

**Azərbaycan Milli Elmlər Akademiyası
İNFORMASIYA TEXNOLOGİYALARI İNSTİTUTU**

Yadigar İmamverdiyev

Lyudmila Suxostat

**NITQ TEXNOLOGİYALARI
terminlərinin izahlı lüğəti**

Bakı -2015

İmamverdiyev Y.N., Suxostat L.V. Nitq texnologiyaları terminlərinin izahlı lüğəti Bakı: “İnformasiya Texnologiyaları” nəşriyyatı, 2015, 112 səh.

Təqdim olunan lüğətdə nitq siqnallarının rəqəmsal emalı və nitq texnologiyaları sahəsində elmi ədəbiyyatda rast gəlinən terminlər toplanmışdır. Lüğətdəki sözlər nitq texnologiyalarının əsas istiqamətlərini əhatə edən müvafiq mövzular üzrə verilmişdir. Nitq siqnallarının rəqəmsal emalı, nitqin avtomatik tanınması və sintezi, səsə görə şəxsin tanınması üzrə terminlər haqqında ətraflı məlumat verilmişdir.

Nitq texnologiyaları üzrə ixtisaslaşan mütəxəssislər, tələbələr və elmi tədqiqat aparan şəxslər üçün nəzərdə tutulmuşdur.

Vəsait AMEA İnformasiya Texnologiyaları İnstitutu Elmi şurasının qərarı ilə çapa məsləhət görülmüşdür.

Elmi redaktor: tex.e.d. Ramiz Alıquliyev

ISBN: 978-9952-434-71-2

© “İnformasiya Texnologiyaları” nəşriyyatı, 2015

MÜNDƏRİCAT

Giriş	4
Səsin fiziki xassələri	5
İnsanın səs traktı və nitqin yaranması modelləri	10
Nitqin fonetik və prosodik elementləri	16
Signalların rəqəmsal təsviri	24
Signalların filtrasiyası	30
Nitq signallarının spektral analizi	40
Nitq signallarının parametrləşdirilməsi	52
Səsin kodlaşdırılması/sıxılması	57
Nitqin avtomatik tanınması	63
Nitqin avtomatik sintezi	75
Səsə görə şəxsin tanınması	82
Nitq bazaları	96
Nitq signallarının emalı üçün proqram təminatı	99
Ədəbiyyat	102
Adlar göstəricisi	105

Giriş

Nitq insanlar arasında ünsiyyətin ən təbii formasıdır və hazırda insan-maşın qarşılıqlı əlaqəsində də getdincə geniş rol oynamağa başlayır. Müasir telekommunikasiya və kompüter sistemlərində istifadə etmək üçün nitq siqnalları bir sıra rəqəmsal emal mərhələlərindən keçirilir. Məsələn, mobil telefonda danışarkən siqnalların rəqəmsal emalı üsullarından istifadə edilməklə səs kodlaşdırılır, sıxılır və modulyasiya edilir.

Son illər nitq texnologiyaları sahəsində sürətli inkişaf məlum səbəblərdən ingilis dilində çox sayda terminin meydana çıxmasına səbəb olmuşdur. Bu lüğət nitq texnologiyaları və nitq siqnallarının rəqəmsal emalı sahəsində azərbaycan dilində terminologiyanın sistemləşdirilməsi üçün bir cəhddir.

Təqdim olunan lüğətin məqsədi oxuculara nitq texnologiyaları üzrə elmi əsərləri oxumaqda kömək etməkdir. Lüğətdəki sözlər nitq texnologiyalarının və nitq siqnallarının rəqəmsal emalının əsas istiqamətlərini əhatə edən müvafiq mövzular üzrə verilmişdir. Digər tərəfdən, lüğətdə şəkillər və düsturlar da verilmişdir və bu baxımdan ona bu sahədəki müvafiq mövzular üzrə qısa sorğu kitabçası (mühazirə qeydləri) kimi də baxmaq olar.

Səsin fiziki xassələri

Acoustic signal Акустический сигнал

Akustik signal – səs signalı.

Acoustics Акустика

Akustika (yunanca *akustikos* – eşitmə) – səs fiziki hadisə kimi öyrənilməsi ilə məşğul olan elm sahəsidir.

Amplitude Амплитуда

Amplitud – müəyyən qanun üzrə rəqs edən dəyişənin orta qiymətdən və ya şərti sıfır qəbul edilmiş müəyyən qiymətdən ən böyük uzaqlaşmasıdır; signalın baxılan anda kəmiyyət qiymətidir.

Bark scale Барк-шкала

Bark şkalası – səs yüksəkliyinin psixofiziki vahididir (1961-ci ildə təklif edilib). Bark şkala kritik eşitmə diapazonları ilə əlaqədardır, bu diapazonların eni qeyri-bərabərdir, səs rəqslərinin tezliyi artdıqca artır. Buna görə, bark şkalası da qeyri-xəttidir. Səs barkla yüksəkliyi (b) ilə tonun herslə tezliyi (f) arasındakı düz və tərs asılılıq uyğun olaraq aşağıdakı düsturlarla verilir:

$$\text{Bark: } b = 13 \arctg(0.00076 \cdot f) + 3.5 \arctg\left(\frac{f}{7500}\right)^2,$$

$$\text{Hers: } f = \frac{52548}{b^2 - 52.56b + 690.39}.$$

Critical band Критический диапазон

Kritik diapazon – eşidilən tezliklər oblastını bir neçə kritik diapazona bölürlər (adətən, 24 kritik diapazona), onlar daha yüksək tezliklər üçün qulağın həssaslığının aşağı düşməsinə bildirir. Kritik diapazonları tezliyə oxşar olaraq, səs daha bir xarakteristikası kimi qəbul etmək olar. Lakin eşitmə orqanlarından asılı olmayan tezlikdən fərqli olaraq, kritik diapazonlar eşitmə qavrayışına uyğun olaraq müəyyən edilir. Nəticədə, onlar tezlik qavrayışının müəyyən

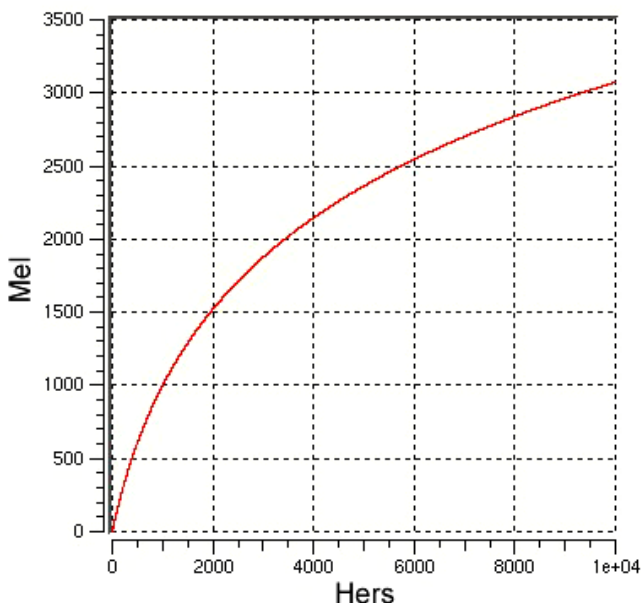
ölçülərini formalaşdırırlar, onlar üçün ölçü vahidləri, məsələn, *mel* və *bark* daxil edilir.

Decibel Децибел

Desibel (dB) – intensivliyin nisbi dəyişməsinin loqarifmik miqyasda ölçü vahididir. Desibel Bel-in onda biridir ($1 \text{ dB}=0.1 \text{ Bel}$). İntensivliyi desibeldə $I_{\text{dB}} = 10 \lg(I_1/I_0)$ düsturu ilə hesablamaq olar, burada I_1 – intensivliyin ölçülmüş qiyməti, I_0 – intensivliyin bazis kimi götürülmüş qiymətidir.

Mel scale Мел-шкала

Mel şkalası – səsin yüksəkliyinin psixofiziki qavrayışının empirik şkalasıdır, onu signalın tezliyi ilə əlaqələndirir.



Səs tonlarının yüksəkliyinin subyektiv qavranılması haqqında böyük həcmdə verilənlərin statistik emalına əsaslanır. Tədqiqatların nəticələri göstərir ki, səsin yüksəkliyi əsas etibarlı ilə rəqslərin tezliyi ilə əlaqədardır, lakin səsin gurluğundan və tembrindən də asılıdır. Səsin mel şkalasında yüksəkliyi (m) ilə tonun herslərlə tezliyi (f) arasında düz və tərs asılılıqlar aşağıdakı kimidir:

$$\text{Mel: } m = 1127.01048 \cdot \ln\left(1 + \frac{f}{700}\right),$$

$$\text{Hers: } f = 700 \cdot \left(e^{\frac{m}{1127.01048}} - 1\right).$$

Tərifə görə, tezliyi 1000 Hs olan səs dalğaları gurluğun səviyyəsi 40 dB olduqda eşitməsi normal müşahidəçiyə öndən təsir etdikdə müşahidəçidə səsini yüksəkliyinin 1000 mel ilə ölçülən qavrayışını yaradır. 3430 Hs tezlikli səs isə 40 dB gurluq səviyyəsində subyektiv səslənmə yüksəkliyi əvvəlkindən 2 dəfə böyük qiymətləndirilir – mel-şkalada bu səsini yüksəkliyi 2000 mel qəbul edilir.

Phone Фона

Fon, fona [yun. phone – səs] – səsini gurluğunun səviyyəsinin fiziki vahididir, təmiz ton üçün desibelə uyğun gəlir.

Pitch Высота звука

Səsin yüksəkliyi – səsini qulağa təsiri ilə müəyyən edilir, subyektiv qiymətləndirilir. Qulağa gələn enerji axını nə qədər böyükdürsə, yüksəklik o qədər böyükdür. Səsin yüksəkliyini ölçmək üçün *səsin intensivliyi* əlverişlidir. Səsin yüksəkliyi desibel (dB) ilə ifadə edilir. Məsələn, yarpaqların xışıltısı 10 dB, pıçıltı 20 dB, küçədəki səs-küy 70 dB, ağrı yaradan hədd 120 dB, ölümcül səviyyə 180 dB olaraq qiymətləndirilir.

Reverberation Реверберация

Reverberasiya – dəfələrlə əks olunması zamanı səsini intensivliyinin tədricən azalması prosesi. Reverberasiya hadisəsi bir səs mənbəyindən olan müxtəlif exo-siqnalların superpozisiyasından ibarətdir. Bəzən *reverberatorların* köməyi ilə bu effektin imitasiyasına da reverberasiya deyirlər. Nitqin, musiqinin yazılması, müxtəlif küy effektlərinin yaradılması zamanı *süni reverberasiyanın* istifadə edilməsi audiosiqnalların ümumi emalının tərkib hissəsidir.

Signal Сигнал

Signal (latınca *sign* – işarə) – geniş mənada “siqnal” məsafəyə ötürülən informasiyanın maddi daşıyıcısını bildirir. Dar mənada “siqnal” termini informasiya sistemində məlumatı ötürmək üçün məxsusi yaradılmış informasiya maddi daşıyıcısıdır. Siqnal parametrləri ötürülən məlumata uyğun olaraq dəyişən istənilən fiziki proses ola bilər. Determinik və ya təsadüfi siqnalın riyazi modeli siqnalın parametrlərinin dəyişməsinə xarakterizə edən funksiya ilə təsvir edilir.

Sound Звук

Səs – yayılan mexaniki dalğalardır (yayılma mühiti havadır). İnsanın qavradığı səs rəqslərinin tezlik diapazonu 20-20 000 Hs təşkil edir. Tezliyi 20-20 000 Hs olan akustik dalğalar *səs dalğaları*, 20 Hs-dən kiçik olanlar – *infrasəs*, 20 000 Hs-dən böyük olanlar *ultrasəs* dalğalar adlanır. Nitqin tezlik diapazonu daha kiçikdir və 300-4 000 Hs təşkil edir. Tezlik diapazonunu 300-2 400 Hs-ə qədər sıxdıqda da nitq aydın eşidilir.

Sound intensity Интенсивность звука

Səsin intensivliyi (səsin gücü) – səs dalğasının dalğanın yayılması istiqamətinə perpendikulyar və sahəsi vahid olan səthdən vahid zamanda daşdığı enerjidir. Rəqslərin amplitudu və rəqs edən cismin sahəsi artdıqda səsin intensivliyi artır.

Speech sounds Речевые звуки

Nitq səsləri – şifahi məlumatın elementləridir, danışanın əhval-ruhiyyəsini, onun haqqında danışdığına, həmsöhbətinə, auditoriyaya və s. münasibətini ifadə edən səs əlamətlərindən ibarətdir.

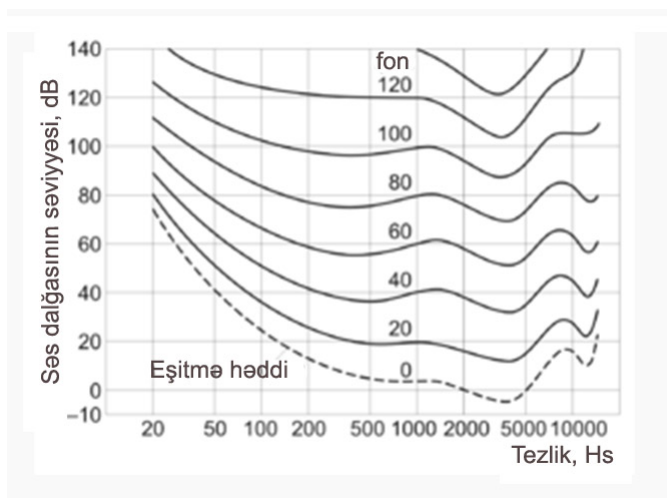
Volume Громкость

Səsin gurluğu – səsin gücünün subyektiv gəvrənilməsidir (eşitmə hissəsinin mütləq ölçüsüdür). Gurluq əsasən səsin təzyiqindən və səs dalğalarının tezliyindən asılıdır. Səsin gurluğuna onun spektral

tərkibi, fəzada lokallaşması, tembr, səs dalğalarının təsirinin müddəti və başqa amillər də təsir edir.

Volume level Уровень громкости звука

Səsin gurluğunun səviyyəsi – nisbi ölçüdür. O, fonla ifadə edilir və ədədi qiyməti ölçülən səsə eyni gurluğa malik, tezliyi 1 kHs olan sinusoidal tonun yaratdığı səs təzyiqinin səviyyəsinə bərabərdir.



