



**AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE**

Voice Activity Detection

Metody, wykorzystanie

Wydział Inżynierii Mechanicznej i Robotyki

Kraków, 26.01.2015r.

VAD – co to jest?

- **Algorytm używany w procesie przetwarzania sygnałów mowy**
- **Umożliwia rozpoznawanie aktywności odbiorcy (głosu lub jego braku)**
- **Umożliwia zaoszczędzenie zasobów sieciowych oraz energii nadajnika**
- **Używany w systemach rozpoznawania, kodowania i kompresji głosu**
- **Pomocny przy procesie rozpoznawania mowy z sygnałów zaszumionych**
- **Zasada decyzyjna**

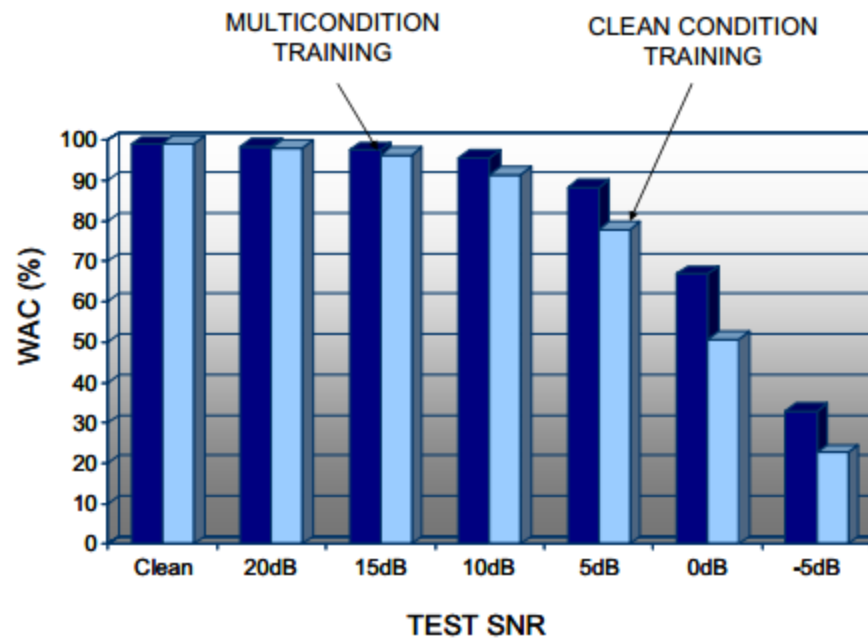
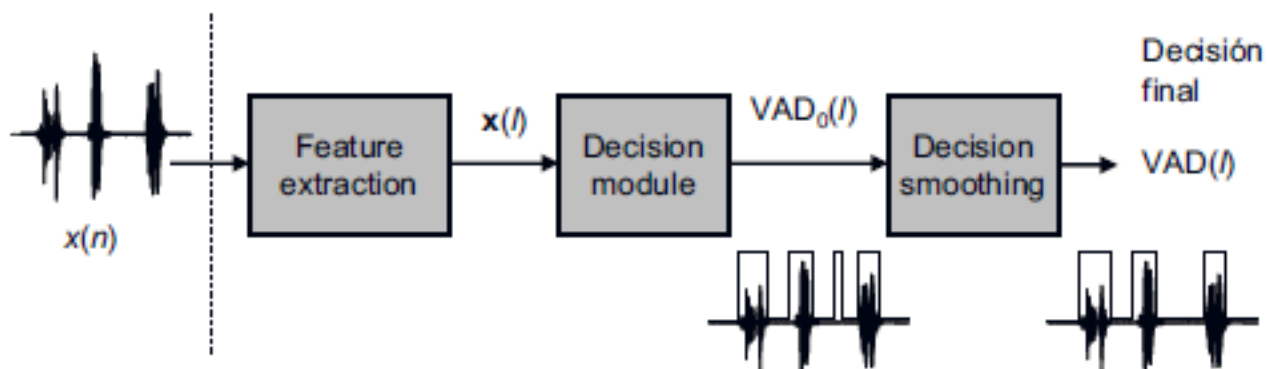


Figure 4. Results obtained for an enhanced feature extraction process incorporating VAD-based Wiener filtering and non-speech frame-dropping.

Schemat działania



VAD

- **PROBLEM: detekcja mowy i ciszy w sygnale zaszumionym**
- **Kodowanie – 2 tryby**
 - Aktywny
 - Pasywny
- **Parametry do rozpoznania mowy:**
 - Energia
 - Widmo
 - Gęstość przejść przez 0
- **Spektrum szumów uzyskiwane podczas okresów nieaktywnych**
- **Usuwanie nieaktywnych ramek**
- **REGUŁA DECYZYJNA:**
 - Przyporządkowanie parametru do danej klasy: mowy/ciszy
 - Likelihood Ratio Test
 - Klasyfikator Bayesa – minimalizacja błędu

Metody stosowane w VAD

- **SED – Simple Energy Detector**
- **AED – Adaptive Energy Detector**
 - Szybsza aktualizacja progu energetycznego
 - Bufor energii nieaktywnych ramek
- **LTSE – Long-Term Spectral Envelope**
 - Używanie długiego okna mowy
 - Bazowanie na obwiedni spektrum
- **Połączenie kilku metod**
- **MO-LRT – Multiple Observation Likelihood Ratio Test**
 - Więcej obserwacji
 - Informacje kontekstualne
- **Odzyskiwanie fonemów o niskiej energii**
 - Detektor przejść przez zero
 - Detektor słabych głosek szczelinowych przy pomocy autokorelacji wektora wariancji

Wykorzystanie VAD

- **Telefony sterowane głosem**
 - Nasłuchiwanie sygnału głosu w trybie czuwania
 - Po wykryciu aktywności głosowej oczekiwanie na słowo-klucz
 - Po usłyszeniu słowa-klucza gotowość do interakcji
 - Słowo – klucz może być zależne lub niezależne od głosu użytkownika
- **UBI**
 - Bezprzewodowy komputer sterowany głosem
 - Możliwość wysyłania wiadomości, sterowania muzyką itp. za pomocą głosu
 - W przyszłości połączenie z technologią Smart House
- **Amazon Echo**
 - Zawsze w stanie czuwania
 - Podawanie informacji np. odnośnie pogody, sterowanie muzyką
 - Wykrywanie głosu z całego pomieszczenia
 - Połączenie z chmurą – samouczenie się systemu



Film o amazon echo

Bibliografia

- **Gokhun Tanyer S., Ozer H.; *Voice Activity in Nonstationary Noise, IEEE Transactions on Speech and Audio Processing*, No.4, Czerwiec 2008**
- **Sangwan A., Chirath M.C., Jamadagni H.S., Sah R., Venkatesha Prasad R., Gaurav V.; *VAD Techniques for Real-Time Speech Transmission on the Internet***
- **Ramírez J., Górriz J. M., Segura J. C.; *Voice Activity Detection. Fundamentals and Speech Recognition System Robustness***



Dziękujemy za uwagę