

T 180 TELEKOMUNIKACIJŲ INŽINERIJA

Balso signalų kalbos ir tylos intervalų modeliai ir jų taikymas

A. Kajackas, A. Šaltis

Vilniaus Gedimino technikos universitetas

Aušros Vartų g. 7a, LT-2600 Vilnius, Lietuva, tel. +370 5 2626586, el. paštas kajackas@el.vtu.lt, arunas.saltis@el.vtu.lt

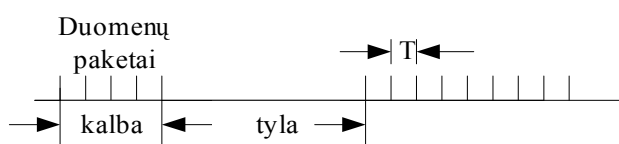
Įvadas

Didelę šiolaikinių telekomunikacinių sistemų laidumo dalį sudaro balso signalų perdavimai. Todėl sistemingai ieškoma būdų, kaip sumažinti balsui perduoti reikalingų duomenų kiekį. Nustatyta, kad kalbant pakaitomis kartojasi trumpi kalbos ir tylos intervalai. Tokią kalbos signalo struktūrą lemia žmogaus kalbos aparato savybės. Žmogus paprastai kalba iškvėpdamas orą ir tyli įkvėpdamas. Telefonu pašnekovai paprastai kalba pakaitomis - vienas kalba, kitas klauso. Tai kita priežastis, kodėl susidaro tylos intervalai. Visų – kalbos ir tylos - intervalų trukmės, be abejo, yra atsitiktinės.

Gerai žinomi balso kodavimo - kompresijos algoritmai, trumpai nusakomi ITU rekomendacijų kodais G.729, G.723.1, bei kiti, kuriuose kartu su kitomis skaitmeniniu apdorojimu pagrįstomis priemonėmis duomenų kiekiui sumažinti taikoma ir balso specifiška pagrįsta priemonė – balso aktyvumo detektoriai, kurie išskiria kalbos (*angl. talkspurt*) laikotarpius, kai žmogus tikrai kalba, ir tylos (*angl. silence*) laikotarpius, kai kalbantysis net trumpam laikotarpiui nutyla.

Balso kompresija balso aktyvumo detektoriais telefono ryšiui pirmą kartą panaudota TASI (*angl. Time Assignment Speech Interpolation*) sistemoje (TASI-A sukurta 1959 m., TASI-B – 1976 m.). Ši balso perdavimo sistema buvo sukurta specialiai tolimajam ryšiui ir taikyta perduodant telefono pokalbių signalus tarpkontinentiniais kabeliais.

Balso aktyvumo detektoriai taikomi ir moderniose paketinio perdavimo priemonėse (1 pav.). Neperduodant balso duomenų tylos laikotarpiais lieka laisva ryšio kanalo laidumo dalis. Šią laidumo dalį galima panaudoti kitų pokalbių signalams bei su balso signalais nesusijusiems duomenims perduoti. Mobiliojo ryšio sistemoje tylos laikotarpiais galima išjungti siūstuvus. Taip taupoma energija ir mažinami trukdžiai kitoms ryšio priemonėms.



1 pav. Koduoto kalbos signalo struktūra

Tobulinant esamas ir kuriant naujas telekomunikacines sistemas, svarbu turėti pakankamai tikslų balso sukuriamo duomenų srauto modelį. Nors pirmųjų balso aktyvumo – tylos intervalų tyrimų rezultatai paskelbti jau 1965-68 m. [1, 2], tačiau šios krypties balso signalų tyrimai tęsiasi. Viena iš priežasčių yra tai, kad sukurtieji modeliai gerokai skiriasi vieni nuo kitų ir priklausomai nuo naudojimo sričių skirtingai interpretuojami gautini rezultatai.

Šiame straipsnyje nagrinėjami balso signalų kalbos ir tylos intervalų modeliai. Aptariamos žinomų modelių savybės, eksperimentiškai tiriami realūs kalbos ir tylos procesai. Tyrimų pagrindu daromi teoriniai apibendrinimai apie sudėtingesnių modelių tikslingumą. Aptariama balso aktyvumo detektorių įtaka vartotojų tinklų charakteristikoms.

