivent pas toucher du matériel informatique avec la main pour éviter tout transfert de germes qui peut être mortel pour le patient [1]. Par contre, pendant l'opération, ils ont besoin de l'accès et l'interaction avec les informations du patient tels que ses images médicaux pour vérifier ou décider les étapes suivantes [2].

Ils utilisent parfois des pads opérés par les pieds, mais ces interfaces ont un espace d'entrée très réduit. Les autres interactions gestuelles nécessitent qu'ils enlèvent leurs mains du patient et du matériel chirurgical pour effectuer les commandes gestuelles [3]. Ils ont donc parfois recours aux commandes vocales qui ne permettent pas d'avoir des interactions directes. Ils nécessitent la mémorisation des commandes tout comme la majorité des interactions gestuelles. Le taux de reconnaissance des mots a été considéré très faible pour un usage professionnel après une vingtaine d'années d'échec de sa démocratisation avec IBM Viavoice.

Avec la nouvelle tendance de l'utilisation des interfaces vocales grâce au deep learning, cette technique a repris de la valeur et la réévaluation des performances de son usage est donc nécessaire surtout pour ce cadre où il n'y a pas beaucoup de choix en terme de modalités d'interaction.

Les problématiques de la vie privée mais surtout de la sécurité des données de santé interdisent l'usage des systèmes connectés à internet. Les services proposés par Google et Amazon sont donc éliminés. Le système doit être isolé complètement pour l'obligation de respecter la loi et il doit fonctionner en local.

À travers cette démonstration, nous montrons un système de commande vocale d'une interface de visualisation des images et des modèles 3D médicaux du patient. Ceci est réalisé grâce aux toutes dernières avancées technologiques de reconnaissance vocale et celles basées sur le deep learning [4] et la compression des modèles [5]. La démonstration permet de tester la fiabilité de la reconnaissance ciblée et évaluer les usages futurs qui peuvent en découler.

2 PRÉSENTATION DU SYSTÈME

Le système de reconnaissance est divisé en deux parties principales : le système de reconnaissance vocale et l'interface de visualisation. Ces deux parties sont reliées entre elles en utilisant une communication moyennant une websocket.

2.1 Système de reconnaissance vocale

Le système de reconnaissance vocale est basé sur les dernières technologies open source disponible sur internet. Contrairement aux systèmes répondus dans les assistants vocaux, notre module de reconnaissance est paramétré à reconnaître une liste prédéfinie de commandes au lieu de tout le lexique possible d'un langage. Ce choix nous a permis de simplifier énormément le travail à réaliser. Nous pouvons se passer d'un sous-module de traitement des langues naturelles nécessaire pour reconstruire la grammaire de la phrase pour comprendre le but de l'utilisateur. L'usage des commandes prédéfinies permet au docteur qui l'utilise de s'adapter au système et d'avoir une interaction quasi-déterministe..

2.2 Interface de visualisation

Nous avons pris le choix d'avoir une interface de visualisation réalisée complètement avec des technologies web HTML5. Ceci permet aussi d'optimiser le développement à faire et d'intégrer les changement en mode prototypage rapide. Nous utilisons le support existant des websockets pour la communication ainsi que les technologies WebGL pour l'affichage et la manipulation des images en hautes performances. La bibliothèque Three.js est utilisée comme surcouche au dessus de WebGL pour ses fonctionnalités d'imports de modèles 3D et de manipulation du contenu des objets et des textures utilisées.

2.3 Commandes reconnues

Les commandes reconnues sont limitées en nombre pour minimiser le risque d'oubli par les docteurs. Nous avons choisi en majorité des mots-clés très répandus lors de la

manipulation des visualisations incluant :

- Zoom [in, out] ;
- Move [Left, Right, Up, Down] ;
- Move [West, East, North, South] ;
- Rotate [Left, Right, Up, Down] ;
- Next Model, Previous Model ;
- Reset position

Cette liste n'est pas exhaustive et d'autres commandes peuvent être activées dans la version de démonstration tels que la capture de la vue actuelle pour la sauvegarde ou l'envoi des détails à un autre docteur distant.

3 PERSPECTIVES

Les mots-clés de reconnaissance sont actuellement en anglais. Ceci est dû à l'accès au grand nombre de bases audios utilisées pour l'entraînement ou qui sont déjà entraînées et disponibles directement pour l'utilisation. Nous sommes en train de travailler sur l'intégration de la reconnaissance des mots clés français que ce soit via la construction d'une base pour l'entraînement, ou via la réutilisation du modèle anglais existant mais après un peaufinage pour reconnaître la prononciation française de seulement les mots-clés choisis en français.

D'autres cas d'usages de la reconnaissance vocale sont à explorer comme ceux pour les malvoyants, les personnes handicapées ou les personnes âgées. L'usage de la voix peut remplacer un déplacement vers l'objet à contrôler ou une phase de recherche de l'objet lui-même. Les personnes âgées qui vivent seules, et qui perdent leurs équilibres et tombent très fréquemment peuvent éviter la mort en prononçant une commande de secours. À ce moment là, ils n'ont que leurs voix pour les sauver et appeler les secours via un périphérique de reconnaissance vocale.

RÉFÉRENCES

- [1] Hatscher, B., Luz, M., Nacke, L. E., Elkmann, N., Müller, V., & Hansen, C. (2017, November). GazeTap: towards hands-free interaction in the operating room. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction* (pp. 243-251). ACM. DOI: <https://doi.org/10.1145/3136755.3136759>
- [2] Mewes, A., Hensen, B., Wacker, F., & Hansen, C. (2017). Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *International journal of computer assisted radiology and surgery*, 12(2), 291-305. DOI: <https://doi.org/10.1007/s11548-016-1480-6>
- [3] Jalaliniya, S., Smith, J., Sousa, M., Büthe, L., & Pederson, T. (2013, September). Touch-less interaction with medical images using hand & foot gestures. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication* (pp. 1265-1274). ACM. DOI: <https://doi.org/10.1145/2494091.2497332>
- [4] Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., & Ng, A. Y. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567v2*
- [5] Strimel, G. P., Sathyendra, K. M., & Peshterliev, S. (2018). Statistical Model Compression for Small-Footprint Natural Language Understanding. *arXiv preprint arXiv:1807.07520*