

dr hab. inż. Przemysław Dymarski
05520 Konstancin - Jeziorna
ul. Bielawska 31

6 września 2010

Recenzja

rozprawy doktorskiej mgr. inż. Łukasza Apiecionka

pt. "Metoda oceny jakości transmisji głosowej w telefonii VoIP"

sporządzona dla Rady Naukowej Instytutu Podstawowych Problemów Techniki PAN

1. Cel badań i teza rozprawy

Temat rozprawy wiąże się z zagadnieniami pomiaru jakości usług telekomunikacyjnych. Zastosowanie sieci IP do transmisji dźwięku i obrazu spowodowało wzrost zainteresowania metodami pomiaru jakości usług typu Voice over IP, audio- i wideokonferencja, Video on Demand, IPTV, etc. Autor rozprawy ogranicza się do badania jakości najbardziej popularnej z tych usług: VoIP. Celem autora jest monitorowanie jakości transmitowanej mowy bezpośrednio u odbiorcy nie znając oryginalnych (pozbawionych zniekształceń i zakłóceń) fraz mowy podlegających transmisji. Taki nieintryzyjny (*nonintrusive*) algorytm jest trudniejszy do skonstruowania i mniej dokładny niż algorytm intryzyjny (*intrusive*), polegający na porównaniu frazy odebranej i nadawanej (np. algorytm PESQ, Zalecenie ITU-T P.862). Ma jednak tę zaletę, że nie wymaga przesyłania z góry ustalonych fraz i może być wykorzystany do stałego monitorowania jakości mowy u odbiorcy.

Jedynym algorytmem tego typu, zalecanym obecnie przez ITU-T, jest algorytm 3SQM (*Single Sided Speech Quality Measure*, Zalecenie P.563). Autor postawił sobie ambitne zadanie poprawy dokładności tego algorytmu, przy czym miarą dokładności ma być odstęp od wyników (MOS - *Mean Opinion Score*) uzyskanych z wykorzystaniem algorytmu intryzyjnego PESQ. Zadanie badawcze jest postawione właściwie, jednak brzmienie głównej tezy rozprawy ("Możliwa jest ocena jakości rozmowy realizowanej w telefonii VoIP za pomocą metody nie ingerującej w badaną sieć") nie jest adekwatne do postawionego zadania - wszak algorytm 3SQM umożliwia taką ocenę jakości. Teza powinna, moim zdaniem, brzmieć: "Możliwa jest poprawa dokładności algorytmu 3SQM, poprzez..." - i tu wymienić najważniejsze cechy proponowanego rozwiązania.

2. Charakter rozprawy - teoretyczny, doświadczalny, konstrukcyjny

Praca ma charakter teoretyczny i doświadczalny. Warstwa teoretyczna obejmuje propozycję algorytmu korekcyjnego, umożliwiającego poprawę dokładności wyników uzyskanych z wykorzystaniem algorytmu 3SQM, natomiast dowód tezy rozprawy przeprowadza się metodami doświadczalnymi, wykonując serie eksperymentów, szczegółowo opisanych w rozdziale 4 i w

Dodatku A.

3. Sposób przeprowadzenia analizy źródeł (w tym literatury światowej i stanu zagadnień w przemyśle). Sposób sformułowania wniosków z analizy

Bibliografia zgromadzona przez Autora liczy niemal 90 pozycji - obejmują one podstawowe zagadnienia poruszane w rozprawie. Rozprawa nie zawiera osobnego rozdziału poświęconego analizie bibliograficznej, jednak elementy tej analizy pojawiają się w rozdziałach 1-3. Rozdział 1 ma charakter przeglądu, nie wnosi wiele do zasadniczego wątku rozprawy, jednak świadczy o dobrej znajomości przez autora zagadnień sieci IP, w szczególności protokołów wykorzystywanych do transmisji głosu.

