

Kalbos signalų apdorojimas

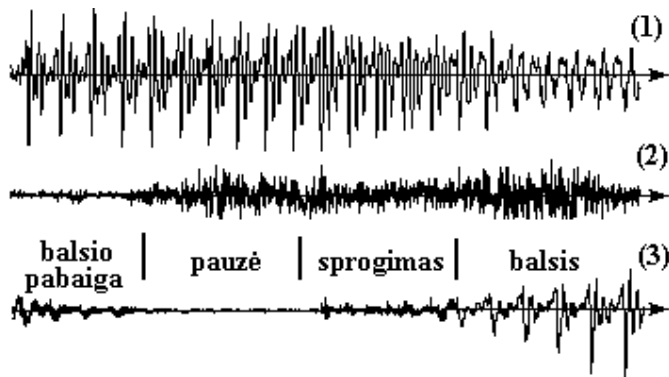
Kai kurios kalbos signalų savybės

1 pav. pateikta keletas kalbos signalų pavyzdžių. Net nežinodami jų kilmės galime pastebėti keletą pagrindinių kalbos signalų charakteristikų:

1. Kalbos signalas kinta laike. Kalbos signalo kitimo būdas ir yra ta informacija, kurią žmonės naudoja komunikavimui. Lyginant su rašytine forma perduotu pranešimu, sakytiniame pranešime yra labai daug informacijos.

2. Atskiruose segmentuose kalbos signalas yra kvaziperiodinis. Kai kuriuose segmentuose jis panašus į triukšmą. Be to, signale yra pauzių, kurios yra daug trumpesnės už tas, kurias kalbėtojas daro tyčia.

3. Lyginat su rašytine forma, kuri yra simbolių grandinė, kalbos signalas yra vientisas, kitimai signale yra laipsniški ir tolydūs (su kai kuriomis išimtimis), o ne staigūs.



1 pav. Kalbos signalų pavyzdžiai

Kalbos signalų sukūrimas

Skyriuje apie fonetiką buvo pateikta žmogaus kalbos organų schema (n pav.). Kalbos signalo formavimą galima išskaidyti į du etapus: sužadinimas ir formos suteikimas.

Sužadinimas galimas trimis būdais:

1. Vokalizuotas sužadinimas. Jis sukuriamas balso stygomis. Taip sukurtas signalas pasižymi periodiškumu ir jis charakterizuojamas pagrindinio tono aukščiu.

2. Nevokalizuotas sužadinimas. Sukuriamas veržiantis oro srovei pro koki nors apribojimą ir taip gaunamas triukšmas.

3. Perėjimo sužadinimas. Oras suspaudžiamas uždariant oro srautą, o po to staiga atidaromas.

Kai nėra jokio sužadinimo, gauname pauzę.

Sužadinimo signale yra nepakankamai informacijos, kad jį būtų galima naudoti kaip šnekamosios kalbos informacijos nešėją. Signalui formą suteikia kalbos traktas, kuriame svarbiausią vaidmenį vaidina liežuvio ir lūpų padėtis, o taip pat liežuvėlis, kuris laiko uždarytą arba atidarytą nosies ertmę.

Signalų vaizdavimas kompiuteryje

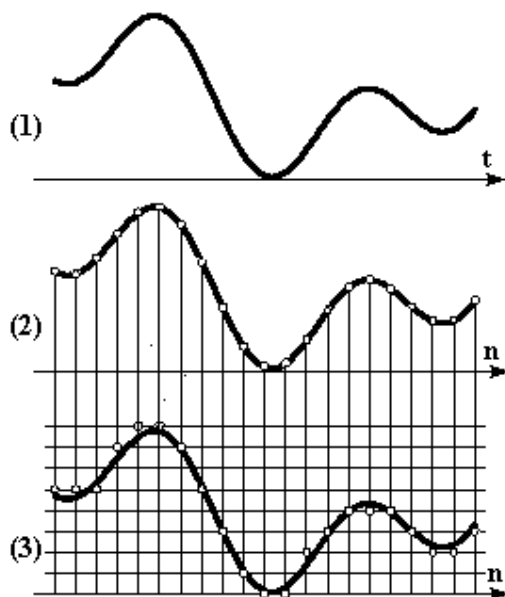
Signalą galima nusakyti keliais būdais:

1. matematine funkcija;
2. tikimybinio pasiskirstymo;
3. empirinių matavimų seka.

