



**ZWIĄZEK MIĘDZY WYRAZISTOŚCIĄ LOGATOMOWĄ
A OBIEKTYWNĄ MIARĄ QE-ARM DLA MOWY
PRZESYŁANEJ PRZEZ INTERNET**

**Relation between logatom intelligibility and objective QE-ARM measure
for speech transmitted via Internet.**

Stefan Brachmański

Politechnika Wrocławska, Instytut Telekomunikacji i Akustyki
stefan.brachmanski@pwr.wroc.pl

STRESZCZENIE

Zaprezentowano obiektywną metodę pomiaru jakości transmisji mowy opartą na technice automatycznego rozpoznawania mowy – metoda QE-ARM. Przedstawiono wstępne wyniki badań jakości transmisji mowy w sieciach wykorzystujących protokoły IP, z uwzględnieniem różnego sposobu kodowania. Pomiary wykonano na rzeczywistych sieciach Internetowych z wykorzystaniem programowych i sprzętowych rozwiązań komunikacji głosowej (VoIP). Wyniki otrzymane z wykorzystaniem obiektywnej metody QE-ARM porównano z wynikami subiektywnych pomiarów wyrazistości logatomowej wykonanych zgodnie z Polską Normą.

1. WPROWADZENIE

Jakość mowy jest jednym z ważnych i złożonych aspektów komunikacji między ludźmi. Rozwijająca się technika komunikacji głosowej poprzez Internet postawiła nowe zadania przed zespołami zajmującymi się problematyką oceny jakości transmisji mowy [2,3,7]. Metody oceny jakości transmisji mowy mogą być podzielone na trzy klasy:

1. subiektywne (percepcyjne) metody pomiaru zrozumiałości lub wyrazistości zdań, wyrazów, sylab lub głosek,
2. subiektywne (percepcyjne) metody oceny jakości (naturalności) transmitowanej lub kodowanej mowy,
3. obiektywne (instrumentalne) metody oceny jakości oparte o pomiar fizycznych charakterystyk systemów transmisji lub kodowania mowy.

Trzecia klasa metod (metody obiektywne) jest wciąż niespełnioną nadzieją na eliminację wysoce kosztownych, pracochłonnych i wymagających wysokich kwalifikacji metod subiektywnych z udziałem ekip słuchaczy i mówców. Metody te zapewniając wysoce powtarzalny, dokładny ilościowy pomiar „jakości” czy też „wyrazistości mowy” są szczególnie użyteczne przy ocenie różnych wariantów urządzeń i systemów do kodowania,

utajniania czy też przetwarzania mowy. Niestety, mimo kilkudziesięciu lat prac nad nimi do dziś nie ma powszechnie akceptowanej i wiarygodnej metody obiektywnej. Stosuje się je powszechnie do wstępnej czy też zgrubnej oceny, a ostateczną weryfikację wykonuje się metodami subiektywnymi [1,4,5,8,9].

W wielu ośrodkach na całym świecie prowadzone są badania nad opracowaniem uniwersalnej, obiektywnej metody oceny jakości transmisji mowy, traktując badany kanał telekomunikacyjny jako „czarną skrzynkę” (pomiaru klasy „end to end”). Jedną z takich metod jest opracowana i testowana w Instytucie Telekomunikacji i Akustyki metoda oparta na technice automatycznego rozpoznawania mowy. Metoda, która została nazwana QE-ARM jest testowana w różnych warunkach transmisji (analogowe kanały telekomunikacyjne, cyfrowe kanały telekomunikacyjne, różne techniki kompresji i kodowania sygnału mowy, transmisja pakietowa).

2. METODA OCENY JAKOŚCI TRANSMISJI MOWY WYKORZYSTUJĄCA TECHNIKI AUTOMATYCZNEGO ROZPOZNAWANIA MOWY (QE-ARM)

Metoda QE_ARM umożliwia obiektywizację i pełne zautomatyzowanie oceny jakości mowy. Algorytm metody stworzono w oparciu o procedury automatycznego rozpoznawania mowy. Zastosowano skończone stany, bezpamięciowy automat rozpoznający. Rozpoznawanych jest 100 izolowanych fraz (np. logatomów). Sygnały wzorcowe, w zależności od ich długości (czasu trwania), dzielone są na $2 \div 5$ zbiorów zamkniętych (klas czasowych). Uzyskuje się w ten sposób wstępną klasyfikację sygnałów testowych – obraz odpowiadający każdemu z nich jest poszukiwany wśród elementów jednego z tych zbiorów.

Cały proces pomiarowy sterowany jest za pomocą programu QE-ARM zainstalowanego na komputerze klasy PC wyposażonym w kartę przetworników A/C i C/A umożliwiającą jednoczesne odtwarzanie i próbkowanie sygnału z rozdzielczością 16 bitów. Program pracuje w środowisku Windows 98 lub wyższym. Wymagania sprzętowe programu odpowiadają minimalnym wymaganiom sprzętowym tego systemu.

