



De evolutie van spraak- en audiogestuurde elektronische apparaten

▲

*Figuur 1: Spraakgestuurde robotassistent.
(Bron: PaO_STUDIO op Shutterstock)*

Mark Patrick (Mouser Electronics)

Spraak is een efficiënte manier voor mensen om ideeën en wensen uit te drukken. Reeds vóór het industriële tijdperk ontdekte de mens dat dieren konden worden getraind op het herkennen van elementaire commando's, bijvoorbeeld om bepaalde taken uit te voeren. De volgende logische stap was het vinden van een manier om met onze machines te communiceren en ze aan te sturen met behulp van onze stem. Het gebruik van spraak en audio als bedieningsinterface voor elektronische apparaten is de laatste jaren steeds populairder geworden en de ontwikkelingen gaan door om te voldoen aan de verwachtingen van gebruikers en de eisen van nieuwe toepassingen. In dit artikel gaan we dieper in op de voordelen van spraak en audio voor het aansturen van elektronische apparatuur en machines, en we bespreken enkele toepassingen. We laten ook zien hoe deze bedieningsinterface nu kan worden ingebed in offline apparaten en hoe de audio-ervaring die ze bieden sterk verbeterd kan worden.

Het gebruik van spraak voor het aansturen van elektronische apparatuur

De interactie tussen spraak en machine biedt een aantal duidelijke voordelen:

- Spraak is voor mensen een intuïtieve vorm van communicatie en dus een makkelijke manier om verbaal opdrachten over te brengen.
- Communicatie door middel van spraak is ook mogelijk wanneer iemands handen en ogen met iets anders bezig zijn. Hoewel aanvullende aansturing met spraak handig kan zijn, is het in bepaalde omstandigheden, zoals tijdens het autorijden, verboden om te proberen een ander apparaat met aanraking aan te sturen.
- Spraak is een efficiënt medium voor het aansturen van slimme machines die kunnen luisteren en reageren zonder de noodzaak van complexe commando's.
- Spraakintegratie maakt een touchscreen op veel apparaten vrijwel overbodig. Dit is vooral wenselijk voor op afstand geplaatste of draagbare, door batterijen gevoede apparatuur, waarbij kleinere afmetingen en minder stroomverbruik een gemeenschappelijke uitdaging vormen voor ontwerpers. Het wegnemen van de noodzaak om een apparaat fysiek aan te raken is ook hygiënischer, zeker voor toepassingen met meerdere gebruikers.
- zoals te zien is in **figuur 1**, kan het een hulpmiddel zijn voor mensen met een lichamelijke beperking voor wie fysiek contact geen haalbare optie is. Spraakcommunicatie met machines kan worden gebruikt voor het uitvoeren van verschillende taken, zoals het openen van deuren, of om updates over iemands gezondheid door te geven op afstand.

Het audio-ingangsdeel (audio front-end, AFE) van een spraakgestuurd apparaat omvat een microfoonarray en signaalverwerkingsblokken. Het AFE verwerkt het signaal van een meerkanaals microfoonarray om interferentie van eventuele achtergrondruis of van geluiden die het apparaat zelf maakt te neutraliseren. Dit signaal wordt vervolgens naar een detectiefunctie voor een activeringswoord ('wake-word') gestuurd die op het apparaat voorgeprogrammeerde woorden herkent, zoals „Alexa” of „OK Google”. Er worden meerdere signaalverwerkingsalgoritmen gebruikt om ongewenste stoorsignalen te elimineren. Een oplossing voor spraakaansturing bevat onder meer de volgende componenten:

Microfoonarray: Systemen voor spraakactivering vereisen één of meerdere microfoons om het audiosignaal op te vangen. Grootte, kosten, prestaties en robuustheid zijn de belangrijkste overwegingen bij de

keuze voor een microfoonarray. Door verschillende signalen van een multi-microfoonarray te combineren, kan de verhouding tussen signaal en ruis (signal to noise ratio, SNR) voor de audiosignaalketen worden verbeterd.