Praktikoje dauguma signalų aprašomi trečiuoju būdu.

Kalbos signalas yra vienmatė laiko funkcija ir paprastai žymimas $s(t)$. Kompiuteryje signalą galima saugoti tik skaitmeniniu pavidalu, t. y. kaip skaičių seką $s(t) \rightarrow \{s_i\}$.

Skaitmeninis signalas yra diskretus dviem prasmėmis: nepriklausomo kintamojo t atžvilgiu ir priklausomo kintamojo s atžvilgiu. Kintamojo t diskretizavimas paprastai vadinamas diskretizavimu (angl. sampling), o s diskretizavimas – kvantavimu (angl. quantization). Signalų diskretizavimas ir kvantavimas pavaizduotas 2 pav.



2 pav. Signalų diskretizavimas ir kvantavimas.

Paprastai diskretizavimas atliekamas vienodais intervalais, t. y.

$s(n) := s(nT)$, kur $n = -\infty \dots -1, 0, 1, \dots +\infty$, T – diskretizavimo intervalas.

Diskretizavimo intervalas T šiame žymėjime paprastai praleidžiamas. Atvirkščias diskretizavimo intervalui dydis vadinamas diskretizavimo dažniu ir žymimas F .

Kaip parinkti diskretizavimo intervalą T (arba diskretizavimo dažnį), nusako Šenono (angl. Shannon) teorema. Diskretizavimo dažnis turi būti bent du kartus didesnis už didžiausią signale esantį dažnį. Šio reikalavimo nesilaikymas labai iškraipo signalą. Geros kokybės kalbos signale yra dažniai iki 8 kHz, todėl pakanka 16 kHz diskretizavimo dažnio. Telefonuose naudojamas iki 4 kHz apribotas kalbos signalas, todėl pakanka 8 kHz diskretizavimo dažnio.

Dėl baigtinio kompiuterio atminties ląstelės ilgio, diskretizuoto signalo atskaitos saugomos kvantuotu pavidalu:

$s(n) \rightarrow [s(n)]_Q = k(n) * q$, kur k – sveikasis skaičius.

Jei kvantuojama vienodais žingsniais, q yra konstanta – kvantavimo žingsnis. Be to, k reikšmė paprastai būna apribota. Ribas nusako vienai atskaitai saugoti skirtų bitų skaičius. Geros kokybės kalbos signalui saugoti paprastai pakanka 12 bitų (4096 amplitudės lygių), telefoninės kokybės signalui – 8 bitų (256 lygių).

Kalboje esanti informacija

Kalba kaip žmonių bendravimo priemonė turi dvi formas: garsinę ir tekstinę. Informacija kalboje egzistuoja kaip diskrečių elementų seka laike ar erdvėje. Šie elementai gali būti nagrinėjami įvairiais lygiais: pasakymo, sakinio, frazės, žodžio, morfemos, skiemens, fonemos. Aukštesnio lygio elementuose paprastai yra labai paprasta priklausomybė tarp garsinės ir tekstinės formos, tik žemiausiame lygyje reikalingos rašybos ar tarimo taisyklės. Vis dėl to garsinė ir tekstinė forma dviem esminiais požymiais:

1. Garsinėje formoje yra informacijos, kurios nėra tekste: akcentai, intonacija, kalbėtojo charakteristikos ir t. t.

2. Tarimo taisyklės nusakomos kaip tam tikra sužadavimo ir artikuliacinių organų tikslinių padėčių seka, tačiau iš tikrųjų garsai veikia vieni kitus, kontekste jie tariami kitaip, nei atskirai.