Rozdziały 2 i 3 poświęcone są przedstawieniu czynników wpływających na jakość mowy w telefonii IP, oraz analizie porównawczej metod oceny jakości mowy. Autor dobrze orientuje się w tych zagadnieniach, choć czasem unika niektórych określeń (np. w tabl. 2.1, 2.2 i 2.3 pokazano wartości MOS w skali porównawczej, degradacyjnej i bezwzględnej), wprowadza własne kryteria podziału metod pomiaru jakości (np. jego definicja metody obiektywnej na str. 59 odnosi się do metody intruzyjnej, z pełnym odniesieniem do oryginalnej frazy). Autor mógłby również wyraźniej zaznaczyć, które wartości MOS odnoszą się do jakości sygnału mowy, a które do jakości usługi VoIP (*call quality*). Bez tego rozróżnienia pojawia się pozorna sprzeczność wyników prezentowanych w tab.2.9 i na rys. 2.8-2.9, np. dla kodera G.711. Dyskusyjne są też uwagi autora nt. związku wyrazistości logatomowej i jakości mowy - moim zdaniem nie jest możliwe jednoznaczne przyporządkowanie wartości MOS wartościom wyrazistości logatomowej. Generalnie jednak wnioski autora są prawidłowe i stanowią punkt wyjścia do rozwiązywania postawionego zadania badawczego.

4. Rozwiązanie postawionego zadania - właściwość przyjętych metod i założeń

W oparciu o analizę bibliograficzną autor słusznie postanowił oprzeć się na algorytmie P.563 (3SQM). Wykonał serię badań symulacyjnych mających na celu porównanie uzyskanych tą metodą wyników (MOS) z wynikami uzyskanymi algorytmem P.862 (PESQ). Wyniki, prezentowane na rys. 3.12-3.19, charakteryzuje duża rozbieżność w sytuacji, gdy przyczyną pogorszenia jakości są przerwy w sygnale. Rozbieżności świadczą raczej przeciwko algorytmowi P.563, gdyż w wielu przypadkach większemu uszkodzeniu sygnału towarzyszy zwiększenie mierzonej tym algorytmem wartości MOS. Autor wysnuł stąd wniosek, że należy skorygować wartość MOS uzyskaną algorytmem P.563 w taki sposób, aby zbliżyć się do wyniku uzyskanego metodą PESQ. Wniosek ten jest słuszny (algorytm PESQ jest intruzyjny, wymaga znajomości frazy oryginalnej, a więc jest dokładniejszy), jednak bardziej przekonujące byłoby odniesienie do subiektywnych (odsłuchowych) pomiarów MOS. Zdaje sobie jednak sprawę z trudności, związanych z przeprowadzeniem sformalizowanych odsłuchów.

Zaproponowany przez autora algorytm korekcyjny (OJTD) został opisany w p.4.3. Ma on charakter heurystyczny i opiera się na obserwacji częstości zmian decyzji mowa-cisza, generowanych w prostym dyskryminatorze VAD, również pomysłu autora. Zmiany te zliczane są w obrębie okien czasowych o różnej długości, oraz zagregowane w dwóch parametrach EB1 i EB2. Korektor działa w oparciu o MOS z algorytmu P.563 oraz te dwa parametry. Algo-

rytm jest heurystyczny, a więc trudno oceniać jego poprawność w oparciu o jego konstrukcję, należy oprzeć się na uzyskanych wynikach.

Do eksperymentów autor wykorzystał 1044 frazy mowy atestowane przez ITU-T. Próbki mowy były kodowane z wykorzystaniem 4 koderów, uszkodzane przez wprowadzanie przerw (41 wariantów), strat pakietów (30 wariantów), echa i zmiennego opóźnienia pakietów. W sumie wykonano dziesiątki tysięcy porównań wartości MOS dla PESQ, P.563 i dla P.563 z korektorem. Wyniki świadczą o zbliżeniu się wartości MOS do wartości uzyskanej algorytmem PESQ dzięki zastosowanemu korektorowi, zwłaszcza gdy przyczyną degradacji sygnału były przerwy lub utrata pakietów. W tekście rozprawy należało jednak zaznaczyć, że korektor projektowano dla innych fraz niż te, które wykorzystano do jego testowania (taką informację uzyskałem od Autora drogą korespondencyjną). Jest to istotne ze względu na tzw. "*data overfitting*". Korektor, doskonale dopasowany do bazy uczącej (*overfitted*), może dać gorsze wyniki dla fraz spoza tej bazy.

W sumie warstwę eksperymentalną rozprawy oceniam wysoko, ze względu na liczbę wykonanych doświadczeń. Wydaje się, że autor nie do końca wykorzystał ten atut, prezentując jedynie wartości średnie, wariancje i histogramy błędów (błędem nazywa się tu odległość MOS od wartości referencyjnej uzyskanej algorytmem PESQ). Można byłoby obliczyć przedziały ufności dla wyników, a także testować hipotezę H_0 (korektor nie zmniejsza średniej wartości bezwzględnej błędu) przeciw hipotezie H_1 (korektor zmniejsza tę wartość).

