

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**INFORMATIKOS FAKULTETAS**  
**PROGRAMŲ INŽINERIJOS KATEDRA**

Mindaugas Bliūdžius

**SKAITMENINIŲ KALBOS ĮRAŠŲ GLAUDINIMO METODAI**

Magistro darbas

Darbo vadovas : doc. Vacius Jusas

**KAUNAS, 2004**

## SANTRAUKA

Per paskutinius tris dešimtmečius buvo stebimas didelis progresas žemų dažnių kalbinių kodekų taikyme civilinėse ir militaristinėse komunikacijose taip pat, kaip ir programinėje įrangoje susijusioje su balso ir kalbos skaitmeninėmis technologijomis. Šio progreso centre buvo naujų kalbinių kodekų kūrimas sugebančių išlaikyti aukštos kokybės kalbos kodavimą prie žemų bitų dažnių. Daugumo šių kodekų realizuoja mechanizmus sugebančius: atvaizduoti spektrinius kalbos parametrus ir “optimizuoti“ kodekų našumą žmogaus ausiai. Dalis šių kodekų jau yra pripažinti tarptautiniais telefonijos standartais.

Šio dokumento analitinės dalies paskirtis supažindinti su kalbos glaudinimo metodais, išskiriant tuos kurie šiuo metu yra populiariausi programinėje įrangoje susijusioje su balso ir kalbos technologijomis. Apžvalga pradedama nuo garso ir garso bangų charakteristikų apžvalgos, toliau apžvelgiamos pagrindinės garso kodekų charakteristikos. Dar toliau apžvelgiami kalbiniai kodekai bei jų klasifikacija, palyginami populiariausi šiuo metu esantys kalbiniai kodekai.

Sekančiose skyriuose pristatoma magistratūros studijų metu sukurta paskirstyta stenografavimo sistema, atliekama jos kokybės analizė ir siūlomi galimi sukurtos programinės įrangos tobulinimo variantai.

## **Abstract**

The past three decades has witnessed substantial progress towards the application of low-rate speech coders to civilian and military communications as well as computer-related voice applications. Central to this progress has been the development of new speech coders capable of producing high-quality speech at low data rates. Most of these coders incorporate mechanisms to: represent the spectral properties of speech, provide for speech waveform matching, and "optimize" the coder's performance for the human ear. A number of these coders have already been adopted in national and international cellular telephony standards.

The objective of this paper is to provide a tutorial overview of speech coding methodologies with emphasis on those algorithms that are part of the recent low-rate standards for voice applications. Although the emphasis is on the new low-rate coders, we attempt to provide a comprehensive survey by covering some of the traditional methodologies as well. The paper starts with a historical perspective and continues with a brief discussion on the speech properties and performance measures. Then I proceed with descriptions of waveform coders, linear predictive vocoders, and analysis-by-synthesis linear predictive coders.

At the end the system for computer-based stenographing is presented. Quality research and ways how to improve this system will be provided.

# TURINYS

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>IVADAS</b>                                | <b>5</b>  |
| <b>2</b> | <b>ANALITINĖ DALIS</b>                       | <b>7</b>  |
| 2.1      | GARSINĖS BANGOS BEI JŲ CHARAKTERISTIKOS      | 7         |
| 2.2      | GARSINIŲ BANGŲ DISKRETIZAVIMAS               | 8         |
| 2.2.1    | Skaitmeninių įrašų dydžio problema           | 9         |
| 2.2.2    | Skaitmeninės informacijos glaudinimo metodai | 10        |
| 2.3      | KODEKŲ PAGRINDINĖS CHARAKTERISTIKOS          | 10        |
| 2.4      | KALBOS KODAVIMO METODAI                      | 11        |
| 2.4.1    | Banginiai kodekai                            | 12        |
| 2.4.2    | Balsiniai kodekai                            | 14        |
| 2.4.3    | Hibridiniai kodekai                          | 16        |
| 2.5      | KALBINIŲ KODEKŲ STANDARTŲ APŽVALGA           | 17        |
| 2.5.1    | G.711  | 18        |
| 2.5.2    | G.722  | 18        |
| 2.5.3    | G.723.1                                      | 18        |
| 2.5.4    | G.726  | 19        |
| 2.5.5    | G.728  | 19        |
| 2.5.6    | G.729  | 19        |
| 2.5.7    | G.729A                                       | 19        |
| 2.5.8    | GSM-AMR                                      | 20        |
| 2.5.9    | Santrauka                                    | 20        |
| <b>3</b> | <b>PROJEKTINĖ DALIS</b>                      | <b>21</b> |
| 3.1      | SISTEMOS APIBŪDINIMAS                        | 21        |
| 3.1.1    | Programų sistemos funkcijos                  | 21        |
| 3.1.2    | Sistemos kontekstas                          | 22        |
| 3.1.3    | Vartotojo problemos                          | 22        |
| 3.1.4    | Sistemos tikslai                             | 23        |
| 3.2      | ARCHITEKTŪROS PATEIKIMAS                     | 23        |
| 3.2.1    | Architektūros tikslai ir apribojimai         | 23        |
| 3.2.2    | Panaudojimo atvejų požiūris                  | 24        |
| 3.2.3    | Paskirstymo požiūris                         | 25        |
| 3.2.4    | Komponentas „Formų menedžeris“               | 26        |
| 3.2.5    | Komponentas „Transkribavimo menedžeris“      | 29        |
| 3.2.6    | Komponentas „Garso menedžeris“               | 30        |
| 3.2.7    | Komponentas „Transkribavimo sistema“         | 32        |
| 3.2.8    | Komponentas „Duomenų menedžeris“             | 33        |
| <b>4</b> | <b>TYRIMO DALIS</b>                          | <b>34</b> |
| <b>5</b> | <b>EKSPERIMENTINĖ DALIS</b>                  | <b>39</b> |
| <b>6</b> | <b>IŠVADOS</b>                               | <b>42</b> |
| <b>7</b> | <b>ŽODYNAI</b>                               | <b>43</b> |
| 7.1      | SANTRUMPŲ ŽODYNAS                            | 43        |
| 7.2      | TERMINŲ ŽODYNAS                              | 43        |
| <b>8</b> | <b>LITERATŪRA</b>                            | <b>44</b> |
| 8.1      | NUORODOS                                     | 44        |

# 1 Įvadas

Sparčiai tobulėjant skaitmeninėms garso technologijoms bei gausėjant jų įvairovei vis plėtėja jų panaudojimo sritis. Daugiau yra naudojamos internetinėmis garso ir vaizdo konferencijomis, atsirado toks terminas kaip IP telefonija, mobiliuose telefonuose vis dažniau pradeda naudotis balso pašto paslaugomis. Visos šios technologijos neišvengiamai turi perduoti arba archyvuoti žmogaus kalbos skaitmeninius garsinius įrašus. Tam, kad juos perduoti efektyviai, reikia kuo stipriau juos suglaudinti (suspausti) bet tuo pačiu išlaikyti įrašo kokybę, kad išplėtus būtų įmanoma jį suprasti.

Programinės įrangos priemonė medijos duomenų srautui suglaudinti ir išskleisti vadinama kodeku (*codec* – pavadinimas kilęs iš dviejų angliškų žodžių „*code*“ ir „*decode*“). Tokių programinių priemonių yra žinoma daug, pvz.: *MPEG*, *QuickTime*, *Windows Media*, *RealAudio* ir t.t. Vienos iš jų geriau tinka kompaktinių plokštelių gamybai, kitos video produkcijos gamybai, treči skirti perdavimo kompiuterių tinklais.

Kalbos garsinių signalų diskretizavimas, glaudinimas ir išplėtimas vadinamas kalbos kodavimu, o tai atliekantys kodekai yra vadinami tiesiog kalbiniais kodekais. Šiame darbe apžvelgiami tik tie kodekai, kurie yra būtent skirti kalbos kodavimui.

Nuo pat pradžių kalbiniai kodekai išsiskyrė į dvi klases: balsinius (*vocoders* – sutrumpinimas *voice coders*) ir banginius (*waveform*) kodekus. Abi kryptys radikaliai skyrėsi savo naudojamais principais bei charakteristikomis. Pirmosios klasės kodekai rėmėsi kalbos analize [2, bandydami atvaizduoti garsinį signalą kelių parametru deriniu. Banginiai kodekai bando, be jokių papildomų žinių apie tai kaip koduojamas signalas buvo sugeneruotas, atkurti signalą, kurio bangos forma būtų kiek galima panašesnė į originalo.

1980-1990 metais kalbos kodavime įvyko revoliucija. Naujuosius kodekus buvo sunku suklasifikuoti į balsinius bei banginius kodekus. Naujos signalo nuspėjimo technologijos bei atskleistos naujos žmogaus ausies savybės leido atsirasti daugybei naujų kodekų, pvz. : *CELP*, *LPC*, *G.721*, *G.726*, *G.727*, *LD-CELP*, *GSM*, *TrueSpeech*, *Speex* ir kiti. Kodekai kurie naudoja tiek balsinius, tiek banginių kodekų principus yra vadinami hibridiniais kodekais.

Pagrindinis šio darbo tikslas – ištirti kalbos kodavimo metodus (kalbinius kodekus).

Darbo uždaviniai :

- apžvelgti skaitmeninių garsinių įrašų dydžio problemą,
- palyginti kalbinius kodekus ir įvertinti jų kokybę,
- nustatyti, kurie iš kalbinių kodekų labiausiai tiktų mūsų kuriamoje paskirstytoje stenografavimo sistemoje.

Pirmajame skyriuje trumpai apžvelgiama garsinės bangos ir jų pagrindinės charakteristikos, diskretizavimo principai, kalbos suglaudinimo metodai (kodekai). Skyriaus pabaigoje palyginami tarpusavyje skirtingų pakraipų kodekai.

Antrajame skyriuje apžvelgiama magistratūros studijų metu sukurtos programinės įrangos architektūra. Pateikiamas pasirinkto sprendimo realizacijos kelias.

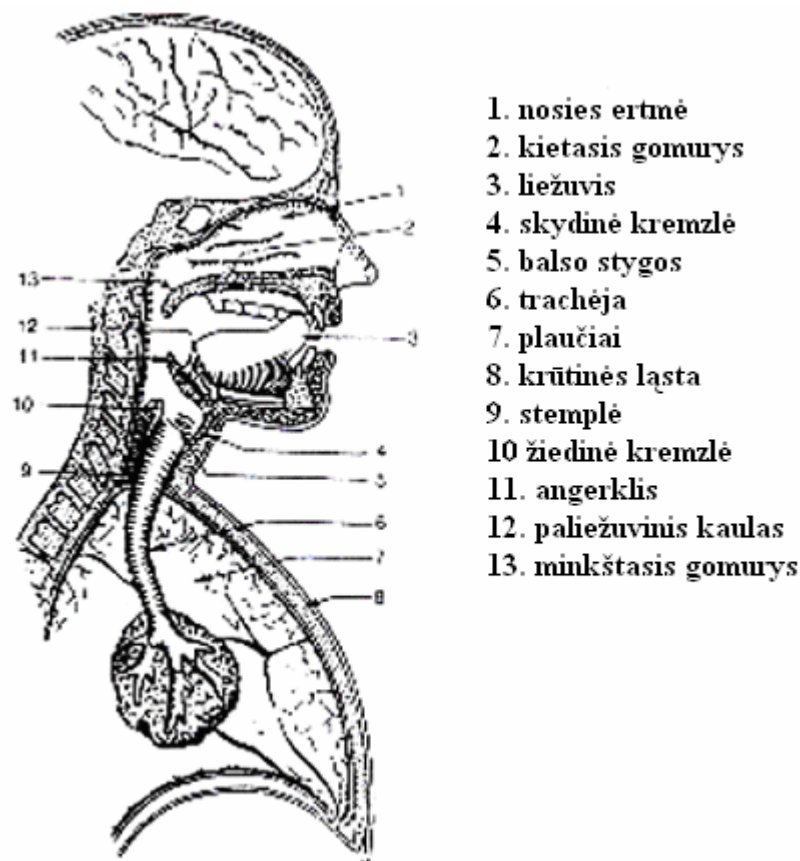
Tiriamoji ir eksperimentinė dalis yra skirta kalbinių kodekų kokybės analizei.

## 2 Analitinė dalis

Šiame skyriuje trumpai apžvelgiama garsinės bangos bei jų pagrindinės charakteristikos, diskretizavimo principai, kalbos kodavimo metodai (kalbiniai kodekai). Skyriaus pabaigoje palyginami tarpusavyje skirtingų pakraipų kodekai.

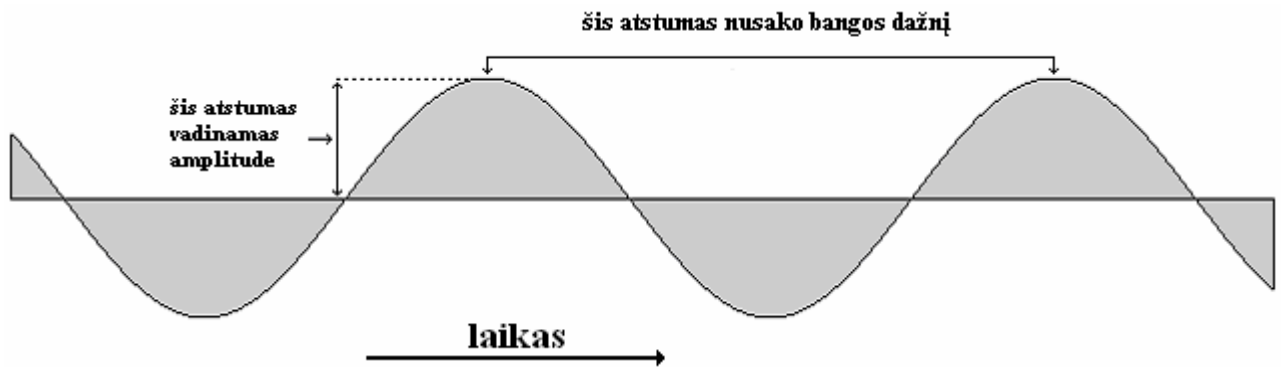
### 2.1 Garsinės bangos bei jų charakteristikos

Garsai yra išgaunami stumiant orą iš plaučių per balso stygas bei toliau per balso traktą. Balso traktą sudaro ryklė (dalis nuo stemplės iki burnos) ir burnos ertmės. Vidutiniškai vyrui bendras balso trakto ilgis yra apie 17 cm. Įtemptos balso stygos gerklose sukelia vibraciją dėl oro srovės. Oro srovė yra paverčiama kvaziperiodiniais impulsais, kurie tada moduluojami pagal dažnį praleidžiant per gerklės ertmę (ryklę), burnos ertmę ir galimai nosies ertmę. Atsižvelgiant į nuo įvairių artikuliacijų (smakro, liežuvio, minkštojo gomurio, lūpų, burnos) padėtį yra sukuriami įvairūs garsai [10].



Pav. 1 Žmogaus kalbinis aparatas

Vibruojančios balso stygos sukelia garsines bangas, kurios keliaudamos oru susiduria su ausies būgneliu ir priverčia jį vibruoti. Mūsų smegenys perima signalus iš ausies būgnelio, tokiu būdu suteikdamos mums galimybę girdėti garsus.



Pav. 2 Garsinės bangos parametrai

Garso skambėjimas priklauso nuo bangos charakteristikų. Bangos dažnis apibrėžia kiek kartų per sekundę banga pereina nuo savo aukščiausio taško iki žemiausio taško ir atgal iki aukščiausio. Ši charakteristika matuojama hercais (Hz) arba ciklų skaičiumi per sekundę. Bangos dažnis apibrėžia jos aukštumą. Dažnesnis virpėjimas suteikia bangai didesnę aukštumą, tankesnis – mažesnę. Vidutinis žmogus gali girdėti dažnius apytikriai nuo 15 arba 20 Hz iki 20,000 (20 kHz) [18].

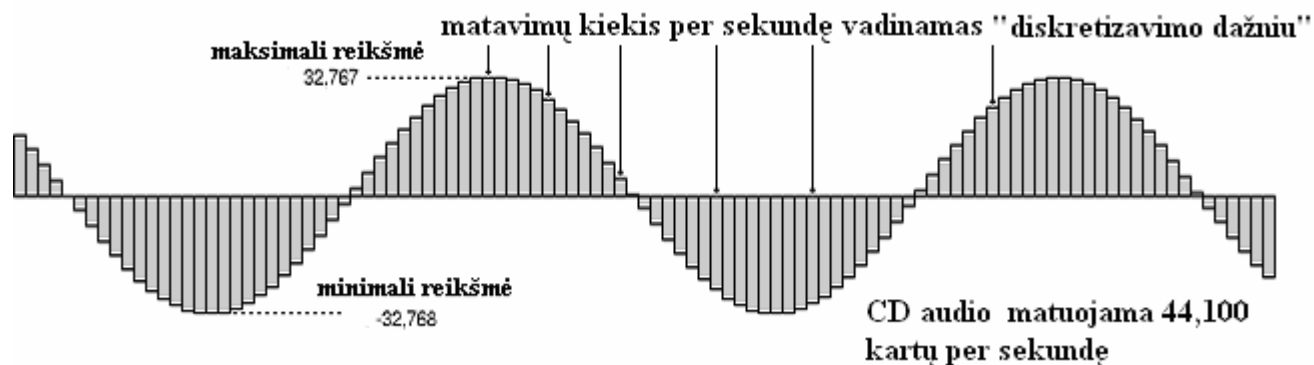
Bangos amplitudė apibrėžia pusę atstumo tarp aukščiausio bango taško ir žemiausio. Kuo didesnė bangos amplitudė, tuo stipresnis garsas, kuris matuojamas decibelais (dB) [18].

## 2.2 Garsinių bangų diskretizavimas

Antrajame pasauliniame kare inžinieriai eksperimentavo su garsu, versdami analogines garso bangas diskrečiomis reikšmėmis. Tai buvo pasiekama matuojant garso bangą daug kartų per sekundę. Kiekviename matavime buvo fiksuojama bangos amplitudė tuo laiko momentu (įskaitant ar banga „kilo“ ar „leidosi“). Pagal *Nyquist Theorem* [18] matavimų (toliau diskretizavimo) dažnis (matavimų kiekis per sekundę) turėtų būti bent jau du kartus didesnis negu didžiausias užfiksuotas bangos dažnis tam, kad būtų išvengta visokių trukdžių.