Dėl šių priežasčių pereinant nuo garsinės formos prie tekstinę prarandama dalis informacijos, o atliekant atvirkščią perėjimą ne visa atstatoma. Galimi ir įvairūs tarpiniai saugojimo būdai, kurie naudojami, pvz., kalbinę informaciją perduodant ryšio linija. Dar reikia pastebėti, kad garsinė informacija užima neproporcingai daug vietos lyginant su tekste. Reikalingas atminties kiekis garsinei ir teksteinei informacijai, bei kai kurie tarpiniai variantai pateikti 1 lentelėje.

Informacijos lygis	Pavaizdavimas	Informacijos srutas bitais per sekundę
Akustinis	Kalbos signalas diskretizuotas 20 kHz, 12 bitų atskaitai	240000
	Žemutinė riba akustinei informacijai perduoti naudojant geriausias suspaudimo technologijas	2400
Parametrinis	Geros kokybės parametrinis pavaizdavimas	4800
	Žemutinė parametrinio pavaizdavimo riba	400
Akustinis /fonetinis	Mišrus parametrinis/fonetinis pavaizdavimas (fonetinis vokoderis)	100 - 500
Fonetinis /fonologinis	Fonemų seka (40 fonemų, 6 bitai fonemai, 10 garsų per sekundę)	60
Fonologinis	Informacija, kurios pakanka pranešimui suprasti naudojant visą lingvistinę, sintaksinę, semantinę, pragmatinę informaciją	10 - 20

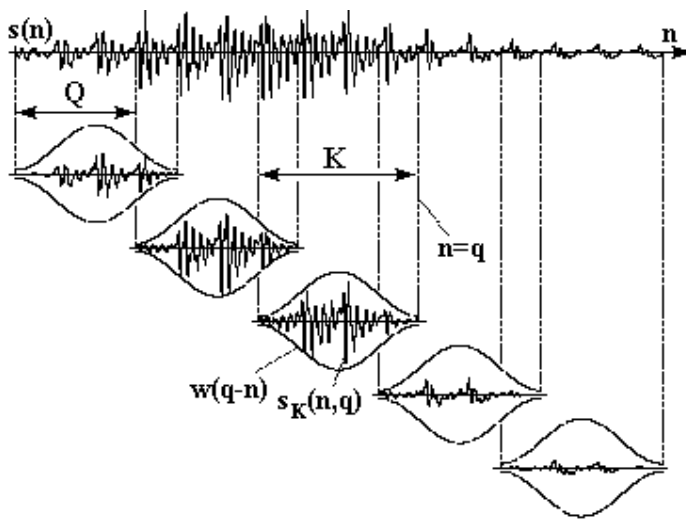
1 lentelė. Informacijos kiekis įvairiuose lygiuose.

Trumpalaikės analizės principas

Laikoma, kad kalbos signalas yra kintantis laike, tačiau esminiai parametrai kinta lėtai lyginant su diskretizavimo intervalu. Tai reiškia, kad išmatuotų parametų kitimas laike laikomas nauju signalu, kurį galima diskretizuoti žymiai mažesniu dažniu lyginant su pradiniu signalu. Taigi parametų signalas P apibrėžiamas kokiam nors taške n q , kitos P atskaitos imamos žingsniu Q nuo taško q . Pačiai P reikšmei apskaičiuoti iš pradinio signalo išskiriama taško q aplinka naudojant lango funkciją $w(k)$. Funkcija $w(k)$ yra baigtinė ir turi baigtinio ilgio K apibrėžimo sritį. Lango funkcijos paveiktas signalas taip pat bus baigtinio ilgio, jį galima užrašyti tokia formule:

$$s_K(n, q) = \begin{cases} s(n)w(q-n), & n = q - K + 1 \\ \text{neapibr.} & \text{priešingu atveju.} \end{cases}$$

Signalų trumpalaikės analizės pavyzdys pateiktas 3 pav.