5. Oryginalność rozprawy - samodzielny dorobek autora - pozycja rozprawy w stosunku do stanu wiedzy i poziomu techniki prezentowanego w literaturze światowej

Badanie jakości usług telekomunikacyjnych jest dziedziną znajdującą się w stanie intensywnego rozwoju. Pojawia się wiele algorytmów, jednak zalecenia ITU-T uzgadniane są opornie, gdyż trudno znaleźć rozwiązanie dostatecznie uniwersalne. Z tego względu opracowanie korektora dla algorytmu P.563, umożliwiającego uzyskanie dokładniejszych wartości MOS w warunkach występowania przerw w sygnale i utraty pakietów, jest sporym osiągnięciem. Jest to oryginalny dorobek autora rozprawy.

6. Poprawność przedstawienia uzyskanych wyników - zwięzłość, jasność, umiejętność przekonywania, poprawność redakcyjna

Lektura rozprawy wywołuje raczej korzystne wrażenie - jej struktura (podział na rozdziały i podrozdziały) jest właściwa, podporządkowana przekazaniu treści. Rozdział 1 można by jednak skrócić a rozszerzyć opis algorytmu P.563 w rozdziale 3 - algorytm ten stanowi podstawę metody proponowanej przez autora, a jego opis jest nieco zdawkowy.

Utrudnieniem w lekturze są pomyłki redakcyjne i nieścisłości, zauważone usterki i sprawy do wyjaśnienia przytaczam poniżej:

- Wzory 4.5 nie odpowiadają schematowi z rys. 4.4, natomiast wyniki (4.7 i 4.8) są poprawne. Podobnie jest ze wzorem 4.9 i wynikiem 4.10. Z tego względu błędy we

wzorach 4.5 i 4.9 uznaję za usterki redakcyjne.

- We wzorze definiującym jitter (2.2) pominięto wartość bezwzględna, dopiero po jej wstawieniu wzór odpowiada definicji zgodnej z dokumentem RFC1881.
- Wzór 4.13 nie opisuje pierwiastka z mocy. Aby tak było pierwiastek musi obejmować także mianownik.
- Jeśli czas trwania ciszy wynosi 0, to porównujemy dwa identyczne sygnały. PESQ powinien wykazać MOS=4.5 i jest tak w istocie dla próbek dźwięku A, B, D, natomiast dla próbki C (rys. 3.14 i 3.18) jest mniejsza wartość. Dlaczego?
- We wzorach 3.2 i 3.3 występuje W_{nk} a w spisie oznaczeń W_{kn} .
- Warunek ciszy, opisany wzorem 4.16 powinien zawierać jakieś zabezpieczenie przed dzieleniem przez zero.
- Występują drobne błędy literowe i gramatyczne, na str 18, 27, 32, 34, 36, 49, 57, 62, 81, 98, 116, 120, 123, 126.

7. Słabe strony rozprawy, jej główne wady

Wymieniłem je już w poprzednich punktach, odnoszą się głównie do redakcji tekstu rozprawy.

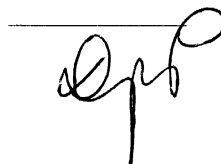
Poza tym, rozprawa byłaby lepsza, gdyby autor wyjaśnił genezę proponowanego przez siebie algorytmu korekcyjnego. Dlaczego przyjęto takie, a nie inne definicje parametrów EB1 i EB2? Jaki mają one sens fizyczny? Autor wykonał bardzo wiele eksperymentów, jednak niektóre z nich odnoszą się do warunków rzadko spotykanych w praktyce. Np. zamiast badać MOS przy utracie pakietów 30% czy 40%, dobrze byłoby zwiększyć liczbę eksperymentów przy stopie utraty pakietów mniejszej niż 10%.

8. Przydatność rozprawy dla nauk technicznych, przemysłu, obronności kraju itp.

Ze względu na niewątpliwą poprawę dokładności pomiarów jakości mowy w warunkach utraty pakietów, algorytm korekcyjny zaproponowany przez autora mógłby być wykorzystany wraz z algorytmem P.563 w systemach monitorowania jakości usługi VoIP u użytkownika lub w dowolnym punkcie sieci IP.

9. Zaliczenie rozprawy do jednej z kategorii

Moim zdaniem pracę można zakwalifikować jako spełniającą wymagania stawiane rozprawom doktorskim.



P.Dymarski