

„Propozycje algorytmów parametryzacji, odpornych na zmienność tonu krtaniowego, dla automatycznego rozpoznawania mowy”

Mgr inż. Stanisław Gmyrek

Katedra Akustyki, Multimediów i Przetwarzania Sygnałów (K76)

Wydział Elektroniki, Fotoniki i Mikrosystemów

Promotor: prof. dr hab. inż. Ryszard Makowski

Promotor pomocniczy: dr inż. Robert Hossa

Streszczenie:

Projektowanie systemów Automatycznego Rozpoznawania Mowy (ARM) wymaga uwzględnienia losowości i zmienności sygnału mowy, wynikającej z różnorodnych czynników, których negatywny wpływ powoduje obniżenie efektywności działania tych systemów. Szczególną uwagę należy poświęcić zmienności wewnątrzosobniczej oraz międzyosobniczej mówców. Systemy ARM bazują na widmach fragmentów mowy, gdyż słuch jest analizatorem widma. Okresowość pobudzenia oraz zmienność częstotliwości tonu krtaniowego prowadzą do degradacji tych widm, a w konsekwencji do zwiększenia wariancji estymatorów współczynników cepstralnych i pogorszenia wyników rozpoznawania. W celu zminimalizowania negatywnego wpływu różnych czynników stosuje się rozmaite techniki kompensacyjne, takie jak: grupowanie mówców, adaptacja modeli akustycznych, normalizacja wartości wektorów obserwacji oraz odporna parametryzacja.

Głównym celem badawczym niniejszej pracy było opracowanie algorytmów odpornej parametryzacji ukierunkowanych na kompensację negatywnego wpływu zmienności częstotliwości tonu krtaniowego i poprawę jakości rozpoznawania dla klasycznych systemów ARM.

W ramach pracy zaproponowano dwie autorskie metody odpornej parametryzacji. Pierwsza z nich, PS-HFCC (*Pitch-Synchronized Human Factor Cepstral Coefficients*), stanowi uogólnienie klasycznych metod parametryzacji cepstralnych poprzez analizę sygnału w ramach o zmiennej długości, synchronizowanej z okresem tonu krtaniowego. Druga metoda bazuje na filtracji odwrotnej IAIF (*Iterative Adaptive Inverse Filtering*), realizowanej z wykorzystaniem algorytmu LPC (*Linear Predictive Coding*). Filtracja odwrotna pozwala na estymację modułu transmitancji kanału głosowego, który staje się podstawą parametryzacji.

W badaniach zastosowano powszechnie wykorzystywany w systemach ARM statystyczny model akustyczny postaci GMM (*Gaussian Mixture Model*). Model ten opisuje wartości współczynników cepstralnych każdej jednostki fonetycznej jako ważoną sumę rozkładów Gaussa, określoną za pomocą wektorów wartości oczekiwanych, macierzy kowariancji oraz współczynników wagowych. Wyznaczanie elementów GMM realizowano w oparciu o klasyczny algorytm Bauma-Welcha, stanowiący algorytm estymacji typu EM (*Expectation-Maximization*).

Jako miary skuteczności zaproponowanych metod przyjęto: (i) odchylenia standardowe współrzędnych wektorów obserwacji, (ii) dywergencję Kullbacka-Leiblera (stanowiącą miarę odległości pomiędzy wielowymiarowymi rozkładami prawdopodobieństw GMM) oraz (iii) wskaźnik błędów rozpoznawania pojedynczych ramek sygnału – FER (*Frame Error Rate*). Wyniki analizy wskazują, że zaproponowane przez autora metody przewyższają pod względem skuteczności klasyczne podejście HFCC, zapewniając lepsze rozpoznawanie najczęściej występujących fonemów dźwięcznych i kompensując negatywny wpływ zmienności częstotliwości tonu krtaniowego na współczynniki parametryzacji.

Praca zawiera pięć rozdziałów. We wstępie omówiono podstawowe zagadnienia związane z funkcjonowaniem systemów ARM oraz określono cel, tezę i założenia badawcze niniejszej pracy. W rozdziale drugim dokonano przeglądu literatury dotyczącej metod parametryzacji sygnału mowy, opisano działanie konwencjonalnych i współczesnych systemów ARM oraz omówiono wspomniane wcześniej techniki kompensacyjne. W rozdziale trzecim przybliżono główny problem badawczy i zaprezentowano autorskie algorytmy odpornej parametryzacji. W rozdziale czwartym przeprowadzono badania i analizę zaproponowanych metod na sygnałach modelowych i rzeczywistych oraz przedstawiono wyniki rozpoznawania. Rozdział piąty stanowi podsumowanie wraz z wnioskami końcowymi.

30.03.2026