3 pav. Signalų trumpalaikė analizė

Lango ilgį K reikia parinkti taip, kad: 1) parametrus galima būtų apskaičiuoti pakankamai tiksliai ir 2) parametrai beveik nesikeistų šiame intervale. Kadangi tai prieštaringi reikalavimai, tai praktikoje kalbos signalams paprastai imama Q nuo 5 iki 25 ms., K nuo 20 iki 50ms., t. y. gretimi langai persidengia.

Priklausomai nuo parametų vertinimo būdo, už lango ribų esantį signalą galima traktuoti keliais būdais: 1) prilyginti jį nuliui; 2) periodiškai pakartoti signalą; 3) laikyti neapibrėžtu.

Kai lango ilgis K lygus pagrindinio tono periodo ilgiui T , turime specialų atvejį, vadinamą analize, sinchroniškai pagrindiniam tonui. Naudojant šį metodą išlaikomas natūralus signalo periodiškumas ir lango funkcija įneša į signalą mažiausius iškraipymus.

Paprasčiausi kalbos signalo parametrai

Trumpalaikis galingumas (angl. power):

$$P_K(q) = \frac{1}{K} \sum_{n=q-K+1}^q s_K^2(n, q).$$

Trumpalaikė energija:

$$U_K(q) = \sum_{n=q-K+1}^q s_K^2(n, q).$$

Modulio vidurkis:

$$|\bar{s}_K(q)| = \frac{1}{K} \sum_{n=q-K+1}^q |s_K(n, q)|.$$

Trumpalaikė amplitudė:

$$A_+(q) = \max[s_K(n, q)], \quad A_-(q) = \min[s_K(n, q)], \quad n = q - K + 1.$$

Dar vienas įdomus parametras, susijęs su dažniu – tai nulio ašies kirtimų skaičius:

$$H_K(q) = \frac{1}{K} \sum_{n=q-K+1}^q x(n), \quad \text{kur} \quad x(n) = \begin{cases} 0, & \text{sgn}[s(n)] = \text{sgn}[s(n-1)] \\ 1, & \text{sgn}[s(n)] \neq \text{sgn}[s(n-1)] \end{cases}.$$

Tiesinės prognozės modelis

Pagrindinė tiesinės prognozės idėja yra prognozuoti kalbos signalo atskaitą remiantis buvusiomis atskaitomis [2]. Tarkime, $\{s_1, \dots, s_N\}$ yra signalo atskaitų seka. Elemento s_n M-tos eilės tiesinė prognozė formuojama kaip prieš jį einančių M atskaitų tiesinė kombinacija. Prognozuotą atskaitą pažymėkime y_n . Tada

$$y_n = - \sum_{i=1}^M a_i s_{n-i}.$$

Koeficientai $a_i, i=1, \dots, M$ šioje sąsūkoje ir yra ieškomieji prognozės koeficientai. Juos reikia parinkti taip, kad minimizuoti prognozės paklaidą. Vienos atskaitos prognozės paklaida e_n lygi

$$e_n = s_n - y_n,$$

o pilna kvadratinė paklaida apibrėžiama tokiu būdu:

$$\alpha = \sum_{n=n_0}^{n_1} e_n^2 = \sum_{n=n_0}^{n_1} \left[\sum_{i=0}^M a_i s_{n-i} \right]^2 = \sum_{n=n_0}^{n_1} \sum_{i=0}^M \sum_{j=0}^M a_i s_{n-i} s_{n-j} a_j,$$

kur n_0 ir n_1 yra sumavimo rėžiai. Pažymėkime

$$c_{ij} = \sum_{n=n_0}^{n_1} s_{n-i} s_{n-j}.$$

Tada kvadratinė paklaida α užrašoma taip:

$$\alpha = \sum_{i=0}^M \sum_{j=0}^M a_i c_{ij} a_j.$$

Norint minimizuoti α , reikia apskaičiuoti α dalines išvestines a_k , $k=1, \dots, M$ atžvilgiu, prilyginti jas nuliui ir išspręsti gautą lygčių sistemą. Atsižvelgiant į tai, kad $a_0=1$, gauname tokią tiesinių lygčių sistemą:

$$\sum_{i=1}^M a_i c_{ik} = -c_{0k},$$

kur $k=1, \dots, M$.

