Изследване на влиянието на продължителността на пакетните загуби върху качеството на звуковите сигнали

Доц. д-р инж. Михаил Благоев Момчеджиков, инж. Ивайло Христов Генчев Технически университет-София, ФКТТ, Радиотехника, мом@vmei.acad.bg

*MomchedjikovM.Bl., I.H. Genchev, **A research over perceived speech quality degradation depending on duration of lost audio blocks.** Real-time voice applications over packet-switched networks suffer specific signal losses. The development of packet-loss recovery techniques needs proper objective estimators of perceived speech quality based on models of human auditory perception. These models allow to properly compare the efficiency of commonly used recovery techniques as well as recently developed ones. They are also helpful for determining the ranges of use of all existing recovery methods.*

1. Въведение.

Традиционните методи за предаване на мултимедийна информация на далечни разстояния подлежат на преосмисляне поради множеството им недостатъци. Днес все по-широко приложение намират комуникационните мрежи с пакетна комутация, като предимствата им се състоят основно в тяхната универсалност и гъвкавост. С развитието на технологиите се оказа, че този тип мрежи могат да се използват и за целите на аудио / видео конференцията, телефония и т. н. Но съществуват някои проблеми, дължащи се на различния първоначален замисъл на пакетно-комутируемите мрежи. Част от най-съществените проблеми са известни като „пакетни загуби". Те включват недоставяне (или прекалено закъснение) на някои пакети (т. е. пропадания във възпроизведения сигнал), разместване, мултиплициране на пакети и т. н.

Разработени са различни методи за компенсиране на ефектите от пакетните загуби. Те изискват различна изчисллителна сложност от крайните системи и имат различна ефективност.

Важна задача е да се оцени качеството на възпроизведения сигнал за различните методи, при еднакви други условия. В случаите на предаване на звукова информация посредством Интернет например (т. нар. IP-телефония), това означава да се даде оценка на влошаването на предавания аудиосигал спрямо оригиналния, за различните похвати за компенсиране на пакетните загуби.

Целта на тази разработка е да направи сравнение между три различни метода в приемната страна на базата на симулационна програма за обективно оценяване на качеството на звуковите сигнали. И трите сравнявани метода изискват доста ниска изчислителна мощност на приемната система, но поне два от тях могат да бъдат практически полезни в редица реални приложения.

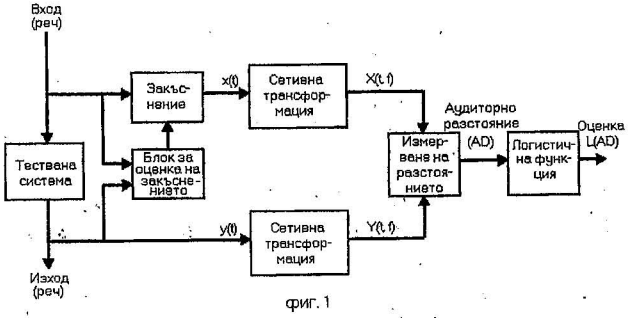
2. Същност на метода за обективно оценяване на качеството на аудиосигналите.

Най-прекият начин за оценяване на качеството на даден аудиосигнал е субективният. Той се състои от възпроизвеждане на сигнала пред широка аудитория и документиране на субективна оценка на слушателите относно качествата на сигнала. Този метод обаче за съжаление отнема твърде много време, средства и т. н. Затова са потърсени методи за обективна (формална, математическа) оценка на качеството на аудио (или поне речевите) сигнали. Тук ще бъде разгледан един доста перспективен такъв метод с претенции за почти пълно съвпадение на резултатите с тези от субективните тестове. Той се нарича „метод с измерващите нормализиращи блокове" ("Measuring Normalizing Blocks" - MNB).

Общата постановка на измерване е показана на фиг. 1. Най-напред се оценява и премахва закъснението, внасяно от изследваната система. ***Сетивните трансформации*** ("Perceptual transformations") съдържат прост модел на слуховото възприятие, а ***измерването на разстояние*** моделира преценката,

Блокът за намиране на разстояния генерира стойности, наричани ***„аудиторни разстояния"("Auditory Distances" - AD)***. Те се увеличават по стойност при увеличаване на субективното усещане за разлика между сигналите на входа и на изхода на системата.

