

MĚŘENÍ A HODNOCENÍ QOS V IP TELEFONII

Doc. Ing. Jan HOLUB, Ph.D.

pracoviště: ČVUT FEL, Katedra měření; mail: jan.holub@mesaqin.com

Abstrakt: V příspěvku je podán stručný přehled vlivů, ovlivňujících výslednou kvalitu hlasového přenosu v IP telefonii. Jsou uvedeny metody pro subjektivní i objektivní měření kvality přenosu hlasu. U objektivních metod jsou popsány typické intrusivní i neinrusivní algoritmy. Závěrem jsou uvedeny některé konkrétní aplikace měření kvality přenosu hlasu.

8 Co je kvalita přenosu hlasu?

Pod pojmem kvalita přenosu hlasu rozumíme subjektivní hodnocení jakosti řečového signálu, přeneseného komunikačním kanálem. Pojem kvalita popisuje degradaci způsobenou přenosem a zohledňuje tedy aspekty lidského vnímání řečového signálu. Díky zvyšující se složitosti a komplexnosti sítí, kdy komunikační řetězce zahrnují stále více a více přenosových technologií, se měření kvality přenosu hlasu stává jednou z mála platforem, pomocí které lze navzájem porovnávat zcela odlišné přenosové technologie, a která je ve své podstatě nejbližší pohledu jednotlivých koncových uživatelů.

Budeme-li posuzovat kvalitu z hlediska lidského vnímání, můžeme za kvalitativní ukazatele považovat např.: hlasitost, srozumitelnost, šum, přeslechy, výpadky a další. Mezi nejvýznamnější složky, které ovlivňují kvalitu hlasu, patří srozumitelnost, „end-to-end“ zpoždění a echo. Jejich význam je následující:

Srozumitelnost (angl. clarity), často také označována jako čistota zvuku, je přímo vázána na vnímání hlasu lidským uchem. Srozumitelností chápeme věrnost, zřetelnost a nezdeformovanost původního signálu. Může být také chápána jako množství informace, kterou jsme schopni ze zprávy získat. Srozumitelnost hlasu je určena frekvenčním pásmem, který je přenesen (resp. deformován). Pro porozumění zprávě nejsou například příliš důležité frekvence nad 1000 Hz, které pouze dokreslují barvu a charakteristiku hlasu mluvčího.

Na čistotu zvuku má vliv převážně působení ztrátových kodeků, časové a amplitudové ořezání, dočasné výpadky a ztráty signálu, změny v okamžitém zpoždění, šumy, útlumy, zesílení a přenosové chyby kanálů.

„End-to-end“ zpoždění – součet všech zpoždění, která působí na přenášený signál po celé přenosové cestě. Tato zpoždění jsou způsobena zejména vzdáleností, na kterou je signál přenášen (vlivem konečné rychlosti šíření signálu) a existencí prvků zpracování signálu. Zpoždění neovlivňuje kvalitu hlasu přímo, ovlivňuje „charakter“ konverzace. Zpoždění do 100 ms nebývají zpravidla postřehnutelná, při zpoždění mezi 100 – 300 ms je patrné jakési „váhání“ v odpovědích. Zpoždění nad 300 ms jsou již naprosto zřetelná, uživatelé mají snahu pomlky přerušovat vlastními vstupy do hovoru. Komunikace se tímto stává nemožnou.

Echo, neboli dozvuk či ozvěna, je zvuk, který se vrací zpět ke zdroji. Dříve byl tento parametr určován zejména u spojení na velmi dlouhé vzdálenosti. U digitálních systémů je však echo zvýrazněno i při komunikaci na krátké vzdálenosti především díky prvkům zpracování signálu a dalším prvkům způsobujícím zpoždění signálu, obsaženým v komunikačním řetězci.

Pokud je mezi původním a odraženým signálem (echem) malý časový rozdíl (do 25-30 ms), je vliv echa zanedbatelný. Při větším časovém rozdílu a dostatečné úrovni echa je jeho existence nežádoucí.

