

Krzysztof Marasek, dr hab.  
Polsko-Japońska Wyższa Szkoła  
Technik Komputerowych  
ul. Koszykowa 86  
02-008 Warszawa

Warszawa, 4 sierpień 2010 r.

Recenzja pracy doktorskiej mgr inż. Łukasza Apiecionka  
pt. „Metoda oceny jakości transmisji głosowej w telefonii VoIP”

## 1. Uwagi wstępne

Rozpowszechnienie pakietowych sieci komputerowych spowodowało także rozwój zastosowań dla których sieci te pierwotnie nie były przeznaczone, a wśród nich także transmisji strumieniowych, np. dźwięku. Stąd rozpowszechnienie rozwiązania VoIP, które staje się pełnoprawną konkurencją sieci telefonicznych, ale także kłopoty związane z jakością transmisji głosu przez takie medium. Z tego punktu widzenia rozprawa doktorska p. Łukasza Apiecionka bardzo dobrze wpisuje się w aktualny trend, jest niewątpliwie na czasie i dotyczy istotnego problemu praktycznego.

## 2. Temat i zakres rozprawy

Przedstawiona do recenzji rozprawa dotyczy pomiaru jakości transmisji głosu w sieci VoIP, a w szczególności autorskiej propozycji metody pomiaru jakości transmisji charakteryzującej się:

- brakiem ingerencji w działanie sieci i brakiem sygnału referencyjnego,
- niską złożonością obliczeniową,
- przedstawianiem wyniku w skali MOS,
- uwzględnianiem przerw w transmitowanym sygnale.

Powyższa charakterystyka metody pomiaru odpowiada z grubsza tezom pracy, które uważam za prawidłowo sformułowane i starannie, i szczegółowo opracowanym w rozprawie. Praca doktorska zawiera szczegółowe przedstawienie zagadnienia badawczego, działania sieci

IP i protokołów transmisji dźwięku w takich sieciach, opis dotychczas stosowanych metod oceny jakości transmitowanego dźwięku, jak i wyniki własnych eksperymentów Autora. Praca ma zatem klasyczną formę, w której można wyróżnić część przeglądową – erudycyjną oraz oryginalną, przedstawiającą osiągnięcia Autora rozprawy.

Rozprawę p. Apiecionka można umiejscowić na styku telekomunikacji, informatyki i psychoakustyki, przy czym ten pierwszy aspekt jest niewątpliwie najsilniej przedstawiony w pracy. Ze względu jednak na konwergencję mediów, można ją odnieść do szeroko pojętej informatyki, a w szczególności technik informacyjnych.

### 3. Ogólna charakterystyka pracy

Praca składa się z 4 podstawowych rozdziałów, wprowadzenia, wniosków i jednego dodatku, podającego wyniki pomiarów proponowaną metodą oceny jakości transmisji głosu w sumie o objętości 145 stron. Bibliografia pracy liczy 89 pozycji, które nie są uporządkowane alfabetycznie, co utrudnia odszukanie pozycji literaturowych. Praca zawiera starannie opracowany indeks stosowanych oznaczeń i skrótów, nie zamieszczono jednak spisu tabel i ilustracji.

Rozdział pierwszy stanowi wprowadzenie do sieci IP. Przedstawiono w nim najpopularniejsze standardy przesyłania w czasie rzeczywistym dźwięku w sieci i protokoły sygnalizacji stosowane przy transmisji dźwięku. Praca dotyczy sieci IP, omówiono w niej w zasadzie wersję 4 protokołu – troszkę szkoda, że Autor nie rozwinął informacji o wersji 6, która ma kilka rozszerzeń ułatwiających zapewnienie ciągłości transmisji (także szyfrowanej, pole Flow Label), czy też wspomagających unifikację transmisji. Następnie, dość pobieżnie, przedstawiono dyskretyzację dźwięku i jego parametryzację (kodowanie). Niewątpliwie zdanie ze str. 27 „Wybór standardu wiąże się z określoną metodą kodowania, co gwarantuje różną ilość danych do przesłania” jest niepełne – brakuje odniesienia do jakości transmisji po zakodowaniu dźwięku, który to proces w oczywisty sposób wpływa na odbiór jakości przez użytkowników (a więc i na parametr MOS). Warto tu wspomnieć o pierwszych telefonach komórkowych w Japonii, w których jakość kodowania była tak słaba, że nie można było rozpoznać po głosie, kto dzwoni (bodaj standard IS-54).