Намирането на разстоянието се извършва на два етапа. Първият е по-груб и включва намирането на обвиващите криви на двата сигнала, тяхното филтриране и крос-корелиране.



фиг. 1

Максимумът на корелационната функция вече дава добра оценка за закъснението, но все още груба (с неточност ±4 ms). След това се преминава към по-финия етап, чрез намиране на спектралните плътности на мощността и отново тяхното крос-корелиране. Това се прави няколко пъти и резултатите се анализират.

Блоковете за сетивна трансформация реализират модел на човешкия слух. В сигнала се подчертават (извличат) онези параметри, които са съществени за човешката преценка за качество. Въз основа на сравненията с резултатите от множество субективни тестове, проучванията сочат два детайла от елементите на модела като най-важни: функцията на честотната разделителна способност на слуховия апарат в зависимост от стойностите на честотата и кривата на субективното възприемане за сила на звука в зависимост от четотата.

Избрана е честотна скала на Барк. В нея вместо самата честота f се използва като независима променлива величината b при следната връзка между двете:

$$b = 6\sinh^{-1}\left(\frac{f}{600}\right) \quad (1)$$

Това заместване се прави с цел съобразяване с нелинейната характеристика на честотната разделителна способност на ухото в зависимост от честотата.

По отношение на втория споменат важен параметър – психоакустичното възприемане на ***силата на звука***, е избрано ***логаритмичното моделиране***.

Блокът за измерване на разстоянието цели моделиране на субективното усещане за ***разлика*** между два перцептуално трансформирани сигнала.

Съставя се йерархична структура от верижно свързани измерващи блокове, всеки от които работи върху избран честотен диапазон (по Барк) и измерва параметри на разликите между всички времеви фрагменти на сигналите, като генерира набор от величини, пряко служещи за оценяване на разликата. Всеки блок премахва от единия от сигналите съответната

част от спектъра, върху която е извършена оценката, и подава сигналите на входовете на следващия блок. Той от своя страна отделя от сигналите следващия честотен обхват и извършва отново оценка по всички времеви интервали и т. н. При такъв принцип на работа блоковете се наричат ***„времеви измерващи и нормализиращи блокове" (TMNB – "Time Measuring Normalizing Blocks").*** Възможно е съставянето на абсолютно аналогична структура, но с ***„честотни измерващи нормализиращи блокове"(FMNB - "Frequency Measuring Normalizing Blocks")***, при които разделянето на честотни и времеви интервали е реверсирано в сравнение с TMNB. Структурната схема на един TMN блок е показана на фиг. 2, а тази на FMN блока е аналогична, но с разменени роли на времевите и честотните операции.

Един TMN блок оперира в честотен обхват $fl \div fu$ и използва времеви интервали, определени от моментите t_i $(i = 1 \div N)$, нормализира $Y(f,t)$ до $\hat{Y}(f,t)$ и генерира $2N$ измервания $m(2i)$ на база на следните зависимости:

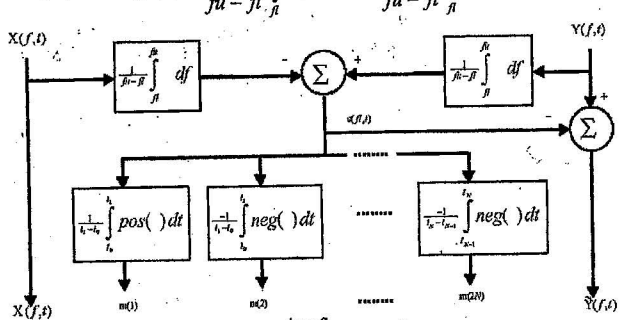
$$\hat{Y}(f,t) = Y(f,t) - e(fl,t);$$

$$m(2i-1) = \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \max(e(fl,t),0)dt,$$

$$m(2i) = \frac{-1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \min(e(fl,t),0)dt, \qquad (2)$$

$$i = 1 \div N,$$

$$\text{където}: e(fl,t) = \frac{1}{fu - fl} \int_{fl}^{fu} Y(f,t)df - \frac{1}{fu - fl} \int_{fl}^{fu} X(f,t)df.$$