İnsanın səs traktı və nitqin yaranması modelləri

Acoustic theory of speech production Акустическая теория речеобразования

Nitqin yaranmasının akustik nəzəriyyəsi – nitqin yaranması zamanı baş verən akustik və aerodimaik prosesləri tədqiq edir. Əsası 1863-cü ildə alman alimi Herman fon Helmhols (1821-1894) tərəfindən qoyulmuşdur. 1960-cı illərdə Q. Hant və C. Flanaqan tərəfindən inkişaf etdirilmişdir.

Helmholsa görə, nitqin yaranması prosesi bir-birindən asılı olmayan iki komponentdən ibarətdir: 1) səs mənbəyində səsin yaranması və 2) artikulyasiya traktının rezonans tezliklərinin hərəkətə gəlməsi hesabına səsin fonetik keyfiyyətinin formalaşması və formalaşmış səsin ağız və burun vasitəsilə fəzaya şüalanması. Səs traktı akustik baxımdan ardıcıl birləşdirilmiş iki Helmhols rezonatorundan ibarətdir və səs traktında dalğanın yayılması xüsusi törəməli diferensial tənliklərlə təsvir olunur.

Alveolar tubercle Альвеолярный бугорок

Alveol qabarığı – yuxarı çənənin zoğudur, özündə dişləri daşıyır.

Articulation Артикуляция

Artikulyasiya – səslərin tələffüzü prosesində nitq orqanlarının hərəkətləridir.

Epiglottis Надгортанник

Qırtlaqüstü – bu qırtlağın üçüncü böyük qığırdağıdır. O, "ləçəkdən" və "saplaqdan" ibarətdir. Birləşdirici toxumadan olan saplaq qalxanşəkilli qığırdağın yuxarı (üst) buynuzlarının arasında yerləşir. Ləçək sərbəstdir və arxada dilin kökünə çevrilir. O, qidanın udulması zamanı qırtlağa girişi bağlamaqla çox əhəmiyyətli funksionu yerinə yetirir.

Glottal stop Смычка (гортанная смычка, затвор)

Sürgü – nitq orqanlarının hava axınının yolunu tamamilə kəsməyə yönəlmiş hərəkətidir. Sürgüdən keçid nəticəsində partlayan samitlər və partlayan-nov samitləri yaranır.

Glottis Голосовая щель

Səs yarığı – qırtlaq ortaya yaxın daralır və ən dar yerdə iki üfüqi qırçın şəklində səs telləri yerləşir. Onların arasındakı boşluq *səs yarığı* adlanır (açıq vəziyyətdə səs yarığının uzunluğu 2 sm və eni 1 sm olur).

Gullet Пищевод

Qida borusu – içiboş orqandır, boğaz ilə mədənin arasından keçən uzun və dar boru şəklindədir.

Larynx Гортань

Qırtlaq – yuxarı tənəffüs yollarına aid olan orqandır və tənəffüs, qoruyucu, səs və danışq kimi bir neçə əhəmiyyətli funksiyanı yerinə yetirir.

Lips Губы

Dodaqlar – ağız boşluğuna girişi əhatə edən dəri-əzələ qırışıqları.

Nasal cavity Носовая полость

Burun boşluğu – burunun iki bölməsindən biridir (ikinci bölmə – xarici burun), kəllənin daxilindədir və aşağıdan ağız boşluğu, yuxarıdan ön kəllə çuxuru (beyinin alın paylarının yerləşdiyi yer), yan tərəflərdən göz yuvaları kimi vacib anatomik törəmələrlə həmsərhəddir.

Nose Нос

Burun – tənəffüs və ya respirator yollarının ilk bölməsini təşkil edən orqandır və bu funksiyaları yerinə yetirir: (1) udlağa (boğaza) havanın keçirilməsi; (2) qismən nəmləmə və havanın qızdırılması; (3) hissəciklərin və patogen mikroorqanizmlərin tozcuqlarından

havanın təmizlənməsi, (4) burunun əlavə cibləri nitqin yaranmasında iştirak edir, çünki səslər üçün yaxşı rezonatorudur.

Oral cavity Полость рта

Ağız boşluğu – həzm yolunun başlanğıc bölməsi olan anatomik törəmədir. Ağız boşluğu: ağız boşluğu dibinin əzələləri, dil, yumşaq və bərk damaq, yanaqlar, dodaqlar və dilin qövsələrindən ibarətdir. Ağız boşluğunun sümük əsası üst və alt çənələrdir.

Palate Həbo

Damaq – bərk damağın davamıdır, ondan fərqli olaraq, sümük əsasına malik deyil, lakin böyük miqdarda birləşdirici və piyli toxuma saxlayır. Onun qatlarında çox sayda xırda tüpürcək vəziləri və toxuma limfoidləri var.

Arxa hissəsində yumşaq damağın kənarı *damaq pərdəsi* şəklində sərbəst sallanır, orada kiçik böyümə – dil yerləşir. Dillə birlikdə yumşaq damaq – udma, səsin formalaşması və fonasiya (tələffüz prosesi) aktlarında iştirak edən orqanlardır.

Pharynx Глотка

Boğaz (Udlaq) – içiboş əzələ orqanıdır, ağız boşluğunu davam etdirir və sonra qida borusuna keçir. Onun başlanğıcı kəllənin əsasına və sonu 6 boyun fəqərəsinə uyğundur. Boğazın ümumi uzunluğu təxminən 12-14 sm təşkil edir, yaşıdan asılı olaraq, onun forması yuxarıya genişlənərək və aşağıya daralaraq qıfı oxşayır.

Psychoacoustics Психоакустика

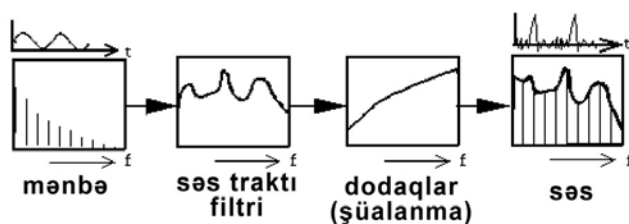
Psixoaqustika – səsin insan tərəfindən qavranılmasının psixoloji və fizioloji xüsusiyyətlərini öyrənən elm sahəsidir.

Soft palate Нёбная занавеска

Damaq pərdəsi – yumşaq damağın hərəkətli hissəsidir.

Source-filter model Модель «источник фильтр»

Mənbə-filtr modeli – İsveç alimi Qunnar Fant 1960-cı ildə “Nitqin yaranmasının akustik nəzəriyyəsi” əsərində təklif etmişdir. Mənbə-filtr modelində Helmhols rezonatoru filtrlə əvəz olunur, yəni səs mənbəyi – səs telləridir, səs traktı yalnız mənbəni modifikasiya (filtrasiya) edir; səs tellərinin və səs traktının akustik xassələri baxımından bir-birindən asılı deyil, onları ayrılıqda öyrənmək olar. Əgər səs mənbəyi və səs traktı bir-birindən asılı deyilsə, onda səs signalından mənbəni müəyyən etmək olar (*tərs filtrasiya*). Bu nitqin sintezi, səsə görə şəxsin identifikasiyası, vokal patologiyaların öyrənilməsi üçün zəruridir.



Teeth Зубы

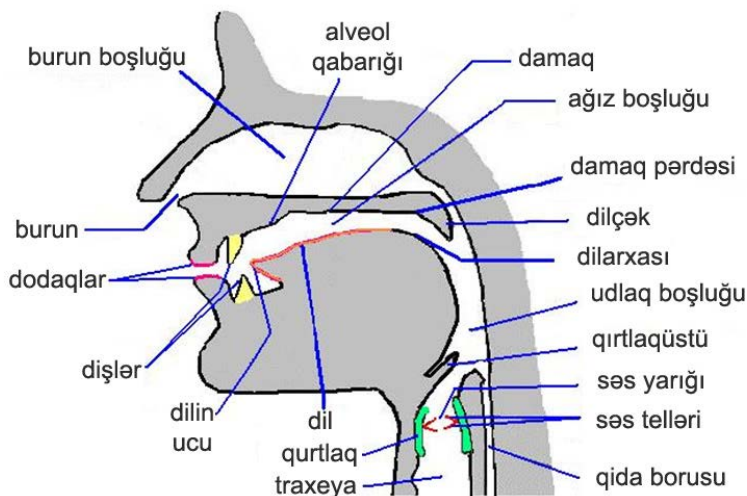
Dişlər – bu üst və alt çənədə yerləşmiş sümük törəmələridir. Onlar çeynəmə funksiyasını yerinə yetirirlər, həmçinin danışq səslərinin formalaşmasında da iştirak edirlər. Bundan başqa, dişlər – xarici mühit və ağız boşluğu arasında təbii fizioloji maneə rolunu oynayırlar.

Tongue Язык

Dil – ağız boşluğunda yerləşmiş əzələ orqanıdır, üzərində müxtəlif dadların qavranılmasına xidmət edən *dad reseptorları* olur.

Trachea Трахея

Traxeya – tənəffüs yollarının altbölməsinin başlanğıcını təşkil edir və içəridən selikli qabıqla örtülmüş qıgırdaq divarına malik içiboş boru şəklindədir. Cinsindən və bədən quruluşundan asılı olaraq yaşlı insanda traxeyanın ümumi uzunluğu 11-13 sm ola bilər.



Uvula Язычок

Dilçək – bu udlağın (boğazın) ən hərəkətli hissəsidir, havanın keçməsi zamanı rəqs edir və tipik səs verir.

Vocal chords Голосовые связки

Səs telləri – qırtlağın sağ və sol tərəfində yerləşmiş və qabaqdan arxaya dartılmış elastiki törəmə. Səs tellərinin arasındakı sahə **səs yarığı** adlanır. Səs ağciyərlərdən təzyiq altında çıxan havanın təzyiqi nəticəsində səs tellərinin periodik rəqsləri nəticəsində yaranır, bu zaman səs telləri yaxınlaşır və dartılır.

Vocal source Голосовой источник

Səs mənbəyi (qırtlaq, səs telləri) – artikulyasiya traktında səsin yaranması üç üsulla reallaşa bilər: 1) ağ ciyərlərdən gələn hava axınının səs telləri vasitəsilə periodik modulyasiyası ilə səsin yaranması (səs mənbəyi); 2) səs aparatının daralmaları nəticəsində hava axınının turbulent burulğanlaşması ilə küyün yaranması (küy mənbəyi); 3) artikulyator traktında sürgünün qəfil açılması zamanı hava axını təzyiqinin sıçrayışla dəyişməsi yolu ilə səsin yaranması (impuls mənbəyi).

Vocal tract Речевой тракт

Səs traktı – səs tellərinin üstündə yerləşmiş nitq orqanlarıdır.

Nitqin fonetik və prosodik elementləri

Accent Акцент

Aksent (latınca *accentus* – “səsi yüksəltmək”)

1) söz və ya bütöv söz birləşmələri tərkibində hecanın fonetik vasitələrlə (səsi gücləndirməklə, tonu yüksəltməklə və s.) fərqləndirilməsi; nitqin vurğu ilə, ifadəli, tonu yüksəltmək və alçaltdırmaqla, səsi gücləndirmək və ya zəiflətməklə tələffüzü.

2) ana dilində danışmayan, digər dil və ya dialekt daşıyıcılarına xas olan və dil normalarından kənaraçıxma ilə özünü göstərən tələffüz xüsusiyyətləri.

Allophone Аллофон

Allofon – fonemin sözdəki yerindən, digər səslərlə qonşu olmasından, saitın vurğulu və ya vurğusuz olmasından asılı olaraq tələffüz variantıdır. Məsələn, ingilis dilində /t/ foneminin tələffüz variantları: 1) «type» sözündə nəfəsli /t/ fonemi, 2) “butter” sözündə birvurğulu /t/ fonemi və 3) «hot» sözündə partlayışlı /t/. Trifonlar və kvinfonlar nitq tanıma sistemlərində geniş istifadə edilən allofon modelləridir.

Aural perception Слуховое восприятие

Eşitmə qavrayışı – əsas xarakteristikalarına görə ətraf mühitin müxtəlif səslərini seçmək bacarığı: gurluq, yüksəklik, tembr, temp.

Bigram Биграмм

Bigram («digram») – uzunluğu 2-yə bərabər *n*-qram. *Bax. n-qram*

Diphone Дифон

Difon – bir səsin ortasından sonrakı səsin ortasınadək olan səs vahididir.

Energy Энергия

Enerji – amplitudun tezliklər üzrə paylanmasıdır.

$$E = 10 \log_{10} \left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \right),$$

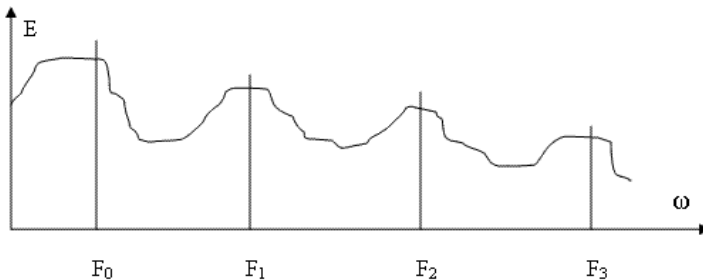
burada x – səs signalının freymi, n – freymdə bölgülərin nömrələri və N – bölgülərin sayıdır.

Formant Форманта

Formant (lat. formans (formantis) – *doğuran, formalaşdırən*) – nitq səslərinin tezliyinin səviyyəsi ilə əlaqəli olan və səsin tembrini əmələ gətirən akustik xarakteristikasıdır. Formantalar – səs tonu spektrinin hissəsidir, ortalama tezlik qiymətinə görə müəyyən edilmiş tezliklər oblastıdır. Səsin tanınması prosesi əsas etibarlı ilə formantların qavranılması ilə şərtlənir.

Formant nitq signalının spektrini (enerjinin və ya amplitudun tezliklər üzrə paylanması) xarakterizə edən vacib parametrdir, formantı məhdud tezlik oblastında enerjinin konsentrasiyası kimi müəyyən etmək olar. Formant tezlik, en və amplitud ilə xarakterizə edilir. Formantın tezliyi olaraq maksimal amplitudun tezliyi götürülür. Başqa sözlə, formant – spektrin qrafikində amplitudun müəyyən kəskin yüksəlişi, onun tezliyi isə bu yüksəlişdə pikin tezliyidir.

Şəkildə tezlikləri F_0 , F_1 , F_2 , F_3 olan dörd formant göstərilib.



F_0 formantı nitq signalının *əsas ton tezliyi* də adlanır.

F_1 - F_6 formantları nitq signalının enerjisinin tezlik üzrə konsentrasiyasını müəyyən edirlər və vokal (adətən, sait) səsləri xarakterizə edirlər. F_3 -dən başlayaraq formantlar danışan şəxs

haqqında məlumat daşıyırlar və diktorun tanınması sistemlərində geniş istifadə edirlər.