Balso aktyvumo tyrimų analizinė apžvalga

Bene pirmieji balso signalo struktūros tyrimai aprašyti darbe [2]. Nustatyta kalbos intervalų vidutinė trukmė $\tau_{k0} = 352$ ms, o tylos intervalų vidutinė trukmė - $\tau_{t0} = 650$ ms. Gerokai vėlesniame darbe [4] gauti visiškai kitokie duomenys: $\tau_{k0} = 2360$ ms, $\tau_{t0} = 606$ ms. Tokio didelio tyrimų rezultatų skirtumo priežastis – skirtingos matavimų metodikos bei skirtingi balso aktyvumo detektoriai, kurie stebimą intervalą priskiria kalbai ar tylai. Taikant paprastus balso aktyvumo detektorius, priskyrimo kalbai ar tylai kriterijus yra artimas energiniam kriterijui. Taip gaunami rezultatai irgi artimi aprašytiems [2]. Pavyzdžiui, ITU rekomendacija P.59 [5] teikia $\tau_{k0} = 227$ ms ir $\tau_{t0} = 596$ ms. Energinis balso aktyvumo detektorių trūkumas tas, kad jie kartais dirbtinai iškarpo kalbos intervalus ir tuo blogina atkurto balso kokybę. Šie detektoriai blogai veikia, kai kalbama triukšmo aplinkoje.

Kita rezultatų grupė, kai kalbos intervalai žymiai ilgesni už tylos intervalus, gaunami taikant gana sudėtingus balso aktyvumo detektorius, kurie kalbai nuo tylos atskirti naudoja inertiškus, nuo signalo atkarpos priklausiančius algoritmus. Sprendimai „kalba“ ar „tyla“ priimami apdorojus tam tikros trukmės signalo atkarpą prieš stebimą laikotarpį. Šis faktorius apibūdinamas kaip „paveldas“ HF (*angl. hangover*). Paprasčiausiai HF realizuojamas tada, kai trumpi tylos intervalai (kurių trukmė trumpesnė už h) neišskiriami [4]. Pavyzdžiui, balso kodavimo protokole G.729B HF trukmė yra 200 ms.

Balso aktyvumo detektoriai su HF funkcija neiškiria labai trumpų tylos intervalų, nes trumpesnius už h priskiria kalbai. Tais laikotarpiais, kai tylos intervalo trukmė ilgesnė už h , HF funkcija kiekvieną kalbos intervalą pailgina dydžiu h .

Šiame straipsnyje su nuoroda į darbą [4] pateikti duomenys gauti naudojant balso aktyvumo detektorius su HF funkcija. Darbe [3] gauti irgi panašūs duomenys: $\tau_{k0} = 1200$ ms ir $\tau_{i0} = 1800$ ms. Tokiomis sąlygomis, kai kalbos bei tylos intervalai nustatomi taikant HF, toje pat ITU rekomendacijoje P.59 [5] nurodomos kalbos vidutinė trukmė $\tau_{k0} = 1004$ ms ir tylos, $\tau_{i0} = 1587$ ms.

Susisteminti kalbos ir tylos intervalų vidutinių trukmių duomenys pateikti 1 lentelėje. Pirmųjų dviejų skilčių duomenys gauti naudojant paprastus – energinius balso aktyvumo detektorius, o kitų – naudojant skirtingus balso aktyvumo detektorius, bet su HF funkcija.

1 lentelė. Kalbos ir tylos intervalų vidutinių trukmių duomenys

Šaltinis	[5]	[2]	[5]	[3]	[4]
τ_{i0} [ms]	596	650	1587	1800	606
τ_{k0} [ms]	227	352	1004	1200	2360
Periodas [ms]	823	1002	2591	3000	2966
q	0,2758	0,3513	0,3875	0,4000	0,7957

Lentelėje taip pat pateiktas *balso aktyvumo rodiklis* q , išreiškiamas kalbos intervalo vidutinės trukmės ir vidutinės periodo trukmės santykiu:

$$q = \frac{\tau_{k0}}{\tau_{i0} + \tau_{k0}}. \quad (1)$$

Kaip matyti, priklausomai nuo tyrimų priemonių bei metodikos vidutinė kalbos ir tylos periodo trukmė ($\tau_{i0} + \tau_{k0}$) kinta nuo 823 ms iki 3000 ms. Balso aktyvumo rodiklis kinta nuo 0,2758 iki 0,7957, tai yra maždaug tris kartus.

Žemesnioji balso aktyvumo rodiklio riba yra sąlygota žmogaus kalbos proceso ypatumu. Deja, šią ribą palaikantys balso aktyvumo detektoriai nėra tobuli, perduoto balso kokybė nėra pakankamai aukšta. Tobulinant detektorius irgi palyginti paprastu būdu (naudojant HF), sumažėja išskiriamų tylos intervalų trukmė, gerokai padidinamas balso aktyvumo rodiklis (jis artėja prie vieneto).