фиг. 2

Дефиницията на един FMN блок е аналогична, с разменени роли на честотата и времето. В интервала от време $(t_0 \div t_0 + \tau)$, използвайки честотни обхвати, дефинирани от f_i, $i = 1 \div N$, такъв блок генерира $2N$ измервания $m(2i)$:

$$\hat{Y}(f,t) = Y(f,t) - e(f,t_0),$$

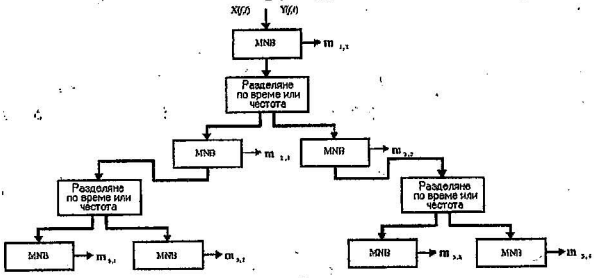
$$m(2i-1) = \frac{1}{f_i - f_{i-1}} \int_{f_{i-1}}^{f_i} \max(e(f,t_0),0)dt,$$

$$m(2i) = \frac{-1}{f_i - f_{i-1}} \int_{f_{i-1}}^{f_i} \min(e(f,t_0),0)dt, \qquad (3)$$

$$i = 1 \div N,$$

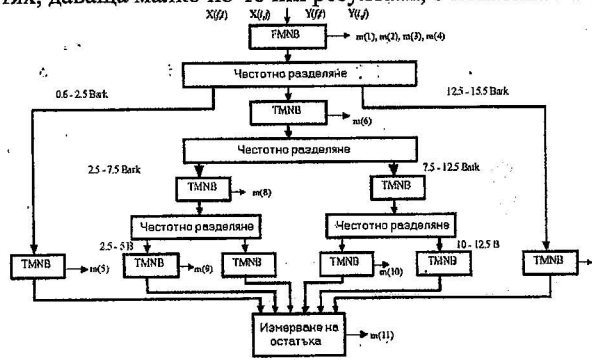
$$\text{където}: e(f,t_0) = \frac{1}{\tau} \int_{t_0}^{t_0+\tau} Y(f,t)dt - \frac{1}{\tau} \int_{t_0}^{t_0+\tau} X(f,t)dt.$$

Дотук бе описано функционирането на TMN и FMN блоковете. Те се свързват в йерархични структури с цел тяхното най-ефективно използване. Обобщена блокова схема на тези структури е показана на фиг. 3.



фиг. 3

Всеки MN блок генерира вектор измервания $m_{i,j}$. За всяка подобна структура е в сила противоречието между сложност и ефективност. При две структури има много добър постигнат компромис в това отношение. Те са наричани *„MNB структура 1 и 2"* съответно. Блоковата схема на втората от тях, даваща малко по-точни резултати, е показана на фиг. 6.



фиг. 4

Някои от получените вектори са линейно зависими, което налага отпадането на някои от тях и избирането на пълната комбинация от

линейно независими вектори. Такъв пълен комплект образуват например векторите с нечетни индекси в (2) и (3). Една линейна комбинация от тях е добър критерий при оценка на качеството на говорните сигнали и тя се нарича *аудиторно разстояние ("Auditory Distance" - AD):*

$$AD = w^T.m \quad (4)$$

Тук w е тегловен вектор. AD има неотрицателни стойности. Ако двата сигнала са еднакви, AD има нулева стойност. При увеличаване на възприеманата разлика между тях, AD расте. Стойностите на аудиторното разстояние AD са подходящи като критерий за качеството на речевите сигнали.

3. Числено описание на метода и симулационни резултати.

Алгоритъмът ще бъде описан последователно в няколко точки чрез т. нар. *„псевдокод"*.