Taigi 1970 m. kai *Philips* ir *Sony* pradėjo ieškoti kelių pagerinti garso įrašų kokybei, jie pasuko garsinių bangų diskretizavimo kryptimi. Diskretizavimo dažnis lygus 44,000 matavimų per sekundę (44.1 kHz) buvo pasirinktas, kadangi jis viršijo nustatytą tikslą – 40 kHz (du kart tiek kirk girdi žmogus, 20 kHz) ir todėl, kad būtent tiek informacijos galėjo būti išsaugojama vaizdo juostoje – pagrindinėje saugojimo laikmenoje tais laikais.





Pav. 3 Garsinės bangos kvantavimas

Kiekvienas matavimas yra fiksuojamas 16 bitų skaičiumi, kintančiu nuo -32,768 iki 32,767. Šis skaičius nusako bangos amplitudę matavimo laiko momentu. Visų galimų skaičių, kuriais gali būti užfiksuotas matavimas, diapazonas vadinamas skaitmeninio signalo rezoliucija. Svyruojanti banga pirmyn ir atgal nuo -32,768 iki 32,767 būtų garsiausia banga, kokią šis formatas galėtų atvaizduoti, banga kintanti nuo -1 iki 1 būtų tyliausia, ir galų gale eilė nulių reikštų visišką tylą. Šios amplitudės reikšmių ribos pakankamos, kad būtų galima užfiksuoti net labai subtilius garso pasikeitimus. Toks skaitmeninis garso matavimas yra žinomas kaip pulso kodo moduliacija ir yra populiariausias diskretizavimo metodas [18]. Tokio tipo skaitmeninė informacija gali būti saugoma plačiau paplitusioje laikmenoje *WAV*.

## 2.2.1 Skaitmeninių įrašų dydžio problema

Yra įmanoma (be to labai paprastai) nukopijuoti garsinę informaciją iš kompaktinės plokštelės ir išsaugoti ją *WAV* formato bylose kompiuteryje, ir esant reikalui, ją galima būtų klausytis. Idealu būtų klausytis tokios kokybės įrašus visur (automobilyje, portatyvinio grotuvo pagalba), kadangi tai geriausia kokybė, kokia įmanoma įsigyti. Deja, tai nėra taip paprasta, kadangi įrašai užima labai daug vietos.

Kiekvienas matavimas (bandinys) yra 16 bitų arba dviejų baitų dydžio. Kadangi matavimai atliekami 44,100 kartų per sekundę ir moderni muzika yra įrašoma dviem kanalais (stereo), gauname, kad reikia  $(2 \cdot 44100 \cdot 2) = 176,400$  baitų tam, kad išsaugotume sekundės ilgio įrašą. Tai reiškia, kad 10,584,000 baitų arba apytiksliai 10 megabaitų reikalinga vienos minutės kompaktinės plokštelės kokybės įrašui išsaugoti. Tai neatrodo labai bauginančiai kadangi šiomis dienomis standieji diskai gali patalpinti savyje daug daugiau informacijos. Deja, portatyviniai diktofonai arba grotuvai neturi tiek daug vietos. Jei imtume kaip pavyzdį balso pašto žinučių saugyklą, tai joms saugoti reiktų labai daug vietos. Taigi, akivaizdu, kad reikia kažkokiais būdais suglaudinti skaitmeninius garso įrašus, norint jų klausytis portatyviniuose grotuvuose ar efektyviai archyvuoti saugyklose.

## 2.2.2 Skaitmeninės informacijos glaudinimo metodai

Egzistuoja du, iš esmės skirtingi, glaudinimo metodai : prarandantis informaciją bei jos neprarandantis.

Neprarandantis informacijos – toks, kai suglaudintas, mažesnio dydžio įrašas gali būti išplėstas atgal neprarandant informacijos. Tai yra jis bus visiškai identiškas originaliam įrašui. Deja, suglaudinti garsinius įrašus neprarandant informacijos yra gana sudėtinga. Bendros paskirties programos, tokios kaip *WinZip* ir *gzip* sugeba suspausti vidutiniškai tik apie 5 proc.. Egzistuoja specialios paskirties programinė glaudinimo įranga (tokia kaip *FLAC* [16]), kuri buvo specialiai sukurta garsiniams įrašams glaudinti neprarandant informacijos. Tokios priemonės sugeba suglaudinti apie 50 proc. palyginti su originaliu įrašo dydžiu. Nors tai gali atrodyti pakankamai, muzikos arba pokalbių įrašus (norint juo padaryti portatyvinius), reikėtų suglaudinti dar daugiau.

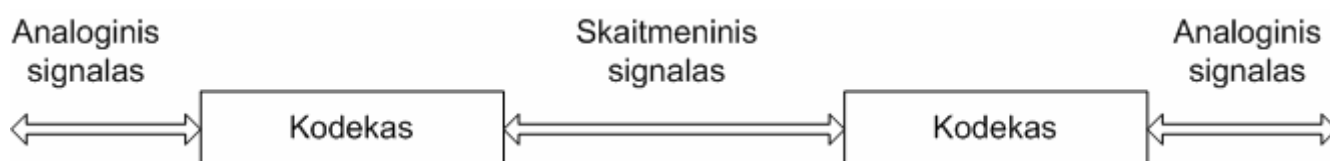
Prarandantis informaciją glaudinimo būdas toks, kuris praranda informaciją. Suglaudinus ir po to išplėtus įrašą bus gaunama kažkas panašaus į originalą, bet ne identiškas. Tai nėra gerai informacijai, kuria manipuliuoja kompiuteris, pavyzdžiui, programinės įrangos vykdomieji failai. Bet tai labai tinka žmogaus manipuluojamai informacijai, tokiai kaip fononutraukos arba garsiniai įrašai. Gudrybė slypi informacijos gabaliukų pašalinime iš ten, kur ji niekada nėra įvertinama.

Tokios priemonės naudoja psichoakustinį modelį. Tai yra, projektuojant kaip žmogaus ausys (bei smegenys) girdi garsus, yra įmanoma surasti vietų kur galima pašalinti informaciją, kurios žmogus vis tiek niekada nesuvoktų.

Naudojant tokias technologijas įmanoma sukurti tokius garsinių įrašų glaudinimo hibridinius kodekus kaip *Speex* [17], *Ogg Vorbis* [18] arba *Mp3* [19], kurios išsaugo garsinio įrašo kokybę, bet yra tik 10 proc. originalaus įrašo dydžio. Suglaudavimo lygis skiriasi nuo to, kokia informacija glaudinama, ar tai, pavyzdžiui, muzika ar pokalbis. Toliau bus apžvelgiami tik kalbai glaudinti skirti metodai.

## 2.3 Kodekų pagrindinės charakteristikos

Programinės įrangos priemonė medijos duomenų srautui suglaudinti ir išskleisti vadinama kodeku (*codec* – pavadinimas kilęs iš dviejų anglišku žodžių „*code*“ ir „*decode*“).



Pav. 4 Bendra kodekų struktūra

Kalbiniai ir garso kodekai gali būti įvertinti tokiomis pagrindinėmis charakteristikomis : bitų dažnis, kalbos kokybė, signalo vėlinimu, sudėtingumu, atsparumu triukšmams, ir atsparumas kanalo klaidoms.

Bitų dažnis nusako kiek reikalinga bitų vienos sekundės ilgio garsiniam signalui išsaugoti. Kuo mažesnis bitų dažnis tuo mažiau vietos užims įrašas ir atvirkščiai. Nesuglaudinto garso įrašo bitų dažnis gali būti suskaičiuojamas dauginant diskretizavimo dažnį iš rezoliucijos ir kanalų kiekio. Pavyzdžiui, kompaktinės plokštelės kokybės įrašo diskretizavimo dažnis - 44,100 Hz, rezoliucija - 16 bitų ir naudojami du kanalai. Bitų dažnis apytiksliai gaunasi 1.4 milijonai bitų per sekundę.

Kalbos kokybė gali būti matuojama subjektyviai ir objektyviai. Subjektyvus įvertinimas yra gaunamas tiesiog klausantis testinių kalbos įrašų. Objektyvus įvertinimas yra išskaičiuojamas iš užkoduotos kalbos parametrų. Detaliau kalbos kokybės vertinimas bus aptariamas tyrimų skyriuje.

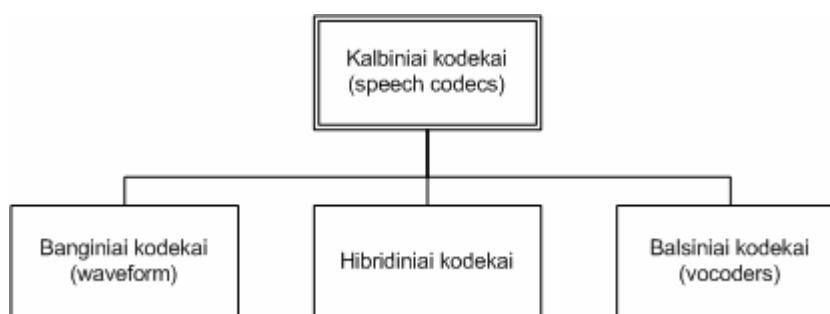
Signalų vėlinimas koderiui apibrėžiamas kaip laiko skirtumas tarp laiko kai signalo matavimas papuola į jo koderį (kodavimo dalis) ir laiko momento kai matavimas išeina iš dekoderio (kodeko – dekodavimo dalis), turint omenyje, kad bitų srautas (matavimų srautas) yra tiesiai paduodamas iš koderio į dekoderį.

Sudėtingumą apibrėžia instrukcijų kiekis (milijonais) atliktas per sekundę (MIPS).

## 2.4 Kalbos kodavimo metodai

Nuo pat pradžių kalbiniai kodekai išsiskyrė į dvi klases: balsinius (*vocoders* – sutrumpinimas *voice coders*) ir banginius (*waveform*) kodekus. Abi kryptys radikaliai skyrėsi savo naudojamais principais bei charakteristikomis. Pirmosios klasės kodekai rėmėsi kalbos analize [1], bandydami atvaizduoti garsinį signalą kelių parametrų deriniu. Banginiai kodekai bando, be jokių papildomų žinių apie tai kaip koduojamas signalas buvo sugeneruotas, atkurti signalą, kurio forma būtų kiek galima panašesnė į originalo.

Taip pat egzistuoja kodekai, kurie naudoja tiek balsinius, tiek banginių kodekų principus yra vadinami hibridiniais kodekais.



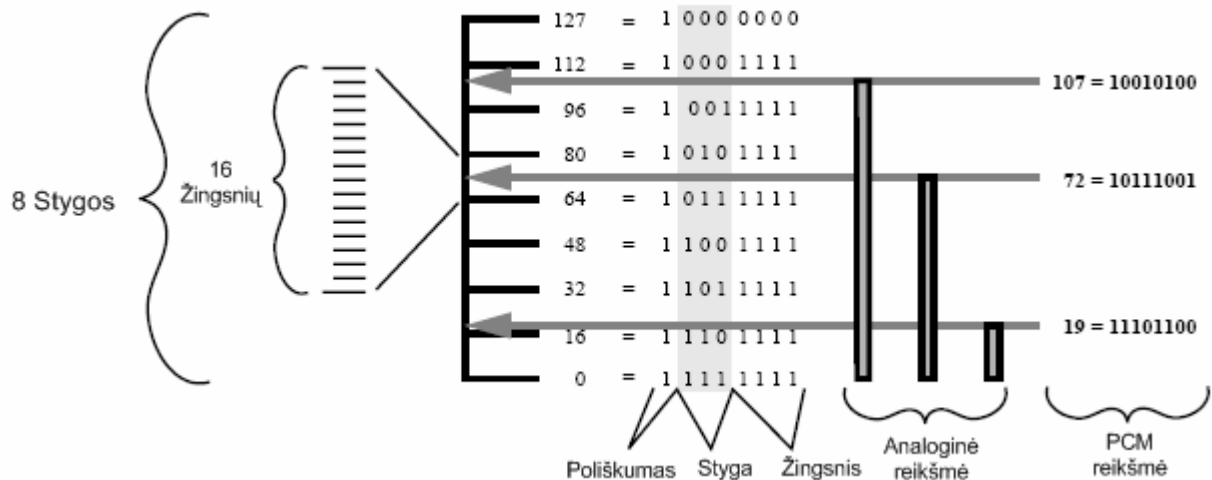
Pav. 5 Kalbinių kodekų klasifikacija

Toliau apžvelgiama kiekviena kalbinių kodekų klasė atskirai. Prie kiekvienos klasės apžvelgiami tos klasės kodekų pavyzdžiai kurie vėliau šiame darbe bus palyginti tarpusavyje.

### 2.4.1 Banginiai kodekai

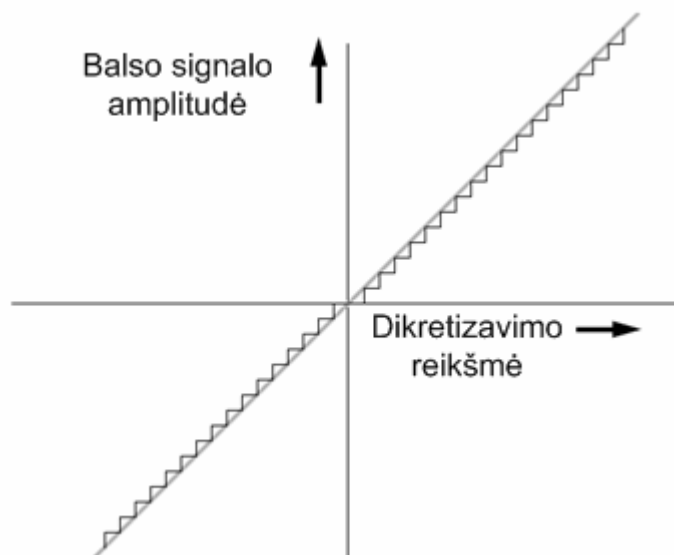
Banginiai kodekai bando be jokių papildomų žinių apie tai, kaip koduojamas signalas buvo sugeneruotas, atkurti signalą, kurio forma būtų kiek galima panašesnė į originalą. Tai reiškia, kad teoriškai jie turėtų būti nepriklausomi nuo signalo ir taip pat gerai tiktų ne tik kalbiniams signalams. Bendrai paėmus, šie kodekai nėra sudėtingi bei išsaugo aukštos kokybės kalbą naudojant 16 kbitų/s ir aukštesnį bitų dažnį. Kai bitų dažnis žemesnis, atgamintos kalbos kokybė gali stipriai sumažėti.

Pulso kodo moduliacijos (*Pulse Code Modulation – PCM*) kodekai yra paprasčiausi šioje klasėje. Kalbos signalas diskretizuojamas 8000 kartų per sekundę ir naudojamas 8 bitų rezoliucija. Naudojant 8 bitų rezoliucija išviso įmanomos 256 reikšmės. *PCM* išskiria šias reikšmes į 127 teigimus bei 127 neigimus užkodavimo lygius, plus nulinės amplitudės lygis. Šie lygiai yra sudalinti į 8 juostas vadinamas stygomis. Kiekvieną stygą sudaro 16 žingsnių. Toliau pavaizduota stygų/žingsnių struktūra tiesinėje diskretizacijoje (kvantavime):



Pav. 6 PCM Diskretizacija lygiai - stygos ir žingsniai

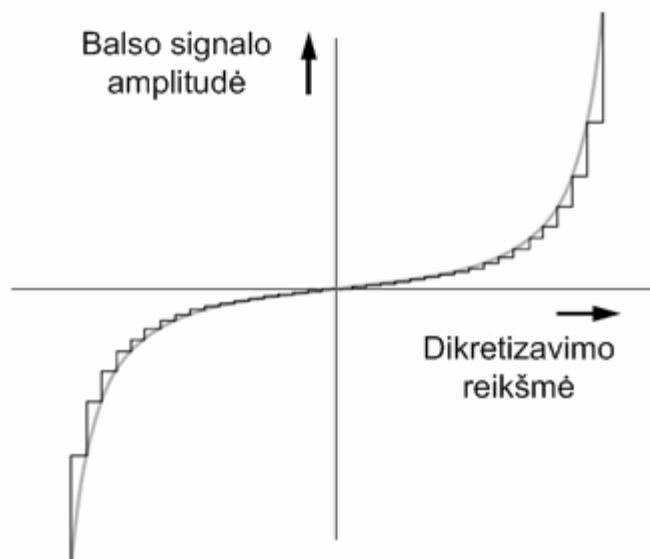
Kiekviena diskretizuojama signalo amplitudės reikšmė papuola į konkrečią stygą ir žingsnį. Pagal tai yra nustatoma skaitmeninė (diskreti) reikšmė. Sekančiame paveiksle parodytas tiesinės diskretizacijos procesas. Diskretizavimo lygis tiesiškai priklauso nuo balso signalo amplitudės padidėjimo. 127 diskretizavimo lygiai yra po lygiai paskirstyti balso signalo dinaminiam diapazone. Taigi, gaunasi, kad garsiems balso signalams yra naudojama ta pati rezoliucija kaip ir tyliems signalams. Taip koduojant analoginį signalą, išlaikant paprastumą, neduoda optimalaus adekvatumo rekonstruojant žmogaus balsą.



**Pav. 7 Tiesinė diskretizacija**

Taip dalinant balso signalo amplitudę į lygias teigiamas ir neigiamas dalis nėra efektyvus kelias užkoduoti balsą panaudojant *PCM*. Tokiu būdu nėra įvertinama natūralios kalbos savybės : balsiniai signalai dažniausiai būna žemos amplitudės. Taigi didžiausia balso signalų dalis sukoncentruota žemesnėje dinaminame balso diapazono dalyje.