Paprasčiausias ir dažniausiai taikomas balso signalo struktūros modelis pagrįstas prielaida, kad kalbos ir tylos intervalų trukmių tikimybių skirstiniai yra eksponentiniai:

$$w_k(\tau) = \frac{1}{\tau_{k0}} e^{-\left(\frac{\tau}{\tau_{k0}}\right)}, \quad (2)$$

$$w_i(\tau) = \frac{1}{\tau_{i0}} e^{-\left(\frac{\tau}{\tau_{i0}}\right)}. \quad (3)$$

Čia τ_{k0} - vidutinė kalbos intervalo trukmė, o τ_{i0} - vidutinė tylos intervalo trukmė.

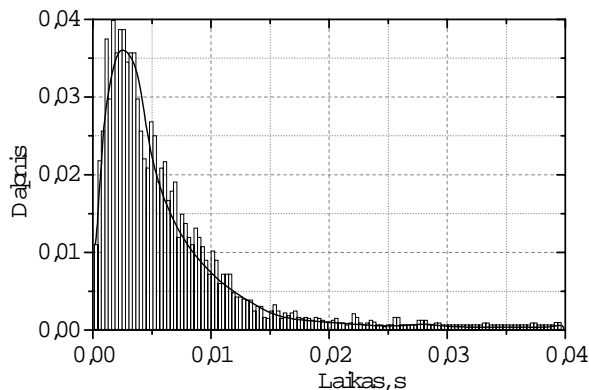
Šio modelio pagrindu kalbos ir tylos proceso kaita imituojama Markovo grandine. Markovo modelio pagrindu darbe [8] parodyta, kad esant n balso šaltinių (n kalbančiųjų) tikimybė, jog vienu metu susigrupuos k kalbos intervalų, išreiškiama binomo formule

$$p_k = \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k}, \quad (4)$$

kurioje q - balso aktyvumo rodiklis (1).

Dar balso signalų tyrimų pradininko P. T. Brady darbe [2] pažymėta, kad kalbos ir tylos intervalų tikimybių pasiskirstymo funkcijų aproksimacija eksponentėmis (2) ir (3) nėra tiksli. Apie eksponentinio modelio netikslumus nurodoma ir kituose darbuose, pavyzdžiui [4] ir [5]. Eksperimentinių duomenų pagrindu patikslinti balso kalbos ir tylos trukmių pasiskirstymo modeliai aprašyti darbuose [6], [7]. Naudojami modeliai: dviejų eksponenčių suma [7] ir logaritminių-normalinių kreivių suma [6]. Mes pabandėme tuos modelius palyginti tarpusavyje. Deja, skirtingų autorių patikslinti modeliai gerokai skiriasi. [6] darbe pateiktos matematinės išraiškos bei nurodyti jų parametrai neatkuria straipsnyje pateiktą grafiką.

Anksčiau pateikti duomenys ir nuorodos apie kalbos ir tylos intervalų statistiką yra gauti analizuojant monologus. Telefonu bendrauja du pašnekovai. Dialogo atveju empirinė tylos intervalų pasiskirstymo funkcija pateikta 2 pav. [6]. Kaip matyti, šiuo atveju tylos intervalų trukmė gali siekti 50 - 100 s ir dar daugiau. Todėl tylos intervalų modelis telefonijos tikslams turėtų būti sudaromas iš dviejų dalių – mikrointervalų, ir makrointervalų.

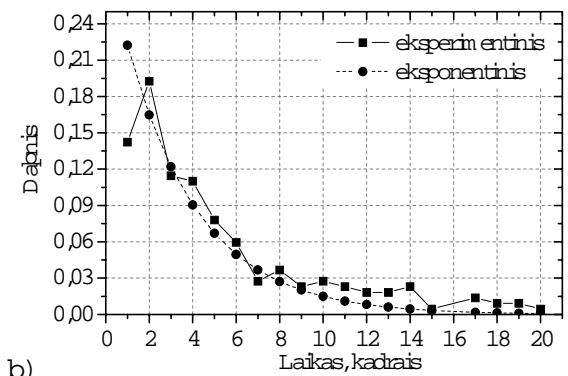
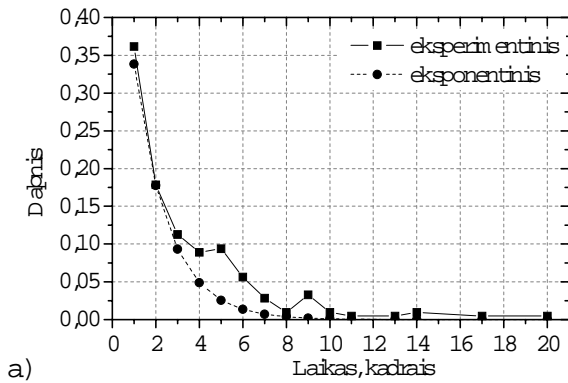


