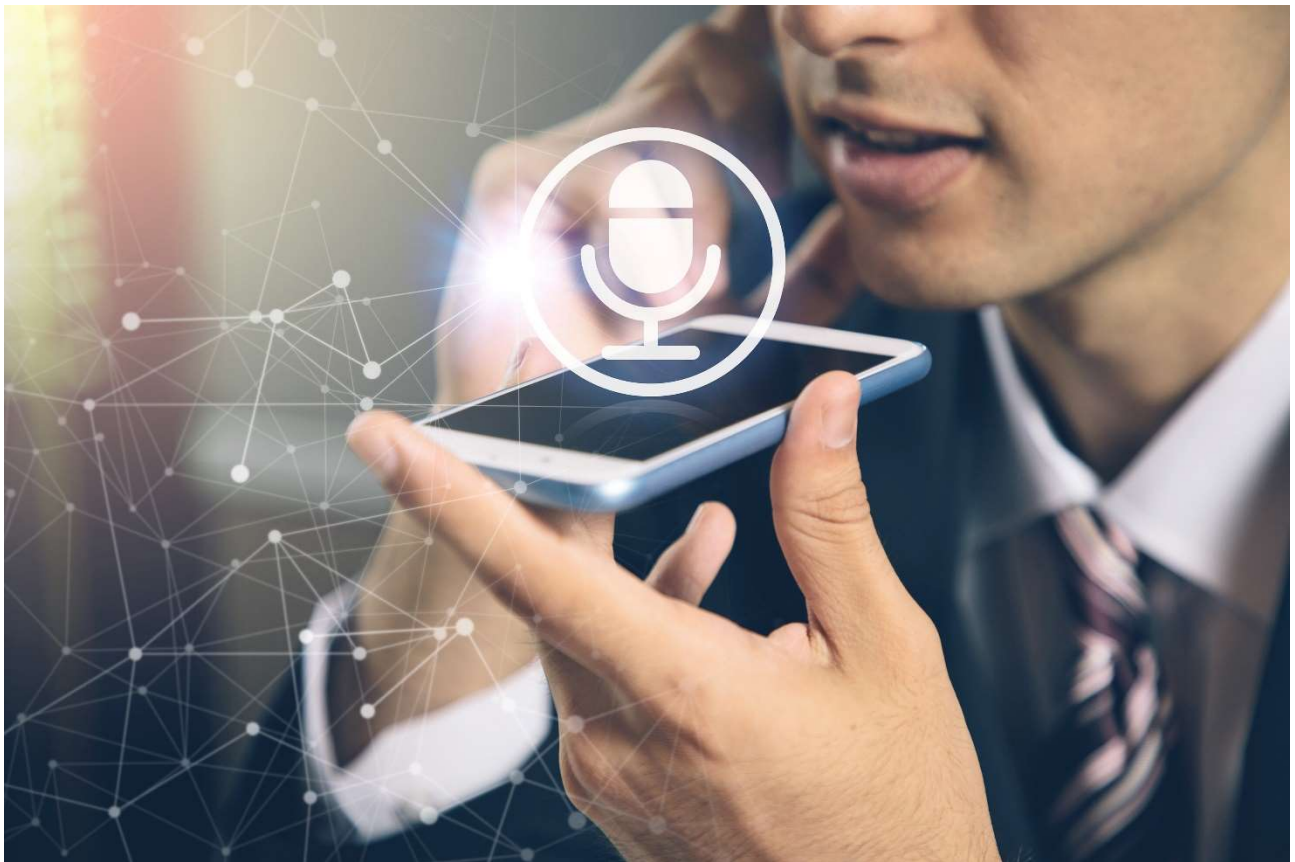


Tout le monde en parle - Ajouter la commande vocale à votre application

Par Cliff Ortmeyer, Global Head of Solutions Development, Premier Farnell and Farnell element14



Depuis presque aussi longtemps qu'existent les ordinateurs, les ingénieurs travaillent à améliorer les moyens d'interaction entre les humains et la technologie. Jusqu'aux dernières itérations d'écrans tactiles, la plupart des interfaces entre les humains et les ordinateurs consistaient en un compromis entre facilité d'utilisation et contrôle granulaire. L'écran tactile offre une méthode de contrôle bien plus naturelle et intuitive, accessible même les enfants.