Do przeprowadzenia badań, należy dysponować wcześniej przygotowanymi, spróbkowanymi sygnałami wzorcowymi (np. listami logatomowymi, wyrazowymi, zdaniowymi, itp.). Szesnastobitowe, monofoniczne pliki dźwiękowe zapisano w standardzie Windows PCM (wav). Przewidziano dwa tryby pracy programu:

1. Praca w czasie rzeczywistym - program QE-ARM pobiera z dysku spróbkowany sygnał wzorcowy (logatom) i poprzez przetwornik C/A oraz układ dopasowujący przesyła daną frazę na wejście badanego kanału telekomunikacyjnego, po przejściu którego zostaje podany na przetwornik A/C. Uzyskane w ten sposób sygnały stają się sygnałami testowymi poddawanych rozpoznawaniu.
2. Praca w trybie offline – przygotowane wcześniej sygnały testowe (sygnały, które zostały przesłane przez badany kanał telekomunikacyjny) są pobierane z dysku.

Algorytm pomiaru jakości transmisji mowy jest następujący. Sygnały wzorcowe wczytywane są z dysku, natomiast sygnały testowe, w zależności od wybranego trybu pracy, uzyskiwane z przetwornika A/C lub wczytywane z dysku. Operacja przetwarzania zaczyna się od poddania sygnałów operacji preemfazy. Następnie sygnały zostają sparametryzowane metodą FFT, BF-FFT lub LPC. Sygnały są wstępnie klasyfikowane na podstawie ich długości. Parametry sygnałów wzorcowych zapisywane są do n tablic wzorców odpowiadających uzyskanym w wyniku klasyfikacji wstępnej zbiorom. Uzyskuje się w ten sposób n skończonych zbiorów wzorców. Jako regułę decyzyjną zastosowano algorytm NN

(najbliższy sąsiad). Każdy z sygnałów testowych jest porównywany z wzorcami z odpowiedniego zbioru a najbliższy z nich uznany zostaje za odpowiedź systemu czyli rozpoznaną klasę. W celu znormalizowania czasowego sygnałów, przed porównaniem każdej pary wzorzec – sygnał badany, zastosowana jest technika dynamicznego dopasowania czasowego (DTW). W charakterze funkcji podobieństwa wykorzystuje się miarę odległości Hamminga. Opcjonalnie można również zastosować odległość Euklidesa, Camberra lub (w przypadku parametryzacji LPC oraz Cepstrum) miarę Itakury

3. EKSPERYMENT

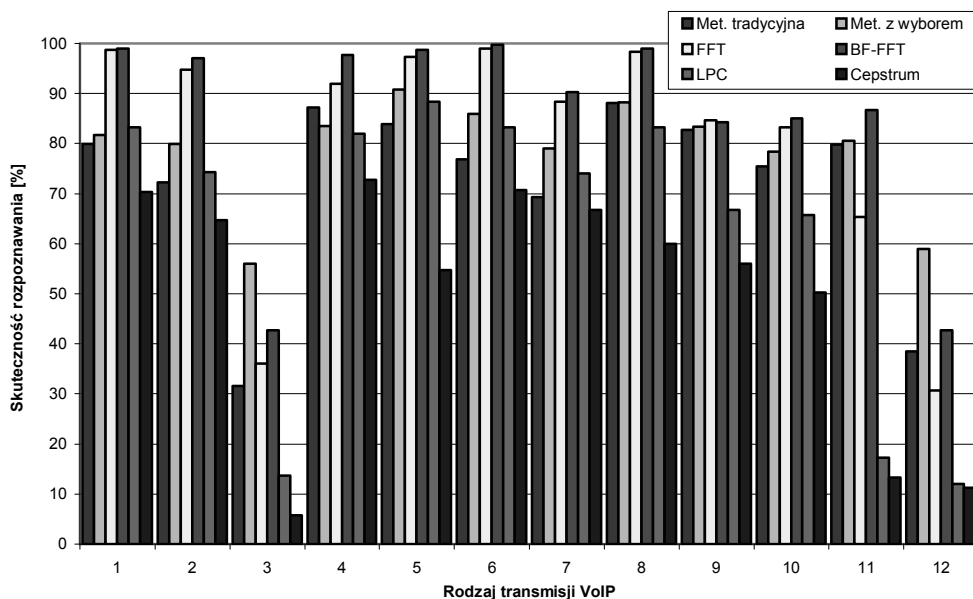
W zrealizowanym eksperymencie, subiektywne pomiary wyrazistości logatomowej wykonano zgodnie z zaleceniami Polskiej Normy PN-V-9002 [9]. Wiek słuchaczy mieścił się w przedziale od 18 do 30 lat. Pomiary przeprowadzono w cichym pomieszczeniu mieszkalnym stosownie do odpowiednich zaleceń. Słuchaczom stworzono warunki odsłuchu zbliżone do tych w jakich wykorzystywane są normalnie badane systemy telefonii IP. Odsłuch przeprowadzono metodą dwuuszną, ponieważ z reguły rozmowy prowadzone przez software'owe rozwiązania telefonów IP wykorzystują mikrofon i słuchawki (dwuuszne) podłączone do karty dźwiękowej komputera. Materiał testowy stosowany w pomiarach wyrazistości logatomowej metodą tradycyjną i z wyborem oraz w pomiarach obiektywnych stanowiły zrównoważone fonematycznie i strukturalnie listy logatomowe. Pomiary wykonano w czasie nie rzeczywistym, tzn. przed wykonaniem pomiarów subiektywnych i obiektywnych listy testowe zostały przesłane przez badane kanały VoIP i zarejestrowane na magnetofonie cyfrowym oraz na twardym dysku komputera wyposażonego w kartę przetwornika A/C. Zapewnione zostały w ten sposób identyczne warunki transmisji dla wszystkich porównywanych metod oceny jakości mowy.