А) Нека имаме векторите $\vec{X}$ и $\vec{Y}$, които представляват съответно входният и изходният сигнал за изследваната система, при услове, че те са синхронизирани, работи се с 8000 Hz дискретизираща честота, дължината на думата на един отчет е 16 бита и във всеки от векторите има поне една секунда продължителен сигнал. Нека броят отчети във всеки от тях е $N1$.

А. 1) Премахване на средната съставка от всеки от тях:

$$x(i) = x(i) - \frac{1}{N1}\sum_{j=1}^{N1} x(j),$$

$$y(i) = y(i) - \frac{1}{N1}\sum_{j=1}^{N1} y(j),$$

$$i = 1 \div N1.$$

А. 2) Нормализиране на векторите:

$$x(i) = x(i)[\frac{1}{N1}\sum_{j=1}^{N1}(x(j))^2]^{-\frac{1}{2}},$$

$$y(i) = y(i)[\frac{1}{N1}\sum_{j=1}^{N1}(y(j))^2]^{-\frac{1}{2}},$$

$$i = 1 \div N1.$$

Б) Трансформация в честотната област

Б. 1) Всеки от векторите $\vec{X}$ и $\vec{Y}$ се разделя на блокове с дължини по 128 отчета и коефициент на припокриване 50 %. След последния пълен блок отчетите до края на вектора се игнорират.

Б. 2) Получени са $N2$ броя пълни блокове с дължини 128. Всеки от тях се умножава (отчет по отчет) с прозореца на Хаминг с дължина 128:

$$h(i) = 0.54 - 0.46\cos(\frac{2\pi(i-1)}{127}),$$

$$i = 1 \div 128.$$

Б. 3) Така получените блокове се подават на FFT. От всеки от получените блокове се получават първите 65 стойности на амплитудата, взети на квадрат. Фазите не са от значение при оценката на качеството.

Образуват се две матрици $\boldsymbol{X}$ и $\boldsymbol{Y}$ с размери 65 x $N2$.

В) Избор на блокове

От матриците $\boldsymbol{X}$ и $\boldsymbol{Y}$ се избират само онези блокове (стълбове), които отговарят на определен енергиен критерий – енергията на стълбът-вектор да бъде не по-малка от следния енергиен праг:

$$xthreshold = 10^{-1,5}\max[xenergy(j)],$$

$$\text{където}: xenergy(j) = \sum_{i=1}^{65} x(i,j),$$

$$j = 1 \div N2,$$

тоест на практика прагът се избира 15 dB под нивото на енергията в стълба с максимална такава. Съответно за $\boldsymbol{Y}$:

$$ythreshold = 10^{-3,5}\max[yenergy(j)],$$

$$\text{където}: yenergy(j) = \sum_{i=1}^{65} y(i,j),$$

$$j = 1 \div N2,$$

като тук избраният праг е с 35 dB под максималната енергия.

Запазва се j-тият стълб, ако:

$$\{xenergy(j) \geq xthreshold\} \wedge \{yenergy(j) \geq ythreshold\},$$

и освен това даденият номер стълб да не съдържа нулеви стойности в никой от двата вектора. Новите матрици $\boldsymbol{X}$ и $\boldsymbol{Y}$, съдържат $N3$ стълба. Ако $N3=0$, записите в първоначалните вектори не са подбрани добре (напр. пауза и др.) и анализът се прекратява.

Г) Апроксимация на възприеманата сила на звука.

$$x(i,j) = 10\lg[x(i,j)],$$

$$y(i,j) = 10\lg[y(i,j)],$$

$$i = 1 \div 65, j = 1 \div N3.$$

Д) FMN блок в началото – работата му се описва по следния начин:

$$f1(i) = \frac{1}{N3}\sum_{j=1}^{N3} Y(i,j) - \frac{1}{N3}\sum_{j=1}^{N3} X(i,j) \rightarrow \textit{измерване},$$

$$Y(i,j) = Y(i,j) - f1(i) \rightarrow \textit{нормализиране},$$

$$f2(i) = f1(i) - f1(17) \rightarrow \textit{привеждане}$$

$$f3(i) = \frac{1}{4}\sum_{j=1}^{4} f2[1 + 4(i-1) + j] \rightarrow \textit{изравняване}, i = 1 \div 16$$

Тук посоченото привеждане е спрямо честота 1 kHz, и по-точно съответстващата й по скалата на Барк – член 17 на получената редица измервания.