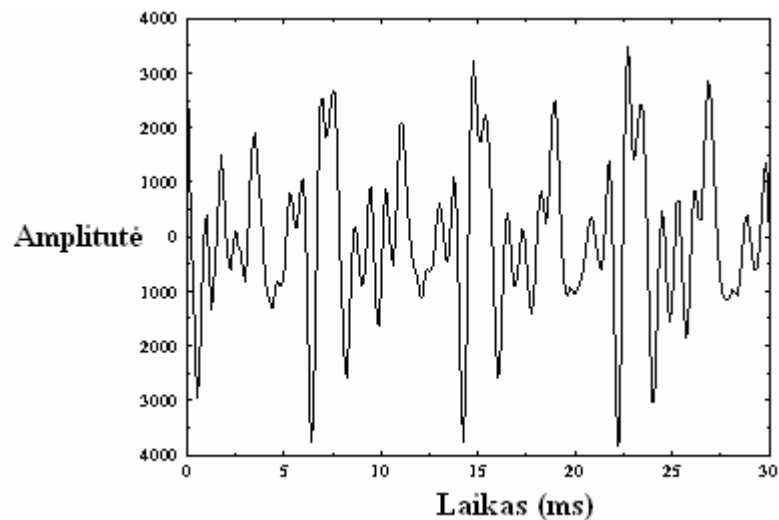
Norint optimaliai atgaminti balsą naudojant *PCM*, diskretizavimo procesas turi įvertinti tai, kad dauguma balsinių signalų yra žemos amplitudės. Norint tai įvertinti atitinkamai yra pertvarkoma stygų bei žingsnių struktūra. Žingsnių dydis yra mažesnis mažesnės amplitudės signalams. Diskretizacijos lygiai paskirstyti pagal logaritminę priklausomybę, duoda mažesnę rezoliuciją žemesnėms signalo amplitudėms.



**Pav. 8 Logaritminė diskretizacija**

Praktikoje, balso diskretizacija tiksliai neatitinka logaritminės kreivės. Šiaurės Amerikoje logaritminė funkcija vadinamas *M-law*. Koduojanti funkcija tiksliai aproksimuoja logaritminę kreivę. Žingsniai keičia dydį tik pereinant iš stygos į stygą. Europoje naudojama truputį kitokia, bet labai panaši *A-law* logaritminė funkcija. Atitinkamai, pagal tai kokią funkciją naudoja, vadinami ir kodekai. Dėl jų paprastumo jie dar yra naudojami ir šiais laikais, nors standartizuoti jau buvo 1960-ais metais *CCITT* komiteto kaip G.711 standartas [11].

Diferencialinė pulso kodo moduliacija (*Differential Pulse Code Modulation - DPCM*) kitas gerai žinomas diskretizacijos metodas [11]. Šis metodas pagrįstas tuo kad koduojama ne pačios signalo amplitudės reikšmės, bet skirtumas tarp tikrosios ir numatomos (nuspėjamos) reikšmės. Reikšmė gali būti nuspėjama pagal ankstesnes reikšmes. Šis technika grindžiama faktu, kad daugumuose balsinių signalų pastebima žymi koreliacija (tarpusavio ryšys) su prieš tai diskretizuotomis reikšmėmis.



Pav. 9 Tipinis žmogaus kalbos garsinis signalas

*DPCM* suglaudinimas priklauso nuo spėjimo technikos, kuo geresnė spėjimo technika tuo daugiau ir kokybiškiau bus suglaudinami įrašai.

Dar vienas metodas – adaptyvi diferencialinė pulso kodo moduliacija (*Adaptive Differential Pulse Code Modulation - ADPCM*) [7][8]. Tai pagerinta *DPCM*, kuri prisitaiko prie koduojamos kalbos charakteristikų. 1980-aisiais *CCITT* komitetas standartizavo 32 kbitų/s *ADPCM*, kuris žinomas *G721* pavadinimu. Šis kodekas atkuria kalbą beveik taip pat kaip ir 64 kbitų/s *PCM*. Vėliau buvo standartizuoti *G726* ir *G727* kodekai. Pastarieji buvo pritaikyti darbui su 40, 32, 24 ir 16 kbitų/s bitų dažniu.

## 2.4.2 Balsiniai kodekai

Įvertina tai, kaip garsinis signalas buvo sugeneruotas, ir bando išgauti iš koduojamo signalo jo parametrus. Šie parametrai ir būna perduodami dekoderiui.

Nesvarbu kokia kalba yra kalbama, visi žmonės naudoja tą patą kalbinį aparatą. Oras yra stumiamas iš plaučių, per balsinį traktą ir burną, priverčia vibruoti balso stygas, ko pasekoje, sukeliama garsiniai signalai. Turint omenyje tokią struktūrą būtų galima plaučius įsivaizduoti kaip garso šaltinį, o balsinį traktą kaip filtrą kuris kuria įvairiausius garsus kurie sudaro kalbą [13].

Norint suprasti kaip balso traktas paverčia orą, stumiamą iš plaučių, garsu, reikėtų apibrėžti kai kuriuos svarbius išsireiškimus. Fonemos yra apibrėžiamos kaip ribota individualių garsų aibė. Yra dviejų kategorijų fonemos skardžios bei duslios, kurias įvertina tiesinė prognozė (*Linear Predictive Coder - LPC*), kai analizuojama ir sintezuojama kalba.

Skardūs garsai dažnai yra balsiai, generuojami stumiamo oro per balso stygas. Pastarosios vibruoja kažkokiu tai periodiškumu sukeldamos serijas oro impulsų vadinamų balsaskylės impulsais. Garso aukštis priklauso nuo to kokių dažnių vibruoja balso stygos. Yra žinoma, kad vaikų ir moterų balsas yra aukštesnis negu vyrų. Taip yra, todėl, kadangi vyrų balso stygos vibruoja lėčiau.

Duslūs garsai dažniausiai yra priebalsiai. Jie turi mažiau energijos bei didesnius dažnius negu skardūs garsai. Duslūs garsai sukuriama kai oras yra išstumiamas per balsinį traktą sukuringa tėkme. Šios proceso eigoje balso stygos ne vibruoja, vietoj to, jos būna atviros kol garas yra skleidžiamas. Aukštis yra svarbus duslių garsų atributas.

Garsų skirstymas į duslius ir skardžius yra svarbus momentas analizės ir sintezės procesuose. Kitas svarbus komponentas, kuris įtakoja kalbėjimą, yra paties balsinio trakto forma. Kai asmuo kalba, balsinis traktas nuolat keičia formą labai ne dideliu dažnumu, ko pasekoje, sukuriama skirtingi garsai.

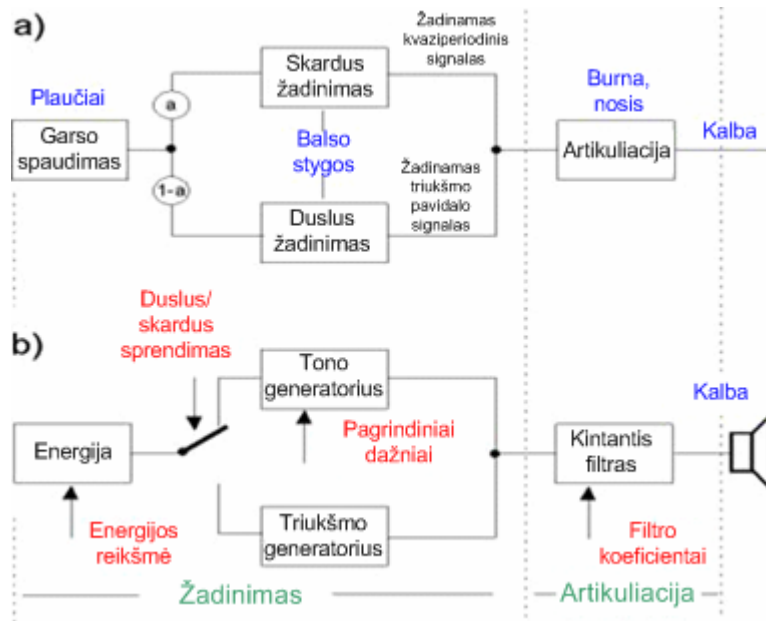
Paskutinis komponentas kuris įtakoja garso kūrimą, tai oro kiekis plaučiuose. Kuo didesnis oro tūris kuris stumiamas per balsinį traktą, tuo stipresnis garsas.

Pati idėja plaučius vadinti garso šaltiniu, o balsinį traktą filtru vadinama šaltinis-filtru modeliu garsui skleisti [10]. Šis modelis naudojamas tiesinėje prognozėje. Jis pagrįstas šaltinio atskyrimu nuo filtro idėja. Metodas naudojamas ir kodavime ir dekodavime ir yra išvestas iš kintančio diametro vamzdžio matematinės aproksimacijos. Keliaujančio oro per balso stygas sužadimas vadinamas šaltinio žadinimu. Šis oras tekėjimas gali būti periodinis, taip sukurdamas skardžius garsus, vibruojant balso stygomis, arba gali būti sukurtinis ir atsitiktinis, taip sukurdamas duslius garsus. *LPC* kodavimo procesas apima parametrų aibės nustatymą, reikalingų balsinio trakto modeliavimui, norint atkurti kalbą. Tokie kodekai persiunčia tik patį balsinio trakto modelį bei kai kuriuos nurodymus kokio tipo garsas turėtų būti atkuriamas.

Tam tikras šaltinis-filtru modelis yra naudojamas tiesinės prognozės kodavimo metode. Jis susideda iš dviejų pagrindinių dalių: analizės arba užkodavimo ir sintezės arba dekodavimo.

Analizės dalis apima kalbos signalų tyrimą ir skaidymą į segmentus arba blokus. Kiekvienas segmentas yra toliau tiriamas tam, kad būtų galima atsakyti į sekančius klausimus:

- Ar segmentas skardus ar duslus?
- Koks segmento aukštis?
- Kokie reikalingi parametrai norint sukurti filtrą kurio pagalba būtų galima atkartoti balsinio trakto garsus duodam segmentui?



Pav. 10 a) Žmogaus kalbos modelis, b) šaltinis-filtras modelis

Pav. matyti kokios kodeko dalys atitinka žmogaus kalbiniam aparatui. Visi balsiniai kodekai bando modeliuoti du dalykus: sužadimą ir artikuliaciją. Sužadimas, tai garso tipas kuris paduodamas į filtrą arba balsinį traktą ir artikuliacija tai sužadinto signalo transformacija į kalbą.

Šios klasės kodekai sugeba operuoti apie 2,4 kbitų/s bitų dažniu ir netgi žemesniu. Deja, jų atkuriamą kalbą labai skiriasi nuo natūralios kalbos. Pagrindiniai šie kodekai yra naudojami militaristinėse struktūrose, kur natūralus skambėjimas nėra toks svarbus kaip žemi bitų dažniai, leidžiantis įgyvendinti sudėtingus šifravimus ir informacijos apsaugos algoritmus. Daugiau informacijos apie balsinius kodekus galima rasti šioje literatūroje [10].

### 2.4.3 Hibridiniai kodekai

Hibridiniais kodekais siekiama užpildyti tarpą tarp banginių ir balsinių kodekų. Banginiai kodekai leidžia pasiekti aukštos kokybės kalbos atkūrimą naudojant 16 kbitų/s dažnius. Deja, tai ir yra jų minimali riba. Kita vertus, balsiniai kodekai sugeba išsaugoti suprantamą kalbą naudojant 2,4 kbitų/s ir mažesnius dažnius, bet negali atkurti natūralios kalbos skambėjimo.



Hibridiniai kodekai naudoja abiejų pastarųjų klasių principus. Jie naudoja tą pačią tiesinę prognozę kaip ir balsiniai kodekai, tik vietoj dviejų būsenų skardus/duslus modelio parametru parinkimo, papildomai bandoma parinkti žadinamą signalą, kurio bangos forma būtų kuo labiau panašesnė į originalą. Šį principą naudojantys kodekai vadinami sintezės-analizės kodekais. Pirmą kartą jie buvo pademonstruoti 1982 metais [14].

Sintezės-analizės kodekai dalina įeinantį kalbos signalą į 20 ms kadrus. Kiekvienam kadru yra nustatomi sintezės filtro parametrai. Po to yra nustatomi žadinimo šaltinio parametrai. Norint tai atlikti reikia surasti žadinimo šaltinį kuris paduotas į sintezės filtrą minimizuotą paklaidą tarp įėjimo kalbos ir atkuriamos kalbos signalų. Iš to ir seka analizės-sintezės pavadinimas – koderis analizuoja įėjimo signalą sintezuodamas daug įvairių jo aproksimacijų. Galų gale kiekvienam kadru koderis perduoda sintezės filtro parametru informaciją ir žadinimo šaltinį dekoderiui. Pastarajame šaltinis yra leidžiamas per filtrą ir atkuriamas kalba.

Panašus į sintezės-analizės kodekus yra *MPE (Multi-Pulse Excited)* kodekas. Skirtumas yra tas, kad *MPE* kodeke žadinimo šaltinis į filtrą paduodamas fiksuoto kiekio ne nuliniiais pulsais kiekvienam kalbos signalo kadru. Dar vienas giminingas kodekas yra *RPE (Regular Pulse Excited)*, kuris tai pat naudojamas ir *GSM* kodeke [9]. Kaip ir *MPE* kodekas, *RPE* kodekas naudoja ne nulinius pulsus žadinimo signalo kūrimui. Tačiau *RPE* kodekuose pulsai yra reguliariai paskirstytas tam tikrais fiksuotais intervalais. Tai reiškia, kad koderis turi nustatyti tik pirmo pulso poziciją ir visų pulsų amplitudę, o *MPE* kodekuose visų ne nulinių pulsų pozicijos kadruose, ir amplitudėse, turi būti nustatytos koderio ir perduotos dekoderiui. Taigi naudojant *RPE* kodekus reikia perduoti mažiau informacijos apie pulso pozicijas ir yra labai svarbus momentas mobiliuose sprendimuose, tokiose kaip *GSM*.

Nors *MPE* ir *RPE* kodekai suteikia kalbai gerą kokybę naudojant 10 kbitų/s ir didesnius dažnius, jie nėra tinkami naudojimui su mažesniais bitų dažniais. Tai yra, todėl, kad reikia perduoti nemažą kiekį informacijos apie žadinimo šaltinio pulsų pozicijas ir amplitudes.

Dabartiniu laiku plačiausiai naudojamas metodas gerai kalbos kokybei užtikrinti naudojant mažesnius bitų dažnius už 10 kbitų/s yra *CELP (Code Excited Linear Prediction)*. Šiame kodeke žadinimo šaltinis yra apibrėžiamas dideliu kvantatoriaus kodų vektoriumi. Vektoriau indeksui bei reikšmėms koduoti reikia mažiau informacijos negu, pavyzdžiui *GSM RPE* kodekuose [7][8].

## **2.5 Kalbinių kodekų standartų apžvalga**

Tarptautinės telekomunikacijų standartizavimo sąjungos (*Telecommunication Standardization Union -ITU*) standartizavimo sektorius yra viena iš organizacijų užsiimanti kokybiškų standartų priėmimu visose telekomunikacijų srityse. Organizacija įkurta 1993 metais kovo 1d. Ji

pakeitė prieš tai egzistavusį Tarptautinį telegrafijos ir telefonijos konsultavimo komitetą (*International Telegraph and Telephone Consultative Committee- CCITT*) kurio šaknys tęsiasi iki 1865 metų.

Apžvelgsiu populiariausius kalbos kodavimo standartus : *G.711, G.722, G.723.1, G.726, G.728, G.729, G.729A, GSM-AMR*.

### **2.5.1 G.711**

*ITU-T G.711* standartas skirtas balso signalams diskretizuotiems 8000 kartų per sekundę, panaudojant 8 bitų pulso kodo moduliacijos (*PCM*) glaudinimo metodą, glaudinimui. *G.711* koderis sukuria 64 kbps bitų srautą. Standartas turi dvi formas: *A-Law* ir *M-Law*. Pirmasis 13 bitų tiesinės *PCM* diskretizacijos matavimus verčia 8 bitų suspaustais logaritminės *PCM* diskretizacijos matavimais. *M-Law* kodekas verčia 14 bitų tiesinės diskretizacijos *PCM* matavimus 8 bitų suspaustais *PCM* matavimais.

### **2.5.2 G.722**

*ITU-T G.722* yra etaloninis kodekas skirtas plačiajuostės kokybės kalbos kodavimui. Kaip ir visi plačiajuosčiai kodekai signalas diskretizuojamas 16000 kartų per sekundę. *G.722* užtikrina kalbos ir garso signalų pralaidumą (dažnių juostos plotis) iki 7 kHz, tuo tarpu kai siaurajuosčiai kalbiniai kodekai tik 3.6 kHz.

*G.722* kodekas tyra pagrįstas *SB-ADPCM* (*Sub Band – Adaptive Differential Pulse Code Modulation*) principais. Signalas yra dalinamas į dvi dažnių juostas ir abiejuose juostose koduojami naudojant *ADPCM* metodą. Sistemos naudojančios *G.722* kodeką gali būti naudojamos darbui trijuose režimuose 64, 56 ir 48 kbitų/s.

### **2.5.3 G.723.1**

*G.723.1* yra *ITU-T* sudėtinio dažnio kodavimo standartas, skirtas, taip vadinamo, ilgo atstumo telekomunikacinio tinklo kokybės (toll quality), kalbos kodavimui (8000 kartų per sekundę). Tipinis šio kodeko taikymas yra paketinio ryšinio telefonijoje, tokiose kaip, Balsas per internetą (*Voice-over-Internet-Protocol – VoIP*). Šis kodavimas yra plačiai taikomas kalbos kodavimo komponentuose vaizdo konferencijų taikomose programose ir yra dalis viso *H.324* šeimos standartų.

Kodekas dirba dviem bitų dažniais, 5.3 ir 6.3 Kbps. Abudu bitų dažniai naudoja tą pačią trumpalaikę analizės techniką kalbos apdorojimui. Ilgalaikeje kalbos analizėje yra naudojami kiti algoritmai. 5.3 Kbps kodeke naudojami *ACELP* (*Algebraic Code Exited Linear Prediction*) principai, tuo tarpu, 6.3 Kbps kodeke, naudojami – *MP-MLQ* (*Multi Pulse-Maximum Likelihood Quantization*).

Kodekas dirba su 30 ms kadrtais. Be to, yra peržiūrima apie 60 matavimų į priekį (7.5 ms). Tokiu būdu susikaupia 37.5 ms signalo vėlinimas.