2 pav. Tylos intervalų pasiskirstymo funkcija dialogų metu

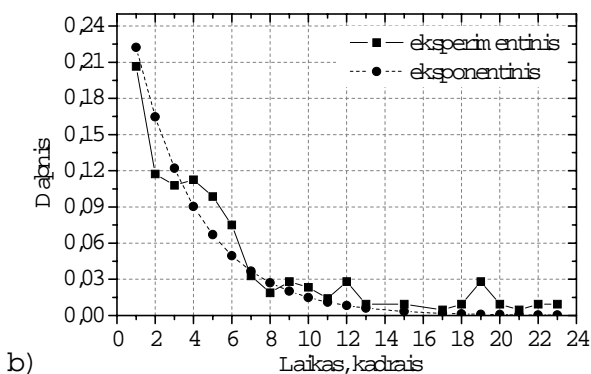
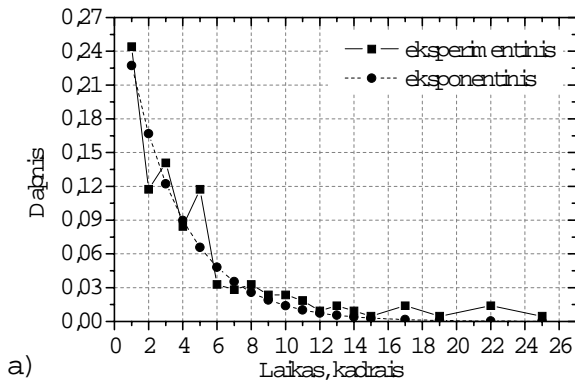
Bendra išvada. Kol kas dar nėra bendrų balso aktyvumą apibūdinančių modelių. Duomenų įvairovė gali būti aiškinama įvairiai: skirtingi kalbos ir tylos atskyrimo kriterijai, skirtingi kalbantieji, skirtinga jų emocinė būseną. Geresnių rezultatų galima tikėtis tik išsamios specifinių savybių analizės būdu. Tokios analizės pavyzdys pateiktas darbe [9].

Eksperimentiniai tyrimai

Atlikome eksperimentinius kalbos ir tylos procesų tyrimus. Kaip pagrindinis balso aktyvumo detektorius panaudotas 3GPP AMR algoritmas [10], kadro dydis 20 ms.



3 pav. Kalbos intervalų pasiskirstymai

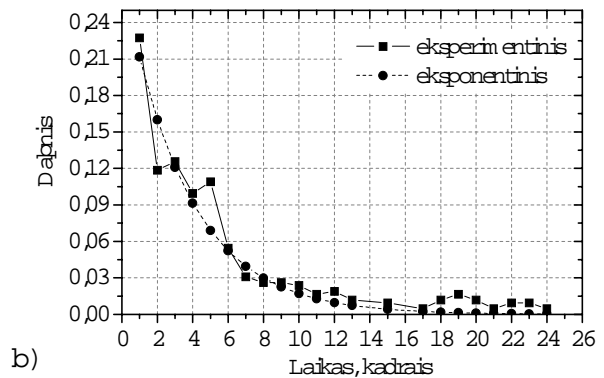
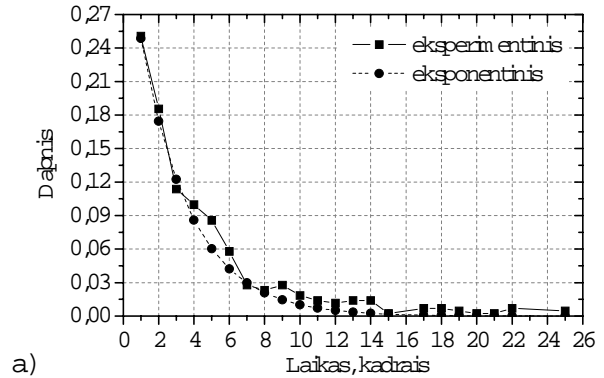


4 pav. Tylos intervalų pasiskirstymai

Tyrimus atlikome neilgų (apie 200 kalbos ir tylos periodų) monologų pagrindu. 3 pav., a ir b, pateikti skirtingų diktorių kalbos intervalų pasiskirstymai, o 4 pav., a ir b - tų pačių diktorių tylos intervalų pasiskirstymai.

Tuose pat paveiksluose pateikti aproksimacijų (2) ir (3) išraiškėmis grafikai. Grafikai sudaryti apdorojant vieno monologo 200 kalbos ir tylos periodų trukmės įrašus.

Tų pačių diktorių ir tų pačių monologų kalbos ir tylos intervalų, kaip vieno duomenų masyvo, pasiskirstymai pateikti 5 pav., a ir b. Padidėjus duomenų kiekiui abu pasiskirstymai tiksliau aproksimuojami eksponentiniais dėsniais.