Detector voor de richting van aankomst (direction-of-arrival, DoA): Bepaalt de positie van de gebruiker ten opzichte van het apparaat, waarna de microfoonarray zich kan richten op de stem.

Bundelvormer (beamformer): Registreert geluiden van de DoA en onderdrukt geluidssignalen uit overige richtingen. De prestaties zijn afhankelijk van de geometrie en van de SNR van de microfoonarray, evenals de bundelbreedte en de hoeveelheid achtergrondruis.

Akoestische echowisser (acoustic echo canceller, AEC): Onderdrukt het playbacksignaal op de speaker van het apparaat (bijvoorbeeld als de speaker van het apparaat muziek afspeelt), zodat het spraakcommando van de gebruiker duidelijk en zonder stoorsignalen binnenkomt.

Adaptieve interferentiewisser (adaptive interference canceller, AIC): Neutraliseert externe ruis van andere geluidsbronnen die moeilijk te onderdrukken is met een traditionele bundelvormer, bijvoorbeeld lawaai dat door andere apparatuur wordt voortgebracht.

Activeringswoorddetector: Vergelijkt het verwerkte spraaksignaal van de AFE met een bibliotheek van activeringswoorden, zoals „Hey Google”. Hierbij wordt gebruikgemaakt van een detectiealgoritme voor activeringswoorden dat onderdeel is van een machine-learning model. Hoe groter het model, des te nauwkeuriger het resultaat. Een getraind model van 1 MB is bijvoorbeeld nauwkeuriger dan een model van 64 kB, maar vergt ook meer processorkracht. Om de activeringswoorden nauwkeurig te detecteren en het aantal valse alarmen te verminderen zijn grote modellen voor activeringswoorden nodig.

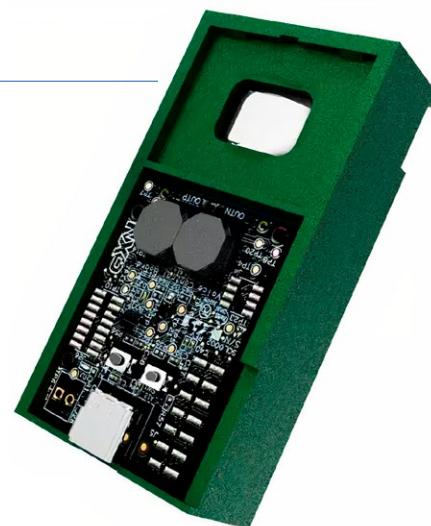
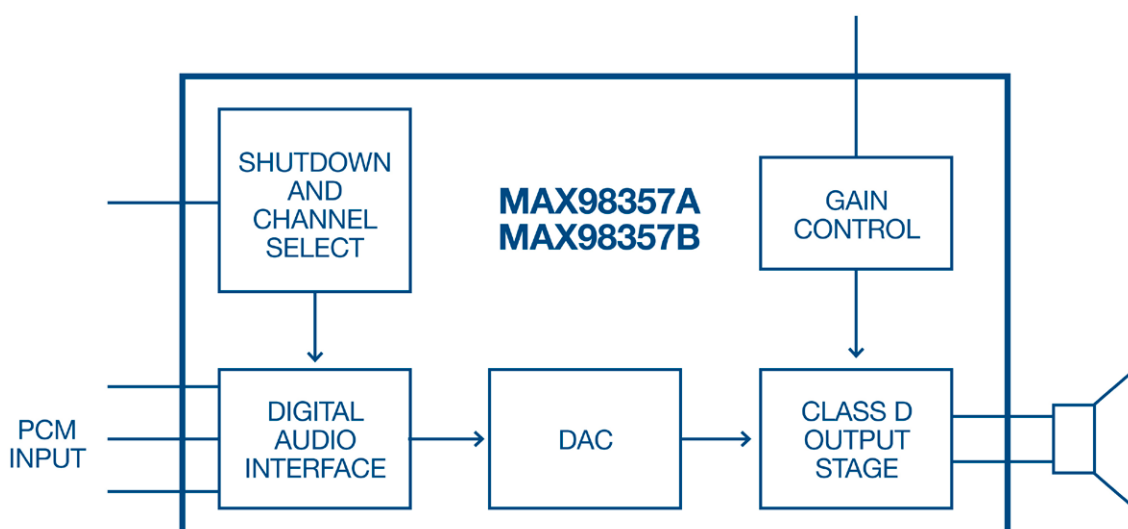
Klasse D audioversterkers