#### **2.5.4 G.726**

*G.726* dirba keturiais bitų dažnio režimais: 16, 24, 32 ir 40 Kbps. Tai banginis kodekas ir yra pagrįstas *ADPCM* kalbos kodavimo metodu. Šis kodekas priima signalą diskretizuotą 8000 kartų per sekundę ir suspaudžia jį, priklausomai koks pasirinktas bitų dažnis.

#### **2.5.5 G.728**

*ITU-T G.728* yra mažo vėlinimo kalbinio kodeko standartas skirtas, ilgo atstumo telekomunikacinio tinklo kokybės (toll quality), kalbos kodavimui (8000 kartų per sekundę). *G.728 Annex G (G.728 G)* yra fiksuoto taško kodeko specifikacija dibančio 16000 bitų per sekundę dažniu. *G.728 Annex I (G.728 I)* yra paketų praradimą slepianti technika, naudojama kartu su *G.728 G*. *G.728* kodekai yra plačiai naudojami taikomiosiose programose kurioms keliami labai dideli signalo vėlinimo reikalavimai. *VoIP* yra tipinis šio kodeko naudojimo pavyzdys. Tai labai patikimas kalbinis kodekas sugebantis atkurti labai geros kokybės kalbą, lyginant su 32 kbitų/s *ADPCM*.

*G.728* kodekų pagrindas, *LD-CELP (Low Delay-Code Excited Linear Prediction)*. Šis kodavimas naudoja tradicinį sintezės-analizės metodą, kuris naudojamas visuose *CELP* klasės kodekuose. Mažas kodeko vėlinimas yra pasiekiamas naudojant atgalinio pritaikymo prognozę bei stiprinimą.

#### **2.5.6 G.729**

*G.729* naudoja 8 Kbps *CS-ACELP (Conjugate-Structure Algebraic-Code-Excited Linear Prediction)* kalbos glaudinimo metodą patvirtintą *ITU-T. G.729 Annex A* yra supaprastinto sudėtingumo versija.

*G.729 AB* kalbinis kodekas buvo sukurtas naudojimui taikomose programose kur reikalinga vienu metu perduoti tiek balsą, tiek duomenis. Koderis apdoroja signalą 10 ms kadrais ir atlieka 5 ms išankstinę peržvilgą, ko pasekoje, sukaupiamas 15 ms algoritminį vėlinimą. Šio metodo įėjimas/išėjimas yra 16 bitų tiesinė *PCM* diskretizacijos reikšmės.

#### **2.5.7 G.729A**

*G.729 Annex A (G.729 A)* yra žemo bitų dažnio *ITU-T* kalbos kodavimo standartas, skirtas, ilgo atstumo telekomunikacinio tinklo kokybės (toll quality), kalbos kodavimui (8000 kartų per sekundę). *G.729A* algoritmas paprastesnis negu *G.729*. Abu kodekai yra suderinami tarpusavyje, pvz., *G.729* paketas gali būti išspausintas *G.729A* dekoderiu ir atvirkščiai. Tipinis panaudojimas – *VoIP*. Tai labai patikimas kalbinis kodekas sugebantis atkurti labai geros kokybės kalbą, lyginant su 32 kbitų/s *ADPCM*.

Kodekas dirba 8000 bitų/sekundę dažniu. Kaip ir G.729 kodekas, G.729A yra taip pat pagrįstas CS-ACELP principais. Dirba su 80 matavimų kadrais (10 msec). Be to, atlieka išankstinę 40 matavimų apžvalga (5 ms). Taigi gaunamas 15 ms vėlinimas.

## 2.5.8 GSM-AMR

GSM-AMR – adaptyvaus multi dažnio (*Adaptive Multi Rate- AMR*) kalbos kodavimo standartas pristatytas 3 Kartos Kompanijos Projekto (*3<sup>rd</sup> Generation Partnership Project – 3GPP*), kuris yra bendras, įvairių organizacijų, projektas skirtas, ilgo atstumo telekomunikacinio tinklo kokybės (toll quality), kalbos kodavimui (8000 kartų per sekundę). Šis kalbinis kodekas pagrindinai naudojamas kalbos glaudinime trečios kartos mobilioje telefonijoje.

Sugeba dirbti net aštuoniais skirtingais bitų dažniais : 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15 ir 4.75 Kbps. Su visais dažniais naudojamas ACELP metodas. Kodekas dirba su 160 matavimų kadrais (20 ms) ir nėra reikalinga išankstinė apžvalga. Algoritminis vėlinimas – 20 ms.

## 2.5.9 Santrauka

Sekančioje lentelėje išvardinami standartiniai kodekai ir glausta informacija apie juos: naudojamas glaudinimo metodas, bitų dažnis, vėlinimas ir standartizavimo metai.

| Standartiniai kalbiniai kodekai |                    |  |                                | Lentelė 1 |
|---------------------------------|--------------------|--|--------------------------------|-----------|
| Standartas                      | Glaudavimo metodas | Bitų dažnis, kbps                              | Kadras/Išankstinė apžvalga, ms | Metai     |
| G.711                           | PCM                | 64   | 0.125 / 0                      | 1972      |
| G.726                           | VBR-ADPCM          | 16, 24, 32, 40                                 | 0.125 / 0                      | 1990      |
| G.722                           | SB-ADPCM           | 48, 58, 64                                     | 0.125 / 1.5                    | 1988      |
| G.728                           | LD-CELP            | 16   | 0.625 / 0                      | 1994      |
| G.729                           | CS-ACELP           | 8  | 10 / 5                         | 1995      |
| G.723.1                         | MP-MLQ             | 5.3, 6.4                                       | 30 / 7.5                       | 1995      |
| G.729 Annex A                   | CS-ACELP           | 8  | 10 / 5                         | 1996      |
| GSM                             | RPE-LTP            | 13   | 20 / 0                         | 1987      |
| GSM-AMR                         | ACELP              | 12.2, 10.2, 7.95, 7.40, 6.70, 5.90, 5.15, 4.75 | 20 / 0                         | 1999      |
| IS-54                           | VSELP (TIA)        | 7.95   | 20 / 5                         | 1990      |
| PDC                             | VSELP              | 6.7  | 20 / 5                         | 1990      |
| IS-96                           | QCELP (TIA)        | 8.5, 4.2, 0.8                                  | 20 / 5                         | 1993      |
| PDC PSI                         | CELP               | 3.45   | 40 / 10                        | 1993      |
| FS – 1015                       | LPC – 10E          | 2.4  | 22.5 / 90                      | 1984      |
| FS – 1016                       | CELP               | 4.8  | 30 / 7.5                       | 1991      |
| MELP                            | MELP               | 2.4  | 22.5 / 23                      | 1997      |

## 3 Projektinė dalis

Magistratūros studijų metu sukurtos programinės įrangos techninė-projektinė dokumentacija. Joje pateikiamas pasirinkto sprendimo realizacijos kelias.

### 3.1 Sistemos apibūdinimas

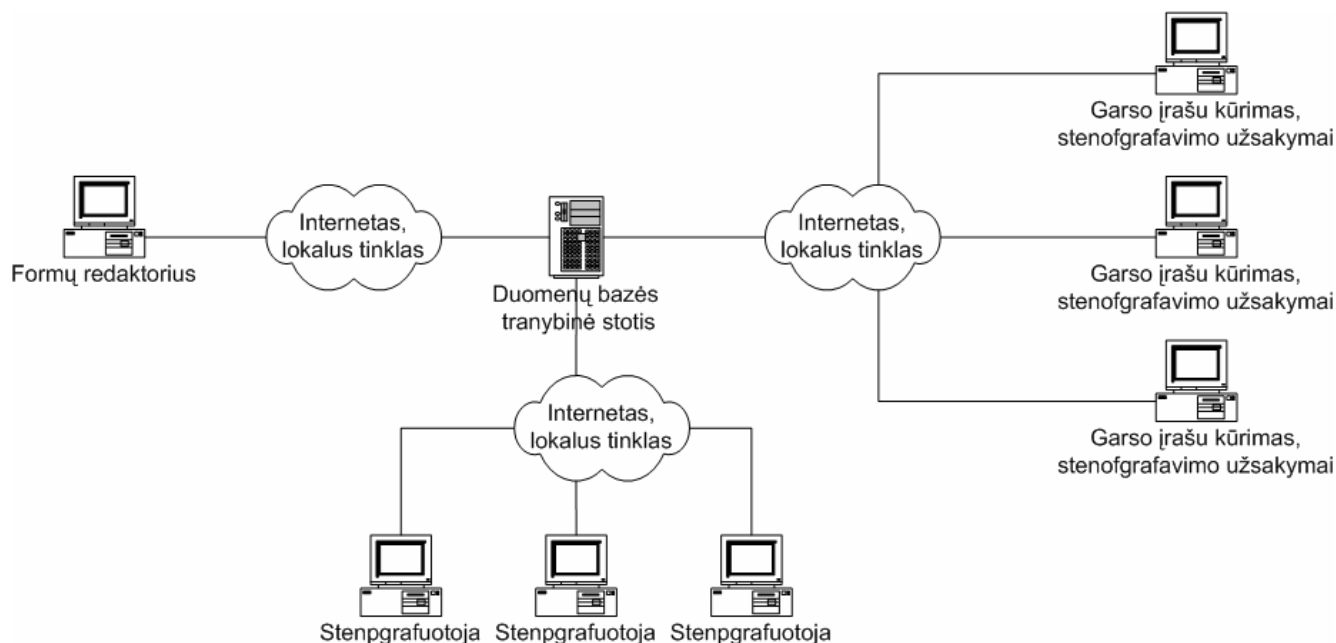
#### 3.1.1 Programų sistemos funkcijos

Pradžioje sistema buvo kuriama kaip elektroninės pacientų įrašų sistemos *MediPas / Siemens Melior* komponentas, kurio viena dalis įdiegiama į daktarų kabinetus, kita dalis į stenografuotojų kabinetą. Abu kabinetai sujungiami kompiuteriniu tinklu. Darbo eigoje buvo nuspręsta atsiskirti nuo *MediPas / Siemens Melior* sistemos ir toliau vystyti kuriamą produktą kaip atskirą *paskirstytą stenografavimo sistemą*.

Pirminis sistemos veikimas - gydytojas įdiktuoja diagnozę į garso bylą, naudodamasis mikrofonu ir įrašymui skirtu pedalu, arba atitinkamai nuspaudus sistemos meniu lange įrašymo mygtuką. Garso įrašas yra glaudinamas realiaame laike. Po to suformuojamas įrašas pagrindinėje duomenų bazėje, kuriame saugoma su pacientu ir gydytoju susijusi tekstinė informacija, o taip pat ir skaitmeninis garso įrašas. Šis įrašas taip pat prieinamas kitiems klientams, t.y. stenografuotojoms jų kabinete. Stenografuotoja, pamačiusi naują įrašą duomenų bazėje naudodamasi specialiu grotuvu, kuris atkuria išsaugotą duomenų bazėje garso įrašą ir grotuvo valdymo pedalu turi suvesti įdiktotą tekstą į reikiamą duomenų bazės "Diagnozės" formą, kurią nurodo gydytojas. Atlikus stenografavimą pažymima, kad šis įrašas yra apdorotas ir patalpinamas duomenų bazėje. Gydytojas turi priėjimą prie bet kurios tiek garsinės tiek stenografuotos diagnozės formos. Stenografuotoja turi priėjimą tik prie tam tikro senumo dokumentų. Laikas nustatomas sutartinai. Jį gali pakeisti sistemos administratorius. Daktaras ir sekretorė turi galimybę peržiūrėti diagnozės formas elektroninėje formoje, bei jas atspausdinti.

Po apsisprendimo atsiskirti nuo *MediPas / Siemens Melior* sistemos, išliko bendras veikimo principas, tik buvo nuspręsta produktą pritaikyti ne tik medicinos srityje. Sukurta programinė įranga turėtų tikti interviu, posėdžių, forumų, sesijų, derybų, paskaitų, mitingų, konferencijų garsiniams įrašams stenografuoti.

### 3.1.2 Sistemos kontekstas



Pav. 11 Sistemos kontekstas

Sistemoje išskiriame tokios dalys :

*Formų redaktorius* - šią dalį naudoja sistemos administratorius. Ji atsakingas už vizualų sistemoje naudojamų įvedimo formų redagavimą. Suteikia galimybę stumdyti laukus esančius formoje, keisti jų parametrus, pridėti naujus laukus, šalinti senus laukus. Tai pat šios dalies pagalba galima sukurti naujas formas arba šalinti senas.

*Stenografuotoja* – šią dalį naudoja organizacijos stenografuotojos stenografavimo darbams atlikti. Jos pagalba yra išklausomi garsiniai įrašai bei suvedama juos atitinkanti tekstinė informacija į sistemos administratorių paruoštus įvedimo formų šablonus.

*Garso įrašų kūrimas, stenografavimo užsakymai* - tikslas suteikti vartotojui galimybę įrašyti/atkurti garsą. Taip pat užsakyti pasirinktam garsiniam įrašui stenografavimą bei atspausdinti gautą rezultatą.

### 3.1.3 Vartotojo problemos

Problemos su kuriomis susiduria vartotojas:

- Daugkartinis kasečių panaudojimas, dėl to prastėja įrašo kokybė.
- Duomenų saugumas. Duomenys perduodami per kurjerį, todėl duomenų saugumas automatiškai mažėja.
- Duomenų judėjimo greitis. Priklauso nuo kurjerio bei atstumo iki stenografuotojų kambario.
- Užsakovo darbo našumas. Dirbant labai daug laiko užima formų pildymas.

### 3.1.4 Sistemos tikslai

Sukurta sistema turėtų:

- Užtikrinti ilgalaikę įrašų kokybę bei ilgalaikį saugojimą.
- Padidinti duomenų perdavimo saugumą.
- Padidinti duomenų apsikeitimo greitį tarp užsakovo ir stenografuotojų.
- Padidinti užsakovo darbo našumą.

## 3.2 Architektūros pateikimas

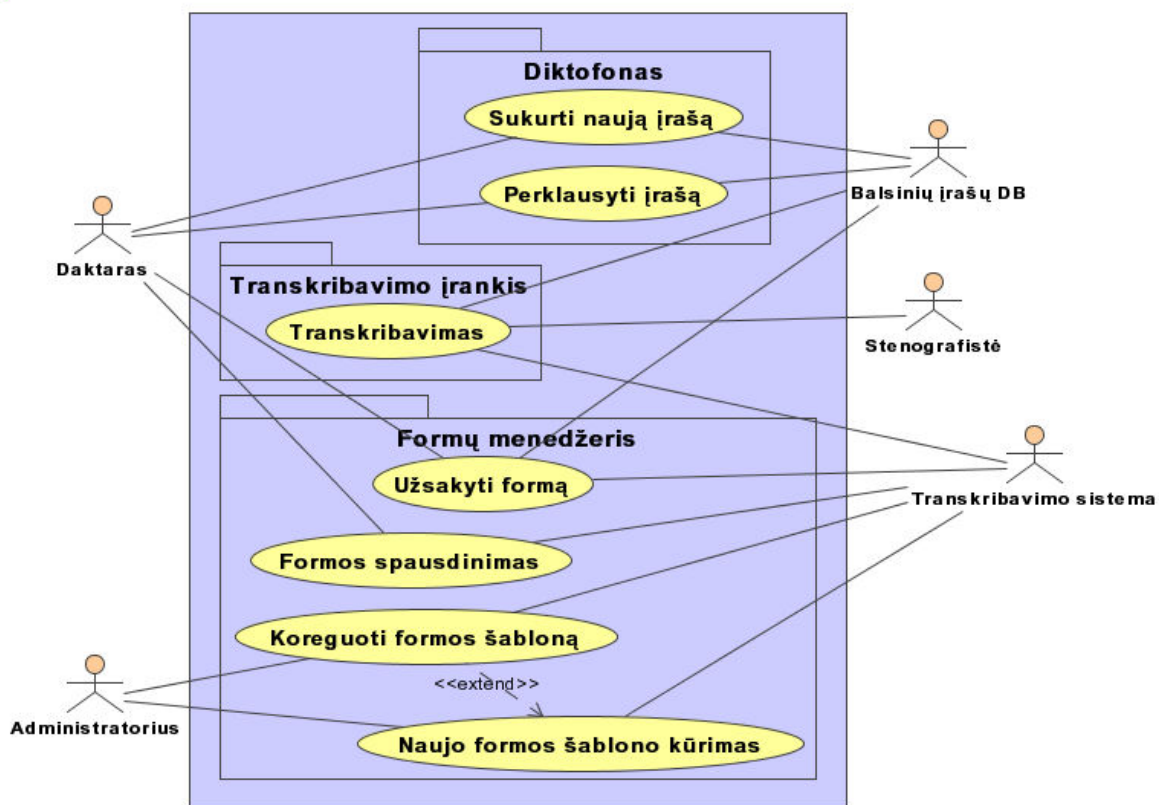
Šis dokumentas pateikia architektūrą keliais skirtingais požiūriais į sistemą : panaudojimo atvejų požiūris, loginis požiūris, procesų požiūris, paskirstymo požiūris. Šie požiūriai pateikti naudojant *MagicDraw 5.5* modelius kurie naudoja *UML (Unified Modeling Language)*.

### 3.2.1 Architektūros tikslai ir apribojimai

Sprendimai buvo priimti remiantis šiais apribojimais:

- Duomenų bazei saugoti bus naudojamas *MS SQL Server 2000*.
- Sistema kuriama *.NET* platformos pagrindu.
- Stenografuotojos ir kiti vartotojai turi galimybę jungtis tiek per *LAN* tiek per *Internet* tinklą.
- Sistema turi būti apsaugota nuo neautorizuoto priejimo.
- Sistema bus įgyvendinama kaip į *Web Service* sus orientuota architektūra.
- Garsiniams įrašams perklausyti bei įrašyti bus naudojama *DirectX* technologijos.
- Sistema kuriama norint apsiginti *KTU* magistrinį darbą.
- Į sistemą turi įeiti tokie nepriklausomi moduliai : diktofonas, stenografavimo įrankis, formų redagavimo įrankis.