Formant extraction algorithm

Алгоритм выделения формант

Formantların müəyyən edilməsi algoritmi. Nitq signalının digər parametrləri kimi, formantlar da zamana görə tez dəyişirlər və onları səs signalının 10-20 ms-ən kəsiyində müəyyən edirlər, hesab olunur ki, belə intervalda signal təxminən stasionardır (sabitdir).

Təbii olaraq, ilk mərhələdə signalın spektrini almaq üçün ilkin signalın Furiye çevirməsi hesablanır. Sonra belə hərəkət edirlər.

Məlumdur ki, birinci formant (əsas ton tezliyi) 50 Hz-dən (yaşlı kişidə) 300 Hz-ə (uşaqlarda və bəzi qadınlarda) olan aralıqda yerləşir. Bundan başqa, hər sonrakı formantın amplitudası əvvəlkindən kiçikdir, onların tezliyi isə əsas ton tezliyinin tam ədədə misillərinə bərabərdir (F1 üçün 2-yə, F2 üçün 3-ə və s. vurulur).

Beləliklə, 50 Hz-dən 300 Hz-ə qədər intervalda signal amplitudunun maksimal qiyməti tapılır, sonra həmin tezliyi 2-yə vuraraq bu nöqtənin müəyyən ətrafında amplitudun maksimal qiyməti tapılır və onun daha böyük ətrafda amplitudun maksimumu olması yoxlanılır. Əgər bu fərziyyə doğru çıxarsa, onda ikinci formant tapılmış olur (F1). Əks halda, signalda ikinci formant yoxdur.

Sonra əsas tonun tezliyi 3-ə vurulur və yuxarıdakı addımlar üçüncü formant üçün təkrarlanır. Sonra digər formantlar üçün də təkrar edilir. Bir qayda olaraq, 4-6 formant hesablanır.

Fricative sound Фрикативные (турбулентные) звуки

Frikativ (turbulent) səslər ([v], [f], [z], [ş], [j]) – nitq traktının tam bağlanması və bu maneənin qarşısında böyük təzyiqin yaranması və maneənin qəfildən götürülməsi nəticəsində yaranırlar.

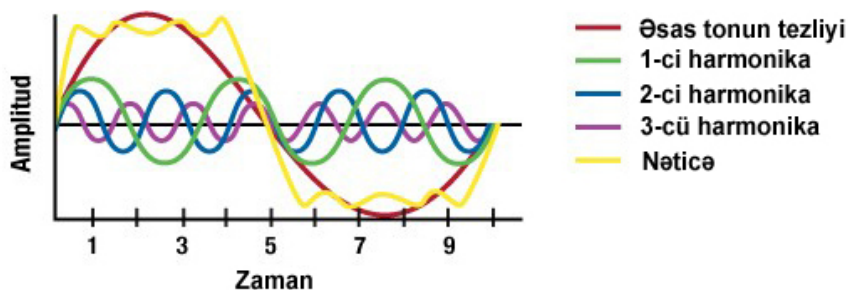
Fundamental frequency Частота основного тона

Əsas tonun tezliyi – səs tellərinin vibrasiyalarının tezliyinə deyilir. Boğazda və ağız boşluğunda yaranan digər yüksək tezlikli

harmonikalar üçün baza rolunu oynayır. Əsas tonun tezliyi nitqin əsas tonuna bütün digər faktorlardan daha çox təsir edir. Əsas ton diapazonu uşaqlar üçün təxminən [170..600] Hs, kişilər üçün [85..200] Hs, qadınlar üçün [160..300] Hs-dir. Praktikada əsas tonun periodunu avtokorrelyasiya funksiyasının maksimumuna görə müəyyən etmək olar.

Harmonics Гармоники

Harmonikalar – tezliklərinin əsas tonun tezliyinə nisbətləri tam ədədlər olan obertonlardır. Əsas ton birinci harmonika hesab edilir.



Intonation Интонация

İntonasiya – səslənən nitqin xarakteristikasıdır, ton (səsin yüksəkliyi), temp, intensivlik (gurluq), həmçinin tembr kimi nitq vasitələri ilə yaranır. İntonasiya cümlənin əhəmiyyətli əlamətidir, onun qrammatik ifadə vasitələrindən biridir.

N-gram N-грамм

N-qram – *N* fonemdən/pauzadan ibarət ardıcılıq.

Overtone Обертон

Oberton (almanca “oberton” – *yuxarı ton*) – harmonik olmayan səs rəqslərini Furiye teoreminə görə tezlikləri müəyyən tezliyin misilləri olan harmonik toplananların cəmi şəklində göstərmək olar. Ən kiçik tezliyə (ν) malik harmonik rəqslər *əsas ton*, digərləri *obertonlar* adlanır (birinci obertonun tezliyi 2ν , ikincinin tezliyi 3ν və s.-dir).

Obertonlar ağız boşluğunda yaranırlar və onların fərqi artikulyasiya zamanı ağızın dəyişən həcmindən və formasından asılıdır.

Phone Фона

Fon – nitqin tanınması və ya nitqin sintezi sistemlərinin lüğətində (leksikonunda) tələffüzün təsviri üçün istifadə edilən simvol. Onların sayı dildəki fonemlərin sayından bir qədər çox və ya az ola bilər. Fon çoxluğu sistemin dəqiqliyini optimallaşdırmaq üçün seçilir.

Phoneme Фонема

Fonem – dildə ən kiçik fonetik vahidlərdir. Məsələn, /d/ və /t/ səsləri ingilis dilində ayrı-ayrı fonemlərdir, çünki onlar «do» və «to» kimi sözlərdə fərqlənirlər. Dillərin əksəriyyətində təxminən 50 fonem olur.

Phonetics Фonetика

Fonetika (yun. phonetike, phone – səs) – dilçiliyin insan nitqinin artikulyasiya və akustik xüsusiyyətlərini öyrənən bölməsidir.

Plosive sound Взрывной (импульсный) звук

Partlayışlı (impulslu) səs ([r], [s], [t], [k]) – səs traktı tam bağlandıqda yaranan səs (adətən, səs traktının başlangıcında). Sürgünün arxasında yüksək dərəcədə sıxılmış hava yaranır və sonra hava qəflətən azad olunur.

Pronunciation Произношение

Tələffüz – konkret dildə nitq səslərinin artikulyasiyasının xüsusiyyətləridir. Xalqlar, tayfalar, yerli sakinlər, cəmiyyətin müxtəlif qatları tərəfindən mənimsənilmiş tələffüz usulu, şəkli, tərzidir.

Prosodic features Просодические признаки

Prosodik əlamətlər – ton, ləhcə və ritmik xarakteristikalardır. Həm qavrayış səviyyəsində (tonun yüksəkliyi, gurluq, müddət), həm də

fiziki səviyyədə (əsas tonun tezliyi, intensivlik zaman) nitq əlamətləri daxildir.

Prosody Просодия

Prosodiya (yunan. prosodikos – *vurğuya aid olan*) – nitqin vurğu, intonasiya, intensivlik, temp və tembr kimi fonetik vahidləri çoxluğudur. Bir qayda olaraq, onun reallaşdırılması üçün nitqin səsdən daha böyük seqmentləri tələb edilir.

Quinphone (pentaphone) Квинфон (пентафон)

Kvinfon və ya pentafon – adətən, iki sol və iki sağ fondan ibarət kontekstdə olan fon.

SAMPA (Speech Assessment Methods Phonetic Alphabet) – *beynəlxalq fonetik əlifbanın* (International Phonetic Alphabet, IPA) işarələrinin ASCII simvollarının (7-bitlik kompüter kodlarının) köməyi ilə yazılışı sistemidir. Keçən əsrin 80-ci illərində təklif edilib.

Senon Сенон

Senon – fonemlərin dayanıqlı ardıcılığıdır.

Speech tempo Темп речи

Nitqin tempi – nitqin elementlərinin (səslər, hecalar, sözlər) tələffüz edilməsi sürətidir. Prosodik elementlərə aiddir, intonasiyanın komponentlərindən biridir. *Normal nitq tempi* – test frazasının orta müddəti 2.4 saniyəyə bərabər olduqda nitqin tələffüz sürətidir. *Sürətli nitq tempi* – test frazasının orta müddəti 1.5-1.6 san olduqda nitqin tələffüz sürətidir.

Stress Ударение

Vurğu – intonasiyanın köməyi ilə çoxhecalı söz tərkibində bir hecanın fərqlənməsidir.

Syllable Слог

Неса – nitq axınının minimal fonetik vahididir, bir qayda olaraq, onun tərkibi bir sait səs və ona qonşu olan samit səslərdən ibarətdir.

Timbre Тембр

Тембр (fran. *timbre* – zəng, zıncırov) – səsin çalarındır, eyni yüksəkliyə malik iki səsi fərqləndirməyə imkan verir. Tembrə görə biz musiqi alətlərinin səslərini fərqləndiririk, tanış adamların səsinə tanıyıırıq. Tembr ikinci subyektiv qiymətləndirmədir. Səsin tembri obertonlar çoxluğu ilə təyin olunur. Bu və ya digər səsə məxsus obertonların müxtəlif sayı ona xüsusi çalar – tembr verir. Bir tembrin digərindən fərqi təkcə obertonların sayı ilə deyil, onların intensivliyi ilə də şərtlənir. Tembr səs tezliklərinin hansı oblastlarında oberton toplularının (formantların) yaranmasından da asılıdır.

Tone Тон

Тон – səs dalğasındakı rəqslərin tezliyi ilə müəyyən edilir. Bəzi səslər haqqında onların *aşağı tonlu* (böyük barabanın yaratdığı səs), digərlərinin *yüksək tonlu* (məsələn, fit səsi) olduğunu deyirik. Sadə ölçmələr (rəqslərin açılışı) göstərir ki, aşağı tonlu səslər səs dalğasında kiçik tezlikli rəqslərdir. Yüksək tonlu səsə böyük tezlikli rəqslər uyğun gəlir.

Trigram Триграмм

Trigram – uzunluğu 3-ə bərabər n -qram.

Triphone Трифон

Trifon – üç fonemdən ibarət olan səs qrupudur.

Unigram Юниграмм

Unigram – uzunluğu 1-ə bərabər n -qram.

Unvoiced sound Невокализованный звук

Qeyri-vokal səs – müəyyən yerdə (adətən, ağzın sonunda) səs traktının daralması və daralmış yerdən havanın turbulent hava axını yaranması üçün kifayət qədər sürətlə itələnməsi zamanı yaranan səsdir. Bu yolla səs traktını həyəcanlandıran genişzolaqlı küy mənbəyi yaranır.

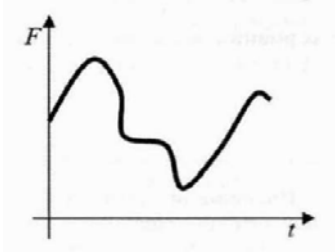
Voiced sound Вокализованный звук

Vokal səs – səs yarığı vasitəsilə havanın itələnməsi ilə yaranan səsdir, bu zaman səs telləri vaxtaşırı gərilir və açılır və hava axını impulslarının səs traktını həyəcanlandıran kvaziperiodik ardıcılığı yaranır.

Signalların rəqəmsal təsviri

Analog signal Аналоговый сигнал

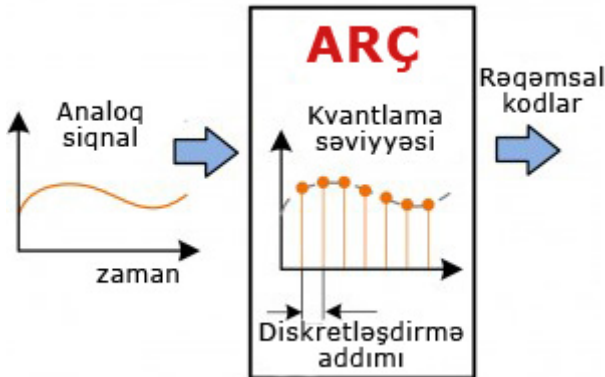
Analoq signal – qiymətləri zamana görə fasiləsiz dəyişən signal. Məsələn, mikrofon səs dalğası təzyiqini qəbul edir və onu elektrik analoq signalına çevirir.



Analog-to-digital converter (ADC) və ya A/D converter

Аналого-цифровой преобразователь (АЦП)

Analoq-rəqəmsal çevirici (ARÇ) – analoq signalı rəqəmsal signala çevirən qurğudur. ARÇ signalın diskretləşdirməsini və kvantlanmasını həyata keçirir.



Bit rate Скорость передачи битов

Bit sürəti – bir saniyə ərzində emal edilən bitlərin sayıdır.

Decimation Децимация

Desimasiya – təkrar (ikinci dəfə) diskretləşdirmə, yəni diskret signalın diskretləşdirmə tezliyinin aşağı salınması. Adətən, desimasiyadan əvvəl signalı aşağı tezlik filtrindən keçirirlər.

Digital signal Цифровой сигнал

Rəqəmsal signal – zamana görə diskret dəyişən ədədi qiymətlər alan signal. Rəqəmsal signal kvantlanmış diskret signaldır.

Digital signal processing (DSP)

Цифровая обработка сигналов (ЦОС)

Signalların rəqəmsal emalı – fiziki signal (gərginlik, cərəyan və s.) ədədlər ardıcılığına çevrilir və sonra hesablama qurğusunda riyazi çevirmə tətbiq edilir. Çevirməyə məruz qalmış rəqəmsal signalı tələb edildikdə yenidən fiziki signala (cərəyan və ya gərginliyə) çevirmək olar.

Digital signal processor (DSP)

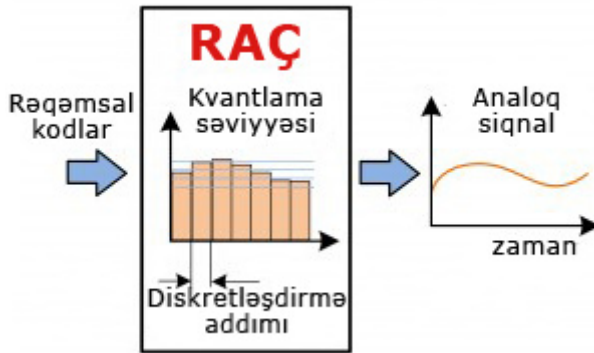
Цифровой сигнальный процессор

Rəqəmsal signal prosessoru – real vaxt rejimində signalların rəqəmsal emalı üçün nəzərdə tutulmuş xüsusi mikroprosessor. Analog Devices, Freescale və Texas Instruments şirkətləri DSP-prosessorların əsas istehsalçılarıdır. Üçşinli Harvard arxitekturalı DSP-prosessorlarda minimum 6-7 müxtəlif şin və 2-3 yaddaş bankı olur.