Priklausomai nuo sumavimo rėžių n_0 ir n_1 parinkimo galima išskirti kovariacinę ir autokoreliacinę parametrų c_{ik} išraišką. Tiesinių lygčių sistemą galima išspręsti bet kuriuo tiesinių lygčių sistemų sprendimo metodu, tačiau yra sukurti specialūs metodai, žymiai pagreitinantys sprendimą remiantis tam tikromis sistemos savybėmis. Pvz. autokoreliacinės realizacijos atveju galima naudoti Levinsono-Durbino metodą.

Tiesinės prognozės koeficientai gali būti naudojami daugeliui tikslų: kalbos signalų suspaudimui, atpažinimui, sintezei.

Kalbos signalo sintezėje kartu su prognozės koeficientais dar naudojami pagrindinio tono periodo ilgis, požymis, nusakantis, ar signalas skardus, ar duslus, kvadratinė šaknis iš signalo atskaitų. Sintezuojant skardžius segmentus filtras sužadinamas kvaziperiodiniais impulsais, atstumas tarp kurių lygus vidutiniam pagrindinio tono periodui, o dusliams sužadinantį signalą sukuria balto triukšmo generatorius. Sužadinantis signalas dauginamas iš daugiklio (stiprinimo koeficiento), nusakančio pradinio ir sintezuoto segmento energijos santykį.

Spektro skaičiavimas naudojant Furje transformaciją

Kalbos signalą patogiau traktuoti kaip sinusoidinių virpesių su įvairiais dažniais sumą. Signalo spektras nusako, kiek energijos tenka kuriam dažniui. Yra įrodyta, kad žmogaus klausos periferijoje atliekama savotiška spektrinė analizė, todėl spektras yra viena dažniausiai naudojamų kalbos signalo charakteristikų. Spektrui skaičiuoti naudojama diskrečioji Furje transformacija (DFT). Tiesioginė ir atvirkštinė DFT nusakomos atitinkamai formulėmis:

$$X(m) := \sum_{n=0}^{N-1} x(n) * e^{\frac{-2\pi jnm}{N}};$$

$$x(n) := \frac{1}{N} \sum_{m=0}^{N-1} X(m) * e^{\frac{2\pi jnm}{N}},$$

kur $x(n)$ – kompleksinė įėjimo signalo seka $x_r(n)+x_m(n)$, o $X(m)$ – kompleksinis signalo spektras $X_r(m)+X_m(m)$.

Nors kalbos signalai paprastai saugomi ne kaip kompleksinių, o kaip realių (ar sveikų) skaičių seka, formaliai juos galima laikyti kompleksinių skaičių seka, kurios menamos dalys lygios 0.

Amplitudės spektras apibrežiamas taip:

$$|X(m)| := \sqrt{X_r^2(m) + X_m^2(m)},$$

o fazės spektras taip:

$$\varphi(m) := \arctan \frac{X_m^2(m)}{X_r^2(m)},$$

nors pastarasis retai naudojamas kalbos signalų apdorojime.
Galingumo spektras apibrėžiamas taip:

$$P(m) := |X(m)|^2.$$

Paprastai (ypač grafiniam pavaizdavimui) naudojamas logaritminėje skalėje decibelais išreikštas spektras