5 pav. Dviejų diktorių kalbos intervalų pasiskirstymas (a) ir tų pačių diktorių tylos intervalų pasiskirstymas (b)

Sudarant balso signalų kalbos ir tylos modelius, atlikto tyrimo sąlygomis aproksimacija eksponentiniais pasiskirstymo dėsniais visiškai priimtina.

Atliktų eksperimentų metu gauti kalbos ir tylos intervalų trukmių vidurkiai gerokai skiriasi nuo pateiktų 1 lentelėje. Kiekvieno diktoriaus kalbos ir tylos trukmių vidurkiai skirtingi.

Modelio normalizacija

Kalbos ir tylos intervalai išsidėsto pakaitomis: kalba, tyla, kalba, tyla... Po kalbos intervalo eina tylos intervalas. Visų kalbos ir tylos intervalų trukmės atsitiktinės.

Pažymėkime j -ojo balso šaltinio kalbos intervalo pradžios laiko žymę t_0^j . Šio kalbos intervalo pabaigą pažymėkime t_1^j . Atitinkamai pažymėję kitų būvio pokyčių laikus $t_2^j, t_3^j, \dots, t_i^j$, gausime laiko žymių seką

$$T^j = t_0^j, t_1^j, t_2^j, \dots, t_i^j, \dots, t_0^j < t_1^j < t_2^j < \dots < t_i^j < \dots, \quad (5)$$

kurioje lyginiais skaičiais pažymėtos kalbos intervalų pradžios, o nelyginiais – kalbos intervalų pabaigos ir tylos intervalų pradžios.

Toliau logiška nagrinėti kalba ir tylos poras. Pirmą porą visiškai apibūdina žymės t_0^j, t_1^j, t_2^j . Šios poros kalbos intervalo trukmė

$$\kappa_0^j = t_1^j - t_0^j,$$

o tylos –

$$\theta_0^j = t_2^j - t_1^j.$$

Apibendrinant j -ojo balso šaltinio k -osios poros kalbos intervalo trukmė išreiškiama

$$\kappa_k^j = t_{2k+1}^j - t_{2k}^j, k = 0, 1, 2, 3, \dots, \quad (6)$$

o k -osios poros tylos intervalo trukmė –

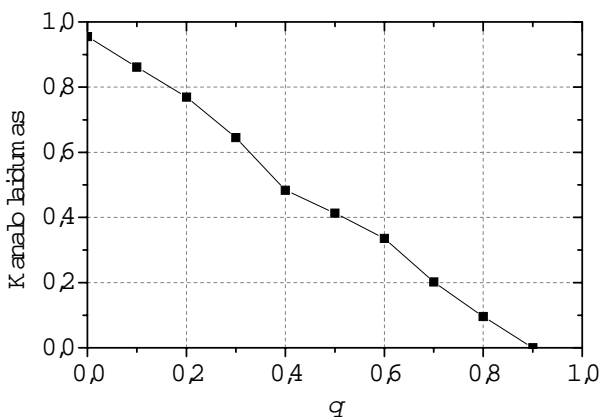
$$\theta_k^j = t_{2k+2}^j - t_{2k+1}^j, k = 0, 1, 2, 3, \dots \quad (7)$$

Kalbos ir tylos intervalų pora $\{\kappa_k^j, \theta_k^j\}$ - dvimatis atsitiktinis dydis, kurio statistinės savybės sietinos su perdavimo sistemos kokybės charakteristikomis. Per laikotarpį $\kappa_k^j + \theta_k^j$ sukuriama balso duomenų srautas proporcingas κ_k^j , o vidutinė per šį laikotarpį kalbos duomenų sparta yra proporcinga momentiniam balso aktyvumo rodikliui

$$q_k^j = \frac{\kappa_k^j}{\theta_k^j + \kappa_k^j}. \quad (8)$$

Užimta ryšio kanalo dalis proporcinga dydžiui q_k^j .

6 pav. pateiktas eksperimentinis dydžio q_k^j pasiskirstymo grafikas. Su tikimybe artima vienetui, užimta tik maža ($< 0,1$) kanalo dalis. Su tikimybe 0,4 užimta tik pusė kanalo.



6 pav. Užimtos ryšio kanalo dalies pasiskirstymas

Laiko sekų T^j sudarymas bei formavimas priklauso nuo balso kodavimo įrangos (algoritmo). Visi balso kompresiją atliekantys algoritmai siejami su tam tikra balso apdorojimo intervalo trukme. Tipiškos apdorojimo trukmės $T = 5, 10, 20$ ar 30 ms. Šios trukmės - tai laiko skalės žingsnis. Kalbos intervalo laikotarpiu kiekvieno žingsnio metu sukuriama vienas duomenų paketas (1 pav.).