Digital-to-analog converter (DAC, D/A converter)

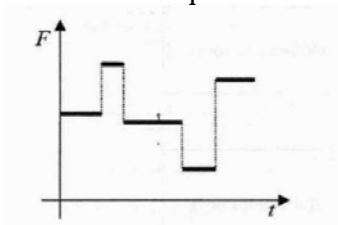
Цифро-аналоговый преобразователь (ЦАП)

Rəqəmsal-analoq çevirici (RAÇ) – rəqəmsal signalı analoq signala çevirən qurğudur.



Discrete signal Дискретный сигнал

Diskret signal – diskret zaman anlarında götürülmüş qiymətlər ardıcılığı şəklində təsvir olunan signal.



Discretization step Шаг дискретизации

Diskretləşdirmə addımı – iki ən yaxın diskretləşdirmə anı arasındakı zaman müddətidir.

Downsampling Даунсемплинг

Daunsemplinq – diskretləşdirmə tezliyinin azaldılması. Məsələn, kompakt diskdəki 44100 Hz tezlikli audionun tezliyi FM radio ilə verilməzdən əvvəl 22050 Hz-ə qədər azaldıla bilər. Beləliklə, audio yazının diskretləşdirmə tezliyi 2 dəfə azalmış olur.

Dynamic range Динамический диапазон

Dinamik diapazon – giriş signalının ən böyük amplitudunun bu signalın emalı üçün nəzərdə tutulmuş sistemin məxsusi küylərinin səviyyəsinə nisbətidir. ARÇ üçün zəruri mərtəbə sayını

qiymətləndirən zaman istifadə edilir. Baxılan termin amplituddan başqa digər parametrlərə də tətbiq edilə bilər (məsələn, *tezlik üzrə dinamik diapazon*).

Frames Кадры (фреймы)

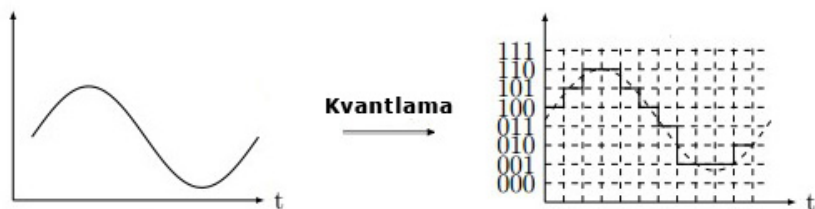
Freymlər (kadrlar) – emal zamanı nitq signalının bölündüyü kiçik hissələr. Hər kadr ilkin emala, məsələn, diskret Furye çevirməsinə məruz qalır. Nəticədə, hər freym *akustik əlamətlər vektoru* adlanan əmsallar yığını ilə səciyyələnir.

Nyquist frequency Частота Найквиста

Naykvist tezliyi – analog signalını onun ən yüksək tezliyindən iki dəfə böyük tezlikdə diskretləşdirdikdə, signalı tezlik spektrindən düzgün bərpa etmək olar. Bu diskretləşdirmə tezliyi Naykvist tezliyi adlanır.

Quantization Квантование

Kvantlama – signal amplitudunun qiyməti ən yaxın tam qiymətə (*kvantlama səviyyəsinə*) yuvarlaqlaşdırılır.



Kvantlama səviyyələrinin sayı çox olduqca yuvarlaqlaşdırmadan yaranan səhv (**quantization noise**, *kvantlama küyü*) kiçik olur.

Quantizing level Уровень квантования

Kvantlama səviyyəsi – verilmiş zaman anında analog signalın ani qiyməti kimi götürülən diskret qiymətdir. Kvantlama səviyyələrinin sayı amplituda ayrılmış bitlərin sayı ilə müəyyən edilir. Əgər amplituda ayrılmış bitlərin sayı n olarsa, onda kvantlama

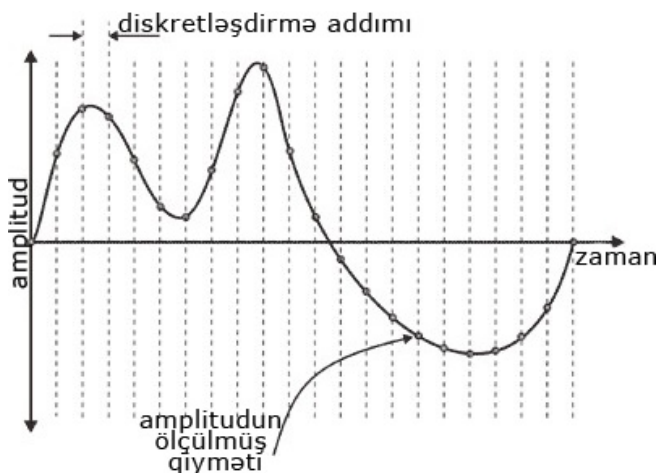
səviyyələrinin ümumi sayı 2^n olur. Onların yarısı amplitudun müsbət, digər yarısı isə mənfi qiymətlərinin yuvarlaqlaşdırılması üçün istifadə edilir.

Resampling Ресэмплирование

Resampling – bir diskretləşdirmə tezliyindən digərinə keçid.

Sample Отсчет

Bölgü – nöqtədə signalı xarakterizə edən müəyyən dəqiqlikdə ədəddir. Təsvirin dəqiqliyi alınan ədədlər diapazonunun enindən asılıdır.



Sampling Дискретизация

Diskretləşdirmə – kəsilməz signalın bir sıra diskret bölgülərə bölünməsi, hər bir bölgü uyğun zaman anında analog signalın qiymətini təsvir edir.

Sampling rate, Sampling frequency Частота дискретизации

Diskretləşdirmə tezliyi – bir saniyədə bölgülərin sayıdır. Məsələn, diskretləşdirmə tezliyi 16 kHs-dirsə, bu o deməkdir ki, 1 san ərzində signalın 16000 bölgüsü yadda saxlanır. Bu həm də o deməkdir ki, qonşu bölgülər arasında zaman intervalı 0.0000625 san-dir (0.0625 msan).

Diskretləşdirmə tezliyi kifayət qədər böyük olmalıdır, əks halda analog formasında olan siqnalın əhəmiyyətli dəyişiklikləri buraxılmış ola bilər. *Kotelnikov teoreminə* görə, diskretləşdirmə tezliyi (F_0) çevrilən siqnalın maksimal tezliyindən (F_{max}) ən azı iki dəfə böyük olmalıdır.

Telefon danışığı 8 kHs tezliyində diskretləşdirilir. Bir qayda olaraq, 16 kHs nitqin tanınması və sintezi üçün yetərli hesab edilir. Audio standartlar 44,1 kHs (kompakt disk) və 48 kHs (rəqəmsal audio lent) diskretləşdirmə tezliyindən istifadə edir. Siqnallar diskretləşdirməyə qədər filtrasiya edilməlidir.

Sampling resolution Разрешение выборки

Bit dəqiqliyi – siqnalın qiymətlərini kodlaşdırmaq üçün istifadə edilən bitlərin sayı. Adətən nitq siqnalları 16 bitdə saxlanır. Telefon danışığı 12 bitlik dinamik diapazonda 8 kHs-də diskretləşdirilir (qeyri-xətti funksiya, yəni A-qanun və ya U-qanun ilə 8 bitdə saxlanır). Qulağın dinamik diapazonu təxminən 20 bit təşkil edir.

Speech frequency range Частотный диапазон речи

Nitqin tezlik diapazonu – bütün toplananların yerləşdiyi oblastla müəyyən edilir. 100-4000 Hs daxilindədir. Nitqin (F_{max}) maksimal tezliyi 4 kHs-ə bərabər olduğundan, (F_0) diskretləşdirmə tezliyi $3 * F_{max} = 12$ kHs-dən bir qədər böyük olmalıdır.

Speech signal Речевой сигнал

Nitq siqnalı – səs mənbəyinin yaratdığı səs siqnalı ilə səs traktının rezonans xüsusiyyətləri hesabına qurşayanın “bükülməsinin” (vurulmasının) nəticəsidir. Məsələn, səs tellərinin rəqsləri zamanı havanın modulyasiyası hesabına yaranır.

Signalların filtrasiyası

1/f noise 1/шум

1/küy – gücünün spektral sıxlığı tezliklə (f) tərs mütənasib olan anomal fluktuasiyalardır (Qeyd edək ki, ağ küyün spektral sıxlığı sabitdir). 1/f küyü elektron lampının katodlarında termoemisiyanın yavaş xaosik dəyişmələri kimi aşkarlanmış və ona ingilis dilində “flicker effect” adı verilmişdi. Sonralar bu effekt digər texniki sistemlərdə, təbiətdə, canlı orqanizmlərdə və sosial sistemlərdə də aşkarlanıb.

Additive noise Аддитивный шум

Additiv küy – küyün növüdür, faydalı siqnalla cəmlənir.

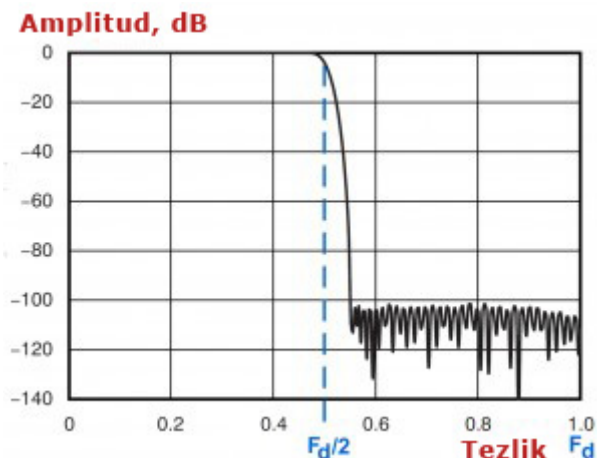
Additive white Gaussian noise

Аддитивный белый гауссов шум

Additiv ağ Gauss küyü – informasiyanın ötürülməsi kanalında maneə təsirinin bir növüdür. Müntəzəm spektral sıxlıq, amplitudun normal paylanmış qiyməti və siqnala additiv təsir üsulu ilə səciyyələnir. Küyün radio-rabitə sistemlərinin modelləşdirilməsi üçün istifadə edilən geniş yayılmış növüdür.

Anti-aliasing filter Антиэлайзинговый фильтр

Antialayzinq filtr – Siqnalın rəqəmsallaşdırılması keyfiyyətini artırmaq üçün ARÇ-nin girişində tətbiq edilən aşağı tezlikli analoq filtridir. Diskretləşdirmə zamanı spektrlərin bir-birini örtməsi təsirlərini aradan qaldırmaq üçün nəzərdə tutulub. Əgər F_d – ARÇ-nin diskretləşdirmə tezliyidirsə, antialayzinq filtrin *kəsmə tezliyi* təxminən $0,5F_d$ -yə bərabərdir. Diskretləşdirmə tezliyinin yarısından böyük bütün tezlikləri söndürür.



Band pass filter (BPF) Полосовой фильтр

Zolaq filtri – yalnız müəyyən tezlik diapazonunda olan tezlikləri buraxır, bütün qalan tezlikləri tutur.

Band stop filter Полосно-заграждающий фильтр

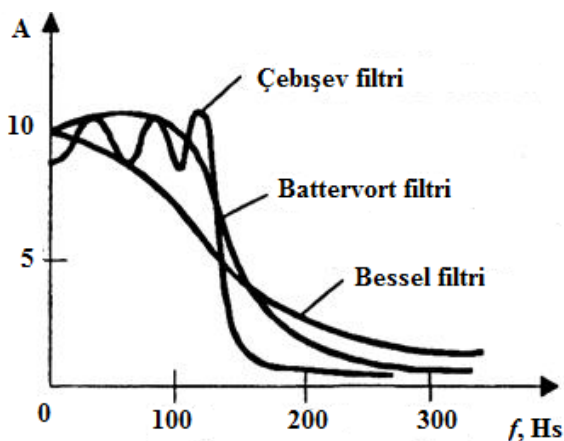
Zolağı bağlayan filtr – verilmiş tezliklər diapazonunu bloklayan filtr.

Bandwith filter *Bax. Band pass filter*

Bessel filter Фильтр Бесселя

Bessel filtri – maksimal xətti tezlik xarakteristikası almağa imkan verən analoq filtridir. Filtrin çıxış signalının faza sürüşməsi giriş signalının tezliyinə mütənəsibdir və bu tələb spektri buraxma zolağında yerləşən signalın gecikmə müddətinin sabit olması şərtinə ekvivalentdir. Bunun nəticəsində Bessel filtri signalın formasını praktik olaraq dəyişməz saxlayır. Zaman oblastında kodlaşdırılmış siqnalların hamarlanması üçün istifadə olunur.

Şəkində bəzi filtrlərin amplitud (A) və tezlik (f) xarakteristikaları göstərilib.

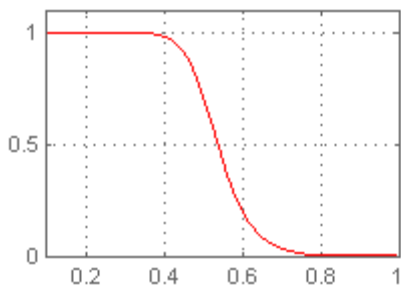


Biquad filter Биквадратный фильтр

Bikvadrat filtr – ötürmə funksiyasının iki polyusu və ən çoxu iki sıfırı olan analoq və ya rəqəmsal filtr. Daha mürəkkəb filtrlərin qurulması üçün istifadə edilir, bunun üçün onları ardıcıl qoşurlar.

Butterworth filter Фильтр Баттерворта

Battervort filtri – siqnalların tezlik seleksiyası üçün nəzərdə tutulmuşdur. Buraxma zolağında kifayət qədər kiçik qeyri-bərabərlik olduqda amplitud-tezlik xarakteristikası (ATX) enişinin maksimal dikliyini təmin edir. Həm analoq, həm də rəqəmsal ola bilər. Battervort filtri *maksimal hamar ATX filtri* də adlandırırlar.



Chebyshev filter Фильтр Чебышёва

Çebişev filtri – xətti analoq və ya rəqəmsal filtrlərin bir növüdür, digər filtrlərlə müqayisədə fərqləndirici xüsusiyyəti amplitud-tezlik xarakteristikasının daha sərt enişi və buraxma zolağındakı tezliklərdə bu xarakteristikanın əhəmiyyətli pulsasiyalar etməsidir. Birinci və ikinci növ Çebişev filtrlərini fərqləndirirlər.