Bien sûr, les écrans tactiles ne sont pas parfaits pour toutes les applications. Ils augmentent les coûts des applications bon marché, leur utilisation devient gênante sur les plus petits appareils, ils entraînent des inconvénients sur les installations extérieures, que cela soit à cause de l'environnement ou du vandalisme, ils peuvent représenter une menace pour la sécurité et, bien sûr, interagir avec eux nécessite une proximité physique. Dans des cas comme ceux-ci, les concepteurs ont besoin d'un moyen plus intuitif de contrôler les technologies, sans les inconvénients. Pour ces applications et beaucoup d'autres, la reconnaissance et le contrôle vocal peuvent représenter la technologie parfaite.

Il n'y a rien de nouveau à l'idée de contrôler la technologie uniquement avec la voix. Depuis l'aube de l'ère de l'informatique, il y a eu de nombreuses tentatives d'exploiter la puissance de la parole. Celles-ci ont rencontré des taux de réussite variables, jusqu'à ce que récemment, grâce à de grandes avancées des algorithmes de reconnaissance de parole combinées avec la puissance de calcul de grande échelle, les concepteurs ont été en mesure de concevoir une technologie qui répond rapidement et précisément à une grande variété de commandes.

La première tentative de création d'un ordinateur qui pouvait comprendre la parole humaine date de 1952 et a été développée par les laboratoires Bell. Le système, nommé Audrey, était très basique et ne pouvait comprendre que quelques chiffres, avec la voix de quelques personnes spécifiques seulement.

La reconnaissance de la parole s'est améliorée progressivement jusqu'aux années 1970, lorsque le département de la Défense des États-Unis a repris le programme SUR (Speech Understanding Research, recherche sur la compréhension de la parole) de la DARPA. Le programme SUR de la DARPA a été un énorme projet qui s'est étalé de 1971 à 1976. La recherche a finalement mené à l'élaboration du système Harpy de Carnegie Mellon, qui disposait d'un vocabulaire de plus de 1 000 mots. Plutôt que d'être seulement une évolution des systèmes antérieurs, Harpy employait une toute nouvelle approche de recherche, beaucoup plus efficace que les systèmes de recherche précédents. La recherche par faisceau peut prédire le réseau d'états finis de phrases possibles.

Les années 1980 ont connu une grande avancée dans la reconnaissance vocale, grâce au développement du modèle Hidden Markov (HMM), une technique de modélisation statistique qui peut prédire si des sons individuels peuvent représenter des mots. Cette avancée a considérablement élargi le nombre de mots qu'un ordinateur peut apprendre, avec plusieurs milliers de mots. La grande avancée technologique suivante est apparue en 1997, avec le lancement du premier système qui pouvait comprendre un discours naturel; Dragon Naturally Speaking qui pouvait traiter environ 100 mots par minute.

Ces progrès ont fourni les blocs de construction de base pour la reconnaissance vocale. Mais pour amener la technologie au grand public, il était nécessaire de disposer d'une informatique de grande échelle, à faible coût et largement disponible, capable de fournir des réponses en temps réel pour le contrôle. Cela a été possible plus récemment, grâce à deux géants de l'industrie : Google et Apple.

En 2011, Apple a lancé Siri, l'assistant personnel numérique intelligent de l'entreprise, sur l'iPhone 4S. Siri a ajouté un niveau de contrôle au système, permettant aux utilisateurs d'appeler des amis, de dicter des messages ou de jouer de la musique à l'aide du contrôle vocal. En 2012, initialement développée pour l'iPhone d'Apple, l'application Voice Search de Google a profité de la connectivité inhérente aux téléphones pour comparer des phrases de recherche avec les données que l'entreprise avait accumulées dans le cloud, provenant des recherches des utilisateurs. La capacité de comparaison avec des données de recherches précédentes a permis d'améliorer grandement la précision, en permettant à l'IA de mieux comprendre le contexte de la recherche. En réalité, Google Search et Siri étaient un moyen de contrôle secondaire, après l'écran tactile. Amazon a alors amené le concept au niveau supérieur avec Echo, qui combine un assistant personnel numérique avec un haut-parleur, tout en abandonnant l'écran tactile

Avec le succès des assistants personnels numériques intelligents, de plus en plus de concepteurs et de passionnés se sont penchés sur la reconnaissance et le contrôle vocal en tant qu'option possible pour leurs futures développements. La tendance est soutenue par le cabinet ABI Research d'analyse industrielle, qui estime que 120 millions d'appareils à commande vocale seront expédiés annuellement d'ici 2021 et que le contrôle vocal sera une interface utilisateur essentielle de la maison intelligente.