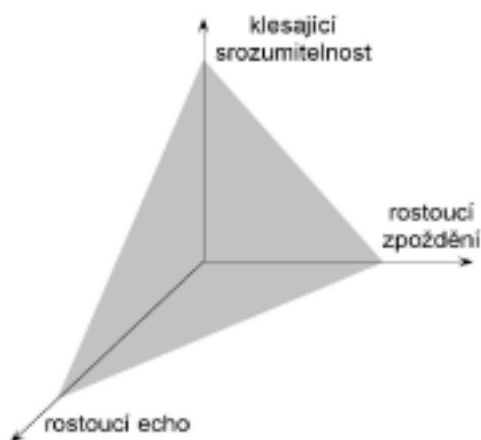
Echo vzniká například akustickou vazbou mezi reproduktorem a mikrofonem (tzv. akustické echo), zakončovacími obvody v síti PSTN apod.

Vztah srozumitelnosti, „end-to-end“ zpoždění a echa je komplexní, viz. obr. 1. Čím blíže se pohybujeme ke středu os, tím je pomyslná kvalita přenosu hlasu lepší. Srozumitelnost a zpoždění je ortogonální hledisko kvality hlasu. Signál může být celkově zpožděn bez ztráty kvality, či naopak za velmi krátkou dobu může být přenesen signál značně zdeformovaný. Echo závisí na zpoždění a ovlivňuje srozumitelnost: Echo je slyšitelné, pokud není zamaskováno užitečným signálem nebo šumem a pokud je zpoždění dostatečně dlouhé.

Uvedené parametry (srozumitelnost, echo a zpoždění) jsou pro hodnocení kvality hlasu zásadní, avšak nikoli jediné (dále např. například přirozenost - stupeň věrnosti hlasu mluvčího, či hlasitost - subjektivní pocit, kterým hodnotíme intenzitu akustického vjemu a další) přičemž lze říci, že zkreslení přenášeného hlasu v GSM je způsobeno především proměnným zpožděním a ztrátou přenesených paketů, ořezáním, šumy, echem a nelineárním kódováním. Význam těchto veličin je následující:

Šum – parametr, který přímo ovlivňuje srozumitelnost. Do přenosové cesty se dostává ještě před konverzí užitečného signálu mikrofonem v podobě rušivých zvuků z okolí, dále pak jako kvantovací šum převodu A/D převodníkem, vlivem kodeků, rádiovým přenosem atd. Šum je více patrný při pasážích ticha než při aktivním hovoru (tzv. maskování signálu signálem o vyšší intenzitě).

Nelineární zkreslení – je způsobeno použitím kodeků se ztrátovou kompresí. Kodek transformuje hlasový signál do digitálního bitového toku, přičemž se vypouštějí redundantní či nepodstatné informace. Cílem kodeků je snížení datového toku, kdy se vychází z fyziologie lidského sluchu. Ideální zpětná rekonstrukce vzhledem k použití ztrátové komprese není možná.



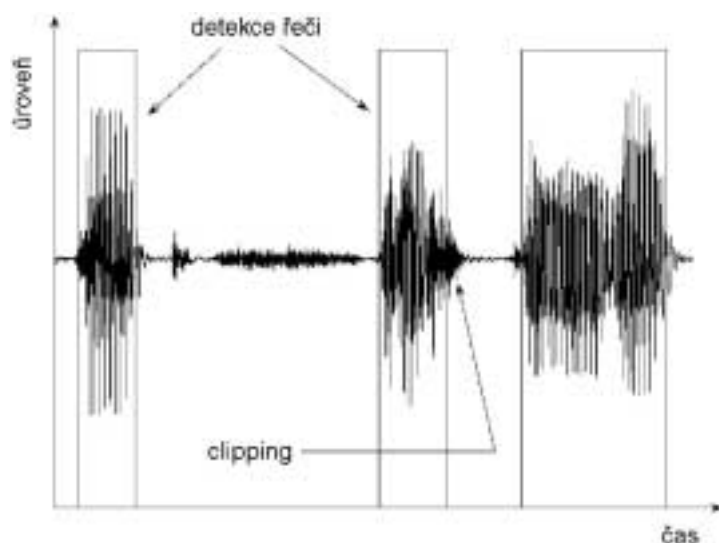
Obr. 1: Znázornění „oblasti kvality“ přeneseného hlasu