Comb filter Гребенчатый фильтр

Daraq filtri – signal bu filtrdən keçdikdə ona signalın özü bir qədər gecikmə ilə əlavə edilir. Nəticədə faza kompensasiyası alınır. Daraq filtrləri, məsələn, əks-səda effekti yaratmaq üçün istifadə edilir. Daraq filtrləri şəkildə nöqtələrin sürüşməsi kimi təhrifləri azaldır. Daraq filtrinə amplitud-tezlik xarakteristikası bərabər paylanmış bir sıra piklərdən ibarətdir və buna görə qəfəsə bənzəyir.

Digital filters Цифровые фильтры

Rəqəmsal filtrlər 1) təhrif olunmuş siqnalın bərpa edilməsi 2) verilmiş signalda müəyyən tezlik diapazonuna aid komponentləri ayırmaq və bu diapazona aid olmayanları buraxmaq üçün istifadə edirlər. Bu problemləri analoq filtrlərin köməyi ilə də həll etmək olar, lakin onların effektivliyi xeyli dərəcədə aşağıdır. Aşağı tezlik filtrləri, yuxarı tezlik filtrləri, zolaq filtrləri və zolaq-bağlayan filtrləri (onları *rejektor* və ya *tıxac filtrləri* də adlandırırlar) ayırırlar.

Filter bank Банк фильтров

Filtrlər bankı – tədqiq edilən bütün tezlik diapazonunu örtən eyni tipli zolaq filtrlərinin toplusudur.

Filter design Разработка фильтров

Filtrlərin yaradılması – Aşağı tezliklər filtrindən yüksək tezliklər filtrinə, onlardan da zolaq və rejektor filtrlərini almaq olar. Bu məqsədlə *spektral inversiya* və *spektral reversiya* metodları istifadə edilir.

Filtration Фильтрация

Filtrasiya – rəqəmsal çevirməyə məruz qalacaq tezliklərin kəsilməz diapazonunun ayrılması prosesidir.

Finite Impulse Response (FIR)

Конечная импульсная характеристика (КИХ)

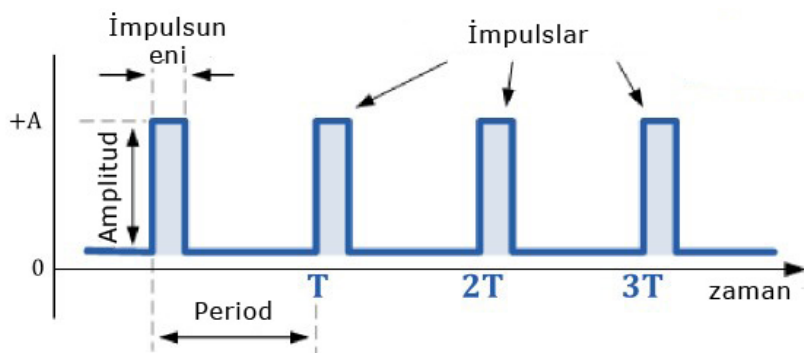
Sonlu impuls xarakteristikası – sonlu sayda sıfırdan fərqli bölgüləri olan impuls xarakteristikası. Çox zaman bu termin filtrasiyanın əsasında rekursiv alqoritmlərin deyil, bükülmə əməliyyatının olduğunu qeyd etmək üçün işlədilir.

High-pass filter Фильтр верхних частот

Yüksək tezliklər filtri – signalın aşağı tezlikli toplananlarını tutur və yalnız yüksək tezlikli komponentlərini buraxır.

Impulse Импульс

İmpuls – çox qısa zaman intervalı istisna olmaqla hər yerdə sıfır qiymət alan signal (*delta-impuls*).



Diskret signal halında delta-impulsda yalnız bir bölgü olur. Kəsilməz signal halında delta impuls müddəti qurğunun bütün daxili qovşaqlarının keçid proseslərinin müddəti ilə müqayisədə olduqca kiçik olan signaldır.

Impulse characteristic Импульсная характеристика

İmpuls xarakteristikası – sıfır başlanğıc şərtlərində sistemin vahid impulsa (delta-funksiyaya) reaksiyası.

Infinite impulse response (IIR)

Бесконечная импульсная характеристика (БИХ)

Sonsuz impuls xarakteristikası – sonsuz sayda sıfırdan fərqli bölgüləri olan impuls xarakteristikası. Eksponent üzrə sönən impuls xarakteristikası misal ola bilər. Çox zaman bu termin filtrasiyanın əsasında bükülmə əməliyyatının deyil, rekursiv alqoritmlərin olduğunu qeyd etmək üçün işlədilir.

Linear digital filter Линейный цифровой фильтр

Rəqəmsal xətti filtr – müəyyən xətti sistemdir, buna görə onu, impuls xarakteristikası kimi anlayışla xarakterizə etmək olar, onu çox zaman *filtrin nüvəsi* də adlandırırlar. Beləliklə, filtrlənmiş signal ilkin signalın filtr nüvəsi ilə *bükülməsi* ilə alınır. Signalın filtrasiyasının digər üsulu *rekursiv filtrasiya*dır.

Low-pass filter Фильтр нижних частот (ФНЧ)

Aşağı tezliklər filtri (ATF) – signalın tezliyi müəyyən qiymətdən kiçik olan komponentlərini buraxır.

MFCC FB-20 filter bank Банк фильтров MFCC FB-20

MFCC FB-20 filtr bankı – eşitmənin psixoakustik təbiətini nəzərə alır. Praktikada aşağıdakı filtrlər bankı daha geniş istifadə edilir:

- **MFCC FB-20** – Davis və Mermelstein tərəfindən 1980-ci ildə təklif edilib, diskretləşdirmə tezliyi 10 kHs, səs zolağı [0; 4600] Hs-dir.
- **MFCC FB-24 HTK** – Kembric HMM Toolkit (HTK), 1995-ci ildə təklif edilib, 24 filtr var, diskretləşdirmə tezliyi 16 kHs, səs zolağı [0, 8000] Hs-dir.
- **MFCC FB-40** – MATLAB üçün Auditory Toolbox (1998-ci il); diskretləşdirmə tezliyi 16 kHs, səs zolağı [133, 6854] Hs-dir.

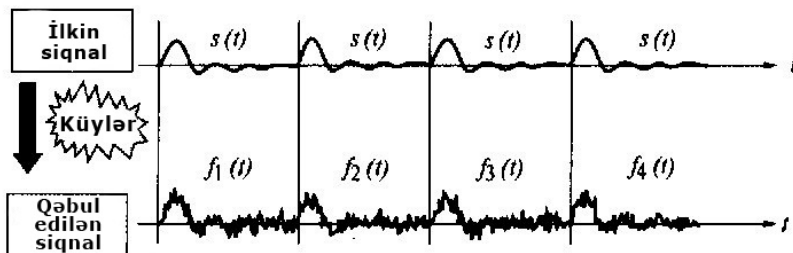
- **HFCC-E FB-29** (Human Factor Cepstral Coefficients, 2004-cü il) – diskretləşdirmə tezliyi 12.5 kHs, səs zolağı [0, 6250] Hs-dir.

Multiplicative noise Мультипликативный шум

Multiplikativ küy – rabitə xəttinin xarakteristikalarının dəyişməsi nəticəsində kanalın ötürmə əmsalının təsadüfi dəyişmələri ilə şərtlənir. Additiv küylər isə aparaturanın komponentlərində istilik prosesləri, atmosfer və sənaye prosesləri, qonşu kanalların işi ilə əlaqəli olan fluktuasiya hadisələri nəticəsində yaranır. Multiplikativ küylərə digər tezliklərlə qarşılıqlı təsir nəticəsində yaranan kombinasiya tezlikləri, parazit modulyasiya və rəqəmsal sistemlərdə kvantlama küyləri aid edilir. Multiplikativ küy siqnalı deformasiya edə – informasiya hissəsinin formasını dəyişə bilər, siqnalın qiymətlərindən və siqnalın müəyyən xüsusiyyətlərindən asılı ola bilər. Multiplikativ küylər küylə siqnalın qarşılıqlı təsiri nəticəsində meydana çıxır. Riyazi olaraq bu siqnal və küyün tezlik komponentlərinin bir-birinə vurulması ilə təsvir olunur.

Noise Шум

Küy – artikulyasiya orqanları tərəfindən yaradılan maneələri keçən zaman hava axınının yaratdığı səslənmədir. Küy impulsu (küy impulsu – partlayış-qovuşuq yerinin açılması zamanı yaranır, məsələn, [p] tələffüz edilərkən) və turbulent (hava axının yolunda ensiz yarıq əmələ gətirən artikulyasiya orqanlarının daralması nəticəsində hava axınlarının sürtünməsi küyü, məsələn, [s] yaranarkən) olur.



Peak signal-to-noise ratio Пиковое отношение сигнала к шуму

PSNR (peak SNR) – signalın pik səviyyəsinin küyə olan nisbəti

Pre-emphasis Предварительное усиление

İlkin gücləndirmə – signalı signal spektrində yüksək tezlikləri gücləndirən filtr tətbiq edilir:

$$H_{pre}(z) = 1 + a_{pre} z^{-1}.$$

İnsan 1 kHs-dən yüksək olan tezlikləri daha yaxşı qavrayır. a_{pre} əmsalının qiyməti, adətən, $[-0.1, -0.4]$ aralığında seçilir.

Signal-to-Noise-Ratio (SNR) Отношение сигнал/шум

"Signal-küy" nisbəti – faydalı signalın gücünün küyün gücünə olan nisbətinə bərabərdir:

$$SNR = \frac{P_{signal}}{P_{noise}} = \left(\frac{A_{signal}}{A_{noise}} \right)^2$$

Adətən, signal-küy nisbəti desibel (dB) ilə ifadə olunur. Bu nisbət nə qədər böyükdürsə, küy sistemin xarakteristikalarına bir o qədər az təsir edir.

$$SNR(dB) = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right)$$

Yüksək dərəcədə nitq aydınlığına nail olmaq üçün signal ən azı 15 dB küy səviyyəsini keçməlidir, eşitmə əngəlli insanlar üçün isə ən azı 20 dB-dan yüksək olmalıdır.

Spectrum inversion method Метод спектральной инверсии

Spektral inversiya metodunda yüksək tezliklər filtrinin nüvəsi eyni dayanma tezliyinə malik aşağı tezliklər filtrinin nüvəsindən alınır, bunun üçün delta funksiyaya aşağı tezliklər filtrinin inversiya edilmiş (minus işarəsi ilə götürülmüş) nüvəsi əlavə edilir.

Fərz edək ki, $x[n]$ giriş signalı yüksək tezlikli $x_1[n]$ signalına və aşağı tezlikli $x_2[n]$ signalına ayrılır:

$$x[n] = x_1[n] + x_2[n].$$

Tutaq ki, aşağı tezlikli signalı almaq üçün nüvəsi $h_2[n]$ olan filtr istifadə edilir, yəni

$$x[n] * h_2[n] = x_2[n].$$

Bükülmənin $x[n] * \delta[n] = x[n]$ xassəsini nəzərə alsaq:

$x_1[n] = x[n] - x_2[n] = x[n] * \delta[n] - x[n] * h_2[n] = x[n] * (\delta[n] - h_2[n])$ alarıq, yəni yüksək tezliklər filtrinin nüvəsi belə olacaq:

$$h_1[n] = \delta[n] - h_2[n].$$

Spectrum reversion method Метод спектральной реверсии

Spektral reversiya metodunda aşağı tezliklər filtrinin nüvəsində hər ikinci bölgünün amplitudu (-1)-ə vurulur. Nəticədə yüksək tezliklər filtrinin nüvəsi alınır. Bu hadisə onunla izah olunur ki, hər ikinci amplitudun (-1)-ə vurulması filtr nüvəsinin tezliyi 0.5 olan sinusoidə vurulmasına ekvivalentdir. Bu zaman f_1 dayanma tezliyi ($0.5 - f_2$) olacaq, burada f_2 – götürülmüş aşağı tezliklər filtrinin dayanma tezliyidir.

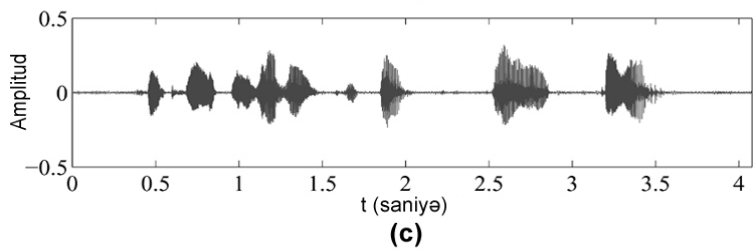
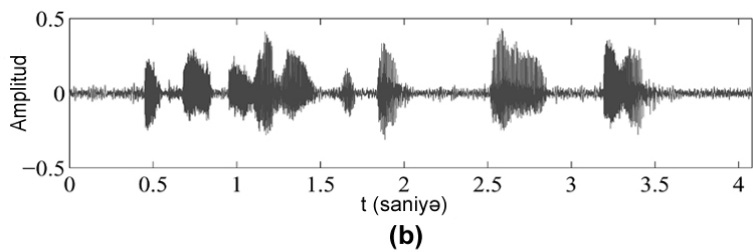
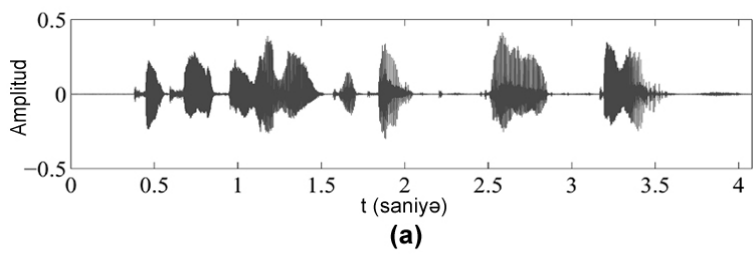
White noise Белый шум

Ağ küy – spektral sıxlığı tezlikdən asılı olmayan təsadüfi küydür, yalnız signalın zaman bölgülərinin qarşılıqlı əlaqəsi olmadıqda mümkündür.

Wiener filter Фильтр Винера

Viner filtri – signal və küyün verilmiş xassələrində signal/küy nisbətinin artırılması baxımından optimal olan filtr (1940-cı illərdə Norbert Viner tərəfindən təklif edilib).

Şəkildə təmiz nitq signalı (a), küylü nitq signalı (b) və Viner filtri ilə küydən təmizlənmiş nitq signalı (c) göstərilmişdir.