Les développeurs qui souhaitent intégrer le contrôle vocal dans un produit doivent cependant prendre en compte certains éléments. S'il est possible de construire un système complet de zéro et de lui permettre de fonctionner hors ligne, celui-ci n'offrirait alors qu'une fonctionnalité relativement limitée. Les algorithmes et les bibliothèques de reconnaissance vocale seraient limités par la mémoire de l'appareil, et l'intégration de nouvelles commandes serait difficile. Cependant, c'est possible. La bibliothèque

PocketSphinx a été développée pour les appareils Android, et la dernière version peut être utilisée en tant qu'application autonome sur tout appareil exécutant Android Wear 2.0.

Cependant, la majorité des développeurs préfèrent offrir un ensemble d'instructions plus complet, et cela nécessite la connexion avec le cloud. La plupart des grands fournisseurs de services dans le

cloud, y compris Amazon et Google, offrent des outils vocaux comme service, et l'intégration de ces derniers dans les conceptions peut être relativement peu coûteuse. Comme pour toute décision de conception, le meilleur choix de service dépendra de vos propres priorités. Par exemple, IBM propose également un service vocal au sein de sa plate-forme cloud Watson. La plate-forme est flexible, mais intéressera peut-être davantage ceux qui veulent profiter de l'expertise analytique d'IBM, plutôt que d'offrir une plate-forme grand public.



Amazon et Google proposent tous deux une plate-forme conçue pour le marché de l'automatisation domestique. Ces deux sociétés ont constitué un écosystème qui comprend certains des développeurs les plus respectés de produits d'automatisation domestique. Les partenaires d'Amazon comprennent Nexia, Philips Hue, Cree, Osram, Belkin et

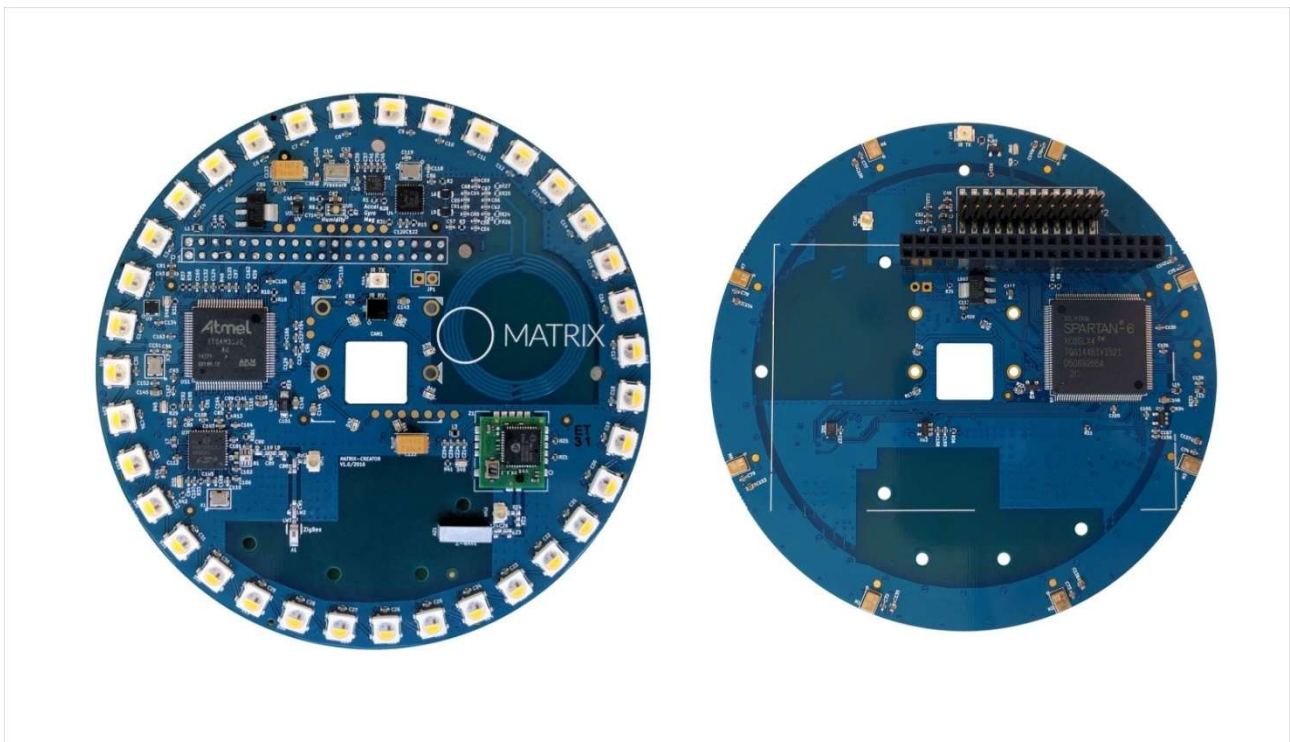
Samsung. Google partage également avec Amazon un grand nombre de ces mêmes partenaires, dont Hive, Nest, Nvidia, Philips Hue et Belkin.