Koreliacija

Kalbos ir tylos intervalų trukmių $\{\kappa_k^j, \theta_k^j\}$ modeliai sudaryti laikant, kad šie dydžiai yra nepriklausomi. Suformulavus hipotezę, kad kalbos ir tylos intervalai gali būti statistiškai priklausomi dydžiai, buvo atlikti balso signalų eksperimentiniai tyrimai.

Eksperimentų duomenų analizė leidžia teigti, kad koreliacija tarp kalbos ir tylos gretimų intervalų yra. Tačiau ji nėra stipri, dažniausiai pasitaikančios koreliacijos koeficiento vertės yra nuo $-0,05$ iki $-0,2$.

Kalbos ir tylos pasiskirstymo dėsnis

Patikslintas balso modelis gaunamas sukuriant statistinį ryšį tarp atsitiktinių dydžių κ_k^j ir θ_k^j . Tarkime, kad k -ojo tylos intervalo trukmė θ_k yra susijusi su prieš jį buvusio k -ojo kalbos intervalo trukme κ_k tokia išraiška:

$$\theta_i - \tau_{i0} = (\zeta_i - \tau_{i0}) + r(\kappa_i - \tau_{k0}). \quad (9)$$

Čia r - koreliacijos koeficientas.

(9) išraiška lengvai pertvarkoma į tokio pavidalo lygtį:

$$\theta_i = \zeta_i + r(\kappa_i - \tau_{k0}). \quad (10)$$

Bendro kalbos ir tylos intervalo trukmė

$$\theta_k + \kappa_k = \zeta_k + (1+r)\kappa_k - r\tau_{k0}. \quad (11)$$

Dydžiai ζ_i ir κ_i čia yra nepriklausomi.

Dviejų atsitiktinių dydžių sumos tikimybių tankio funkcija $W(y)$, kaip žinoma, išreiškiama atskirų dydžių dvimate funkcija $w_2(x, y)$:

$$W(y) = \int_{-\infty}^{\infty} w_2(u, y-u) du. \quad (12)$$

Norint šią formulę pritaikyti (11) išraiška apibūdintam dydžiui, tenka atlikti tam tikrą korekciją - atmesti vidurkio pataisą „- $r\tau_{k0}$ “. Modifikuota tylos ir kalbos ciklo išraiška atrodys taip:

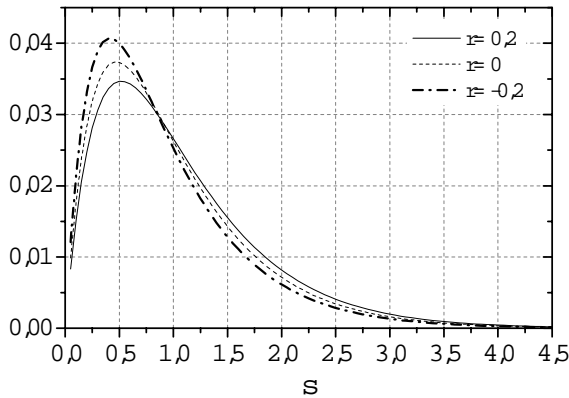
$$\chi_k = \zeta_k + (1+r)\kappa_k. \quad (13)$$

Laikant, kad dydžių ζ ir κ tikimybių tankio funkcijos yra atitinkamai (2) ir (3) (kalbos intervalo trukmės vidurkis yra $\tau_{k0} = 1/\mu$, o tylos ζ vidurkis yra

$\tau_{i0}=1/\lambda$) atsitiktinio dydžio χ tikimybių tankio funkcija išreiškiama šia formule:

$$\begin{aligned}
 W_c(t) &= \int_0^t \lambda \exp\{-\lambda u\} \frac{\mu}{1+r} \exp\left\{-\frac{\mu}{1+r}(t-u)\right\} du = \\
 &= \frac{\lambda\mu}{1+r} \exp\left\{-\frac{\mu}{1+r}t\right\} \int_0^t \exp\{-\lambda u + \frac{\mu}{1+r}u\} du = \\
 &= \frac{\lambda\mu}{1+r} \exp\left\{-\frac{\mu}{1+r}t\right\} \frac{1}{\lambda - \mu/(1+r)} \times \\
 &\times \{1 - \exp\{-\lambda t + \frac{\mu}{1+r}t\}\} = \\
 &= \frac{\lambda\mu}{(1+r)\lambda - \mu} \left\{ \exp\left(-\frac{\mu}{1+r}t\right) - \exp(-\lambda t) \right\}.
 \end{aligned} \tag{14}$$

Šios funkcijos grafikai, kai $\tau_{i0} = 650$ ms ir $\tau_{k0} = 352$ ms, pateikti 7 pav. Kaip matyti, jie primena ilgalaikių tylos intervalų pasiskirstymo grafikus (2 pav.). Standartinė deviacija: $\sigma = 1,5069$ s, kai $r = 0$, $\sigma = 1,643$ s, kai $r = 0,2$, $\sigma = 1,3896$ s, kai $r = -0,2$. Neigiama kalbos ir tylos intervalų tarpusavio koreliacija mažina periodo trukmės sklaidą.