Tabela 1.1 w kontekście Rys.1.3 powinna być uzupełniona o informacje o długości pakietów dla poszczególnych typów kodowania (patrz np. [1]). Z kolei omówienie protokołów sygnalizacji H.323, SIP, bram mediów MGCP i innych jest w zupełności wystarczające do zrozumienia głównych zagadnień pracy.

Rozdział drugi stanowi przegląd zagadnień związanych z jakością dźwięku w telefonii IP. Trzeba tu od razu zastrzec, że Autor rozumie pod tym określeniem jakość przesyłania mowy w sieci i wszystkie zaprezentowane metody oceny dotyczą tylko tego zagadnienia. Autor słusznie wskazuje na podstawowe aspekty poprawności przesyłania dźwięku i prezentuje perceptualną skalę oceny jakości transmisji – MOS. Warto wspomnieć, że skala ta wywodzi się z eksperymentów Likerta i w różnych wariantach jest stosowana w eksperymentach psychologicznych i percepcyjnych, głównie ze względu na brak konieczności kalibracji słuchaczy, stosowania tylko określonych pobudzeń (stimuli) i względną łatwość przeprowadzania eksperymentów. Trzeba jednak zauważyć, że sam proces oceny jakości dźwięku przez słuchacza jest bardzo złożony i silnie zindywidualizowany (dla różnych słuchaczy różne cechy są istotne, skąd konieczne jest w ocenie statystycznej stosowania techniki MDS – *multidimensional scaling*). Istnieją inne skale, które generalnie lepiej oddają istotne zniekształcenia głosu – np. DAM – Diagnostic Acceptability Measure, czy nawet skala GRBAS stosowana w ocenie głosów patologicznych. Trochę mi takiej dyskusji w rozprawie zabrakło, Autor skoncentrował się tylko na materiałach ITU, więc dla porządku odsyłam do pracy [2].

Na str. 41-42 Autor podaje zależności pomiędzy wartością MOS a współczynnikiem R. Niestety, zabrakło wytłumaczenia co to jest to R, a szkoda, bo jest tu zawarty istotny element pracy, a mianowicie powiązanie subiektywnej miary MOS z obiektywnym parametrem R wyliczanym wg E-modelu ITU (opisanego w rozdziale 3.6). Ta zależność (wzór (2.1)) jest tylko pewną estymatą opinii słuchaczy i wynika z korelacji R i MOS, która jednak nie jest 100% (patrz np. [3] gdzie w najlepszym przypadku sięga ona 80%). Tak więc nie można z całkowitą pewnością twierdzić, że kombinacja obiektywnych parametrów sieci jest dokładnym odwzorowaniem opinii słuchaczy (coś na ten temat znalazłem na str. 59 rozprawy). Także zdanie ze str. 41 – „Ponieważ najczęściej stosowaną skalą jest skala MOS, można przyjąć, że najlepszą metodą pomiaru systemów cyfrowej transmisji dźwięku jest metoda, która przedstawia swój wynik właśnie [brak w:] skali MOS” jest zbyt daleko idącym uogólnieniem. Mam też wrażenie, że cennym uzupełnieniem rozprawy byłoby wprowadzenie oznaczeń rekomendowanych przez ITU P.800.1 rozróżniających wyniki w skali MOS otrzymane poprzez zastosowanie różnych rodzajów pomiaru jak w poniższej tabeli:

	Listening-only	Conversational
Subjective	MOS-LQS	MOS-CQS
Objective	MOS-LQO	MOS-CQO
Estimated	MOS-LQE	MOS-CQE

Kolejne podrozdziały opisują czynniki wpływające na jakość transmisji w sieci pakietowej oraz metody zapobiegania problemom, np. mechanizmy QoS. Ta część jest generalnie dobrze napisana, jako recenzent zwracam tylko uwagę na brak wyjaśnienia w Tabelach 2.6 i 2.8 dlaczego takie a nie inne wartości opóźnienia i jitteru są akceptowalne (mechanizmy percepcji dźwięku) oraz brak podanego źródła w Tabeli 2.9.

Rozdział trzeci rozprawy poświęcono metodom oceny jakości transmisji w sieciach IP.