Ořezání v čase – součástí většiny mobilních telefonů je detektor řečové aktivity VAD (Voice Aktivita Detektor). Jde o zařízení, které vyhodnocuje, zda účastník hovoří nebo je v hovoru pauza a tudíž není zapotřebí přenášet signál. Tímto lze uspořit až 50 % přeneseného objemu dat (hovoří většinou jeden z účastníků, není tedy zapotřebí přenášet hlas nepřetržitě oběma směry). To přináší úsporu energie, omezení vyzařování mobilní stanice a snížení zatížení sítě. Detektor je nutné nastavit tak, aby spolehlivě rozpoznal řeč od intervalů ticha, tzv. „promlka“. Chybným nastavením dochází k ořezání řečových intervalů – (clipping)

Ztráta paketů – ke ztrátám paketů dochází například při poruchách na rádiovém kanálu, při neprůchodnosti některých linek či zaplnění zásobníku směrovače, nebo dojde-li k tak velkému zpoždění daného paketu, že jej není možné zařadit do vygenerované řeči. Velmi záleží na tom, zda chybí souvislá sekvence paketů nebo několik osamocených.

Výsledná kvalita hlasového přenosu tedy závisí na množství parametrů, které zahrnují okamžité fyzikální vlastnosti všech prvků, které tvoří komunikační řetězec. V minulosti byla kvalita přenosu po analogové lince determinována převážně velikostí odstupu signálu od šumu, útlumem trasy a u spojení na velmi dlouhé vzdálenosti (zejména u satelitních spojení) také dozvukem či ozvěnou (echo). S nástupem digitální technologie se význam šumu a útlumu trasy snížil. U bezdrátových telekomunikačních systémů se však objevily zcela nové fenomény, degradující kvalitu hlasového přenosu, jako je impulsní šum, krátkodobé výpadky, ořezávání signálu (clipping) a nelineární zkreslení (vlivem použití kodeků se ztrátovou kompresí). Další prvky zpracování signálu v komunikačním řetězci navíc prodlužují dobu šíření signálu a mohou tak mimo jiné přispět ke zvýraznění echa i při komunikaci na krátké vzdálenosti. U paketově orientovaných přenosů se navíc objevuje proměnné zpoždění (jitter delay), které je způsobeno směrováním paketů odlišnými trasami.

Z výše nastíněné analýzy vyplývá množství a různorodost parametrů, které mají vliv na kvalitu přenosu hlasu. Znalost hodnot výše uvedených parametrů pro určité spojení je však pouze prvním stupněm pro zhodnocení kvality tohoto spojení, neboť dalším nezbytným krokem je vhodné shrnutí těchto hodnot do závěrečného zhodnocení kvality jediným souhrnným indexem.



Obr. 2: Znázornění vzniku jevu ořezání v čase (temporal clipping) vlivem VAD

9 Přehled metod měření kvality přenosu hlasu

Z hlediska objektivity je samozřejmě optimální metodou pro hodnocení kvality statistické vyhodnocení názorů dostatečně rozsáhlé skupiny osob. Takovéto testy jsou standardizovány, např. v doporučení ITU-T P.82, „Metody pro hodnocení služeb z hlediska kvality přenosu hlasu“. V sadě doporučení P.830 je popsán způsob výběru respondentů, proces přípravy řečových vzorků, provádění vlastních poslechových testů a vyhodnocení výsledků poslechových testů. Obvykle se používá stupnice MOS (Mean Opinion Score) v hodnotách 1 (nejhorší kvalita) až 5 (nejlepší kvalita), viz tab. 1.

Škála MOS	
5	Excelent – vynikající kvalita
4	Good – dobrá kvalita
3	Fair – přijatelná kvalita
2	Poor – špatná kvalita
1	Bad – velmi špatná kvalita

Tab. 1: Stupnice MOS

Pro dostatečně objektivní hodnocení je zapotřebí dostatečně velký počet účastníků takového výzkumu, neboť na subjektivní hodnocení jednotlivce mohou mít vliv takové faktory jako obsah hovoru či doba, která uplynula od posledního účtu za telefon. Také je třeba brát v úvahu, že uživatel si obvykle pamatuje nízkou kvalitu jediného hovoru více než množství hovorů více kvalitních. Tato úskalí byla v minulosti řešena použitím školených osob. Pro provádění testů v požadovaném rozsahu je však časová i finanční náročnost takového postupu evidentní. Proto se obvykle používá objektivních měřicích metod, jejichž cílem je náhrada poslechových testů vhodným algoritmem, zpracovávajícího přenesený (případně též originální) hlasový vzorek a generující odhad výsledného subjektivního dojmu. Je zřejmé, že přístupy, vycházející z pouhého měření sady parametrů přenosu (např. zpoždění a jeho změny, bitová chybovost či počet nesprávně přenesených a ztracených paketů), zpravidla nepostihuje dostatečným způsobem výslednou subjektivní kvalitu přenosu.