Оттук се получават вече някои стойности на търсеният вектор – резултат от измерването:

$$[m(1)\ m(2)\ m(3)\ m(4)] = [f3(1)\ f3(2)\ f3(13)\ f3(14)].$$

Е) За структура 2 се дефинират векторите:

$$\boldsymbol{u}=[2\ 7\ 43\ 7\ 19\ 7\ 12\ 19\ 29]^T,$$

$$\boldsymbol{v}=[6\ 42\ 65\ 18\ 42\ 11\ 18\ 28\ 42]^T.$$

Това са индексите, по които ще се извършва сумирането при работа на останалите девет TMN блока, която се описва с помощта на следния псевдокод:

- за $k = 1$ до 9:

$$t0(j) = \frac{1}{v(k)-u(k)+1}\sum_{i=u(k)}^{v(k)} Y(i,j) - \frac{1}{v(k)-u(k)+1}\sum_{i=u(k)}^{v(k)} X(i,j),$$

$$Y(i,j) = Y(i,j) - t0(j); (i = u(k) \div v(k), j = 1 \div N3),$$

$$m0(k) = \frac{1}{N3}\sum_{j=1}^{N3} \max[t0(j),0],$$

- край на цикъла за k.

От вектора ***m0*** се избират следните стойности за изходния вектор ***m***:

$$[m(5)\ m(6)\ m(7)\ m(8)\ m(9)\ m(10)] =$$
$$[m0(1)\ m0(2) m0(3) m0(4) m0(5) m0(8)].$$

Ж) Измерване на остатъка:

$$t1(i,j) = Y(i,j) - X(i,j),$$

$$m(11) = \frac{1}{64.N3}\sum_{i=2}^{65}\sum_{j=1}^{N3} \max[t1(i,j),0]$$

З) Линейна комбинация на компонентите на вектора ***m***:

$$\boldsymbol{AD = w^T.m}$$

Тук тегловият вектор има следните компоненти (за структура 2):

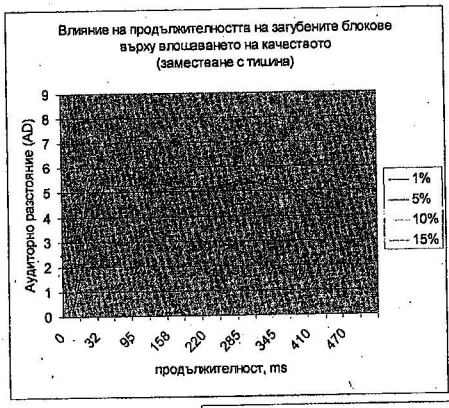
$$w = [0, -0.0837, -0.1199, 0.126, 0.166, 0.6387, 0.2195, 0.0122, 1.5544, 0.0954, 0.172]$$

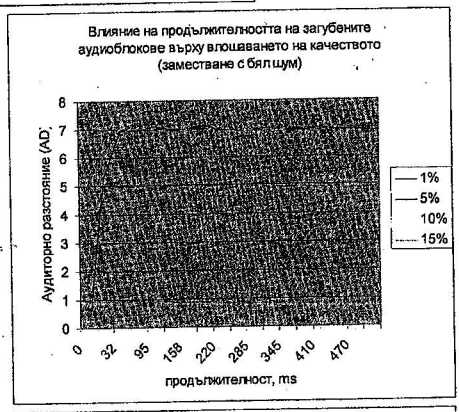
4. Симулационни резултати. Сравнение между резултатите за различните методи на възстановяване.