Nitq signallarının spektral analizi

AMDF (Average magnitude difference function) algorithm
Алгоритм вычисления среднего значения разностной функции

Fərq funksiyasının orta amplitudunun hesablanması algoritmi
– əsas ton periodunun hesablanması üsullarından biridir. AMDF metodunda ilkin nitq signalı ilə zaman oxu üzrə sürüşdürülmüş signalın fərqi hesablanır:

$$A(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} |s(n) - s(n+k)|, k = 0, 1, \dots, N-1$$

Amplitude method for pitch detection

Амплитудный метод обнаружения высоты

Əsas ton periodunun hesablanması amplitud metodu – küy səviyyəsi kiçik olduqda vokal səsin stasionar hissəsində səs rəqslərinin forması demək olar ki, əsas tonun hər bir periodunda təkrarlanır. Qlobal maksimumlar arasındakı məsafəni səs signalının əsas ton periodunun təxmini qiymətinə bərabər hesab etmək olar. Amplitud metodlarının əsas çətinliyi yalançı lokal maksimumları söndürməkdir. Lakin bu zaman həqiqi maksimumu buraxmaq ehtimalı da artır.

Amplitude spectrum Амплитудный спектр

Amplitud spektri – tezliklərin şkalasında bütün amplitudların məcmusudur.

Autocorrelation Автокорреляция

Avtokorrelyasiya – signalın özü ilə korrelyasiyası. Avtokorrelyasiya funksiyasının Furiye çevirməsi ilkin signalın güc (enerji) spektrini verir. Analıq signallar üçün avtokorrelyasiya funksiyası $s(t)$ signalı ilə onun zamana görə sürüşdürülmüş $s(t - \tau)$ surətinin sonsuz intervalda skalyar hasilindən ibarətdir:

$$\Psi(\tau) = \int_{-\infty}^{\infty} s(t)s(t-\tau)dt .$$

Diskret $s[n]$, $n=1,2,\dots,N$ signalı üçün avtokorrelyasiya ardıcılığı $r_{ss}[m]$ aşağıdakı kimi müəyyən edilir:

$$r_{ss}[m] = \frac{1}{N} \sum_{n=1}^{N-m+1} s[n]s[n+m-1] \text{ (Biased autocorrelation)}$$

və ya

$$r_{ss}[m] = \frac{1}{N-|m|} \sum_{n=1}^{N-m+1} s[n]s[n+m-1] \text{ (Unbiased autocorrelation)}$$

burada $m=1,2,\dots, M+1$ – gecikmələr, M – gecikmələrin sayı, N – bölgülərin sayıdır.

Autocorrelation method for pitch detection

Автоткорреляционный метод обнаружения высоты

Əsas ton periodunun hesablanması avtokorrelyasiya metodu – metodun əsasında belə bir teorem durur: periodik signalın avtokorrelyasiya funksiyası da periodikdir və onların periodları üst-üstə düşür. Avtokorrelyasiya funksiyası aşağıdakı düsturla verilir:

$$R_n(k) = \sum_{m=0}^{N-k-1} x(n+m)x(n+m+k),$$

burada N – analiz edilən freymin (kadrın) uzunluğu, n – analiz edilən freymin başlanğıcının bütün signalda cari koordinatı, k – avtokorrelyasiya funksiyası əmsalının nömrəsidir.

Blackman window Окно Блэкмана

Blekman pəncərəsi – Hanning və Hemming pəncərələrinə oxşardır, amma o, döyünləri daha güclü azaldan əlavə kosinus komponentinə malikdir. Onu belə müəyyən etmək olar:

$$w[n] = 0.42 - 0.5 \cos(2\pi n/N) + 0.08 \cos(4\pi n/N), n = 0,1,\dots, N-1.$$

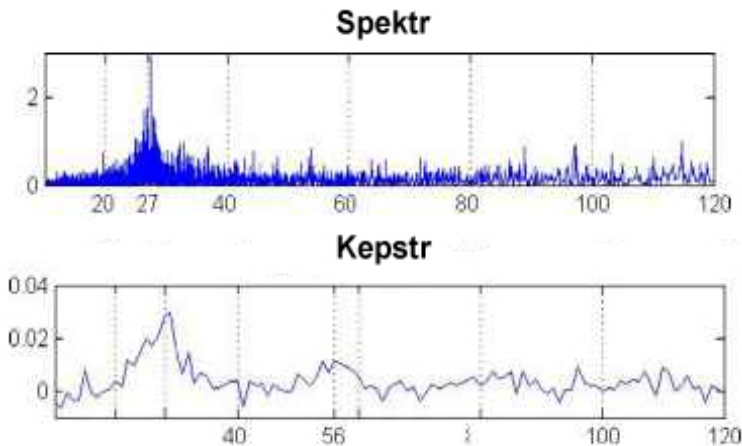
Blackman-Harris window Окно Блэкмана-Харриса

Blekman-Harris pəncərəsi – pəncərə funksiyasıdır, aşağı səviyyəli toplananların, məsələn, giriş signalında böyük təhriflərin ölçülməsi üçün istifadə olunur. O, üç toplanandan ibarətdir:

$$w[n] = 0.422323 - 0.49755 \cos(2\pi n/N) + \\ + 0.07922 \cos(4\pi n/N), n = 0, 1, \dots, N-1$$

Cepstrum Кепстр

Кепстр – signalın spektrinin loqarifmik miqyasda təsviridir (“spektrin spektridir”). Burada loqarifmləmə əməli ilkin spektrdə fluktasiyaları müəyyən qədər hamarlamağa yönəlib ki, keptrdə “parazit” komponentlərin sayı azalsın. Termin «spektr» sözündə hərflərin yerini dəyişməklə alınıb (B.P.Bogert, M.J.R.Healy və J.W.Tukey tərəfindən 1963-cü ildə təklif edilib, frequency→quefrequency, filter→lifter, harmonics→rahmonics və s. kimi terminlər də mövcuddur). Kepstr spektral informasiyanı daha yığcam təsvir etməyə imkan verir. Kepstrin həqiqi, kompleks, enerji (ing. power cepstrum) və faza kepstri kimi növlərini müəyyən etmək olar.



Cepstrum method for pitch detection

Кепстральный метод обнаружения высоты

Əsas tonun hesablanması keptral metodu – analiz edilən hər bir freym üçün alınmış keptr əsas tonun mümkün qiymətlər oblastında pik axtarmaq məqsədi ilə tədqiq edilir. Əgər keptrdə pik keçid qiymətindən böyükdürsə, freym vokal freym kimi təsnif edilir

və pikin yeri əsas ton periodunun qiyməti kimi götürülür. Kepstrin köməyi ilə analiz edilən freymin uzunluğu, adətən, kişi səsləri üçün 512 bölgüyə və qadın səsləri üçün 256 bölgüyə bərabər götürülür. Freymdən-freymə sürüşmənin ölçüsü analiz edilən freymin uzunluğunun $1/16$ - $1/4$ -i kimi seçilir.

Continuous Wavelet Transform (CWT)

Непрерывное вейвлет-преобразование

Kəsilməz veyvlet çevirməsi – $f(t) \in L^2(R)$ funksiyasının kəsilməz veyvlet çevirməsi onun $\psi_{a,b}(t) = a^{-1/2} \psi\left(\frac{t-b}{a}\right)$, $a \in R^+$, $b \in R$ bazis funksiyalarına skalyar hasilinə deyilir:

$$W_f(a,b) = \int_{-\infty}^{\infty} f(t) \cdot \psi_{a,b}(t) dt.$$

$\psi_{a,b}(t) \in L^2(R)$ bazis funksiyaları müəyyən oblastda təyin olunurlar (orta qiymətləri 0-a bərabərdir). Bu funksiylara veyvlet deyilir. Onlara *ana veyvlet* (ing. mother wavelet) adlanan müəyyən $\psi(t)$ funksiyasının sürüşməsi və miqyaslanması kimi baxmaq olar. b – parametri sürüşməni, a isə miqyası idarə edir. a -nın böyük qiymətləri aşağı tezliklərə, kiçik qiymətləri isə yüksək tezliklərə uyğun gəlir.

Cosinus window Косинусное окно

Kosinus pəncərəsi – pəncərə funksiyasıdır, X giriş ardıcılığından aşağıdakı düsturla verilən ümumi kosinus pəncərəsi ilə emal edilmiş Y çıxış ardıcılığı formalaşdırır:

$$y_i = x_i(a_0 - a_1 \cos(2\pi i/n) + a_2 \cos(4\pi i/n) - a_3 \cos(6\pi i/n) + \dots + (-1)^k a_k \cos(2k\pi i/n))'$$

burada a_k – pəncərənin kosinus əmsalları; $k = 0, 1, \dots, m$; n – X ardıcılığında elementlərin sayıdır.

Discrete cosine transform (DCT) Дискретное косинусное преобразование (ДКП)

Diskret kosinus çevirməsi – Furiye çevirməsinə oxşar çevirmədir. Sıqnalı kosinusoidal komponentlərinə ayırmağa imkan verir.

Düz çevirmə:

$$X(m) = \lambda_m \sum_{n=0}^{N-1} x(n) \cos \left(\pi \frac{(n + \frac{1}{2})m}{N} \right), m = 0, \dots, N-1,$$

$$\text{burada } \lambda_m = \begin{cases} \frac{1}{\sqrt{N}}, & m = 0, \\ \sqrt{\frac{2}{N}}, & 1 \leq m \leq N-1 \end{cases}.$$

Tərs çevirmə:

$$x(n) = \sum_{m=0}^{N-1} \lambda_m X(m) \cos \left(\pi \frac{(n + \frac{1}{2})m}{N} \right), n = 0, \dots, N-1.$$

Discrete Fourier transform (DFT) Дискретное преобразование Фурье (ДФФ)

Diskret Furiye çevirməsi (DFÇ) – Furiye çevirməsinin diskret sıqnalı tətbiqidir:

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-\frac{2\pi j}{N} kn}, \quad k = 0, \dots, N-1, \quad j = \sqrt{-1}.$$

Sıqnalı görə spektri (harmonik toplananları) hesablamağa imkan verir, sıqnalın emal vaxtını əhəmiyyətli dərəcədə azaldır.

Fast Fourier Transform (FFT) Быстрое преобразование Фурье (БПФ)

Sürətli Furiye çevirməsi (SFÇ) – diskret Furiye çevirməsinin sürətli hesablanması alqoritmidir. SFÇ-nin əsasını diskret sıqnalın verilmiş bölgülər ardıcılığının bir neçə aralıq ardıcılığa bölünməsi prinsipi təşkil edir. N bölgü üçün SFÇ-də təxminən $N \cdot \log_2 N$ sayda əməliyyat olur. Məsələn, 256 bölgü üçün əməliyyatların sayı 2048-dir (DFT-də 65536-dir).

Fourier transform Преобразование Фурье

Furye çevirməsi – spektral analizin riyazi əsasıdır, zaman və ya məkan signalını (və ya bu signalın müəyyən modelini) signalın tezlik oblastındakı təsviri ilə əlaqələndirir. İki integral çevirməsi cütüdür:

$$\text{Düz çevirmə: } F(\nu) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i \nu x} dx;$$

$$\text{Tərs çevirmə: } f(x) = \int_{-\infty}^{\infty} F(\nu)e^{2\pi i \nu x} d\nu.$$

Frequency spectrum Частотный спектр

Tezlik spektri – tezliklər şkalasında bütün fazaların məcmusudur.

Gaussian window Окно Гаусса

Qauss pəncərəsi – pəncərə funksiyasıdır, aşağıdakı kimi təyin edilir:

$$w[n] = e^{-\frac{1}{2}\left(\alpha \frac{n}{N/2}\right)^2}, \quad -\frac{N}{2} \leq n \leq \frac{N}{2}.$$

Pəncərənin eni α parametrinə tərs mütənasibdir: onun artması ilə zaman pəncərəsinin eni azalır.

Hamming window Окно Хемминга

Hemminq pəncərəsi – pəncərə funksiyasıdır, Hanning pəncərəsinin dəyişilmiş versiyasıdır. Onun forması da kosinusoidal signalın formasına oxşardır və bu ifadə ilə təyin edilir:

$$w(nT) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), & 0 \leq n \leq N-1 \\ 0, & n < 0, \quad n > N-1 \end{cases}$$

Hanning window Окно Ханнинга

Hanning pəncərəsi – pəncərə funksiyasıdır, kosinusoidal signalın yarım perioduna oxşar formaya malikdir və bu tənliklə təyin edilir:

$$w[n] = 0.5 - 0.5 \cos(2\pi n/N), n = 0, 1, \dots, N-1.$$

Inverse Discrete Fourier Transform (IDFT)

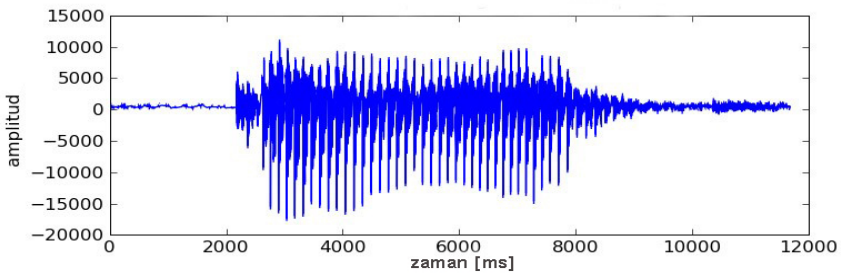
Обратное дискретное преобразование Фурье (ОДПФ)

Tərs diskret Furiye çevirməsi – spektrə görə signalı hesablamağa imkan verir:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{\frac{2\pi}{N} kn}, \quad n = 0, \dots, N-1.$$

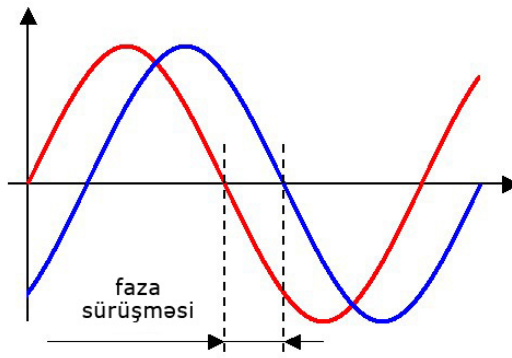
Oscillogram Осциллограмма

Oscilloqram – səs dalğası rəqslərinin amplitudundakı dəyişmələrin zamana görə açılmasıdır. O, səs fraqmentlərinin gurluğunu, müddətini və onların arasında intervalları müəyyən etməyə imkan verir.



Phase Фаза

Faza – rəqslərin zamana görə yerdəyişməsidir.



Rectangular window Прямоугольное окно

Düzbucaqlı pəncərə – bütün zaman intervalında vahidə bərabər olan pəncərə funksiyasıdır:

$$w[n] = 1, n = 0, 1, \dots, N - 1,$$

burada N – pəncərənin enidir. Düzbucaqlı pəncərənin tətbiqi, ümumiyyətlə, hər hansı pəncərənin istifadəsinə ekvivalent deyil, çünki düzbucaqlı funksiya sonlu zaman intervalında signalı kəsir. Belə pəncərə spektrin ən böyük yaygınlaşma dərəcəsini verir.