Pierwsza część tego rozdziału przedstawia taksonomię metod oceny jakości. Jest ona poprawna, ale różne metody oceny stosuje się także w zależności od przewidywanego poziomu degradacji sygnału (np. MUSHRA dla sygnałów o małym lub średnim poziomie zniekształceń). Omówiono metody PAMS, PESQ, PSQM, E-model, P.563, K-factor.

Szczególnie obszernie przedstawiono E-model – i słusznie, wydaje się bowiem, że ma on największe znaczenie praktyczne, przede wszystkim ze względu na możliwość wykorzystania go już na etapie projektowania sieci. Zabrakło przedstawienia metody P.564, która jako parametryczna pozwala prognozować jakość głosu w sieciach IP w zależności od konfiguracji kanałów transmisji.

Na str. 84 rozprawy opisano test metody PESQ. Niestety, opis przygotowania drugiego z sygnałów jest dość niezrozumiały i trudno dociec, dlaczego wyniki MOS-LQE są takie, a nie inne.

Ostatnia część rozdziału zawiera krytykę metody P.563. Wykazano w niej rozbieżności w wynikach tej metody i PESQ oraz błędy oszacowania jakości w przypadku krótkotrwałej utraty dźwięku.

Rozdział czwarty pracy doktorskiej mgr Apiecionka przedstawia autorską metodę oceny jakości transmitowanego dźwięku (OJTD) w telefonii VOIP. W p.4.1 przedstawiono założenia dotyczące proponowanej metody oraz sygnały testowe, będące częścią bazy ITU P.Sup 23. Jak można domniemywać, nie posłużono się niestety nagraniami języka polskiego, choć rozumiem, że celem prac było otrzymanie wyników porównywalnych z innymi pracami oraz tego, że baza ITU zawiera tylko kilkanaście próbek naszego języka. W kolejnych krokach przygotowano sygnały testowe do badania wpływu jednego z czterech standardów kodowania dźwięku na jakość transmisji, do badania wpływu zaniku sygnału na percepcję

jakości transmisji dźwięku (różne długości przerw i sygnałów oraz wypełnienie przerw szumem otoczenia lub ostatnią wartością oryginalnego sygnału), wpływu rozmiaru i strat pakietów IP, echa oraz opóźnień sieci. W pracy nie podano w jakim języku były wypowiedziane próbki testowe i jaki był ich dobór od strony językowej oraz mówców.

Trochę mnie dziwi użyte wypełnianie przerw w transmisji dźwięku: w trakcie nagrywania bazy LUNA (dialogi telefoniczne) zauważyliśmy częste przerwy w sygnale przesyłanym poprzez GSM – przerwy były wtedy wypełnione de facto ciszą (amplituda bliska zeru). Choć o tym nie wspomniano, zakładam że szum otoczenia był podany z naturalną, tzn. nie modyfikowaną głośnością. Stwierdzenie ze str. 97, że ucho ludzkie nie jest w stanie rozpoznać przerw krótszych niż 20 ms nie jest niestety prawdziwe – człowiek jest w stanie usłyszeć nawet krótsze niż 5 ms zaniki sygnału, głównie poprzez zmianę widma sygnału (*gap detection in noise*) i zależy to też od głośności sygnału [4].

Rozdziały 4.2 – 4.4 prezentują autorski wkład doktoranta do opisywanej dziedziny wiedzy. Pan Łukasz Apiecionek przeprowadził w nich analizę wymagań jakie powinna spełniać miara jakości transmisji dźwięku VoIP, tak aby mogła być stosowana do monitorowania stanu pracy sieci realizującej usługę VoIP, przedstawił własną propozycję metody oraz zbadał wpływ szeregu czynników (standard kodowania dźwięku, przerwy w sygnale, echo, opóźnienia pakietów, wielkość i straty pakietów IP). Jak wynika z przeprowadzonych badań autorski algorytm pozwala na uzyskanie wyników bardziej zbliżonych do wyników oceny metodą PESQ (wymagającej sygnału referencyjnego) niż metoda P.563. Proponowany algorytm sprowadza się do rozszerzenia obliczania MOS wg metody P.563 o dwa parametry uwzględniające relacje sygnał/cisza w odbieranym sygnale i uwzględnienie ich wpływu w przeskalowaniu wyników metody P.563. Przeskalowanie to podano w Tabelach 4.3 – 4.5 i wynika ono z odwzorowania pomiarów PESQ i P.563. Wpływ czynników zakłócających podano w odniesieniu do metody PESQ. Wykonano wiele pomiarów, ich wyniki przedstawiono w czytelny i jasny sposób; wykazano, że rzeczywiście wyniki OJTD bliższe są oszacowaniu PESQ. Muszę jednak wytknąć brak analizy statystycznej otrzymanych wyników, a w szczególności choćby oszacowania statystycznej istotności różnic pomiędzy wynikami OJTD i P.563. Chciałbym też zwrócić uwagę na dyskusję na temat detekcji mowy (VAD) i zastosowania przez Autora własnego kryterium wykrywania przerw w sygnale. Otóż to bardzo dobrze, że p. Apiecionek nie skorzystał z VAD w tym celu – algorytmy te generalnie mają słabą rozdzielczość czasową i małą precyzję lokalizacji. Wiele osób twierdzi, że prawidłowe wykrywanie mowy przy zakłóceniach w tle jest równie trudne jak jej rozpoznawanie.