7 pav. Kalbos ir tylos periodų pasiskirstymo funkcija. $\sigma = 1,5069$ s, kai $r = 0$; $\sigma = 1,643$ s, kai $r = 0,2$; $\sigma = 1,3896$ s, kai $r = -0,2$

Dydžio χ vidurkis

$$\begin{aligned}
 M\{\chi\} &= \int_0^{\infty} t W_c(t) dt = \\
 &= \frac{\lambda\mu}{(1+r)\lambda - \mu} \left\{ \frac{(1+r)^2}{\mu^2} - \frac{1}{\lambda^2} \right\} = \\
 &= \frac{1}{\lambda\mu} [(1+r)\lambda + \mu]
 \end{aligned} \tag{15}$$

Tylos ir kalbos intervalų vidutinės trukmės, kaip minėta, nagrinėtos daugelyje šaltinių ir šiame darbe pažymėtos atitinkamai $\tau_{i0} = 1/\lambda$ ir $\tau_{k0} = 1/\mu$. Atsitiktinio dydžio χ vidurkis išreiškiamas naudojantis τ_{i0} ir τ_{k0} ir lygus

$$M\{\chi\} = (1+r)\tau_{k0} + \tau_{i0}. \tag{16}$$

Įvertinus (13) formulės korekciją „- $r\tau_{k0}$ “ kalbos ir tylos periodo vidutinė trukmė yra

$$M\{\Theta + \kappa\} = M\{\chi\} - r\tau_{k0} = \tau_{k0} + \tau_{i0} \tag{17}$$

Tai akivaizdus rezultatas, nes atsitiktinių dydžių koreliacija vidurkio nekeičia.

Išvados

Išnagrinėjus balso aktyvumo intervalų modelius, atlikus eksperimentinius tyrimus, nustatyta, kad prielaidos apie eksponentinius kalbos bei tylos intervalų trukmių skirstinius yra gana pagrįstos. Tai svarbu tais atvejais, kai vartotojų gali būti nedaug.

Nustatyta, kad gretimos kalbos ir tylos intervalų trukmės yra koreliuotos. Be to, dažniausiai koreliacijos koeficientas yra neigiamas, dėl to sumažėja kalbos ir tylos periodo dispersija.

Taikant paketinio perdavimo technologijas ir išmetant balso tylos intervalus, išlaisvinama didelė ryšio kanalo laidumo dalis, kuri gali būti panaudota kitiems duomenims perduoti.

Literatūra

1. **Brady P. T.** A technique for investigating on-off patterns of speech // Bell System Technical Journal, 1965.- Vol. 44(1).- P.1-22.
2. **Brady P. T.** A statistical analysis of on-off patterns in 16 conversations // Bell System Technical Journal, 1968.- Vol. 47(1).- P. 73-91.
3. **Brady P. T.** A model for generating on-off speech patterns in two way conversation // Bell System Technical Journal, 1969.- Vol. 48(9).- P. 2445-2472.
4. **Gruber J. G.** A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection // IEEE Trans. Commun, 1982. - Vol. COM-30(4).- P. 728-738.
5. Telephone transmission quality objective measuring apparatus: Artificial conversational speech // ITU Recommendation.- Geneva, 1993.- P.59
6. **Barceló F.** Statistical properties of silence gap in public mobile telephony channels with application to data transmission // Proc. IEEE ICC, 2001.- P. 2011-2015.
7. **Lee H. H., Un C. K.** A Study of On-Off Characteristics of Conversational Speech // IEEE Trans. Communications, 1986.- Vol. COM-34.- P. 630-637.
8. **Weinstein C. J.** Fractional speech loss and talker activity model for TASI and for packet-switched speech // IEEE Trans. Commun., 1978.- Vol. COM-26.- P. 1253-1257.
9. **Rudzionis A., Rudzionis V.** Noisy speech detection and endpointing // Voice operated telecom services. Do they have a bright future?, Workshop Proceedings, May 11-12, 2000, Ghent, Belgium.- P. 79 – 82.
10. 3rd Generation Partnership Project; Adaptive Multi-Rate (AMR) speech codec; Voice Activity Detector (VAD) (Release 5) http://www.3gpp.org/ftp/Specs/archive/26_series/26.094/26094-500.zip

Pateikta spaudai 2003 03 14

A. Kajackas, A. Šaltis. Balso signalų kalbos ir tylos intervalų modeliai ir jų taikymas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2003. – Nr. 5(47). -P. 7-12.

