

AGH Intelligent Voice Enhancement Interface - Interfejs audio do aplikacji głosowych

Opis rozwiązania

AGH Intelligent Voice Enhancement Interface (AGH-IVEI) to technologia przetwarzania sygnałów mikrofonowych poprawiająca jakości mowy nagranej przy pomocy mikrofonów wbudowanych w inteligentne urządzenia mobilne lub urządzenia dedykowane inteligentnym asystentom takie jak 'smart speakery' czy 'smart soundbary'. Oprogramowanie może stanowić interfejs audio do komunikacji głosowej z komputerem (Human Computer Interaction - HCI), komunikacji głosowej w systemach konferencyjnych działających w trybie głośnomówiącym oraz komunikacji głosowej w systemach samochodowych.

Sygnałem wejściowym do każdego z systemów kontrolowanych głosem, w telefonach komórkowych, urządzeniach wirtualnych asystentów czy samochodach, jest sygnał audio nagrany mikrofonami wbudowanymi w urządzenie. W warunkach akustycznych, w których takie systemy są z reguły używane, jakość nagranej mowy może być niewystarczająca by zagwarantować odpowiednią skuteczność analizowanej mowy. Dużym wyzwaniem warunkującym skuteczność systemów rozpoznawania mowy w wirtualnych asystentach oraz zrozumienie transmitowanego sygnału mowy przez odbiorcę jest uzyskanie użytecznego sygnału o wysokiej jakości. Jest to szczególnie trudne w sytuacji gdy mówca znajduje się w znacznej odległości od urządzeń wyposażonych w mikrofony, często zmienia swoje położenie, a dodatkowo sygnał mikrofonowy zawiera sygnały niepożądanych (zakłócających) mówców, innych źródeł dźwięku, szumy oraz pogłos pomieszczenia. Opracowany interfejs audio umożliwi wirtualnym asystentom zrozumienie komend głosowych wydawanych z odległości poprzez usunięcie sygnałów zakłócających z nagrań mikrofonowych, usunięcie szumów i redukcję negatywnego wpływu pogłosu pomieszczenia na analizowany sygnał mowy.

Oprogramowana technologia umożliwi lokalizację źródeł dźwięku, separację sygnału mowy od sygnałów pozostałych mówców oraz innych sygnałów zakłócających, usunięcie szumu, a także redukcję pogłosu z nagrań w pomieszczeniach. W celu uzyskania znacznej poprawy w doświadczeniu użytkownika UX (User eXperience) rekomendowane jest zastosowanie wielu mikrofonów, z reguły ustawionych liniowo lub na okręgu.

W inteligentnych domach lub inteligentnych biurach, w których znajdują się rozproszone mikrofony lub rozproszone macierze mikrofonowe, możliwe jest rozszerzenie do tzw. rozproszonego interfejsu audio. Poza poprawą jakości mowy z sygnałów rozproszonych macierzy mikrofonowych, technologia Audio Processing using Distributed Acoustic Sensors (APDAS-IVEI) umożliwi lokalizację źródeł dźwięku i pozycji urządzeń nagrywających w przestrzeni oraz jednoczesną synchronizację ich sygnałów.

Obszary zastosowania

AGH Intelligent Voice Enhancement Interface (AGH-IVEI) to rozwiązanie elastyczne, które może być stosowane w urządzeniach wyposażonych w mikrofony takich jak urządzenia mobilne, 'smart speakery', jak również w systemach konferencyjnych. Opracowana technologia znajdzie zastosowanie w głośnomówiących systemach komunikacyjnych, głośnomówiących systemach telekonferencyjnych, interfejsach głosowych człowieka z komputerem, inteligentnych asystentach, komunikacji głosowej w samochodach.

Rozszerzenie technologii do systemów rozproszonych zwane Audio Processing using Distributed Acoustic Sensors (APDAS-IVEI) nadaje się do zastosowania w inteligentnych domach, inteligentnych biurach oraz aplikacjach rozszerzonej rzeczywistości.



KARTA UŻYTECZNOŚCI PRAKTYCZNEJ „KUP Technologię” w ramach projektu „Inkubator Innowacyjności 2.0”,

Innowacyjność i zalety stosowania

1. Inteligentne przetwarzanie sygnałów mikrofonowych
2. Poprawa jakości mowy z sygnałów mikrofonowych
3. Elastyczne rozmieszczenie mikrofonów w zależności od urządzenia (wspierane są macierze liniowe, kołowe, itp. jak również mikrofony i macierze rozproszone)
4. Ekstrakcja mowy - odseparowanie od pozostałych sygnałów
5. Redukcja szumów
6. Redukcja wpływu pogłosu pomieszczenia
7. Dla systemów rozproszonych, jednoczesna lokalizacja wielu źródeł dźwięku i macierzy mikrofonowych
8. Znacząca poprawa skuteczności analizy mowy na wyjściu interfejsu audio IVEI oraz APDAS-IVEI

Gotowość wdrożeniowa

- podstawowe badania nad technologią (TRL1)
- koncepcja technologii i jej przyszłych zastosowań (TRL2)
- laboratoryjne potwierdzenie krytycznych elementów technologii (TRL3)
- potwierdzenie technologii w skali laboratoryjnej (TRL4)
- zweryfikowano podstawowe elementy technologii w warunkach zbliżonych do rzeczywistych (TRL5)
- demonstracja prototypu lub modelu systemu w warunkach zbliżonych do rzeczywistych (TRL6)
- demonstracja prototypu technologii w warunkach operacyjnych (TRL7)
- zakończono badania i demonstracje ostatecznej wersji technologii (TRL8)
- potwierdzono skuteczność technologii w warunkach przemysłowych (TRL9)

Status IP	Forma komercjalizacji
<input type="checkbox"/> zgłoszenie do ochrony <input type="checkbox"/> patent <input checked="" type="checkbox"/> know-how <input type="checkbox"/> wzór użytkowy <input checked="" type="checkbox"/> program komputerowy, baza danych <input type="checkbox"/> utwór chroniony prawem autorskim	<input type="checkbox"/> sprzedaż patentu <input checked="" type="checkbox"/> licencja <input checked="" type="checkbox"/> spin-off <input checked="" type="checkbox"/> umowa wdrożeniowa <input type="checkbox"/> usługa
Kontakt	
Centrum Transferu Technologii tel.: +48 12 617 46 42 e-mail: ctt@agh.edu.pl	 Krakowskie Centrum Innowacyjnych Technologii INNOAGH SP. z o.o. e-mail: innoagh@agh.edu.pl 

KARTA UŻYTECZNOŚCI PRAKTYCZNEJ „KUP Technologie” w ramach projektu „Inkubator Innowacyjności 2.0”,