Kvalitu hlasového přenosu lze v zásadě měřit dvěma způsoby: intrusivním a neintrusivním. Podstatou intrusivního měření kvality přenosu hlasu (případně obecného audiosignálu) je následující (obvykle automatizovaný) postup:

- navázání spojení mezi dvěma stanicemi, které jsou obě obvykle součástí měřicího systému (výjimkou je např. testování hlasových schránek),
- vyslání vzorku signálu a přijetí jeho přenesené verze protistanicí,
- porovnání původního a přeneseného vzorku pomocí vhodného algoritmu, napodobujícího způsob vnímání a zaujímání názoru na kvalitu přenosu průměrným posluchačem,
- znázornění výsledků vhodným způsobem a jejich statistické zpracování.

Neintrusivní metody jsou naproti tomu založeny na pasivním monitorování probíhajících spojení. Obvykle jsou implementovány v rámci monitorovacích či dohledových systémů a umožňují tak měřit kvalitu přenosu hlasu na mnoha spojeních zároveň. Nevýhodou tohoto postupu, nazývaného v příslušných standardech INMD (In-service Non-intrusive Measurement Device) je skutečnost, že hodnotící algoritmus nemá k dispozici originální (původní) verzi řečového signálu, takže je velmi obtížné detekovat určité typy zkreslení signálu při přenosu (např. harmonické zkreslení signálu, způsobené některými typy kodeků, zejména ADPCM).

10 Používané metody a normy

Pro automatizované vyhodnocení kvality hovoru bylo mezinárodními institucemi standardizováno několik modelů hodnocení přenosu. Jejich účelem je predikce celkového subjektivního dojmu uživatele na základě kombinace dílčích (měřitelných) rušivých vlivů. Např. [1] uvádí čtyři starší v telekomunikacích prakticky používané modely:

„Transmission Rating“ – Spojené státy a Kanada

„Catnap83“ – Velká Británie (British Telecom)

„Information Index“ – Francie

„Opine“ – Japonsko

V následujícím textu se zaměříme na příklady standardizovaných algoritmů pro měření kvality přenosu hlasu s odkazem na jejich vhodnost pro použití v IP telefonii.

3.1 P.861 – PSQM

Doporučení ITU-T P.861 (PSQM – Perceptual Speech Quality Measurement) byl donedávna nejpoužívanější algoritmus pro posuzování kvality přenosu hlasu. Skládá se ze dvou částí, kdy první je určena pro měření kvality hlasových kodeků a druhá pro měření kvality přenosu hlasu v celém telekomunikačním řetězci. Princip P.861 spočívá v porovnání amplitud výkonových spekter sobě odpovídajících úseků původního a přeneseného signálu. Délka rámců, na které je řečový signál dělen, je přibližně 16 ms (s časovým překryvem sousedních rámců 50%). Jsou přitom rozlišovány okamžiky aktivního hovoru (speech periods) a úseky ticha (odmlky mezi slovy ap. – silent periods). Obě skupiny jsou hodnoceny odlišně při posuzování vlivu šumu a na závěr výpočtu je toto hodnocení kombinováno do jediného výsledného parametru.