Pour permettre l'intégration du contrôle vocal d'Amazon et Google dans leurs produits, les deux entreprises offrent l'accès à leurs plates-formes pour un coût relativement faible. Alexa Voice Service (AVS) d'Amazon permet aux développeurs d'intégrer Alexa directement dans leurs produits. AVS fournit une suite complète de ressources, y compris des API, des kits de développement logiciel (SDK) et matériel, ainsi que de la documentation.

Google permet également aux développeurs d'utiliser les fonctions de l'assistant numérique intelligent de l'entreprise, Google Assistant, à travers un SDK. Le SDK Google Assistant fournit deux options pour l'intégration de l'assistant - la bibliothèque Google Assistant et l'API Google Assistant gRPC. La bibliothèque Google Assistant est écrite en Python et est prise en charge sur les appareils avec les architectures linux-ARM v7l et linux-x86_64 (telles que le Raspberry Pi 3 B et Ubuntu). La bibliothèque offre une API de haut niveau, basée sur les événements, facilement extensible. L'API Google Assistant gRPC donne accès à une API de bas niveau. Les liaisons pour cette API peuvent être générées dans des langages comme Node.js, Go, C++ et Java, pour toutes les plates-formes qui prennent en charge gRPC.

Pour ceux qui préfèrent éviter ces services, ou utiliser une interface open source, d'autres options sont également disponibles. Par exemple, Mycroft est un assistant personnel intelligent gratuit et open source pour les systèmes d'exploitation basés sur Linux, qui utilise une interface utilisateur vocale. Mycroft est également une application modulaire permettant aux utilisateurs de modifier ses composantes. Jasper est une autre option open source qui permet aux développeurs d'ajouter facilement de nouvelles fonctionnalités au logiciel.

Quant au matériel, l'hôte le plus probable est un ordinateur monocarte, comme un Raspberry Pi. Certaines cartes ont été conçues spécifiquement pour les applications de contrôle vocal, comme la carte Matrix Creator (voir image ci-dessous), qui peut fonctionner comme carte d'extension du Raspberry Pi, ou comme unité autonome. La carte dispose d'une série de sept microphones MEMS pour apporter un champ d'écoute à 360°. La carte est alimentée par un processeur ARM Cortex M3 avec 64 Mo de SDRAM. Divers capteurs sont également intégrés pour permettre aux développeurs d'ajouter des fonctionnalités. Des services prêts à l'emploi peuvent être utilisés avec la carte, y compris Amazon AVS, Google Speech API et Houndify.



De toute évidence, les microphones représentent un facteur important de la conception. Plusieurs microphones, souvent organisés en réseau, sont généralement utilisés pour capturer une représentation plus précise du son. Si la technologie permettant de consolider le son provenant des différents microphones n'est pas intégrée au réseau, un travail de conception et une puissance de traitement supplémentaires peuvent être nécessaires. La technologie de réduction du niveau

sonore est aussi extrêmement importante pour s'assurer que les instructions soient reçues avec précision.

Conclusion

Le Saint Graal pour les développeurs est de créer l'interface la plus intuitive possible entre les humains et les machines. Aucune n'est encore comparable avec la méthode instinctive dont se servent les individus pour communiquer entre eux. Le contrôle vocal en est maintenant au stade où le processus devient presque aussi naturel que de parler à un autre humain. Bien que la technologie ait été lancée par certains des plus grands noms de l'industrie, il est maintenant possible à n'importe quel développeur de l'utiliser dans ses propres conceptions. Parce que la grande partie du traitement est effectuée dans le cloud, le matériel requis n'est pas aussi exigeant que vous pourriez le penser. Il existe aussi des cartes spécialisées, des outils et des services qui sont largement disponibles pour simplifier le processus de façon spectaculaire, ce qui signifie que le contrôle vocal peut être ajouté à presque n'importe quel projet.