Het spraakverwerkingsgedeelte van deze bedienings-interface heeft een grote ontwikkeling doorgemaakt, zodat nu zelfs goedkope apparaten nauwkeurige spraakherkenning kunnen bieden. Het audiogedeelte van de interface heeft echter aanzienlijk minder aandacht gekregen, waardoor de geluidskwaliteit van veel oudere slimme speakers en andere audioapparatuur voor het Internet of Things (IoT) inferieur was aan de geluidskwaliteit van duurdere audioapparatuur. Het nieuwtje van de spraakbesturing zelf werd misschien beschouwd als voldoende afleiding om de aandacht af te leiden van deze tekortkoming.

Maar nu slimme apparaten steeds breder worden toegepast, worden de verwachtingen van de consument m.b.t. de audio-ervaring ook steeds hoger. De lage efficiëntie van traditionele klasse AB audioversterkers maakt ze onpraktisch voor gebruik in IoT-apparaten met een laag vermogen. Daarom hebben verschillende chipfabrikanten onlangs een reeks geavanceerde klasse D audioversterkers geïntroduceerd die een aanzienlijke verbetering zijn ten opzichte van de eerder beschikbare versterkers. Veel van de nieuwe versterkers zijn speciaal ontwikkeld voor hoogwaardige audio in slimme technologie en IoT-apparatuur.

De TAS2770 [1] 15-W ingangsversterker van Texas Instruments [2] verhoogt het volume en de audiokwaliteit. Zijn grotere vermogen om spraak op te vangen betekent een eenvoudigere en natuurlijkere bediening van spraakgestuurde apparaten. De TAS2770 is het eerste audio-ingangsdeel dat een digitale microfooningang combineert met een krachtige I/V-versterker, waardoor spraak en omgevingsgeluiden kunnen worden geregistreerd voor echo- of ruisonderdrukking in spraakgestuurde toepassingen. Maxim Integrated [3] (nu onderdeel van Analog Devices [4]) biedt met zijn MAX98357 [5] en MAX98358 [6] klasse D versterkers een rendement van 92% en een klasse AB audioprestatie van 3,2 W. **Figuur 2** toont een vereenvoudigd blokschema voor deze versterkers. De PAM8106 [7] van Diodes Incorporated [8] heeft een laag stroomverbruik, waardoor deze uitermate geschikt is voor apparaten met 1,5 V loodbatterijen en 3,5 V lithium-ion batterijen. Deze versterker heeft een efficiëntie van 92% en kan zonder omvangrijk koellichaam worden geïmplementeerd in een compacte ontwerpstructuur.

Figuur 2: Vereenvoudigd blokschema klasse D audioversterkers van Maxim Integrated. (Bron: Maxim Integrated)



Figuur 3: SLN-LOCAL2-IOT offline spraakaansturingsooplossing van NXP. (Bron: NXP [10])