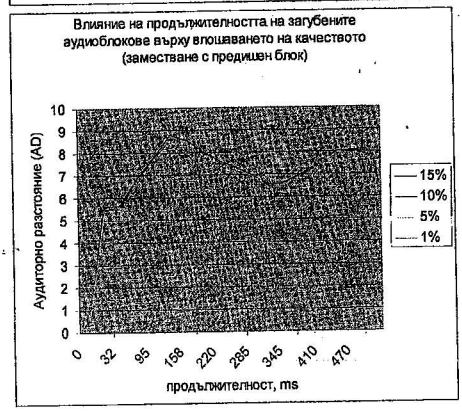
На трите графики са дадени резултатите от симулацията при методи за възстановяване съответно заместване с тишина, с бял шум и с предишни пакети. Като аргумент е избрана продължителността на пропадане в милисекунди, а като параметър – вероятността за загуба.

Методът на заместване с тишина има добра ефективност само при много малки вероятности за настъпване на загубите, както и при малки техни продължителности. За вероятности за загуба до около 1 %, дори по-продължителни загуби не предизвикват усещане за влошено качество, тъй като се случват доста рядко. Но при по-големи вероятности за пропадане този метод е най – неефективен от трите разглеждани.

Методът за заместване с бял шум води до по-добри резултати от първия за стойности на продължителността от 15 до около 100 ms. Дори над тази стойност влошаването на качеството е ограничено в сравнение с метода със заместване с мълчание. Това се дължи на психичните свойства на човешкия слух за подсъзнателно заместване на малки липсващи части от сигнала с правилни такива, което свойство се проявява при заместване на пропаданията с шум, но не и с тишина.







Най-ефективен е методът на заместване с предишни части от сигнала (предишни правилно приети пакети аудиоданни). Тук характерно е силно изразената периодичност на влошаването на качеството с увеличаване на

продължителността на загубените аудиоблокове, която се наблюдава при всички ненулеви стойности на вероятността за загуба. Това се обяснява със структурата на говорните сигнали, състоящи се от фонеми с продължителност от порядъка на (150 - 350) ms. Тези фонеми имат определена структура и характерна форма на автокорелационната функция на сигнала (и неговата обвиваща). Това означава, че отделни части на фонемата си приличат и ако се извърши заместване на една такава част с предходната, се получава най-висока ефективност на метода. При други дължини на заместваните (и заместващите) блокове структурата на фонемата се променя и говорът не звучи естествено, откъдето идва и ниската оценка на качеството при някои по-кратки пропадания в сравнение с малко по-продължителни.

Общо взето, може да се каже че за решаването на практическите проблеми най-ефективен е методът на заместване с предишни извадки от аудиосигнала. Той изисква доста ниска изчислителна сложност и има лесна реализуемост. Това го прави широко използван в редица приложения за предаване на звукова информация през Интернет.

5. Изводи.

Резултатите от изследванията показват, че за създаването на ефективен метод за компенсиране на пакетните загуби при предаването на звукова информация през пакетно-комутируемите мрежи, е необходима комплексна оценка въз основа на много фактори, свързани с особеностите на човешкия слухов и говорен апарат. Максимална ефективност на създадените методи при сравнително ниска изчислителна мощност се постига при отчитането на най-важните фактори, влияещи върху човешката оценка на качеството. От комуникационна гледна точка това представлява съгласуване на комуникационната система с източника на звукова информация от една страна, и с получателя (консуматора) й от друга. Това е съобразено изцяло с поставената цел – постигане на възможно по-високо качество на предоставяните комуникационни услуги в реално време в Интернет.

Използвана литература:

1. Leinonen J., Real-Time Voice over Packet-Switched Networks, Seminar on Real-time And Embedded Systems, Helsinki, Nov. 1999

2. Perkins C., Orion Hodson, Vicky Hardman, A Survey of Packet-Loss Recovery Techniques for Streaming Audio, London, University College, Aug. 1998.

3. Voran S., Objective Estimation Of Perceive Speech Quality Using Measuring Normalizing Blocks, US Department Of Commerce, April 1998

4. *http://www.all-nettools.com*

5. *http://www.packeteer.com/solutions/resources/udp.cfm*