$$L(m) := 20 * \log_{10} |X(m)| := 10 * \log_{10} P(m).$$

DFT buvo apibrėžta baigtinio ilgio sekai. Už šio intervalo ribų laikoma, kad signalas kartojamas begalinį skaičių kartų. Nagrinėjant realius kalbos signalus taip nebūna, kad signale būtų tik dažniai, kurių periodas tilptų į nagrinėjamą intervalą sveiką skaičių kartų, todėl ties intervalo ribomis atsiranda iškraipymai. Iškraipymams sumažinti signalas nagrinėjame intervale dauginamas iš lango funkcijos, kuri lygi 1 intervalo viduryje ir artėja į 0 intervalo galuose. Populiariausia yra Hemingo funkcija:

$$w_H(k) := 0,54 + 0,46 * \cos \frac{2\pi(k - N/2)}{N}.$$

Tiesioginiu būdu skaičiuojant sekos iš N atskaitų DFT reikia sugaišti laiką, proporcingą N^2 . Skaičiavimams pagreitinti buvo sukurti greitosios Furje transformacijos (GFT) algoritmai, sumažinantys laiko sąnaudas iki $N * \log_2 N$. Šie algoritmai remiasi sekos skaidymu ir kai kuriomis kompleksinių skaičių savybėmis. Dauguma GFT algoritmų naudoja sekas, kurių ilgiai $N = 2^r$, r – natūrinis skaičius. Pačių algoritmų čia nepateiksime, GFT algoritmų, realizuotų paprogramės pavidalu, galima rasti literatūroje, pavyzdžiui, [3]. Į tokių paprogramių įėjimą paprastai paduodama sekos $x(n)$ realių ir menamų dalių masyvai, o išėjime tuose pačiuose masyvuose gaunamos spektro taškų realios ir menamos dalys. Įėjime menamų dalių masyvas paprastai užpildomas nuliais.

GFT gali būti naudojama juostiniam signalų filtravimui. Tarkime, kad norime signalą nufiltruoti taip, kad jame liktų tik dažniai, priklausantys intervalui $[w_1, w_2]$. Tada reikia apskaičiuoti signalo GFT, prilyginti 0 spektro koeficientus, nepriklausančius intervalui $[w_1, w_2]$, ir apskaičiuoti atvirkštinę GFT.

Spektro diskretizavimo intervalas susijęs su nagrinėjamo intervalo ilgiu sąryšiu:

$$\Delta f = \frac{F}{N},$$

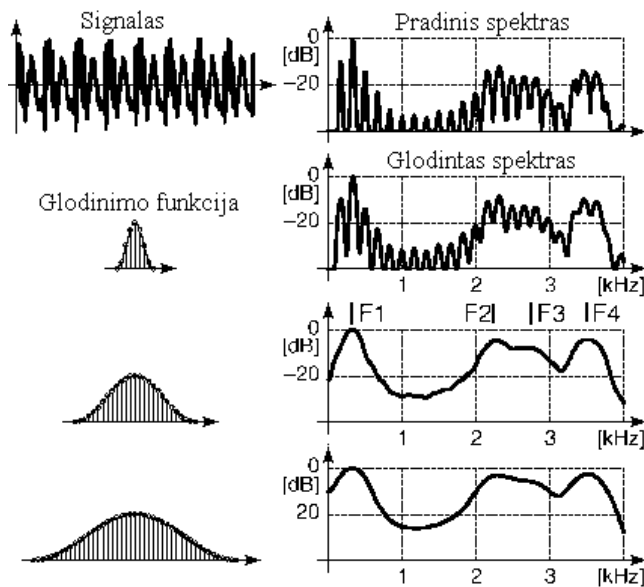
kur F – signalo diskretizavimo dažnis hercais, o Δf – atstumas tarp dviejų spektro reikšmių hercais. Kadangi ne visada norima, kad spektro diskretizavimo intervalas būtų susijęs su signalo nagrinėjimo intervalu, tai galima signalo nagrinėjimo intervalą praplėsti prirašant nulių ir taip sumažinti spektro diskretizavimo intervalą.