Short-Time Fourier Transform (STFT) Кратковременное преобразование Фурье

Qısamüddətli Furiye çevirməsi – signalın tezlik-zaman təsvirini almağa imkan verir.

$$S_u(\omega, b) = \int_{-\infty}^{\infty} u(t) \cdot h(t-b) \cdot e^{-i\omega t} dt,$$

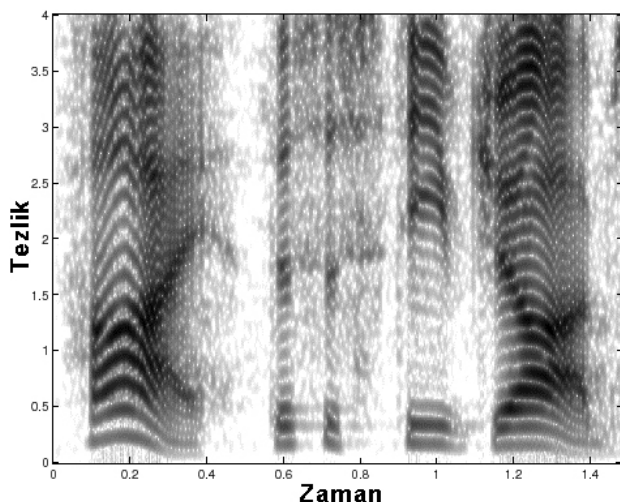
burada $h(t)$ – zaman pəncərəsi, b – zaman oxu üzrə sürüşmə, j – xəyali vahiddir, $j^2 = -1$.

Sound spectrum Спектр звука

Səsin spektri – səsin keçməsi zamanı yaranan hava dalğası rəqslərinin amplitud və tezlik koordinatlarında təsviridir.

Spectrogram Спектрограмма

Spektrogram – akustik signalın gücünün tezliklər üzrə zamana görə paylanmasını əyani göstərir. İkiölçülü diaqram spektrogramın ən geniş yayılmış təsvir formasıdır: üfüqi oxda zaman, şaquli oxda tezlik göstərilir; üçüncü ölçü konkret zaman anında müəyyən tezlikdə amplitudu göstərməklə şəkildə intensivliklə və ya nöqtənin rəngi ilə təsvir edilir. *Sonoqram* da adlanır.



Spectrum analysis Спектральный анализ

Spektral analiz – siqnalların emalı metodlarından biridir, ölçülən siqnalın tezlik tərkibini xarakterizə etməyə imkan verir.

Spectrum method for pitch detection

Спектральный метод обнаружения высоты

Əsas ton periodunun hesablanması spektral metodu – səs traktının vokal həyəcanlanması zamanı siqnalın spektrində əsas ton tezliyinin misilləri olan tezliklərdə piklər olur. Əgər tezlik üzrə kifayət qədər kiçik diskretləşdirmə addımı ilə diskret Furiye çevirməsinə baxsaq, spektrin enerjisinin maksimum qiymətinə uyğun tezliyi əsas ton tezliyinin qiyməti kimi götürmək olar. Maksimumların axtarışını 50-350 Hz intervalında aparmaq lazımdır. Lakin tez-tez elə situasiya olur ki, bu intervalda əsas tonun ikinci harmonikası da yerləşir və bəzən onun enerjisi daha böyük olur. Bu halda o səhvən əsas ton periodu kimi götürülür. Bundan qaçmaq üçün spektrin maksimumuna deyil, spektrin hər hansı qeyri-xətti funksiyasının (çox vaxt loqarifmin) maksimumuna baxırlar.

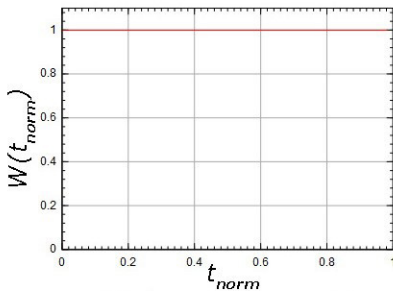
Wavelet Transform (WT)

Вейвлет-преобразование

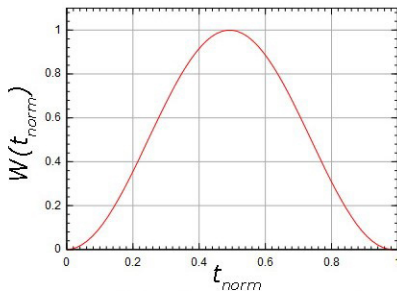
Veyvlet çevirməsi (ing. *wavelet* – *kiçik dalğa* deməkdir) – zaman və tezlik üzrə dəqiqliklə bağlı problemləri aradan qaldırmaq üçün qısamüddətli Furiye çevirməsinə alternativ kimi təklif edilmişdir. (Veyvlet termini 1982-ci ildə A.Grossman və J.Morle tərəfindən daxil edilmişdi.) Praktikada rast gəlinən siqnallarda çox vaxt yüksək tezlikli komponentlərin müddəti kiçik, aşağı tezlikli komponentlərin müddəti isə kifayət qədər böyük olur. Qısamüddətli Furiye çevirməsində zaman və tezlik oxu üzrə dəqiqlik bütün çevirmə üçün sabit kəmiyyətdir. Veyvletlərin köməyi ilə yüksək tezliklər oblastında zaman üzrə yüksək və tezlik üzrə aşağı dəqiqlik, aşağı tezliklər oblastında isə tezlik üzrə yüksək və zaman üzrə aşağı dəqiqlik əldə etmək olar.

Window function Оконная функция

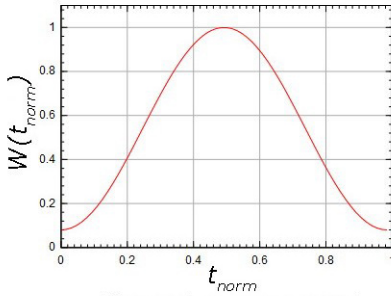
Pəncərə funksiyası – düzbucaqlı impulsdur. Siqnalın pəncərədə emalının məqsədi freymlərə bölmə (*segmentləmə*) nəticəsində meydana çıxan sərhəd effektlərinin azaldılmasıdır. Arzuolunmaz sərhəd effektlərinin qarşısını almaq üçün siqnalı $w(n)$ pəncərə funksiyasına vurmaq qəbul olunmuşdur: $x(n) = E(n)w(n)$.



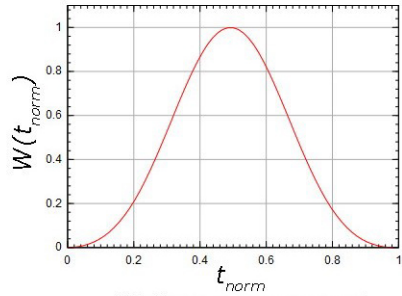
Düzbucaqlı pəncərə



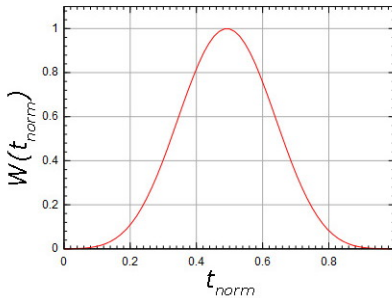
Hanning pəncərəsi



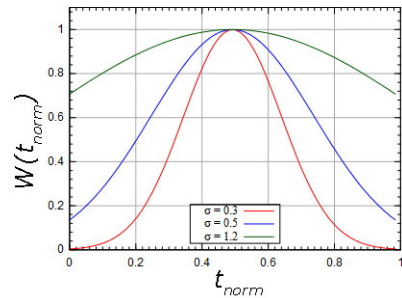
Hemming pəncərəsi



Blekman pəncərəsi



Blekman-Harris pəncərəsi



Qauss pəncərəsi

YIN algorithm Алгоритм YIN

YIN algoritmi – əsas ton periodunun hesablanması üsullarından biridir. YIN algoritmi avtokorrelasiyaya və fərq funksiyasına əsaslanır, hasilə maksimallaşdırmaq əvəzinə nitq signalı funksiyası ilə zaman oxu üzrə sürüşdürülmüş funksiya arasındakı fərqi minimallaşdırmağa çalışırlar (dinamik proqramlaşdırma istifadə edilir).

Z-transform Z-преобразование

Z-çevirmə – diskret siqnalların analizində geniş istifadə edilən üsuldur. Z-çevirmə $\{x(k)\}$ ardıcılığına z kompleks dəyişəninə aşağıdakı düsturla müəyyən edilən funksiyasını qarşı qoyur:

$$X(z) = \sum_{k=-\infty}^{k=\infty} x(k)z^{-k}.$$

Aydındır ki, $X(z)$ funksiyası z -in yalnız yuxarıdakı sıranın yığıldığı qiymətləri üçün təyin olunur.

Z-çevirmə diskret siqnalların analizində Laplas çevirməsinin analoq siqnallar üçün oynadığı rolu oynayır. Burada mühüm fakt ondan ibarətdir ki, diskret sistemin impuls xarakteristikasının Z-çevirməsi z dəyişəninin kəsr-rasional funksiyasıdır.

Nitq siqnallarının parametrləşdirilməsi

Acoustical spectrum Акустический спектр звука

Akustik spektr – mürəkkəb səsin sinusoidal komponentlərinin toplusudur, komponentlər amplitud və tezliklərin köməyi ilə verilir.

Amplitude-frequency characteristic

Амплитудно-частотная характеристика (АЧХ)

ATX (Amplitud-tezlik xarakteristikası) – siqnalın amplitudasının tezlikdən asılılığıdır

Cepstral analysis Кепстральный анализ

Kepstral analiz – spektral analizin bir formasıdır, bu zaman $x(t)$ siqnalına aşağıdakı çevirmə tətbiq olunur:

$$c(t) = F^{-1}[\log(F(x(t)))],$$

burada F – diskret Furiye çevirməsidir.

Cepstral coefficients Кепстральные коэффициенты

Kepstral əmsallar – spektr maqnitudunun loqarifminə tərs Furiye çevirməsinin tətbiqi nəticəsində alınmış kepsstral əlamətlər (məsələn, MFCC, LPCC və PLP).

DCC (Delta-Cepstral Coefficients)

Дельта-кепстральные коэффициенты

Delta kepsstral əmsallar – tanıma dəqiqliyini artırmaq üçün birinci tərtib (delta-əmsallar, Δ -əmsallar) və ikinci tərtib törəmələr (delta-delta-əmsallar, $\Delta\Delta$ -əmsallar) istifadə edilir; onlar uyğun olaraq, kepsstral əmsalların dəyişmə sürətini və təcilini əks etdirən parametrlərdir. Delta-kepsstral əmsal (Δc_k) aşağıdakı düsturla hesablanı bilər (digər düsturlar da mövcuddur):

$$\Delta c_k = c_{k+2} - c_{k-2},$$

burada c – kepsstral əmsallar vektoru, N – freymdə bölgülərin sayıdır.

DDCC (Delta-Delta-Cepstral Coefficients) və ya Double Delta Cepstral Coefficients

Дельта-дельта-кепстральные коэффициенты

Delta-delta cepstral əmsallar – ikinci tərtib törəməni approksimasiya edir, delta cepstral əmsallar əsasında aşağıdakı kimi hesablamaq olar:

$$\Delta\Delta C_k = \Delta C_{k+1} - \Delta C_{k-1}.$$

Frequency-domain Features

Спектрально-временные признаки

Spektral-zaman əlamətləri – müxtəlif şəxslərdə səs impulslarının spektrinin və zaman sırasının formasının özünə məxsusluğuna və onların nitq traktlarının filtr funksiyalarının xüsusiyyətlərini əks etdirməyə imkan verir. Nitq axının danışan şəxsin artikulyasiya orqanlarının hərəkət dinamikası ilə əlaqəli xüsusiyyətlərini xarakterizə edir və nitq axının danışanın artikulyasiya orqanlarının hərəkətinin qarşılıqlı əlaqəsinin və ya sinxronluğunun özünə məxsusluğunu əks etdirən integral xarakteristikasıdır.

High-level features

Признаки высокого уровня

Yüksək səviyyə əlamətləri – hər bir diktorun cepstral səviyyədə əks olunmamış dil və davranış aspektlərini xarakterizə edən əlamətlərdir.

Levinson-Durbin algorithm Алгоритм Левинсона-Дарбина

Levinson-Darbin alqoritmi – xətti proqnoz əmsallarını tapmaq üçün iterativ metoddur:

1. $E_0 = r_{ss}[0]$

for $i=1$ **to** p

2. $k_i = \frac{1}{E_{i-1}} \left(r_{ss}[i] - \sum_{j=1}^{i-1} a_{j,i-1} r_{ss}[i-j] \right)$

3. $a_{j,i} = a_{j,i-1} - k_i a_{i-j,i-1}, j = \{1, \dots, i-1\}$

$a_{ii} = k_i$

$$4. E_i = (1 - k_i^2) E_{i-1}.$$

Linear Prediction (LP) Линейное предсказание

Xətti proqnoz – nitq signalının cari bölgüsü ondan əvvəlki müəyyən sayda bölgülərin xətti kombinasiyası ilə approksimasiya edilir:

$$x_n \approx \sum_{k=1}^p a_k x_{n-k},$$

burada $a_1, \dots, a_p$ *xətti proqnoz əmsalları* adlanır.

Xətti proqnoz insanın səs traktını nitq signalını yaradan *sonsuz impuls reaksiyası* (IIR) sistemi kimi modelləşdirir. LP nitq signallarının sıxılması və sintezi üçün populyar üsuldur.

Linear Prediction Cepstral Coefficients (LPCC)

Кепстральные коэффициенты линейного предсказания

Xətti proqnoz keprstral əmsalları

Keprstral əmsallar xətti proqnoz əmsalları əsasında hesablanır:

$$c_n = \begin{cases} a_n + \sum_{k=0}^{n-1} \frac{k}{n} c_k a_{n-k}, & 1 \leq n \leq p; \\ \sum_{k=n-p}^{n-1} \frac{k}{n} c_k a_{n-k}, & n > p, \end{cases}$$

burada p – xətti proqnoz əmsallarının sayıdır.

Eksperimentlər göstərir ki, əlamətlər vektorunun formalaşdırılması üçün 12-20 LPCC əmsalı kifayətdir.

Linear Prediction Coefficients (LPC)

Коэффициенты линейного предсказания

Xətti proqnoz əmsalları – xətti kombinasiyada istifadə edilən çəki əmsallarıdır. Onlar nitq signalının bölgüləri ilə onların proqnoz edilmiş qiymətləri arasındakı fərqi kvadratinin orta qiymətinin minimallaşdırması ilə birqiymətli təyin edilir (sonlu intervalda).