3.2 P.862 - PESQ

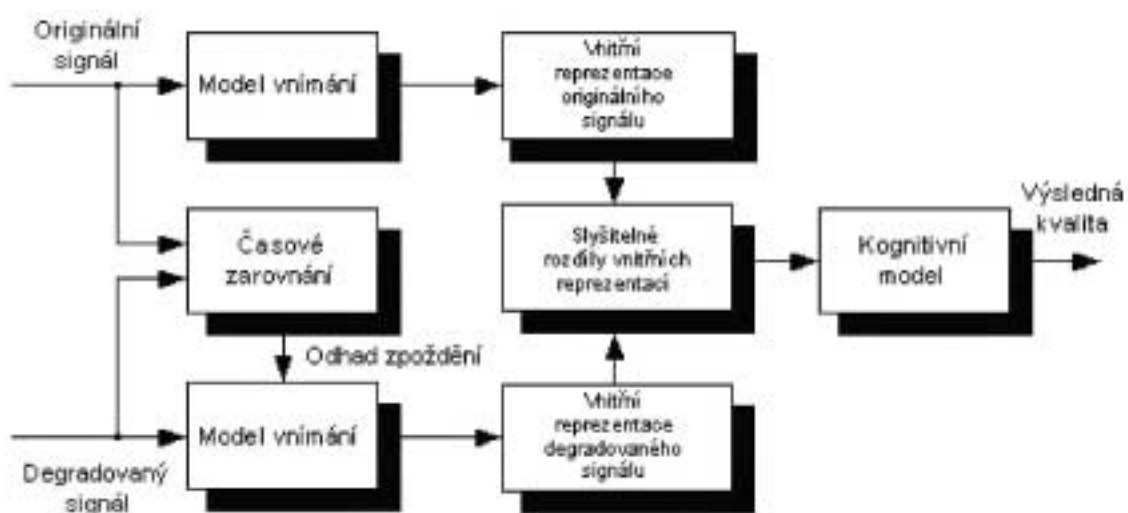
Jeden z problémů IP telefonie spočívá ve skutečnosti, že IP přenosy jsou ve své podstatě paketově orientované, tzn. přenášejí se skupiny bitů, které mohou do místa určení putovat navzájem odlišnými cestami s odlišným zpožděním. To vede k proměnlivému zpoždění rekonstruovaných úseků signálu (které lze do určité míry kompenzovat ukládáním dat do pomocného registru na přijímací straně hloubku tohoto registru však nelze libovolně zvětšovat vzhledem k neúměrně se prodlužujícímu zpoždění).

Žádný z předešle popsaných standardů bohužel neumožňuje měření kvality těchto hlasových přenosů, neboť i nepatrné změny ve zpoždění během přenosu jednoho hlasového vzorku vedou k selhání časové synchronizace (která se u P.861 provádí pro celý - obvykle několikasekundový - vzorek řečí najednou). Navíc u těchto typů datových přenosů vzrůstá počet krátkodobých výpadků, které výše uvedené standardy také nehodnotí v souladu s lidským vnímáním – toto hodnocení je navíc velmi obtížně algoritmizovatelné. Tyto nedostatky vedly ke zvýšenému úsilí o návrh nového standardu pro měření kvality hlasových přenosů. Existuje více

proprietárních algoritmů, např. PAMS (Perceptual Analysis Measurement System), navržený expertní skupinou při British Telecom, vedenou M. Hollierem a A. Rixem.

Referenční (původní) a zkreslený (přenesený) signál jsou nejprve časově a amplitudově korelovány. Tato korelace je však realizována na dílčích blocích, nikoli na celém záznamu najednou. Tím je umožněno zachycení proměnného časového zpoždění přeneseného signálu. Oba signály jsou potom zpracovány algoritmem, jehož výsledkem je potom dvourozměrné (časově-frekvenční) pole hodnot (tzv. poslechová plocha), který odpovídá přibližně lidskému vnímání tohoto signálu. Rozdíl sobě odpovídajících hodnot těchto dvou souborů potom tvoří tzv. poslechovou odchylku (chybovou plochu), která je použita pro odhad hodnotících parametrů. PAMS používá dva takové výsledné parametry, a to odhad poslechového úsilí (Y_{le} – listening effort) a kvality poslechu (Y_{lq} – listening quality). Tyto parametry nabývají hodnot 1 až 5.

Algoritmus PAMS se stal zárodkem nejnovější normy ITU-T P.862, která používá metodu rekurentní časové korelace, převzatou z PAMS a doplňuje ji hodnotícím aparátem, obdobným P.861. Koeficient korelace s poslechovými testy se pohybuje u běžných typů zkreslení v rozmezí 0.85- 0.93.



Obr. 3: Blokové schéma algoritmu PESQ

3.3 P.561 a P.562 (INMD a CCI)

Toto doporučení je typickým příkladem tradičního neintrusivního měření kvality přenosu hlasu. Standard P.561 (INMD In-Service Non-intrusive Measurement Device) obsahuje seznam parametrů, které je třeba vyhodnocovat na přenesené verzi hlasového vzorku (např. úroveň signálu, odstup signál šum, řečová aktivita, echo) a P.562 (CCI – Call Clarity Index) obsahuje návod, jak hodnoty těchto parametrů zkombinovat do jediného výsledného parametru.