Podrozdziały 4.5 i 4.6 odnoszą się do złożoności obliczeniowej algorytmu oraz możliwości zastosowań proponowanego algorytmu. Rozdział czwarty zakończony jest sensownym podsumowaniem odnoszącym proponowany algorytm do oceny jakości transmisji innymi metodami.

Końcowe wnioski wskazują na możliwość praktycznego wykorzystania algorytmu OJTD do monitorowania stanu sieci i usługi VoIP.

W Dodatku A oraz na płycie CD zaprezentowano szczegółowe wyniki pomiarów OJTD w odniesieniu do metody P.563.

#### 4. Ocena edytorska pracy

Praca jest starannie zredagowana i poprawna pod względem językowym. Liczne tabele i ilustracje dobrze podsumowują przekazywany materiał, starannie złożono wzory matematyczne, a drobne literówki i nieliczne błędy stylistyczne nie wpływają w żadnej mierze na jakość rozprawy doktorskiej. Jako recenzent nie mam też zastrzeżeń co do układu pracy i kolejności rozdziałów. Jedyne drobne zastrzeżenia to nieuporządkowana alfabetycznie bibliografia i brak spisów tabel i ilustracji.

#### 5. Podsumowanie

Rozprawa doktorska p. Łukasza Apiecionka jest niewątpliwie interesującą próbą zmierzenia się z problemem oceny jakości transmisji mowy w sieciach IP, co jest zagadnieniem istotnym dla współczesnej techniki informatyczno-telekomunikacyjnej. Opracowany algorytm pomiaru jakości OJTD może być wykorzystywany w czasie rzeczywistym do monitorowania stanu sieci VoIP i mam nadzieję, że znajdzie on praktyczne zastosowanie. Uważam, że rozprawa wykazuje zdolność doktoranta do samodzielnego rozwiązania złożonego problemu badawczego przy użyciu nowoczesnych metod badawczych i własnych osiągnięć (odpowiedni dobór wymagań, istotna modyfikacja algorytmu P.563, solidne przebadanie zaproponowanej metody).

W konkluzji stwierdzam, że opiniowana rozprawa spełnia wymogi stawiane przez ustawę o stopniach i tytułach naukowych dla prac doktorskich, a jej autor zasługuje na przyznanie stopnia naukowego doktora nauk technicznych w dyscyplinie informatyka. Stawiam więc wniosek o dopuszczenie tej rozprawy do publicznej obrony.

## Bibliografia

- [1] <http://www.eetimes.com/electronics-news/4139026/Sorting-Through-GSM-Codecs-A-Tutorial>
- [2] Polkosky M., Lewis J., Expanding the MOS: Development and Psychometric Evaluation of the MOS-R and MOS-X, INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY 6, 161–182, 2003
- [3] Hall T., Objective Speech Quality Measures for Internet Telephony, 2001 — in Proceedings of SPIE Voice over IP VoIP Technology
- [4] Plack Ch., The sense of hearing, Erlbaum Pub., 2005, ale też:  
[http://www.audiologynow.org/pastfuture/Presentation\\_pdf/PP175%20Bradley%20Hess.pdf](http://www.audiologynow.org/pastfuture/Presentation_pdf/PP175%20Bradley%20Hess.pdf)