### 3.2.2 Panaudojimo atvejų požiūris



Pav. 12 Architektūrai reikšminiai panaudojimo atvejai

#### 3.2.2.1 Naujo įrašo kūrimas

Šis panaudojimo atvejis leidžia vartotojams sukurti naujus garsinius įrašus bei patalpinti juos į sistemos duomenų bazę. Norint patalpinti įrašus į duomenų bazę vartotojas turi turėti autorizuotą priėjimą.

#### 3.2.2.2 Įrašo perklausymas

Šis panaudojimo atvejis leidžia vartotojams perklausyti garsinius įrašus. Norint perklausyti garsinius įrankius iš sistemos duomenų bazės, vartotojas turi turėti autorizuotą priėjimą.

#### 3.2.2.3 Transkribavimas

Šis panaudojimo atvejis apima transkribavimo procesą. Transkribavimas suprantamas kaip garsinio įrašo pervedimą į tekstą pagal atitinkamą tekstinio dokumento šabloną. Stenografuotojoms užduotis yra automatiškai paskirstomos pagal jų užimtumo požymį. Norint prieiti prie transkribavimo užsakymų reikalingas autorizuotas priėjimas.



### 3.2.2.4 Formos pildymo užsakymas

Šis panaudojimo atvejis suteikia galimybę vartotojams užsakyti garsinio įrašo transkribavimą į juos dominantį tekstinio dokumento formą. Jeigu toks užsakymas buvo jau vykdytas tai duomenys gražinami iškart.

### 3.2.2.5 Formos spausdinimas

Suteikia galimybę vartotojams atsispausdinti užpildytą formą (tekstinį dokumentą).

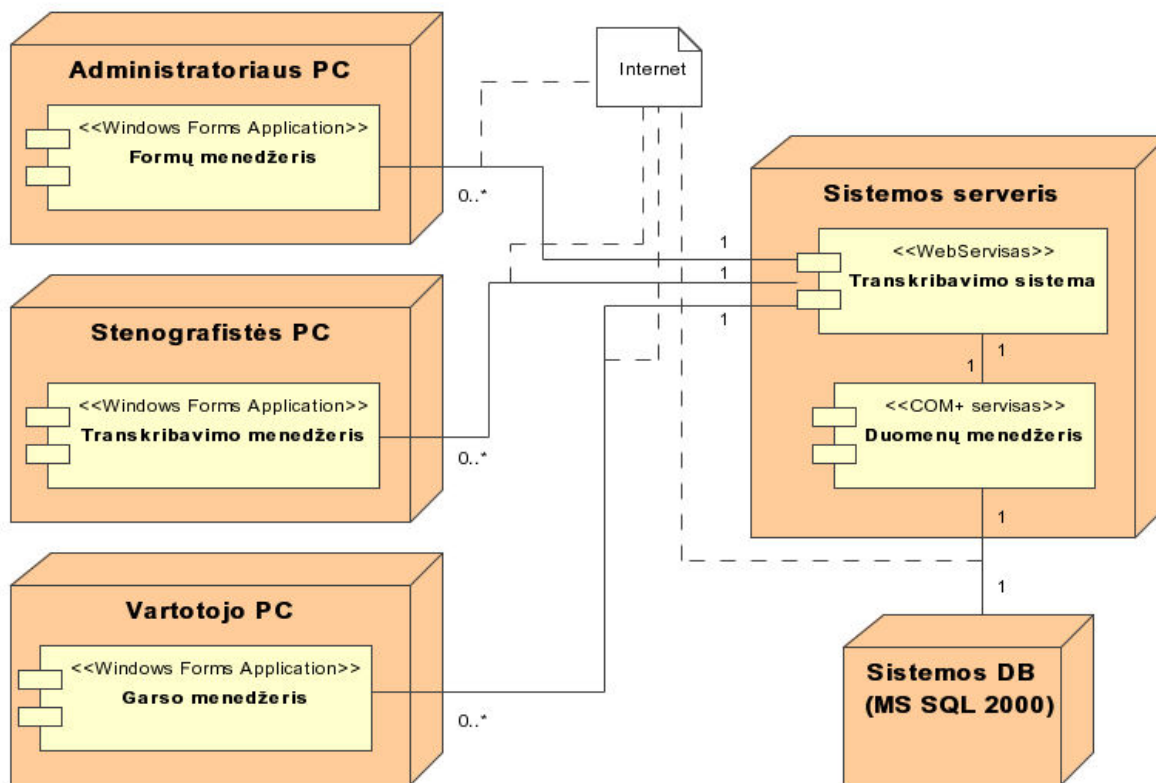
### 3.2.2.6 Formos šablono koregavimas

Šis panaudojimo atvejis suteikia galimybę administratoriui pakoreguoti sistemoje naudojamus formos šablonus.

### 3.2.2.7 Naujo formos šablono sukūrimas

Šis panaudojimo atvejis suteikia galimybę administratoriui sukurti naujus formų šablonus kurie bus naudojami sistemoje.

### 3.2.3 Paskirstymo požiūris



Pav. 13 Bendras sistemos komponentų vaizdas

*Formų menedžeris* - šį komponentą naudoja sistemos administratorius. Komponentas atsakingas už vizualų sistemoje naudojamų įvedimo formų redagavimą. Komponentas suteikia galimybę stumdyti laukus esančius formoje, keisti jų parametrus, pridėti naujus laukus, šalinti senus laukus. Tai pat šio komponento pagalba galima sukurti naujas formas arba šalinti senas.

*Transkribavimo menedžeris* – šį komponentą naudoja organizacijos stenografuotojos stenografavimo darbams atlikti. Jo pagalba yra išklausomi garsiniai įrašai bei suvedama juos atitinkanti tekstinė informacija į sistemos administratorių paruoštus įvedimo formų šablonus.

*Garso menedžeris* - komponento tikslas suteikti vartotojui galimybę įrašyti/atkurti garsą. Taip pat užsakyti pasirinktam garsiniam įrašui stenografavimą ir atspausdinti gautą rezultatą.

*Transkribavimo sistema* – tai pagrindinis sistemos komponentas kuris realizuoja priėjimo prie sistemos duomenų bazės. Visi komponentai norintys pateikti / paimti duomenis iš sistemos turi kreiptis į šį komponentą.

*Duomenų menedžeris* – tai komponentas atliekantis tiesioginius veiksmus su duomenų bazėje esančiais duomenimis. Jis yra naudojamas „*Transkribavimo sistema*“ komponento. Pasikeitus duomenų bazės struktūrai, tereiktų pakeisti tik šį komponentą. Tokiu būdu užtikrinama , kad vartotojams duomenų bazės struktūros pasikeitimai lieka nepastebimi bei nėra trukdomas jų darbas.

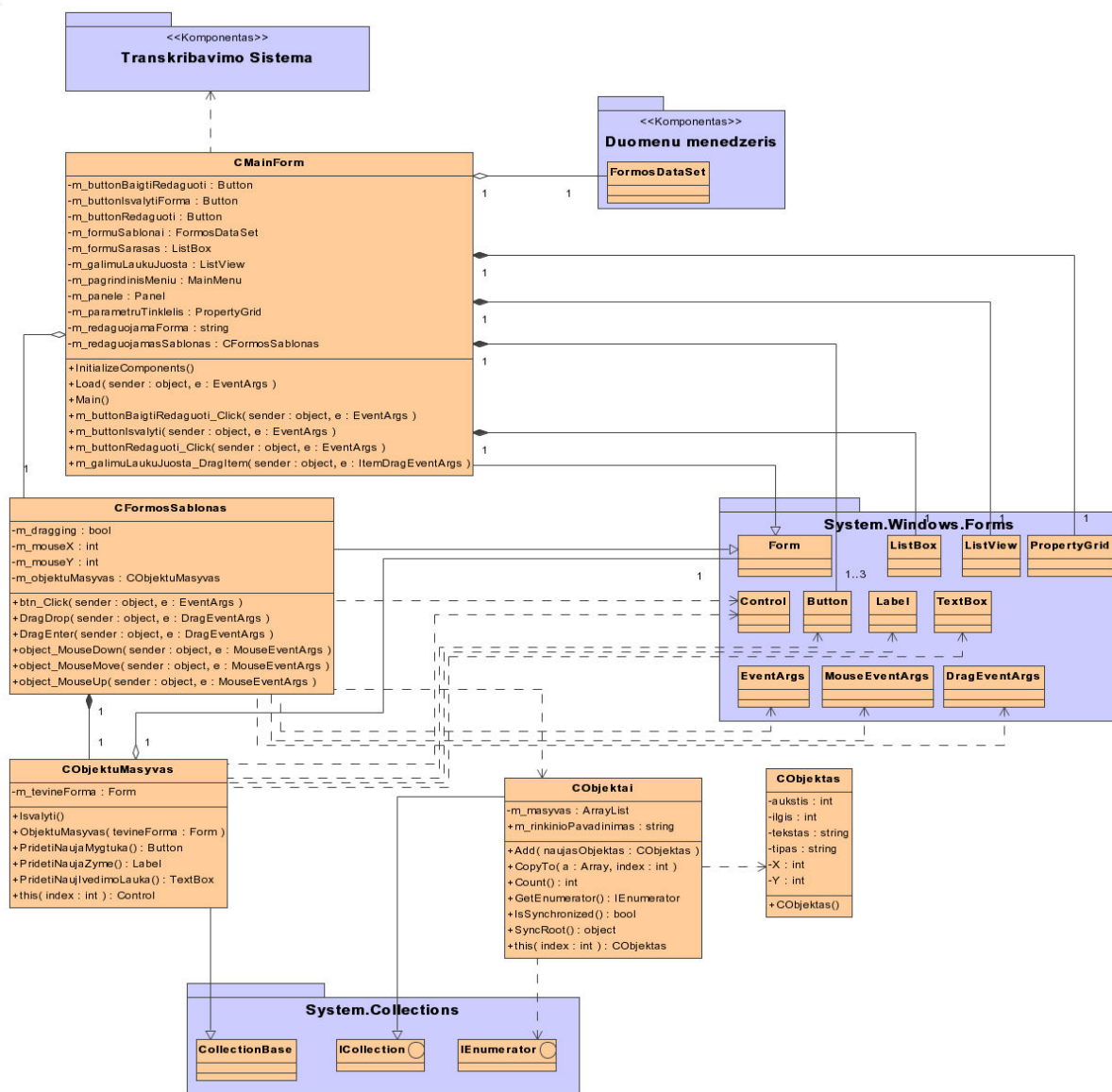
### **3.2.4 Komponentas “Formų menedžeris“**

#### **3.2.4.1 Apribojimai**

Komponentui turi būti suteiktas autorizuotas priėjimas prie sistemos duomenų bazės. Darbui su šiuo komponentu nėra reikalaujami dideli sistemos resursai. Įvedimo formos saugomos sistemos duomenų bazėje pasinaudojant *XML* standartu. Vienu momentu vieną formą gali redaguoti tik tai vienas administratorius. Tuo momentu kai forma redaguojama komponentai kuriems reikalingos šios formos naudoja senus formų šablonus.

#### **3.2.4.2 Struktūra**

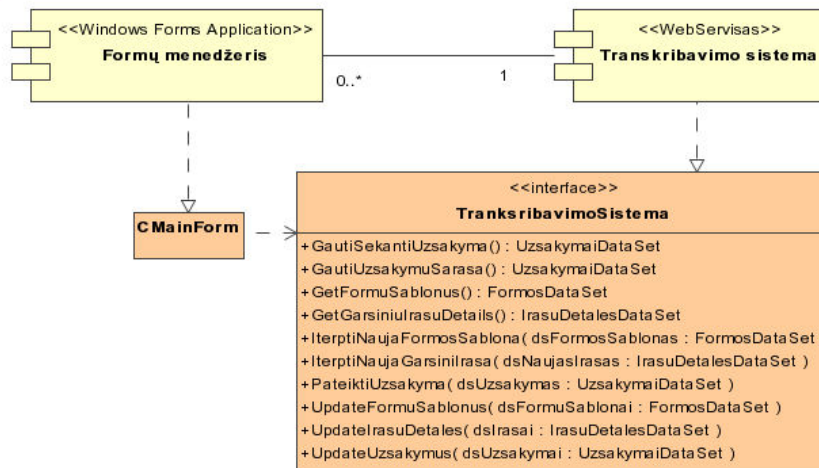
Šis komponentas naudojasi *.NET Windows Forms* klasių paketais. Taip pat komponentas naudoja „*Transkribavimo sistema*“ komponento teikiamomis paslaugomis. Visas vizualizavimas atliekamas pasinaudojant *Windows Forms* klasių paketais. Formų duomenys pasiekiami kreipiantis į „*Transkribavimo sistemos*“ komponentą panaudojant *.NET Enterprise Services* klases.



Pav. 14 „Formų menedžerio“ detali klasių diagrama

### 3.2.4.3 Sąveikavimas

Šis komponentas sąveikauja su kitu sistemos komponentu „Transkribavimo sistema“. Sąveikavimas vyksta *HTTP* protokolu. Sekančiame paveikslėlyje matyti koks naudojamas interfeisas tarp šių komponentų:

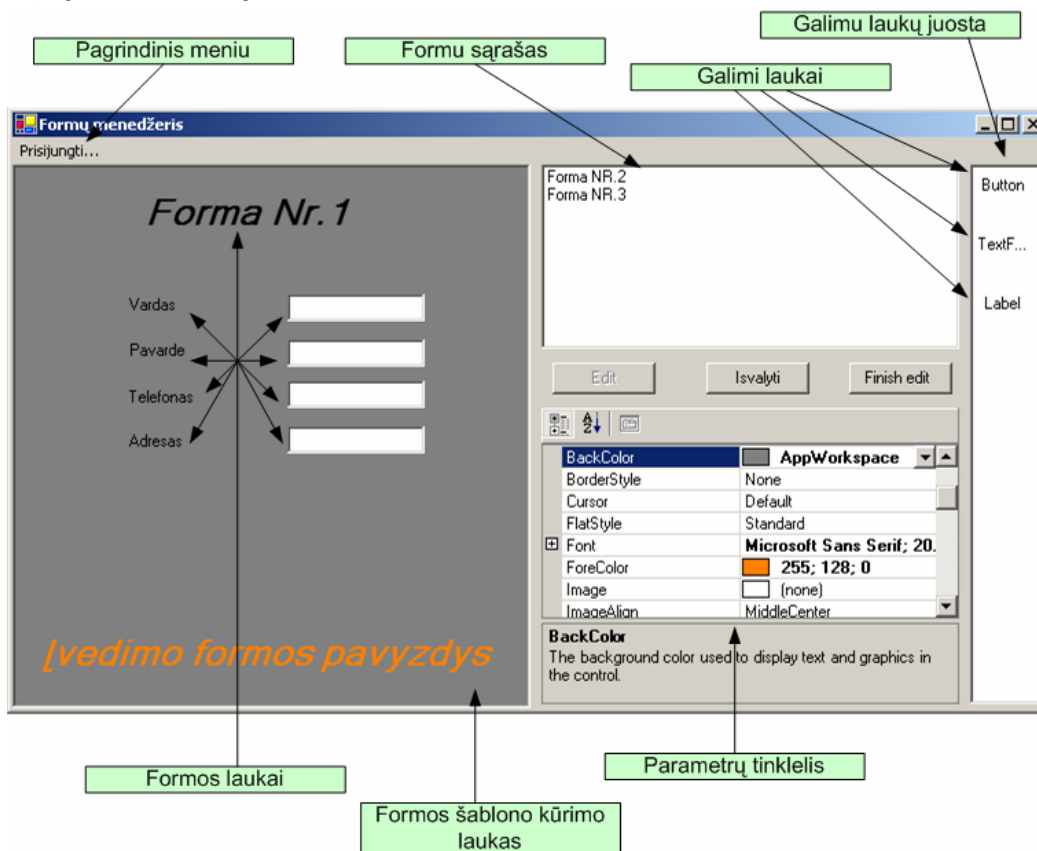


Pav. 15 Sąveikavimo tarp "Transkribavimo sistema" bei "Formų menedžeris" diagrama

### 3.2.4.4 Resursai

Naudoja sistemos duomenų bazę. Tuo momentu kai yra naudojama kažkokia tai pasirinkta įvedimo forma duomenų bazėje ji yra pažymima redagavimo būseną. Tokiu būdu uždraudžiama vienu metu dviem vartotojams redaguoti tą pačią formą.

#### 3.2.4.4.1 Sąsaja su vartotoju



Pav. 16 Grafinė vartotojo sąsaja formos šablono koregavimui

Norint pridėti naują lauką į formos šabloną reikia nutempti norimą lauką iš *galimų laukų juostos* į *formos šablono kūrimo lauką*. Norint keisti lauko parametrus reikia pelės pagalba spragtelėti ant norimo *formos lauko*. Lauko parametrai bus atvaizduoti *parametru tinklėlyje*, kuriame yra keičiami norimi *formos lauko* parametrai. Bet kuris laukas gali būti vizualiai stumdomas pelės pagalba.

### 3.2.5 Komponentas „Transkribavimo menedžeris“

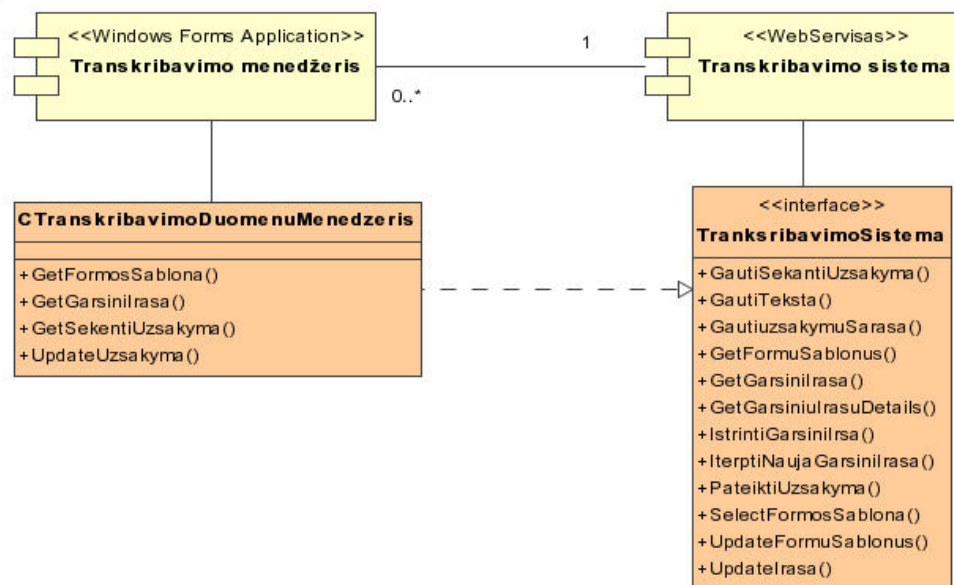
#### 3.2.5.1 Apribojimai

Darbo vietoje turi būti įdiegtas *DirectX* bei garso plokštė. Vienu metu prie sistemos gali jungtis daug organizacijos stenografisčių.