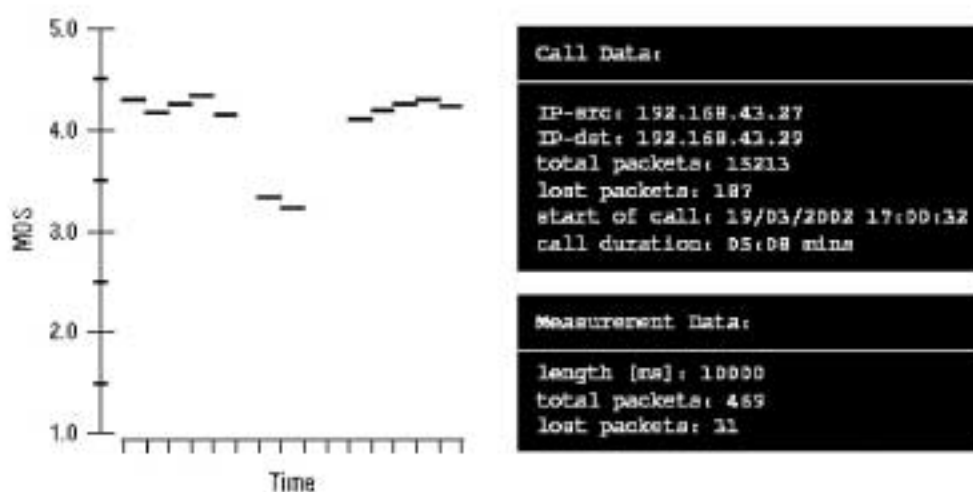
3.4 P.563 (P.SEAM)

P.563 je nejnovějším počinem ITU-T (duben 2004) v oblasti neintrusivních měření. Zpracovává přenesený hlasový vzorek pomocí třech odlišných algoritmů a kombinuje jejich výsledné hodnocení do jediného výstupního parametru. Podrobný popis vnitřní struktury není zatím zveřejněn. Koeficient korelace s poslechovými testy se pohybuje u běžných typů zkreslení v rozmezí 0.78- 0.86.

3.5 PsyVoIP

Algoritmus PsyVoIP je proprietární metoda, vyvinutá firmou Psytechnics Ltd., pro odhad výsledné kvality přenosu hlasu v IP přenosech. Nepracuje s hlasovými vzorky, nýbrž pouze s parametry přenosu jako je zpoždění a jeho změny, typ kodeku či ztráta paketů. Pomocí neuronové sítě je získán odhad výsledné kvality ve stupnici MOS (viz. obr. 4)