Dla rozwiązań software'owych pomiary jakości transmisji mowy przeprowadzono realizując połączenie w sieci lokalnej oraz w sieci miejskiej opartej na telewizji kablowej. W ramach eksperymentu przebadano tory telefonii IP zrealizowane za pomocą oprogramowania: Microsoft NetMeeting (kodowanie CELP i G.723.1), ICQ, LabNet Phone, Paltalk, Buddyphone (przepływność 13 kb/s i 64kb/s) oraz Tlen. Natomiast wersją sprzętową była sieć korporacyjna grupy kapitałowej Howell S.A., oparta na rozwiązaniu i sprzęcie firmy CISCO. Dla tej sieci wykonano dwa rodzaje badań, a mianowicie transmisję w obrębie jednej centrali oraz transmisję z czterokrotną zmianą kompresji

Do celów eksperymentu losowo przypisano numery poszczególnym rozwiązaniom telefonii IP. Dla tych samych rozwiązań telefonii IP wykonano pomiary jakości transmisji mowy obiektywną metodą QE-ARM. W eksperymencie przeanalizowano wpływ metody parametryzacji (FFT, BF-FFT, LPC, Cepstrum) na uzyskiwaną zgodność wyników pomiarów subiektywnych i obiektywnych. Wyniki zostały zamieszczone na rys.1.

4. PODSUMOWANIE

Analizując otrzymane wyniki, można zauważyć, że skuteczność rozpoznawania w dużej mierze zależy od metody parametryzacji. Najwyższą skuteczność rozpoznawania uzyskano dla metody BF-FFT (transformata Fouriera w pasmach barkowych). Z drugiej strony badając dobór metody parametryzacji pod kątem zbieżności z pomiarami subiektywnymi okazuje się, że najlepszy efekt daje parametryzacja LPC.



Rys 1. Wpływ rodzaju parametryzacji na skuteczność metody QE-ARM.

LITERATURA

1. BAŚCIUK K., BRACHMAŃSKI S., The automation of the subjective measurements of logatom intelligibility., The 102nd Convention AES, Munich 1997, Preprint 4407 (A5).
2. BRACHMAŃSKI S., Effect of Disturbances on Speech Transmission over Internet, Proc. of the 4th European Workshop on Image Analysis for Multimedia Interactive Services, World Scientific Publishing, London 2003, s.310-313
3. BRACHMAŃSKI S., Assessment of Quality of Speech Transmitted over IP Networks, IFIP TC6/WG6.4 Workshop on Internet Technologies, Applications and Societal Impact (WITASI 2002), Kluwer Academic Publishers, Wrocław 2002, s. 1-14
4. BRACHMAŃSKI S., Subiektywne metody oceny jakości transmisji mowy w cyfrowych kanałach telekomunikacyjnych, Krajowe Sympozjum Telekomunikacji, Bydgoszcz 1999, Tom B, s.333-342.
5. DECINA M., MODENA G., CCITT Standards on digital signal processing, IEEE Selected Area in Comm., vol.6, nr 2, Feb.1988, s. 227-233.
6. OLACZEK A. BRACHMAŃSKI S., Wpływ szumu na skuteczność rozpoznawania logatomów przesłanych analogowym kanałem telekomunikacyjnym., XLV Otwarte Seminarium z Akustyki, Poznań-Kiekrz, 1998, s. 469-474
7. SCHULZ T., Voice over IP, (white paper), Eicon Technology Corporation, 2000 <http://www.eicon.com/disvpri/whtpap4.htm#conventional> telephony.
8. WANG S., SEKEY A., GERSHO A., An objective measure for predicting subjective quality of speech coders, IEEE Journal on Selected Areas in Communications, vol.10, nr 5. June 1992, s. 819-829
9. PN – V-90002 Cyfrowe łańcuchy telefoniczne. Wymagania i metoda pomiaru wyrazistości logatomowej.