



l'évolution de la commande vocale et audio des appareils électroniques

Mark Patrick (Mouser Electronics)

La parole est un moyen efficace pour les gens d'exprimer des idées et des désirs. Bien avant l'ère industrielle, les humains avaient déjà découvert que les animaux pouvaient être entraînés à reconnaître et à répondre à des ordres de base pour accomplir une tâche.

Il est donc apparu comme logique de développer un moyen de communiquer par la voix avec nos machines pour les commander.

L'utilisation de la voix et de l'audio comme interface de commande des appareils électroniques a gagné en popularité ces dernières années et cette méthode continue d'évoluer afin de répondre aux attentes des utilisateurs et aux exigences des nouvelles applications.

Dans cet article, nous décrirons quels sont les avantages de la commande vocale et audio des équipements et des machines électroniques ainsi que la façon dont cette technologie est mise en œuvre. Nous verrons également comment cette interface de commande peut désormais être intégrée dans des appareils hors ligne et quelles sont les possibilités d'amélioration de l'expérience audio fournie par ces appareils.

Utilisation de la voix pour commander les appareils électroniques

L'interaction voix-machine présente quelques avantages évidents :

- Pour les humains, la parole est une forme de communication intuitive qui permet de transmettre facilement des commandes par voie verbale.
- Il est toujours possible de communiquer verbalement, même lorsque les yeux ou les mains de la personne sont occupés par d'autres choses. La commande vocale simultanée s'avère dès lors très pratique dans certaines circonstances, notamment au volant, où il est illégal de commander un appareil à commande tactile tout en conduisant.
- La voix est un moyen efficace de commander des machines capables d'écouter et de répondre sans nécessiter de commandes complexes.
- L'intégration vocale permet dans de nombreux cas de se dispenser d'un écran tactile. Cela s'avère tout particulièrement utile pour commander des équipements à distance ou dans le cas d'appareils portatifs alimentés par batterie, dont la conception est souvent soumise aux contraintes de la réduction de la taille et de la consommation d'énergie. De plus, la commande vocale est manifestement plus hygiénique que la commande tactile pour les appareils commandés par plusieurs utilisateurs.
- Comme le montre la **figure 1**, cet avantage peut être utilisé dans des applications destinées à assister des personnes handicapées se trouvant dans l'incapacité de toucher un écran. La commande vocale permet dès lors de donner des ordres à la machine – par exemple, ouvrir une porte – ou de communiquer à distance des informations sur l'état de santé d'une personne.

Le système audio frontal (AFE) d'un dispositif à commande vocale comprend un réseau de microphones et des blocs de traitement de signal. L'AFE traite le signal d'un réseau de microphones multicanaux pour annuler les interférences provenant des bruits de fond ou de la lecture de l'appareil lui-même. Ce signal est ensuite envoyé à un moteur de détection de « mot de réveil », lequel a pour fonction de reconnaître les mots préprogrammés qui activent l'appareil (par exemple « Alexa » ou « OK Google » pour les plus connus). Plusieurs algorithmes de traitement du signal sont utilisés pour annuler les signaux d'interférence indésirables. Voici les principaux composants d'une solution à commande vocale.

Un réseau de microphones. Les systèmes à activation vocale nécessitent un ou plusieurs microphones pour

capter le signal de commande audio. Les principaux critères dictant le choix d'un réseau de microphones sont sa taille, son prix, ses performances et sa robustesse. Le fait de combiner différents signaux provenant d'un réseau à multiples microphones permet d'améliorer le rapport signal sur bruit (SNR) de la chaîne de signal audio.

Un détecteur de direction d'arrivée (DoA) permet de déterminer la position de l'utilisateur par rapport à l'appareil contrôlé afin que le réseau de microphones oriente son faisceau en direction de la voix. Un formeur de faisceau (« beamformer ») accepte les sons provenant de la direction d'arrivée et rejette ceux provenant d'autres directions. Son efficacité dépend de la géométrie et du SNR du réseau de microphones, de la largeur du faisceau et du niveau de bruit de fond.

Un compensateur d'écho acoustique (AEC) rejette le signal émis depuis le haut-parleur de l'appareil afin de capturer plus clairement la commande vocale de l'utilisateur, par exemple lorsque celui-ci émet une commande vocale pendant que l'appareil diffuse de la musique.

Un éliminateur d'interférences adaptatif (AIC) supprime les bruits de sources externes qu'un formeur de faisceau traditionnel ne parviendrait pas à supprimer entièrement. Il peut s'agir, par exemple, d'un bruit fort produit par un autre équipement.

Un détecteur de mots de réveil (ou mots d'activation). Un algorithme de détection de mots de réveil compare le signal vocal traité par l'AFE aux entrées d'une bibliothèque de mots de réveil comme « Hé Google ». L'algorithme repose sur un modèle d'apprentissage automatique. Plus le modèle est grand, plus il est précis. Par exemple, un modèle entraîné de 1 Mo sera plus précis qu'un modèle de 64 Ko. En revanche, il sollicitera davantage le processeur. Quand il s'agit de mots de réveil, il faut disposer de grands modèles pour assurer une détection plus précise des mots de réveil et réduire le nombre de faux positifs.