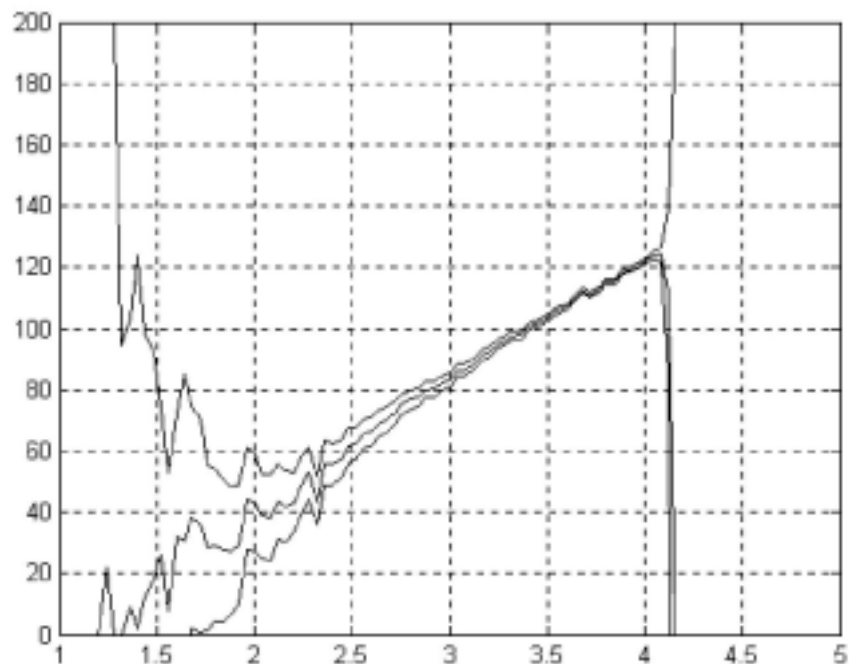
psyVoIP: Call Analysis



Obr. 4: Příklad použití algoritmu PsyVoIP

11 Aplikace měření kvality přenosu hlasu

Měření kvality přenosu hlasu nacházejí celou řadu aplikací. Tradiční skupinou aplikací je přímá náhrada poslechových testů, např. v rámci výběru nového kodeku, nových terminálů či nové síťové technologie. Další významnou skupinou aplikací jsou aplikace typu „early warning“, které jsou založeny na skutečnosti, že při trvalém neintrusivním monitorování kvality na jednotlivých linkách / trasách lze detekovat zhoršení celkové kvality dříve, než jej zaregistruje podstatná skupina uživatelů. Včasným odstraněním poruchy lze proto předejít reklamám, vedoucím případně až k trvalému odlivu uživatelů.



Obr. 5: Příklad závislosti doby hovoru (svíslá osa, s) na průměrné kvalitě přenášeného hlasu během spojení (vodorovná osa, MOS). Vlastní závislost (prostřední křivka) je ohraničena křivkami 95% konfidenčních mezí.

Aplikace typu "QoS based carrier selection" využívá (opět obvykle neintrusivní) metody měření okamžité kvality přenosu hlasu pro výběr optimálního způsobu směrování či sestavení hovoru (pod pojmem optimální se v tomto kontextu rozumí poskytující nejvýhodnější poměr cena / kvalita).

Jednou z nejzajímavějších a v současné době velmi populární aplikací je měření závislosti průměrné délky hovoru v síti na kvalitě přenosu hlasu. Tento psychologický efekt, podaří-li se jej pomocí neintrusivních měření v dané síti kvantifikovat, lze potom využít pro predikci návratnosti investic do nových technologií, jejichž primárním účelem je zvýšení kvality přenosu hlasu. Přírůstek kvality se zjistí pomocí intrusivních měření na modelu sítě v laboratoři a tento se potom extrapoluje do výsledků neintrusivních měření výše uvedené závislosti (viz obr. 5).

Literatura

- [1] Holub, J.: Hodnocení kvality hlasových přenosů v telekomunikačních sítích, Sdělovací technika 6/2001, červen 2001
- [2] Holub J., Street M.D., Šmíd R.: Intrusive Speech Transmission Quality Measurements for Low Bit-Rate Coded Audio Signals, AES 115th Convention, Paper 5954. New York, Audio Engineering Society, listopad 2003.
- [3] Holub J., Beerends J. G., Šmíd R.: A Dependence between Average Call Duration and Voice Transmission Quality: Measurement and Applications, WTS2004, Kalifornie, USA, květen 2004
- [4] Holub, J., Očenášek, J., Šmíd, R.: A Novel Intrusive Voice Transmission Quality Test System for Mobile Networks, IEEE 9th International Workshop on Systems, Signals and Image Processing (IWSSIP), Recent Trends in Multimedia Information Processing. London: World Scientific Publishing Company, Manchester, UK, listopad 2002
- [5] Stránky <http://wireless.feld.cvut.cz/mesaqin>
- [6] Stránky www.mesaqin.com
- [7] ETR 250 - Transmission and Multiplexing (TM), Speech Communication Quality from Mouth to Ear for 3.1 kHz Handset Telephony Across Networks, technical report, ETSI, červen 1996
- [8] ITU-T. Methods for Subjective Determination of Transmission Quality. Series P: Telephone Transmission Quality, Recommendation P.800, ITU, Geneva, 1996
- [9] ITU-T. Objective Quality Measurement of Telephone-band (300- 3400 Hz) Speech Codecs. Series P: Telephone Transmission Quality, Recommendation P.861, ITU, Geneva, 1996
- [10] ITU-T. Subjective Performance Assessment of Telephone-band and Wideband Digital Codes. Series P: Telephone Transmission Quality, Recommendation P.830, ITU, Geneva, 1996.
- [11] ITU-T. Perceptual Evaluation of Speech Quality. Series P: Telephone Transmission Quality, Recommendation P.862, ITU, Geneva, 2001