Offline spraakaansturing

Er zijn cloudgebaseerde oplossingen, zoals Alexa van Amazon en Google Assistant, verkrijgbaar voor apparaten met een stabiele internetverbinding, maar voor apparaten met een slechte of geen verbinding is offline spraakaansturing een betere oplossing. Als een product bijvoorbeeld moet reageren op eenvoudige commando's van één woord, zoals go, stop, reset, enz. (vaak aangeduid als keyword spotting), dan is het zinvol om de verwerking lokaal op het apparaat zelf uit te voeren. Een goedkope ingebedde microcontroller, zoals de EdgeReady MCU-gebaseerde oplossing van NXP voor offline lokale spraakaansturing, maakt de implementatie van een eenvoudig systeem voor commando's met keywords mogelijk. Deze maakt gebruik van de i.MX RT crossover MCU en stelt ontwikkelaars in staat om snel spraakbesturing toe te voegen aan hun producten. De op i.MX RT106S gebaseerde oplossing van NXP omvat de SLN-LOCAL2-IOT development kit [9]. (Zie **Figuur 3**).

Deze oplossing wordt geleverd met volledig geïntegreerde software, draait onder FreeRTOS en is voorzien van een software-ontwikkelingskit (software development kit, SDK) voor een snelle proof of concept. Met deze oplossing kunnen fabrikanten hun producten in het kader van smart home, smart appliance, smart building en smart industrial, eenvoudig voorzien van spraakbesturing, zonder de noodzaak van WiFi of cloud-connectiviteit. Offline spraakaansturing pakt tevens de privacyzorgen van veel consumenten aan die vrezen dat hun systeem kwetsbaar is voor online hackers. De kit heeft ook een op Windows™ gebaseerde tool waarmee spraakmodellen kunnen worden gemaakt voor meer dan honderd commando's op maat en voor meerdere activeringswoorden uit tekstinvoer in meer dan veertig talen.

Conclusie

Spraak en audio worden snel de favoriete interfaces bij uitstek voor veel slimme apparaten. Omdat ze geen dure en energieverlindende digitale displays vereisen, zijn ze namelijk bijzonder geschikt voor gebruik in draagbare IoT-apparaten met een laag stroomverbruik. Veel oudere systemen hadden een inferieure audiokwaliteit en konden alleen worden geïmplementeerd met behulp van een met de cloud verbonden oplossing.

Dankzij een nieuwe generatie zeer efficiënte klasse D audioversterkers kunnen fabrikanten de consument apparatuur aanbieden met een hoogwaardige audio-kwaliteit. Er zijn nu ook oplossingen verkrijgbaar om apparaten met een slechte of geen internetverbinding aan te sturen. Deze innovaties tonen aan dat

deze technologie eenvoudig aan de nieuwe eisen die telkens ontstaan is aan te passen. Naarmate mensen meer gewend raken aan deze bedieningsinterface, zal deze trend zich waarschijnlijk voortzetten. ◀

220620-03



Over de auteur

Als Technical Marketing Manager voor EMEA bij Mouser Electronics is Mark Patrick verantwoordelijk voor het maken en verspreiden van technische content in die regio – content

die essentieel is voor de strategie van Mouser om zijn technische publiek te ondersteunen, te informeren en te inspireren.

Voordat hij leiding gaf aan het Technical Marketing-team, maakte Patrick deel uit van het EMEA Supplier Marketing-team en speelde hij een cruciale rol bij het aangaan en ontwikkelen van relaties met belangrijke productiepartners. Naast een verscheidenheid aan technische en marketingfuncties, heeft Patrick onder andere acht jaar bij Texas Instruments in Applications Support en Technical Sales gewerkt. Als 'hands-on' engineer in hart en nieren, met een passie voor vintage synthesizers en motorfietsen, draait hij zijn hand er niet voor om die te repareren. Patrick heeft een eerstegraads Honours Degree in Electronics Engineering van Coventry University.

WEB LINKS

- [1] TAS2770 Audio Amplifier Evaluation Module : <http://elektor.link/MouserTAS2770>
- [2] Texas Instruments @ Mouser: <http://elektor.link/MouserTexasInstruments>
- [3] Maxim Integrated @ Mouser: <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 10W Audio Amplifier : <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>