Les amplificateurs audio de classe D

L'aspect traitement de la voix de ce type d'interface de commande a connu de nombreuses améliorations, de sorte que même des appareils à faible coût peuvent être dotés d'une fonction de reconnaissance vocale précise. En revanche, l'aspect audio a longtemps été délaissé et la plupart des enceintes intelligentes et autres appareils audio connectés plus anciens offrent une qualité audio médiocre par rapport à des équipements audio haute fidélité.

Il est possible que les constructeurs aient considéré que l'engouement pour la commande vocale allait suffire à

faire oublier le défaut de qualité audio. Or, les consommateurs se sont aussi montrés plus exigeants sur le plan de l'expérience audio à mesure que ce type d'appareils intelligents se démocratisait. Les amplificateurs audio traditionnels de classe AB sont trop peu efficaces pour être utilisés dans des appareils connectés de faible puissance. Cependant, plusieurs fabricants de puces ont récemment mis sur le marché une gamme d'amplificateurs audio de classe D de haute technologie qui constituent une amélioration considérable par rapport aux générations d'amplificateurs précédentes. Nombre de ces nouveaux amplificateurs ont été spécialement conçus pour permettre à des appareils intelligents ou connectés de délivrer un son de qualité supérieure.

Le TAS2770 [1] de Texas Instruments [2] est un amplificateur audio à puissance d'entrée de 15 W qui améliore la qualité audio. Il est en outre doté d'une excellente capacité à capturer la voix, ce qui permet de commander plus facilement et plus naturellement les appareils à commande vocale qui en sont équipés. C'est le premier système audio frontal à combiner une entrée de microphone numérique avec un puissant amplificateur de détection I/V, ce qui lui permet de capturer la voix et les bruits ambients pour l'annulation d'écho ou la réduction du bruit dans les applications vocales. Maxim Integrated [3] (qui fait désormais partie du groupe Analog Devices [4]) propose pour sa part les MAX98357 [5] et MAX98358 [6], des amplificateurs de classe D avec un rendement de 92 % et qui délivrent une puissance de 3,2 W de classe AB. On trouvera un schéma fonctionnel simplifié de ces amplificateurs en **figure 2**. Économie en énergie, le PAM8106 [7] de Diodes Incorporated [8] est parti-

Figure 2. Schéma fonctionnel simplifié des amplificateurs audio de classe D de Maxim Integrated. (Source : Maxim Integrated)

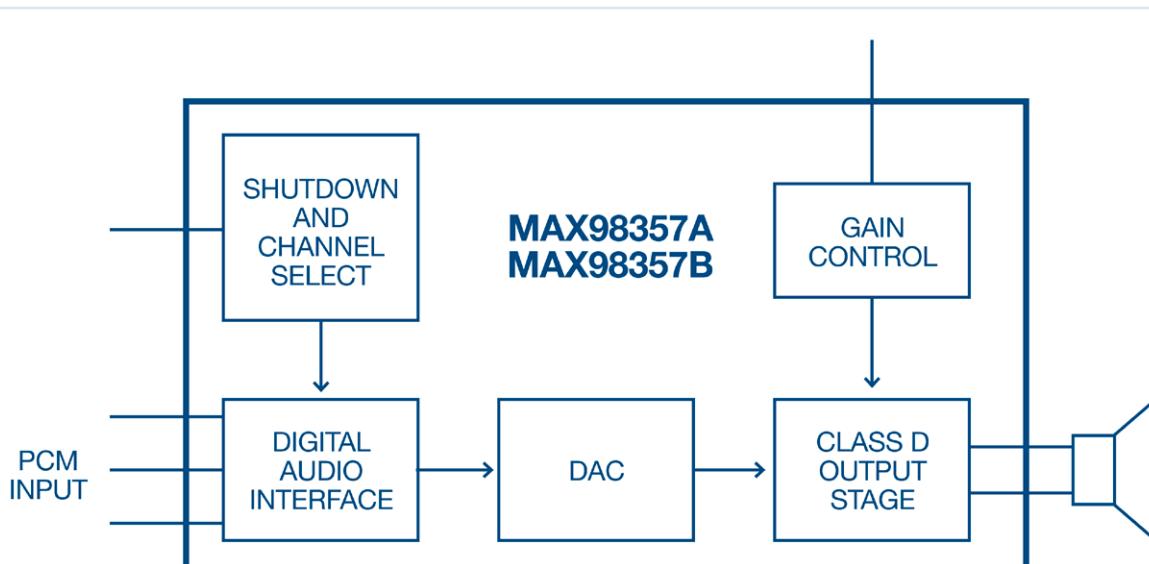


Figure 3. Solution de commande vocale hors ligne SLN-LOCAL2-IOT par NXP. (Source : NXP [10])

culièrement adapté aux appareils alimentés par des batteries plomb-acide de 1,5 V ou lithium-ion de 3,5 V. Il affiche un rendement de 92 % et peut être intégré dans une architecture compacte en se passant d'un encombrant dissipateur thermique.