#### 3.2.5.2 Struktūra

Komponentas pakartotinai panaudoja klases garso įrašymui / atkūrimui bei formų atvaizdavimui kurios atitinkamai buvo naudojamos „*Garso menedžeris*“ ir „*Formų menedžeris*“ komponentuose.

#### 3.2.5.3 Sąveikavimas



Pav. 17 "Transkribavimo menedžerio" bei "Transkribavimo sistemos" sąveikavimas

#### 3.2.5.4 Resursai

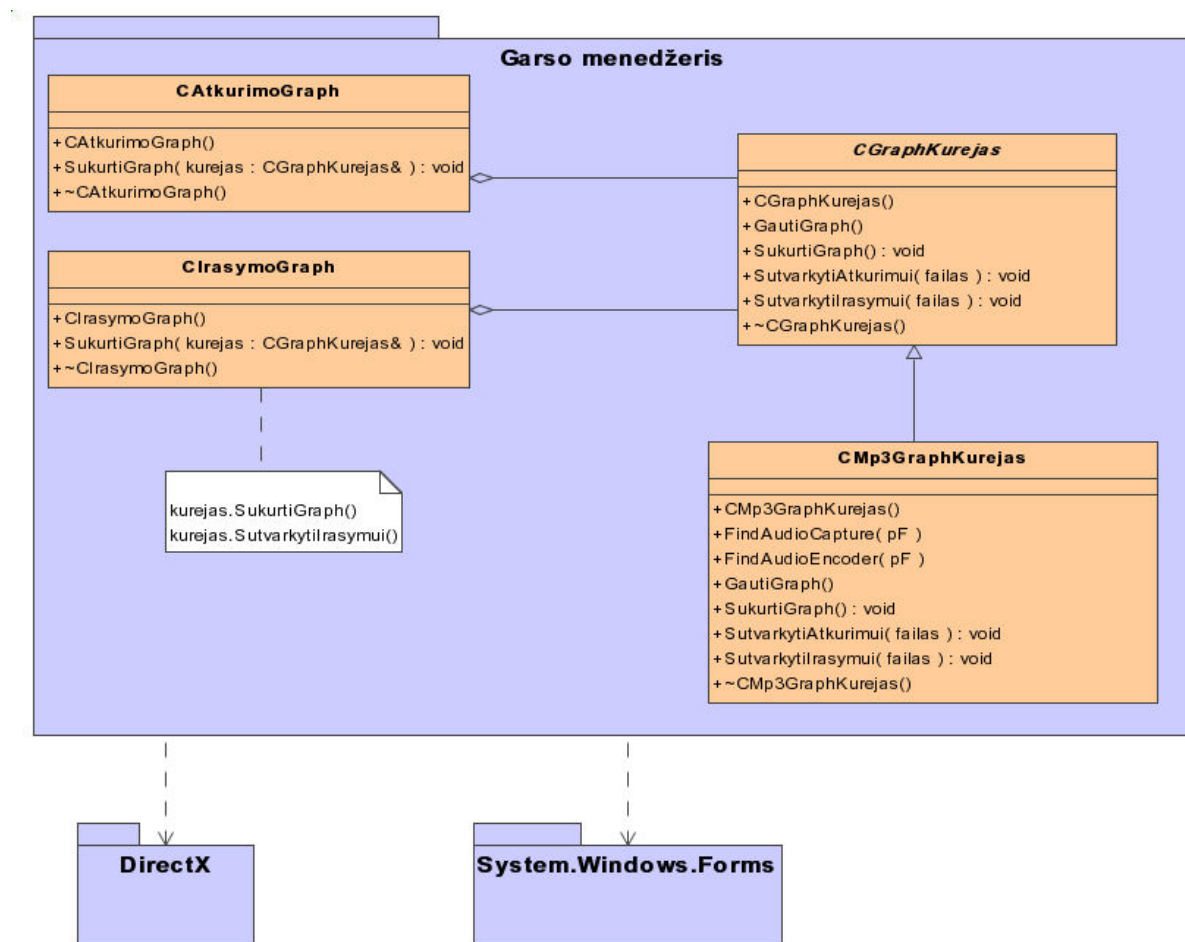
Komponentas reikalingus duomenis gauna iš sistemos duomenų bazės „*Transkribavimo sistema*“ komponento pagalba. Tuo momentu kai yra transkribuojamas įrašas duomenų bazėje jis pažymimas kaip transkribuojamas. Jeigu dėl kažkokių priežasčių transkribavimo pabaigti nepavyko

užsakymas yra atnaujinamas. Baigusi darbą su vienu įrašu stenografistė praneša apie tai sistemai ir gauna sekantį eilėje laukianti užsakymą.

### 3.2.6 Komponentas „Garso menedžeris“

#### 3.2.6.1 Struktūra

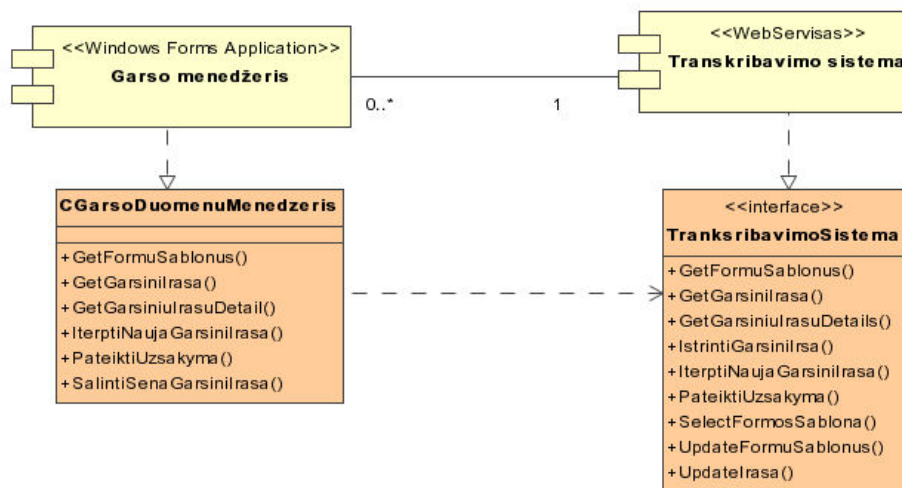
Komponentui realizuoti naudojamos *.NET Windows Forms* klasių paketo klasės, *Microsoft DirectX* paketo klasės. Taip pat komponentas naudoja „*Transkribavimo sistema*“ komponento teikiamomis paslaugomis. Visas vizualizavimas atliekamas pasinaudojant *Windows Forms* klasių paketais. Formų duomenys pasiekiami kreipiantis į „*Transkribavimo sistemos*“ komponentą panaudojant *.NET Enterprise Services* klases.



Pav. 18 Garso menedžerio pagrindinių klasių diagrama

#### 3.2.6.2 Sąveikavimas

Šis komponentas sąveikauja su kitu sistemos komponentu „*Transkribavimo sistema*“. Sąveikavimas vyksta *HTTP* protokolu. Sekančiame paveikslėlyje matyti koks naudojamas interfeisas tarp šių komponentų:

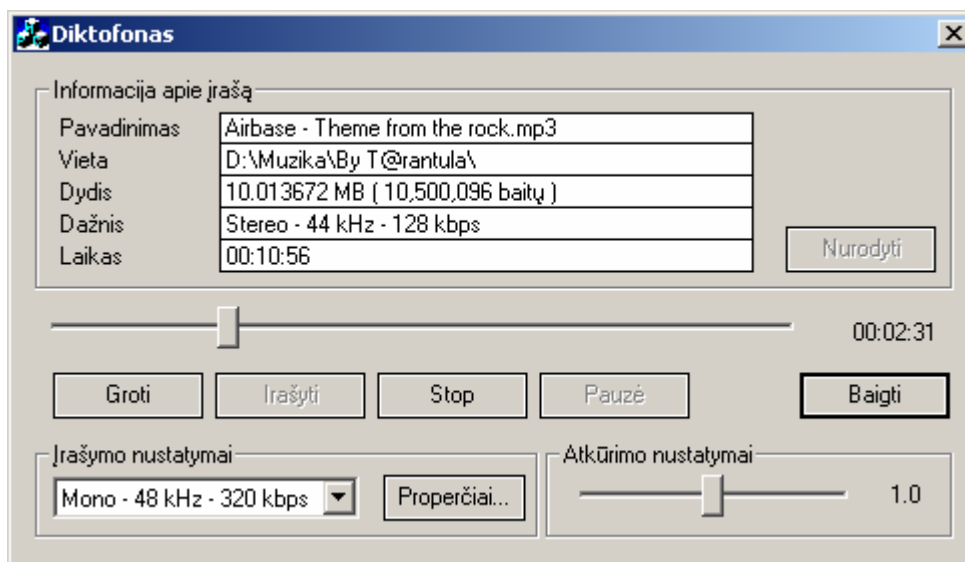


Pav. 19 Sąveikavimo tarp "Trankribavimo sistema" bei "Garso menedžeris" diagrama

### 3.2.6.3 Resursai

Komponentas naudoja sistemos duomenų bazę. Iš jos gauna garsinius įrašus bei išsaugo naujus arba pakeistus garsinius įrašus. Taip pat pateikia transkribavimo užsakymus.

#### 3.2.6.3.1 Sąsaja su vartotoju



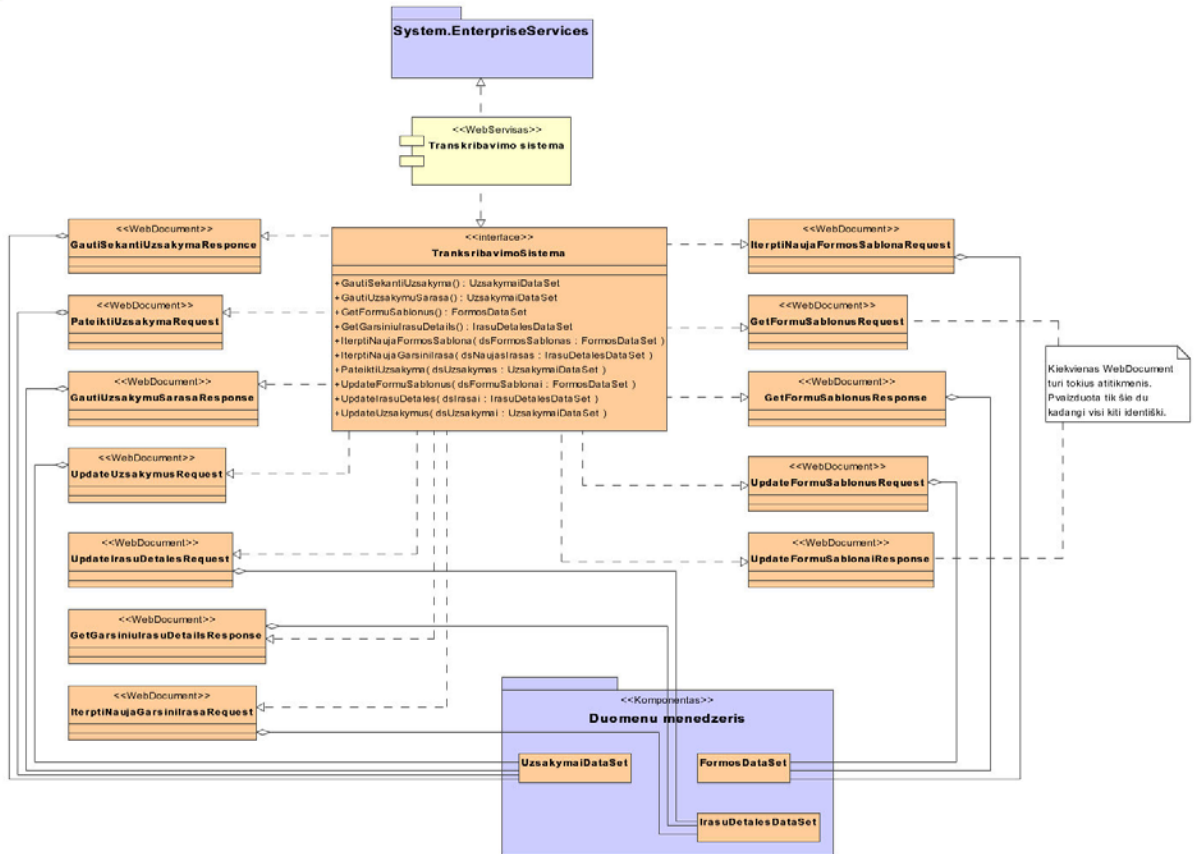
Pav. 20 "Garso menedžeris" komponento grafinė sąsaja su vartotoju Garsinio įrašo atkūrimas.

Darbo pradžioje mygtuko pagalba „Nurodyti“ nurodomas garsinis įrašas su kuriuo norima dirbti. Toliau mygtukų *Groti*, *Įrašyti*, *Stop*, *Pauzė* pagalba yra valdomas garsinis įrašas. Jeigu dirbama atkūrimo režime, tai galima atkūrimą sulėtinti arba pagreitinti. Tas atlieka darant pakeitimus *Atkūrimo nustatymų* grupėje. Jeigu dirbama įrašymo režime, galima pasirinkti norimą įrašymo kokybę. Pasirinkimas atliekamas *Įrašymo nustatymų* grupėje.

### 3.2.7 Komponentas „Transkribavimo sistema“

#### 3.2.7.1 Struktūra

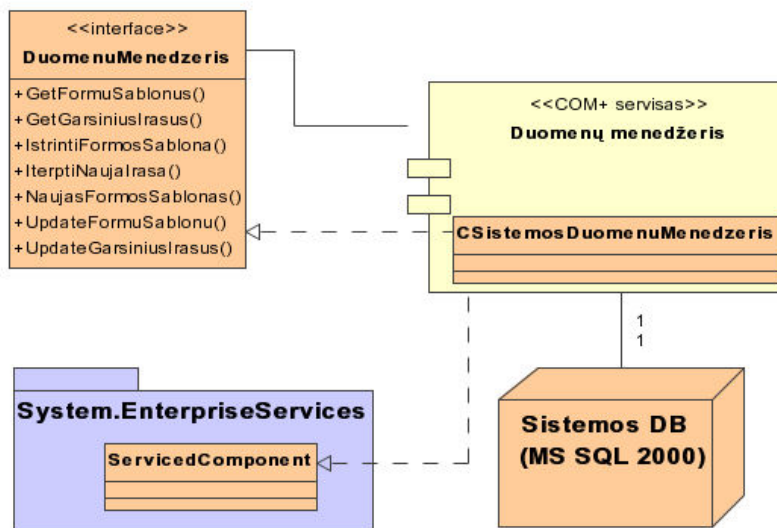
Komponentas bus realizuotas kaip *.NET Web Servisas*. Toliau pateikta komponento bei pranešimų naudojamų duomenims apsiukeisti struktūra :



Pav. 21 "Transkribavimo sistema" komponento struktūra.



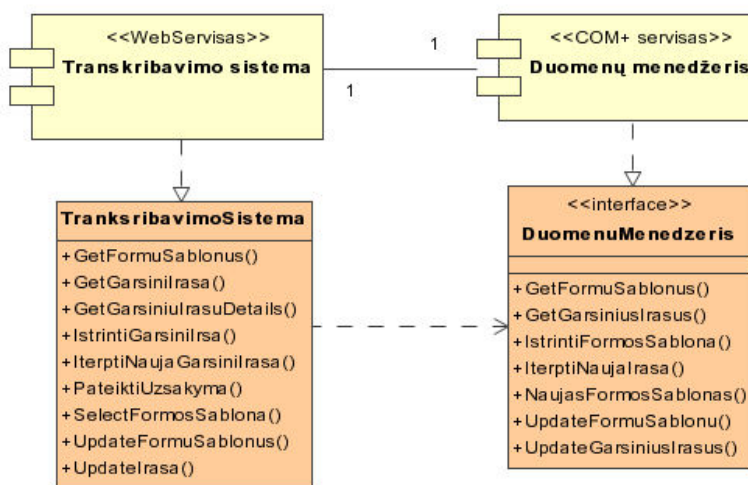
### 3.2.8 Komponentas „Duomenų menedžeris“



Pav. 22 "Duomenų menedžeris" komponento struktūra

#### 3.2.8.1 Sąveikavimas

Komponentas sąveikauja su kitos sistemos komponentu „*Transkribavimo sistema*“. Sekančioje diagramoje pavaizduota sąveikavimo struktūra:



Pav. 23 "Duomenų menedžerio" ir "Transkribavimo sistema" komponentų sąveika

## 4 Tyrimo dalis

Šiame skyriuje apžvelgiami kalbos kokybės vertinimo kriterijai bei atliekamas populiariausių šiuo metu kalbinių kodekų atkuriamos kalbos kokybės vertinimas.

Kalbos kokybė gali būti matuojama subjektyviai ir objektyviai. Subjektyvus įvertinimas yra gaunamas tiesiog klausantis testinių kalbos įrašų. Objektyvus įvertinimas yra išskaičiuojamas iš užkoduotos kalbos parametru.