Spektro glodinimas

Vienas iš pagrindinių kalbos signalo analizės uždavinių yra iš signalo atstatyti kalbos trakto charakteristikas. Kalbant apie spektrinę analizę, tai šiam tikslui reikia atskirti sužadinimo ir signalo formavimo komponentes. Sužadinančio signalo periodiškumas spektre pasireiškia tuo, kad spektre matomi maksimumai, atstumas tarp kurių lygus pagrindinio tono periodo ilgiui. Signalo formavimas atskiriamas nuo sužadinimo glodinant spektrą. Tai gali būti atliekama trimis būdais:

1. tiesiogiai glodinant spektrą;
2. parenkant signalo nagrinėjimo intervalo ilgį mažesnę už vieną pagrindinio tono periodą. Šiuo atveju informacija apie periodinę sužadinimą atmetama jau trumpalaikės analizės metu;
3. panaudojant kepstrą;
4. skaičiuojant spektrą iš tiesinės prognozės koeficientų.

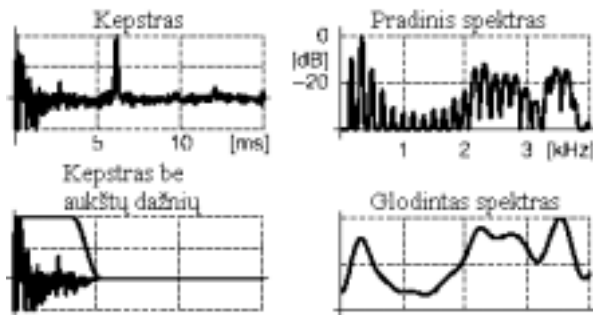
Tiesioginis spektro glodinimas atliekamas vidurkinant spektro taškus panaudojant tam tikrą glodinimo funkciją. Funkciją reikia parinkti taip, kad spektre išsiskirtų formantiniai dažniai ir neliktų jokių kitų maksimumų. Tiesioginio spektro glodinimo naudojant kelias funkcijas pavyzdys pateiktas pav.



4 pav. Tiesioginis spektro glodinimas.

Tarkime, kad signalo spektras yra paprasčiausias signalas. Atstumas tarp maksimumų, atsiradusių dėl pagrindinio tono įtakos, paprastai būna mažesnis už atstumą tarp formantių, todėl galime naudodami Furje transformaciją apskaičiuoti spektro spektrą (vadinamą kepstru), atmesti pagrindinį toną atitinkančius koeficientus, o jei pagrindinio tono periodo ilgis nežinomas – visus aukštesnius dažnius, ir apskaičiuoti atvirkštinę Furje transformaciją. Šis procesas pavaizduotas 5 pav.