Commande vocale hors ligne

Des solutions basées sur le cloud comme Alexa d'Amazon et Google Assistant sont simples à mettre en œuvre sur des appareils qui disposent d'une connexion Internet stable, mais lorsque la connexion est faible ou inexistante, on leur préférera une solution de commande vocale hors ligne. Prenons par exemple un produit qui ne doit répondre qu'à des commandes simples d'un seul mot comme « marche », « arrêt », « réinitialisation », etc. – ce qu'on appelle couramment le repérage de mots-clés – il paraît plus logique

que le traitement s'effectue localement, c'est-à-dire sur l'appareil même. Il est en effet possible de doter l'appareil d'un système de commande par mot-clé de base en y intégrant un microcontrôleur à faible coût. Avec EdgeReady, NXP propose une solution de commande vocale traitée localement et hors ligne reposant sur le microcontrôleur Crossover i.MX RT qui permet aux développeurs d'intégrer rapidement le contrôle vocal à leurs produits. La solution basée sur le microcontrôleur i.MX RT106S inclut le kit de développement SLN-LOCAL2-IOT [9] (voir **figure 3**).

Cette solution s'accompagne d'un logiciel entièrement intégré sous FreeRTOS livré avec un kit de développement logiciel (SDK) afin de réaliser rapidement la preuve de concept. Elle permet aux fabricants d'intégrer facilement de la fonctionnalité de commande vocale à tous leurs produits intelligents (appareils, maisons, bâtiments et produits industriels intelligents) sans qu'il soit besoin de disposer d'une connexion wifi ou au cloud. La commande vocale hors ligne répond ainsi en grande partie aux craintes de nombreux consommateurs concernant le piratage de leurs données accessibles en ligne. Le kit comprend enfin un outil compatible Windows™ servant à créer, à partir de la saisie de texte dans plus de quarante langues, des modèles vocaux pour plus d'une centaine de commandes personnalisées et plusieurs mots de réveil.

Conclusion

La voix et l'audio sont rapidement devenus les interfaces privilégiées dans de nombreux appareils intelligents. Elles remplacent avantageusement les affichages numériques – coûteux et énergivores – pour les appareils connectés portatifs et à faible consommation d'énergie. Les premiers systèmes de ce type souffraient pour la plupart d'une qualité audio médiocre et nécessitaient une connexion au cloud.

Mais une nouvelle génération d'amplificateurs audio de classe D à haut rendement permet désormais aux

fabricants de proposer des équipements qui offrent aux consommateurs une expérience audio de haute qualité. Par ailleurs, il existe aujourd'hui des solutions qui permettent de commander des appareils sans connexion Internet ou avec une connexion restreinte. Ces innovations démontrent que cette technologie est capable de s'adapter aux nouvelles attentes des consommateurs qui apparaissent à mesure qu'ils s'habituent à l'utilisation de cette interface de commande. C'est là une tendance qui devrait se poursuivre. 

220620-04



À propos de l'auteur

En tant que directeur marketing technique de Mouser Electronics pour la région EMEA, Mark Patrick est responsable de la création et de la diffusion du contenu technique - un contenu essentiel à la stratégie de Mouser visant à soutenir, informer et inspirer son public d'ingénieurs.

Avant de diriger l'équipe de marketing technique, Patrick faisait partie de l'équipe de marketing achat de la région EMEA et jouait un rôle essentiel dans l'établissement et le développement des relations avec les principaux partenaires et fournisseurs. En plus d'avoir occupé divers postes dans les départements techniques et marketing, Patrick a travaillé pendant huit ans chez Texas Instruments, dans les services support et ventes techniques.

Ingénieur expérimenté, passionné de synthétiseurs vintage et de motos, il n'hésite pas à les réparer. Patrick est titulaire d'un diplôme d'ingénieur en électronique avec mention très bien de l'université de Coventry.

LIENS

- [1] TAS2770 Audio Amplifier Evaluation Module : <http://elektor.link/MouserTAS2770>
- [2] Texas Instruments @ Mouser : <http://elektor.link/MouserTexasInstruments>
- [3] Produits présentés par Maxim Integrated : <http://elektor.link/MouserMaxim>
- [4] Analog Devices @ Mouser : <http://elektor.link/MouserAnalogDevices>
- [5] MAX98357 : <http://elektor.link/MouserMAX98357>
- [6] MAX98358 : <http://elektor.link/MouserMAX98358>
- [7] PAM8106 amplificateur audio 10 W : <http://elektor.link/MouserPAM8106>
- [8] Diodes Incorporated @ Mouser : <http://elektor.link/MouserDiodesIncorporated>
- [9] SLN-LOCAL2-IOT Solution for Local Voice Control : <http://elektor.link/MouserSLNLOCAL2IOTDevkit>
- [10] SLN-LOCAL2-IOT Product : <http://elektor.link/MouserSLNLOCAL2IOT>