Nagrinėjami balso signalų kalbos ir tylos intervalų modeliai. Tokie modeliai yra svarbūs kuriant modernias radijo priėmų sistemas. Išnagrinėjus žinomus modelius, pastebėta, kad kalbos ir tylos intervalų trukmės laikomos nepriklausomomis, tų trukmių skirstiniai dažnai aproksimuojami eksponentėmis, o jų parametrai labai priklauso nuo balso aktyvumo detektorių ir įvairiuose šaltiniuose yra gana skirtingi. Todėl darytina išvada, kad šiuo metu nėra visuotinai priimtų balso aktyvumą apibūdinančių modelių.

Atlikus eksperimentinius tyrimus nustatyta, kad, nagrinėjant mažas kalbančiųjų grupes, prielaidos apie eksponentinius kalbos bei tylos intervalų trukmių skirstinius yra gana pagrįstos. Eksperimentiniais tyrimais taip pat nustatyta, kad gretimų kalbos ir tylos intervalų trukmės yra koreliuotos, nors koreliacija nėra stipri. Šiuo pagrindu daromi teoriniai apibendrinimai apie sudėtingesnių balso aktyvumo detektorių pagrįstų balso specifinių savybių analize, kūrimo tikslingumą.

Tylos intervalų panaikinimas gerai derinasi su paketinėmis radijo ryšio technologijomis, tačiau WLAN tinkluose, kur vienu metu kalbančių žmonių nėra daug, negalima tikėtis šiuo būdu gerokai padidinti vienu metu perduodamų balso signalų kiekį. Kokybės lygio reikalavimai sąlygoja, kad išmetus tylos intervalus išlaisvintą ryšio kanalo laidumo dalį reikia skirti kitiems duomenims perduoti. Il. 7, bibl. 10 (lietuvių kalba; santraukos lietuvių, anglų rusų k.).

A. Kajackas, A. Šaltis. The Models of Voice Signals’ „Voice” and „Silence” Intervals and their Use // Electronics and Electrical Engineering. – Kaunas: Technologija, 2003. – No.5(47).-P. 7-12.

There is given an analysis on the models of voice signals’ “voice-silence” intervals. These models are important for creating the systems of modern radio systems. After examining the well-known models, it was discovered that the durations of “voice-silence” intervals are interpreted as independent, the distribution of these durations are often approximated by exponential and that the parameters of these durations depend a lot on the voice activity detectors. These parameters at different sources differ from each other very much. Based on the mentioned above, the conclusion was made that at this moment there are no characteristic models of voice activity that could be unanimously confirmed.

The results of this examination confirm the presumptions regarding the duration distribution of exponential voice and silence intervals, when the small talking groups are involved. Also it can be contended that the durations of near-by voice-silence intervals are correlated, though this correlation is not very considerable. The theoretical generalizations regarding more complex voice activity detections methods, which can be substantiated by analysis on specific voice features, are based on these grounds.

The elimination of silence intervals goes well with technologies of packet radio connection. However, it is not recommended to expect that this method would markedly increase the amount of at the same time transmitted voice signals in WLAN networks, where amount of speaking voices at the same time is not numerous. The requirements of quality level stipulate by allowing the channel released after excluding the intervals of silence, to the transmission of other data. Ill. 7, bibl. 10 (in English; summaries in Lithuanian, English, Russian).

А Каяцкас, А Шальтис. О моделях интервалов „речь – пауза“ сигналов голоса // Электроника и электротехника. – Каунас: Технология, 2003. – № 5(47).-С. 7-12.

Рассматриваются модели интервалов „речь – пауза“ в сигналах голоса. Такие модели имеют существенное значение при разработке современных систем радиодоступа. Установлено, что в известных моделях длительности интервалов „речь – пауза“ считаются независимыми величинами. Законы распределения этих величин часто аппроксимируют экспоненциальными распределениями. Параметры распределений в существенной степени зависят от примененных детекторов активности речи и в разных источниках они значительно различаются. Поэтому напрашивается вывод, что в настоящее время нет общепринятых моделей активности речевого сигнала.

На основе нами проведенных экспериментов можно утверждать, что при анализе небольших групп абонентов предположение об экспоненциальных распределениях длительностей интервалов „речь – пауза“ вполне оправдано. Экспериментально установлено, что длительности соседних интервалов „речь – пауза“ коррелированы, хотя коэффициент корреляции не велик. Ил. 7, библи. 10 (на литовском языке, рефераты на литовском, английском и русском яз.).