Tipiškas subjektyvaus įvertinimo būdų sąrašas būtų toks: nuomonių vidurkio įvertis (*Mean Opinion Score – MOS*), diagnostinis ritmo testas (*Diagnostic Rhyme Test – DRT*) ir palyginimo poromis testas (*Paired Comparison Test – PCT*) [12]. Daugiausiai paplitęs ir dažniausiai naudojamas yra nuomonių vidurkio įvertis. Šis įvertis gaunamas išvedus testų rezultatų vidurkį. Testų rezultatai gaunami grupei klausytojų įvertinus jų pačių įspūdį (klausantis įrašų) sekančioj penkiabalėj sistemoj :

|   | Nuomonių vidurkio įvertinio vertinimo sistema            | Lentelė 2 |
|---|--|-----------|
| 5 | Puiki kokybė, nėra pastebimų pablogėjimų                 |           |
| 4 | Gera kokybė, labai silpnas pablogėjimas                  |           |
| 3 | Pakenčiama kokybė, pastebimi, bet priimtini pablogėjimai |           |
| 2 | Silpna kokybė, stiprūs pablogėjimai                      |           |
| 1 | Bloga kokybė, stipriai iškreipta kalba                   |           |

Diagnostiniame ritmo teste patyrę klausytojai bando atskirti pora trumpų, panašiai tariamus žodžius, pvz.: „meat“ ir „beat“, „sheet“ ir „cheat“. Palyginimo poromis teste patyrę klausytojai paprašomi iš dviejų koderių (kodavimui skirta kodeko dalis) pasirinkti vieną kuriam jis teiktų pirmenybę.

Plačiai naudojami objektyvaus įvertinimo būdai yra pagrįsti vidutine kvadratine paklaida. Populiariausias iš jų signalas-triukšmas proporcija (*signal-to-noise ratio – SNR*) [12] :

$$SNR = 10 \log_{10} \left\{ \frac{\sum_{n=0}^{M-1} s^2(n)}{\sum_{n=0}^{M-1} [s(n) - \hat{s}(n)]^2} \right\}, \quad (1)$$

kur  $s(n)$  – originalios kalbos signalo duomenys,  $\hat{s}(n)$  – sugludintos kalbos duomenys ir  $M$  – matavimų kiekis. Šis būdas vadinamas ilgalaikiu rekonstruojamos kalbos tikslumo įverčiu ir yra linkęs paslėpti laikinus rekonstravimo triukšmus. Laikini svyravimai gali būti lengviau aptinkami ir įvertinti naudojant taip vadinama trumpalaikį įvertį kiekvienam  $N$  matavimų ilgio kalbos segmentui.

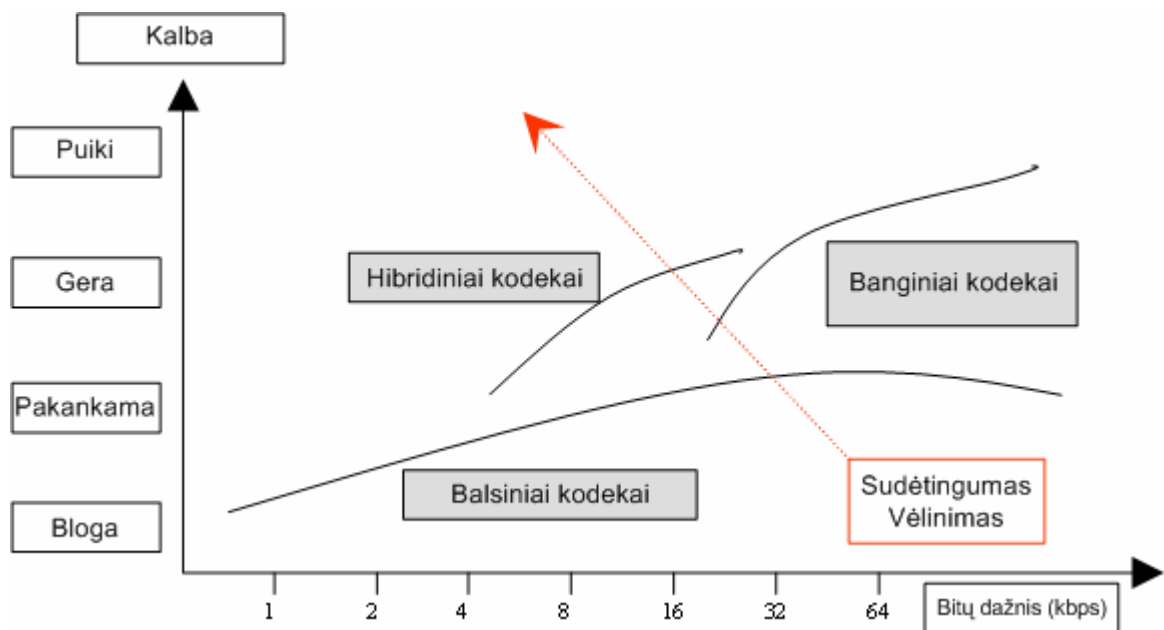
Segmentinis įvertis apibrėžiamas kaip šių trumpalaikių įverčių dB-vidurkis. Sekančioje lygybėje segmentai yra N matavimo ilgio bei išviso L tokių segmentų:

$$SEGSNR = \frac{10}{L} \sum_{i=0}^{L-1} \log_{10} \left\{ \frac{\sum_{n=0}^{N-1} s^2(iN + n)}{\sum_{n=0}^{N-1} [s(iN + n) - \hat{s}(iN + n)]^2} \right\}, \quad (2)$$

Daugiau informacijos apie subjektyviu ir objektyvius įvertinimo metodus galima rasti čia [12].

24 Pav. matosi kalbinių kodekų palyginimas pagal tris pagrindinius parametrus: kalbos kokybę, bitų dažnis ir sudėtingumas (vėlinimas).

Baginiai kodekai paprastai suteikia aukštesnę kalbos kokybę, mainais į didelius bitų dažnius. Tuo tarpu balsiniai kodekai stipriai sumažina bitų dažnį, mainais į kalbos kokybę.

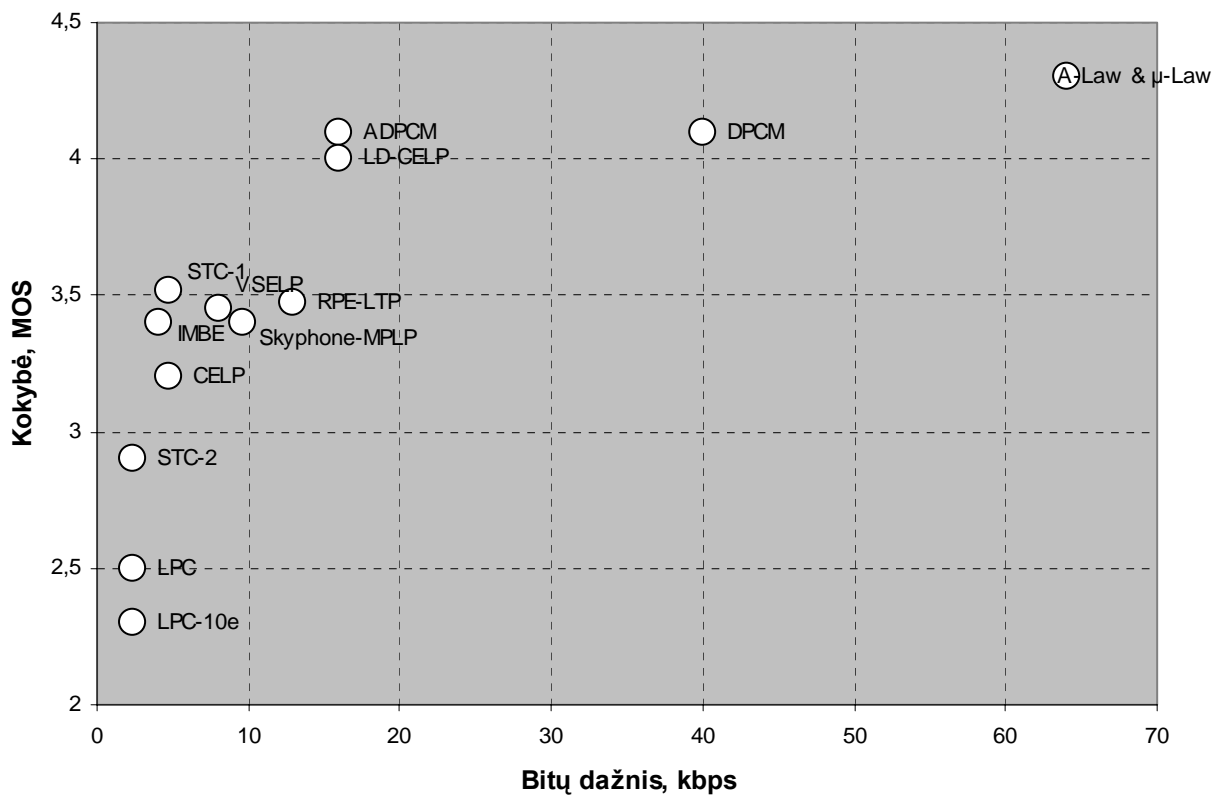


Pav. 24 Kalbinių kodekų glaudinimo palyginimas

Sekančioje lentelėje išvardinami standartiniai kalbos glaudinimo metodai bei jų našumo parametrai. Matyti, kad balsiniai metodai, tokie kaip *LPC*, sumažina bitų dažnį iki 2,4 kbitų/s mainais į kalbos kokybę. Hibridiniai metodai sugeba nežymiai pagerinti kalbos kokybę mainais į metodo sudėtingumą.

Glaudavimo metodų našumas ir sudėtingumas Lentelė 3

| Kodavimo metodas                        | Bitų dažnis, kbps | Kokybė, (MOS) [0 – 5] | Sudėtingumas, MIPS |
|---|-------------------|-----------------------|--------------------|
| <i>A-Law &amp; <math>\mu</math>-Law</i> | 64                | 4,3                   | 0.01               |
| <i>DPCM</i>                             | 40                | 4,1                   | 6                  |
| <i>ADPCM</i>                            | 16 – 32           | 4,1                   | 10 – 16            |
| <i>LPC</i>                              | 2,4 – 4,8         | 2,5                   | 12-20              |
| <i>CELP</i>                             | 4,8               | 3,2                   | 13 – 25            |
| <i>LD-CELP</i>                          | 16                | 4,0                   | ~19                |
| <i>RPE-LTP</i>                          | 13                | 3,47                  | 6                  |
| <i>Skyphone-MPLP</i>                    | 9,6               | 3,4                   | 11                 |
| <i>VSELP</i>                            | 8                 | 3,45                  | 13,5               |
| <i>STC-1</i>                            | 4,8               | 3,52                  | 13                 |
| <i>IMBE</i>                             | 4,15              | 3,4                   | 3                  |
| <i>STC-2</i>                            | 2,4               | 2,9                   | 13                 |
| <i>LPC-10e</i>                          | 2,4               | 2,3                   | ~7                 |
| <i>LPC-LSP</i>                          | 0,8               | -                     | ~20                |



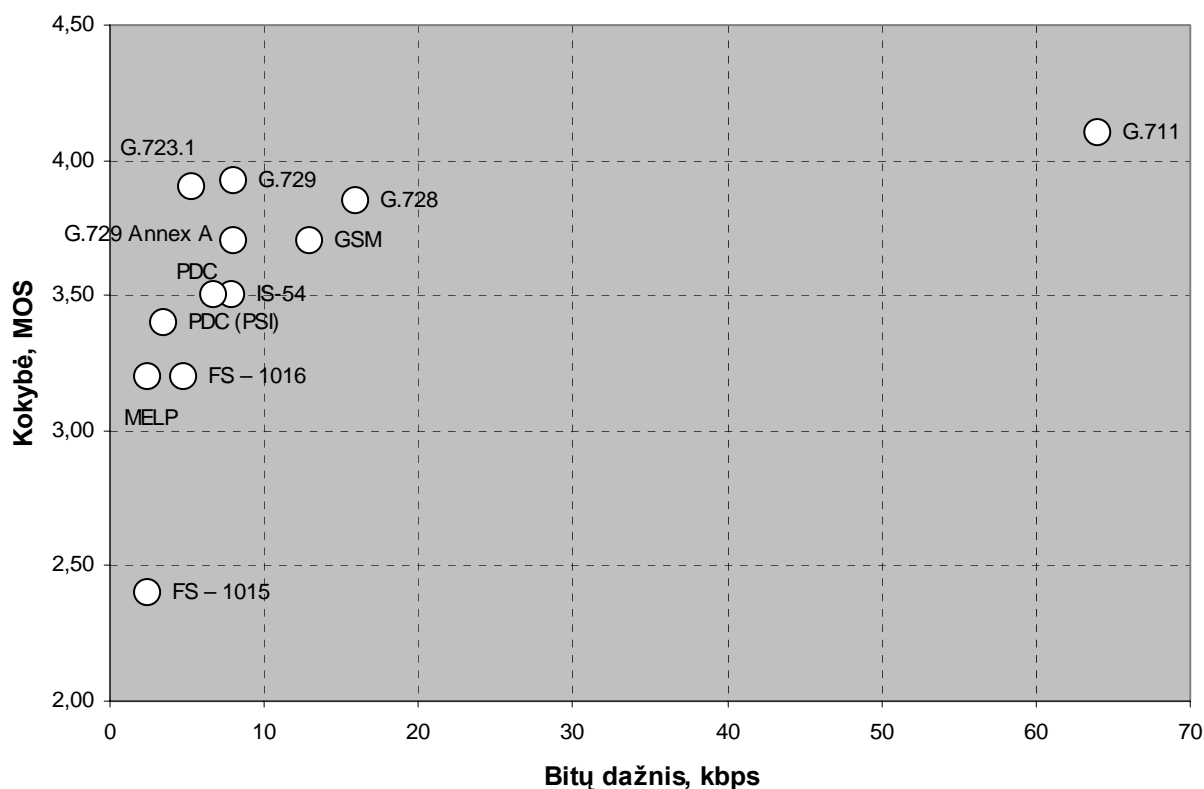
Pav. 25 Glaudinimo metodų bitų dažnio ir kokybės palyginimas

Sekančioje lentelėje ir grafike išvardijami standartinių kalbinių kodekų kokybės įverčiai.

Kalbinių kodekų kokybės įverčiai

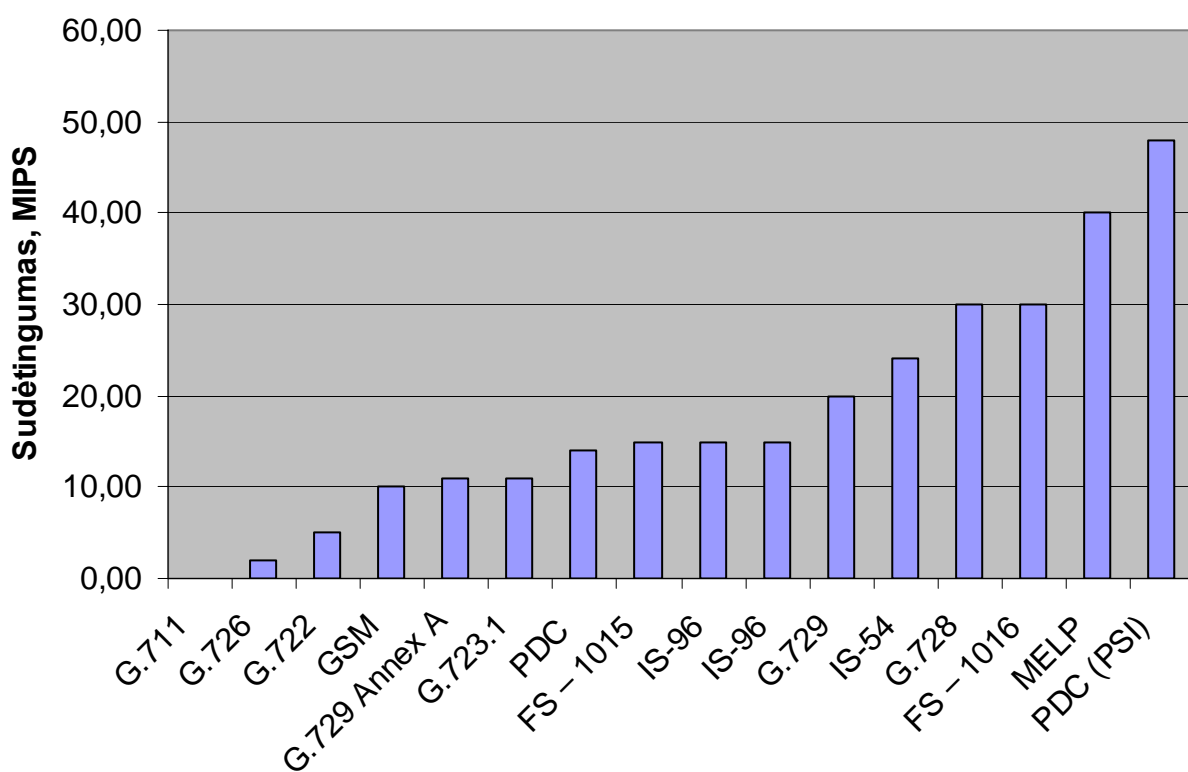
Lentelė 4

| Standartas    | Glaudavimo metodas | Bitų dažnis, kbps | Kadras/Išankstinė apžvalga, ms | Sudėtingumas, MIPS | MOS, [0-5] |
|---------------|--------------------|-------------------|--------------------------------|--------------------|------------|
| G.711         | PCM                | 64                | 0.125 / 0                      | 0.01               | 4.10       |
| G.726         | VBR-ADPCM          | 16, 24, 32, 40    | 0.125 / 0                      | 2                  | 3.85       |
| G.722         | SB-ADPCM           | 48, 58, 64        | 0.125 / 1.5                    | 5                  | 4.10       |
| G.728         | LD-CELP            | 16                | 0.625 / 0                      | 30                 | 3.85       |
| G.729         | CS-ACELP           | 8                 | 10 / 5                         | 20                 | 3.92       |
| G.723.1       | MP-MLQ             | 5.3, 6.4          | 30 / 7.5                       | 11                 | 3.90       |
| G.729 Annex A | CS-ACELP           | 8                 | 10 / 5                         | 11                 | 3.70       |
| GSM           | RPE-LTP            | 13                | 20 / 0                         | 10                 | 3.70       |
| IS-54         | VSELP (TIA)        | 7.95              | 20 / 5                         | 24                 | 3.50       |
| PDC           | VSELP              | 6.7               | 20 / 5                         | 14                 | 3.50       |
| IS-96         | QCELP (TIA)        | 8.5, 4.2, 0.8     | 20 / 5                         | 15                 | 3.30       |
| PDC           | PSI – CELP         | 3.45              | 40 / 10                        | 48                 | 3.40       |
| FS – 1015     | LPC – 10E          | 2.4               | 22.5 / 90                      | 15                 | 2.40       |
| FS – 1016     | CELP               | 4.8               | 30 / 7.5                       | 30                 | 3.20       |
| MELP          | MELP               | 2.4               | 22.5 / 23                      | 40                 | 3.20       |



Pav. 26 Standartinių kalbinių kodekų palyginimas pagal bitų dažnį ir kalbos kokybę

Sekančiame grafike atspindi standartinių kalbinių kodekų sudėtingumas.