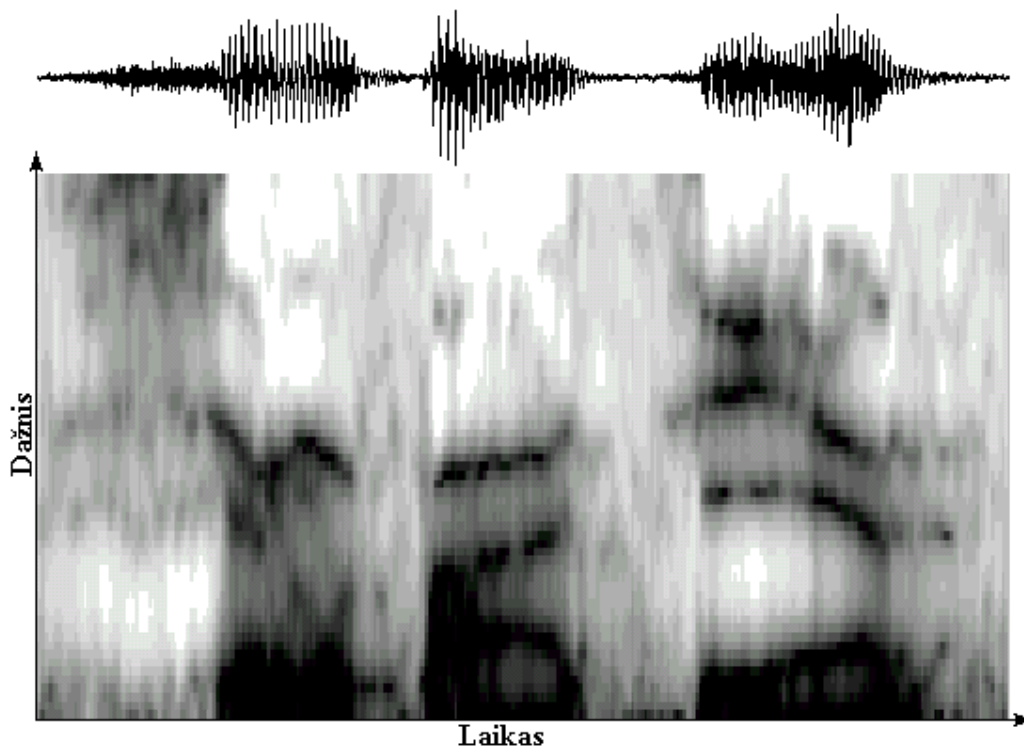
GFT pagalba galima gauti signalo spektrą ne tik iš signalo atskaitų, bet ir iš tiesinės prognozės koeficientų. Tam tikslui vietoje pirmųjų atskaitų imami tiesinės prognozės koeficientai, o vietoje likusių – nuliai. Tokiu būdu gaunamas pakankamai glodus grafikas.



5 pav. Spektro glodinimas naudojant keprą.

Spektrograma

Spektro kitimui laike pavaizduoti dažniausiai naudojamos spektrogramos. Jos gaunamos išdėstant laiko ašyje trumpalaikius spektrus. Kadangi toks pavaizdavimas reikalauja trijų matavimų, tai horizontalioje ašyje paprastai atidedamas laikas, vertikalioje – dažnis, o amplitudė vaizduojama įvairiais pilkumo lygiais. Žodžio “simpatija” spektrograma pavaizduota 6 pav. Trumpalaikiai spektrai skaičiuoti iš tiesinės prognozės koeficientų.



6 pav. Žodžio “simpatija” spektrograma.

Gali būti naudojami ir trimačiai pavaizdavimai.

Jei spektras skaičiuojamas iš nepersidengiančių signalo intervalų, tai naudojant trumpus intervalus gaunamas geras diskretizavimas laiko ašyje, tačiau grubus dažnio

ašyje, o jei ilgus intervalus – atvirkščiai. Gali būti naudojami ir persidengiantys intervalai (kaip 6 pav.).

Kalbos signalų apdorojimo programos

Kalbos signalų apdorojimo programos paprastai suteikia vartotojui tokias galimybes: įvesti kalbos signalą į kompiuterį, pamatyti ekrane signalo grafiką, redaguoti signalus, pasiklausyti signalą ar norimą fragmentą, pamatyti ekrane spektrogramą, formantinių dažnių trajektorijas, pagrindinio tono periodo grafiką ir pažymėto segmento spektrą, filtruoti signalą, užrašyti į failą signalo charakteristikas (segmentus, jų ilgius, formantinius dažnius ir pagrindinio tono periodo reikšmes).

Literatūra

- [1] http://www.ikp.uni-bonn.de/dt/lehre/materialien/grundl_ssv/index.html;
- [2] Маркел, Дж. Д., А. Х. Грей (1980). *Линейное предсказание речи*. Москва. Связь.
- [3] Рабинер Л. Р., Р. В. Шафер (1981). *Цифровая обработка речевых сигналов*. Москва, Радио и связь.