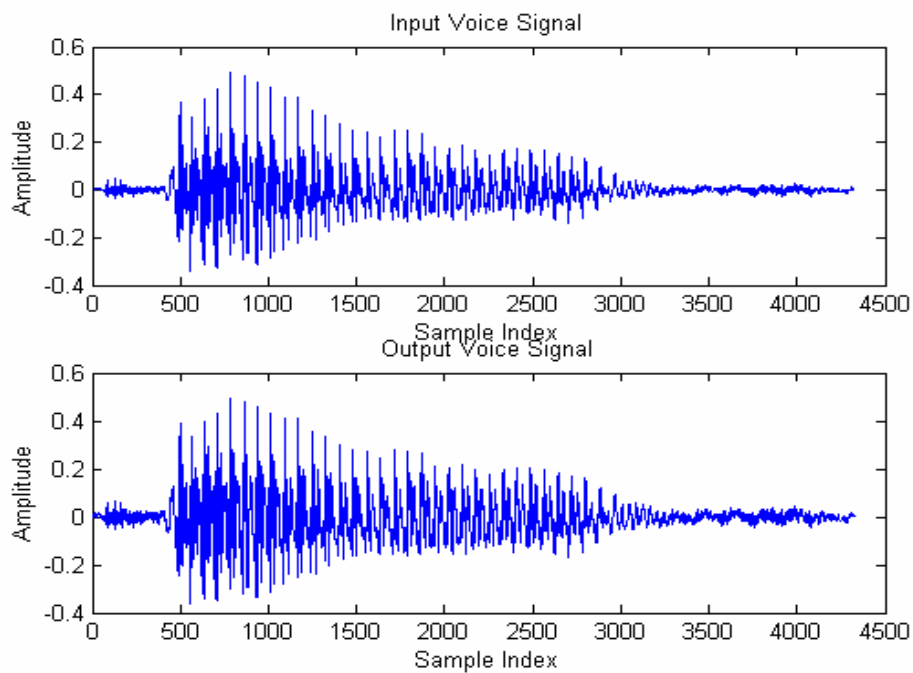


Pav. 27 Standartinių kalbinių kodekų sudėtingumas

## 5 Eksperimentinė dalis

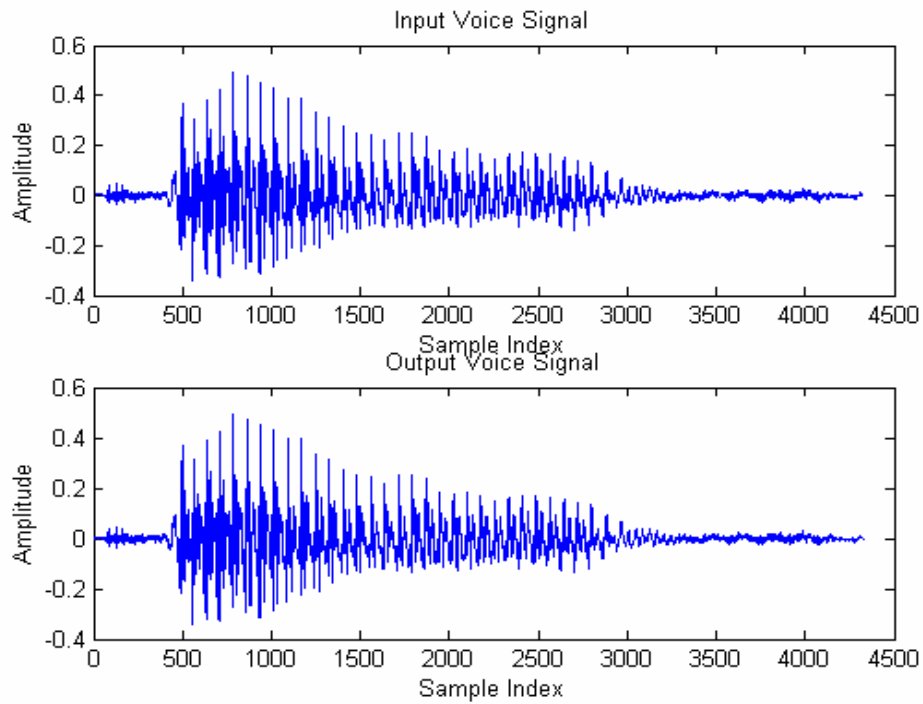
Lyginant kalbinių kodekų kokybę buvo atliekami eksperimentai leidžiantis vizualiai sulyginti originalios kalbos signalo formą su kodekų rekonstruojamos kalbos signalo forma. Šiame skyriuje pateikiami atliktų eksperimentų rezultatai. Originalus signalas buvo skaitomas skaitmeninis įrašas dikretizuotas 8000 kartų per sekundę (8kHz), naudojant 8 bitų rezoliuciją, sudauginus gauname 64 kbitų/s bitų dažnį. Toks skaitmeninis įrašas kalbos kodavime laikomas labai aukštos kokybės.

Visi eksperimentai buvo atliekami naudojantis *MatLab* programine įranga.



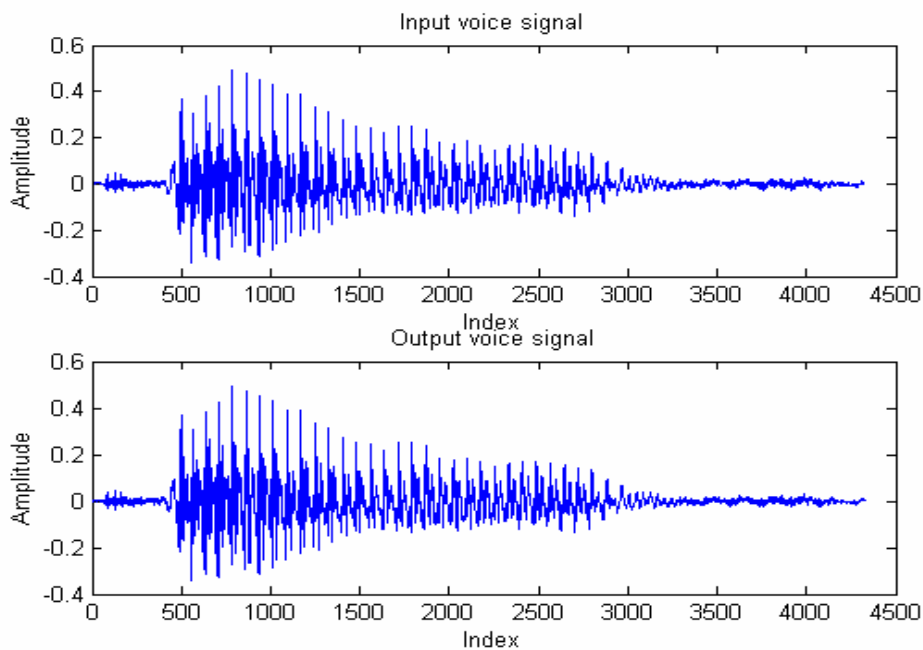
**Pav. 28** Signalų formos palyginimas (A-Law)

Iš palyginimo matyti, kad *A-Law* kodeko atkuriamos kalbos garsinio signalo forma labai panaši į originalų signalą.



**Pav. 29** Signalų formos palyginimas (M-Law kodekas)

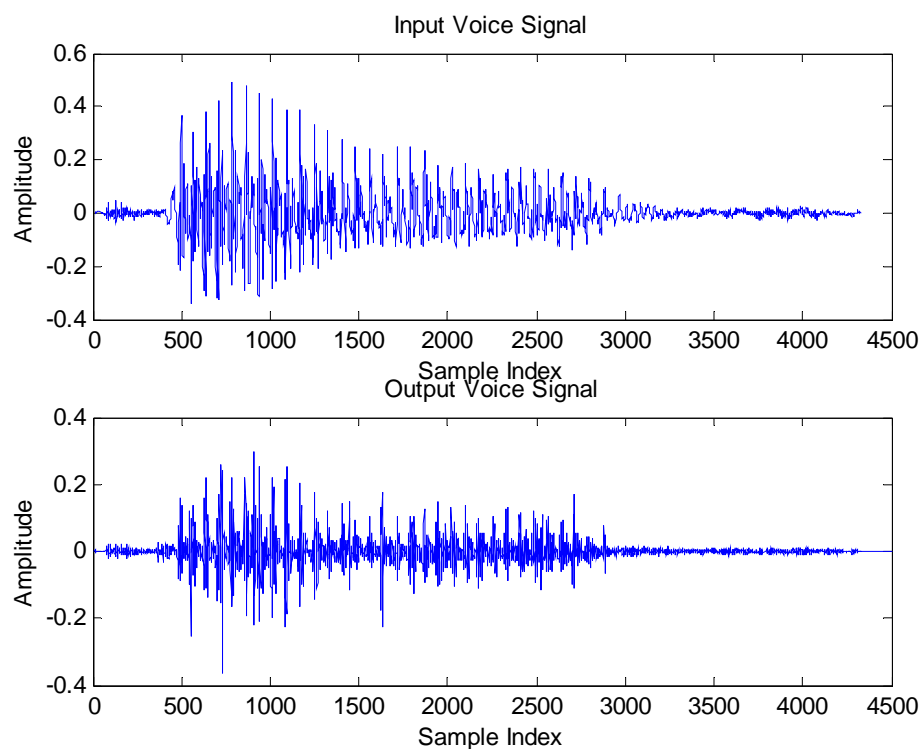
Iš palyginimo matyti, kad *M-Law* kodeko atkuriamos kalbos garsinio signalo forma labai panaši į originalų signalą.



**Pav. 30** Signalų formos palyginimas (DPCM kodekas)

Iš palyginimo matyti, kad DPCM kodeko atkuriamos kalbos garsinio signalo forma labai panaši į originalų signalą.





**Pav. 31** Signalų formos palyginimas (ADPCM kodekas)

Iš palyginimo matyti, kad *ADPCM* kodeko atkuriamos kalbos garsinio signalo forma ganėtinai skiriasi nuo originalaus. Tas duoda pagrindą manyti, kad atkuriamos kalbos kokybė turėtų būti prastesnė negu *M-Law*, *A-Law* ir *DPCM* kodekų.

Prieš tai atlikti eksperimentai buvo susiję su banginiais kodekais. Jų algoritmų sudėtingumas ganėtinai nedidelis. Tuo tarpu balsinių ir hibridinių kodekų sudėtingumas labai stipriai išaugo ir atlikti eksperimentus ganėtina sunku, todėl šiame darbe jie nebuvo atliekami.

## 6 Išvados

Kalbos glaudinimo technologijos tapo svarbiausiu telekomunikacijų etapu jau pačioje telekomunikacijų vystymosi pradžioje. Programinė įranga skirta kalbos suglaudinimui ir išplėtimui vadinama – kalbiniu kodeku. Išsiskyrė trys pagrindinės kalbinių kodekų klasės: balsiniai, banginiai ir hibridiniai kodekai. Pirmosios klasės kodekai rėmiasi kalbos analize, bandydami atvaizduoti garsinį signalą kelių parametrų deriniu. Banginiai kodekai bando, be jokių papildomų žinių apie tai kaip koduojamas signalas buvo sugeneruotas, atkurti signalą, kurio bangos forma būtų kiek galima panašesnė į originalo. Hibridiniai kodekai naudoja abiejų pirmųjų dviejų klasių principus.

Šiame darbe buvo pristatyta pagrindinės garso ir kalbinių kodekų kūrimo idėjos ir koncepcijos. Pristatyti pagrindiniai kalbos glaudinimo metodai: *PCM, DPCM, ADPCM, M-Law, A-Law, LPC, CELP*. Įvertinta ir palyginta tarpusavyje jų atkuriamos kalbos kokybė. Apžvelgta ir palyginta kalbos glaudinimo metodų pasauliniai standartai.

Pristatyta sukurtos paskirstytos stenografavimo sistemos programinės įrangos architektūra, kuri lengvai gali būti praplečiama. Lengvai galima papildyti sistemos palaikomų garso kodekų sąrašą.

Galima teigti, kad daugiausiai dėmesio yra skiriama mažų bitų dažnių balsinių kodekų tyrimui ir kūrimui. Sukurta ir standartizuota daug, mažai glaudinimo principais tesiskiriančių, kalbinių kodekų. Plačiausiai jų panaudojimas yra mobiliųjų ryšių sistemose. Dirbant su kompiuteriniais tinklais, o ypač vietiniais, reiktų įvertinti, kad galima naudoti: 1) žymiai platesnį dažnių juostos diapazoną; 2) įrašų archyvavimui nėra keliami dideli algoritminio vėlinimo reikalavimai. Todėl taikomosiose programose, kurios susiduria su kalbos skaitmeninių įrašų archyvavimo problemomis, geriausiai būtų rinktinis hibridinius (vidutinius) kodekus, kurie pakankamai gerai suspaudžia įrašus tuo pačiu metu išlaikydami aukštą kalbos kokybę. Rekomenduojama naudoti tokius lygio kodekus kaip *G.729, G.723.1, Speex*.

## 7 Žodynai

### 7.1 Santrumpų žodynas

1. PCM – pulso kodo moduliacija (*Pulse Code Modulation*).
2. DPCM – diferencialinė pulso kodo moduliacija (*Differential Pulse Code Modulation*).
3. ADPCM – adaptyvi diferencialinė pulso kodo moduliacija (*Adaptive Differential Pulse Code Modulation*).
4. LPC – tiesinė prognozė (*Linear Prediction Coding*).

### 7.2 Terminų žodynas

1. Kodekas – programinės įrangos priemonė medijos duomenų srautui suglaudinti ir išskleisti.
2. Koderis – kodeko dalis atliekanti suglaudinimą.
3. Dekoderis – kodeko dalis atliekanti išplėtimą.
4. Kalbinis kodekas – kodekas skirtas kalbos įrašams glaudinti.
5. Bitų dažnis – bitų kiekis reikalingas vienai kalbos sekunde suglaudinti.
6. Stenografavimas – procesas, kurio metu kalbos garso įrašai verčiami atitinkamu tekstu.
7. Kalbos kodavimas – žmogaus kalbos garsinių signalų glaudinimas/išplėtimas.

## 8 Literatūra

1. A. Dumčius. *Garso ir vaizdo technika ir technologija*: mokomoji knyga. – Kaunas, Technologija, 2001.
2. B. Khasnabish. *Implementing voice over Ip*. – Wiley-Interscience, A Jhon Wiley & Sons, Inc. Publication, 2003.
3. K. Ratkevičius, A. Rudžionis, V. Rudžionis, P. Kasparaitis. *Kai kurie balso technologijų taikymo pavyzdžiai*: mokslinės techninės konferencijos pranešimų medžiaga [Informacinės technologijos 2002], Kaunas, Kauno technologijos universitetas, 2002, p. 283-288.
4. N. Areška. *Audio-video srautų organizavimas bei pralaidumo įvertinimas e-studijų aplikacijose*: mokslinės techninės konferencijos pranešimų medžiaga [Informacinės technologijos 2003], Kaunas, Kauno technologijos universitetas, 2003, p. IV-20-23.
5. A. Rudžionis. *Balso technologijos šiuolaikiniame pasaulyje*: mokslinės techninės konferencijos pranešimų medžiaga [Informacinės technologijos 2004, sausio 28-29 d], Kaunas, Kauno technologijos universitetas, 2004, p. 228-233.
6. D. Raišutis. *Vaizdo ir garso perdavimo internetu technologiniai aspektai*: mokslinės techninės konferencijos pranešimų medžiaga [Informacinės technologijos 2004, sausio 28-29 d], Kaunas, Kauno technologijos universitetas, 2004, p. 228-233.
7. S. Wichman. *A Comparison of Speech Coding Algorithms: ADPCM vs CELP*. Department of Electrical Engineering, The University of Texas at Dallas, 1999.
8. W. Li, A. Sridhar, T. Teng. *Comparison of Speech Coding Algorithms: ADPCM, CELP and VSELP*. Project for EE 6390, 1999, Instructor: Dr. Murat Torlak, University of Texas at Dallas.
9. K. Lehtonen 55788E. *T-61.246 Digital Signal Processing and Filtering: GSM Codec*.
10. A. Pascal Bernard. *Source-Channel Coding of Speech*. University of California, Los Angeles, 1998.
11. N. Jayant, P. Noll. *Digital Coding of Waveforms*. Englewood, NJ, Prentice-Hall, 1984
12. S. Quackenbush, T. Barnwell, M. Clements. *Objective Measure for Speech Quality*. Englewood Cliffs, NJ, Prentice-Hall, 1988
13. W. T. K. Wong. *Low rate speech coding for telecommunications*. BT Technol J, 14, Nr. 1, p. 28—44 (Sausis 1996).
14. I. Boyd *Speech coding in telecommunications*. Westall F A and Ip S F A (Eds): *Digital signal processing in telecommunications*. Chapman & Hall, p. 300—325 (1993).

### 8.1 Nuorodos

15. Vilniaus universitetas. Matematikos ir informatikos fakultetas. <http://www.mif.vu.lt/~bastys/>
16. Nemokamas informacijos neprarandantis kodekas FLAC. <http://flac.sourceforge.net/>
17. Nemokamas kalbinis kodekas Speex. <http://www.speex.org>
18. Nemokamas atviro kodo garso kodekas Ogg Vorbis. <http://www.vorbis.com>