Mel Mel

Mel (qədim-yunanca μέλος – səs) – səsin yüksəkliyinin psixofiziki vahididir. Mel şkalası insan qulağının həssaslığını aproksimasiya edir. İnsan qulağına "yaxınlaşan" bir çox tezlik şkalası vardır (məsələn, Bark şkalası).

Mel Frequency Cepstrum Coefficients (MFCC)

Мел частотные кепстральные коэффициенты

Mel tezlik keptral əmsallar – keptrin filtrlər bankından istifadə edilməklə mel-şkala üzrə paylanmış qiymətləridir. MFCC əlamətlərin çıxarılması alqoritminə aşağıdakı mərhələlər daxildir: seqmentlərə bölmə, sürətli Furiye çevirməsi, dəyişilmiş spektrin loqarifmi, diskret kosinus çevirməsi.

Adətən, əlamətlər vektorunu formalaşdırmaq üçün MFCC əmsalların sayı 12-yə bərabər seçilir. Enerjinin loqarifmi də 13-cü əmsal kimi əlavə olunur. Ən relevant məlumat ilk 6 əmsalda olur. Əlavə əmsalların götürülməsi konkret hal və diktorla müəyyən edilir.

Delta- və delta-delta əmsalları da əlavə etməklə MFCC əlamətlər vektorunun ölçüsü 39-a bərabər ola bilər:

- enerji, Δ -enerji, $\Delta\Delta$ -enerji;
- 12 keptral əmsal;
- 12 Δ -əmsal;
- 12 $\Delta\Delta$ -əmsal.

Perceptual Linear Prediction (PLP) PLP анализ

Perseptual xətti proqnoz – audio verilənlərin ilkin emalı alqoritmidir, siqnalın amplitudlar ardıcılığını əlamət vektorları ardıcılığına çevirir. O, insanın müxtəlif tezlikləri qavraması xüsusiyyətlərini nəzərə alır. Perseptual xətti proqnoz aşağıdakı qaydada həyata keçirilir: gücün perseptiv spektral sıxlığı hesablanır (Bark şkalası); yüksəkliyin ilkin emalı yerinə yetirilir və intensivliyin kub kökü hesablanır; avtokorrelyasiya funksiyasının ekvivalentini almaq üçün IDFF (tərs diskret Furiye çevirməsi) tətbiq

edilir; xətti proqnoz modeli (LP) qurulur və kepsral əmsallar hesablanır.

Prosodic features Просодические признаки

Prosodik əlamətlər – ton, ləhcə və ritmik xarakteristikalardır. Onların fiziki analoqları nitq signalının əsas tonunun tezliyi, enerjisi və müddətidir.

Short-term spectral features

Кратковременные спектральные признаки

Qısa müddətli spektral əlamətlər – nitq signalının bir-birini örtən qısa zaman intervallarına bölünməsi ilə alınan əlamətlərdir. Onlar danışanın səsinin çalarını xarakterizə edirlər.

Short-time magnitude Кратковременная магнитуда

Qısamüddətli maqnitud – düzbucaqlı pəncərəylə aşağı tezlik filtrinin tətbiqi zamanı nitq signalının qurşayanını xarakterizə edir.

Vocal source features Признаки голосового источника

Səs mənbəyi əlamətləri – səs yarığının cingiltili səsləri həyəcanlandıran signalı xarakterizə edən əlamətləri, məsələn, səs yarığının rəqslərinin forması və əsas tezlik.

Voice Activity Detector (VAD) Детектор речевой активности

Səs detektoru – nitq aktivliyi detektorudur, bu blok giriş nitq signalında nitqi pauzalardan ayırır və pauzalar növbəti analiz blokuna verilmir.

Zero-Crossing Rate Частота переходов через нуль

Sıfırdan keçidlərin sayı – signalın amplitudunun işarəsinin dəyişmələrinin sayıdır.

$$ZCR = \frac{1}{2} \sum_{n=1}^{N-1} |\text{sgn}(x[n]) - \text{sgn}(x[n-1])|.$$

Səsin kodlaşdırılması/sıxılması

Analog audio signals Аналоговые аудио сигналы

Analoq audio siqnalları – səs verilənlərinin telefon xətləri ilə ötürülməsi üçün istifadə edilir. Bu səsən yüksəkliyini dəqiq əks etdirmək üçün səs dalğalarının tezliyinin dəyişməsi və ya modulyasiyası yolu ilə həyata keçirilir. Bu texnologiya radiodalğaların ötürülməsi üçün də istifadə edilir.

Audio codec (enCOder/DECOder) Аудио кодек

Audio kodek (audio Koder/DEKoder) – rəqəmsal audio siqnalları sıxan/açan proqram təminatı. MP3, Windows Media Audio (WMA) rəqəmsal audio verilənlərinin sıxılması/açılması üçün populyar kodeklərdir.

A-law А-закон

A-qanun – rəqəmsallaşdırma öncəsi analoq nitq siqnalının dinamik diapazonunu optimal modifikasiya etmək üçün sıxma alqoritmidir, Avropa 8-bit rəqəmsal rabitə sistemlərində istifadə edilir. U-qanuna oxşardır. x siqnalı üçün A-qanun üzrə kodlaşdırma tənliyi belədir:

$$F(x) = \operatorname{sgn}(x) \begin{cases} \frac{A|x|}{1 + \ln(A)}, & |x| < \frac{1}{A} \\ \frac{1 + \ln(A|x|)}{1 + \ln(A)}, & \frac{1}{A} \leq |x| \leq 1 \end{cases},$$

burada A – sıxma parametridir ($A=87.6$).

A-qanun siqnalın dinamik diapazonunu effektiv azaldır və bu zaman kodlaşdırmanın effektivliyini artırır. A-qanun ilə müqayisədə U-qanun daha geniş dinamik diapazon təmin edir, lakin qısa siqnallarda mütənasib təhriflərə səbəb olur.

CELP (Code Excited Linear Prediction)

Линейное предсказание с мультикодовым управлением

Multi-kod idarəli xətti prediktor – nitq siqnallarının kodlaşdırılması alqoritmidir, 1985-ci ildə Manfred Şröder və Bişnu Atal tərəfindən təklif edilmişdir. Həmin dövrdə bu alqoritm RELP

və LPC audiokodekləri (məsələn, FS-1015) kimi aşağı bit sürətli alqoritmlərlə müqayisədə daha yaxşı keyfiyyət təmin edirdi. ACELP, RCELP, LD-CELP və VSELP kimi variantları ilə birlikdə hazırda ən geniş istifadə edilən nitq kodlaşdırma alqoritmidir. Hazırda CELP konkret kodek üçün deyil, alqoritmlər sinfi üçün ümumi termin kimi istifadə edilir.

Codec Кодек

Kodek – Compressor-DECompressor və ya enCOder/DECoder prosesini əks etdirən termdir. Verilənlər axını kodlaşdıran/dekodlaşdıran proqram və ya aparat təminatını bildirir. Məsələn, ötürücüdə ötürməni, saxlamayı və şifrələri asanlaşdırmaq üçün kodeklər verilənlər axını kodlaşdırır. Qəbul edərkən isə signal müvafiq istifadə üçün dekodlaşdırılır. Kodeklər video-konfrans və axın media həlləri üçün daha çox istifadə edilir.

Companding Компандирование

Kompanding (ing. companding – compression + expanding – sıxma + genişləndirmə) – məhdud dinamik diapazonlu kanalların təsirlərinin azaldılması metodudur. Kvantlama intervallarının sayını giriş signalının kiçik qiymətləri oblastında artırmağa və maksimal qiymətlər oblastında azaltmağa əsaslanır.

$A=87.6$ qiymətində A-qanun üzrə kompanding *signalların çoxkanallı ötürülməsi sistemlərində* standart kimi qəbul edilmişdir.

Compression Компрессия (сжатие)

Kompressiya (sıxma) – gurluğu düzləndirməklə səs dinamikasının dəyişməsi prosesidir, uca səsi yavaş edir. Faktiki olaraq, kompressor – gurluğun avtomatik tənzimləyicisidir.

DAM (Diagnostic Acceptability Measure)

Диагностическая мера приемлемости

Diaqnostik məqbulluq ölçüsü – səslənmənin keyfiyyətini qiymətləndirmək üçün seçilmiş frazalar kişilər və qadınlar olmaqla

bir neçə diktör tərəfindən oxunur, səslənmə ekspertlər tərəfindən dinlənilir və 5-ballıq sistemlə qiymətləndirilir. Qiymətləndirmənin nəticəsi olaraq orta subyektiv qiymət – orta rəy qiyməti (Mean Opinion Score, MOS) hesablanır.

Decoder Декодер

Dekoder – nitq tanıma sisteminin əsas blokudur. Bu komponent girişdəki nitq axınını akustik və dil modellərində saxlanan informasiya ilə müqayisə edir və sözlərin ən ehtimallı ardıcılığını müəyyən edir və bu tanımanın son nəticəsi olur.

DRT (Diagnostic Rhyme Test)

Диагностический рифмованный тест

Diagnostic qafiyə testi – nitqin aydınlığını qiymətləndirmək üçün geniş istifadə edilən metoddur. Samitlərdə fərqlənən və səslənməsi yaxın olan iki söz seçilir (məsələn, “dot-tot”, “kol-qol”), onlar bir neçə diktör (adətən, 10-20) tərəfindən bir neçə dəfə tələffüz edilir və testin nəticələrinə görə təhriflərin nisbi sayı müəyyən edilir.

H.261 – geniş zolaqlar (≥ 64 Kbps) üçün istifadə edilən video kodek. H.263 ensiz zolaq (< 64 -kbps) üçün istifadə edilir. Hər ikisi geniş dəstəklənir.

H.264 – ensiz zolaqlar üçün kodekdir, H.263 ilə müqayisədə daha yüksək keyfiyyətli nəticə verir və onun əvəzinə tövsiyə olunub. H.264 kodeki ISO 14496-10, MPEG-4 part 10 və MPEG-4 AVC (Advanced Video Coding) kimi də tanınır.

H.323 – İnternetdə real rejimdə səs və video-konfrans utilitləri üçün qaydaları müəyyən edən ITU standartıdır. H.323 standartı səs, video, verilənlər və proqramların ortaq istifadəsini dəstəkləyir və onları paketlərə çevirmək üçün media girişləri (portalları – ing. gateway) müəyyən edir.

MOS (Mean Opinion Score)

Средняя субъективная оценка

Orta subyektiv qiymət – ekspertlərin müxtəlif nitq mənbələri tərəfindən səsləndirilən nitq signallarına verdiyi qiymətlərin ortalama qiyməti. Hər ekspert signalı 5-ballıq sistemdə qiymətləndirir: 1 – pis , 2 – kafi, 3 – aydın, 4 – yaxşı, 5 – əla.

MRT (Modified Rhyme Test)

Модифицированный рифмованный тест

Təkmilləşdirilmiş qafiyə testi – sözün əvvəlində və sonunda olan samitlərin aydınlığını yoxlayır, samitlərin mümkün variasiyalarını nəzərə almağa imkan verir.

Perceptual audio coding Перцептивное аудио кодирование

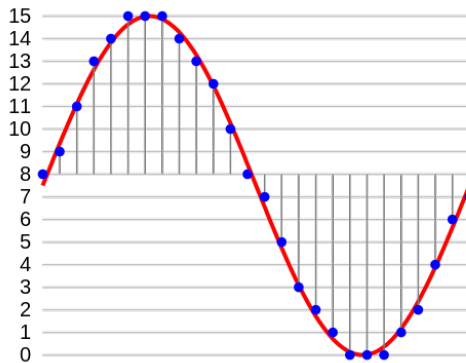
Persepsion kodlaşdırma – rəqəmsal audio signalların səsini insan tərəfindən qavrayışına (psixoaakustika) əsaslanan sıxma üsulları. Çox alçaq və çox yüksək olan tezliklər uca səslərlə boğulurlar. Məsələn, bir neçə musiqi aləti birlikdə çalındıqda baxılan anda tezlik və gurluqdan asılı olaraq bir alətin səsi digər alətin səsini yox edə bilər.

Pulse Code Modulation (PCM)

Импульсно-кодовая модуляция

İmpuls kod modulyasiyası – analoq audio signalların rəqəmsal signallara çevrilməsi üsuludur, diskretləşdirmə tezliyi 8-192 kHs götürülür və hər bir bölgü 8-24 bit uzunluqda ədədlə kodlaşdırılır. PCM verilənləri təmiz rəqəmsal audio nümunələrdir və bir sıra musiqi fayl formatlarında saxlanan verilənləri təşkil edirlər (WAV, FLAC, AIFF).

PCM 1938-ci ildə ixtira olunub, həmin vaxta kimi nitq signalları analoq signallar kimi emal olunurdu. PCM-in ixtirası, rəqəmsal sxemlərin və kompüterlərin inkişafı nitq signallarının rəqəmsal emalına imkan verdi və xüsusilə 1960-cı illərdən sonra nitq verilənlərinin emalında nəzərə çarpan inkişafa səbəb oldu.



4-bit PCM üçün siqnalın (qırmızı) diskretləşdirilməsi və kvantlanması

Speech codec/Voice codec or vocoder

Речевой кодек или вокодер

Səs kodeki və ya vokoder – şifahi nitq rəqəsal koda və tərsinə çevirən aparat təminatı. Ona A/D və D/A çevirmə və sıxma üsulu daxildir. Əgər musiqini nitq kodeki ilə kodlaşdırsaq dekodlaşdırmadan sonra əvvəlki kimi yaxşı səslənməyəcək.

Speech intelligibility Разборчивость речи

Nitqin aydınlığı – qəbul edilmiş nitq elementlərinin (səslərin, hecaların, sözlərin, frazaların) nisbi sayı, ötürülən elementlərin ümumi sayına nisbətə hesablanır.

Speech naturalness Естественность речи

Nitqin təbiiyyəti – sistemin ötürülən nitqin yalnız mənasını deyil, onun tembrini, danışanın fərdi nitq xüsusiyyətlərini də vermək imkanındır.

Subband Coding Поддиапазонное кодирование

Altdiapazon kodlaşdırılması – rəqəmsal siqnalda tezlik diapazonlarının seçilib ayrılması. Altdiapazon kodlaşdırılması xətti PCM musiqi nümunələrinin MP3 və digər formatlara sıxılması üçün

geniş istifadə edilir. Bu zaman müxtəlif tezliklər ayrılır və yaxınlığındakı daha yüksək tezliklərlə “söndürülən” aşağı tezlik diapazonları atılır (nəzərə alınmır).

U-law (μ -law) U-закон (μ -закон)

U-qanun (μ -qanun) – rəqəmsallaşdırmadan öncə analoq nitq signalının dinamik diapazonunu modifikasiya etmək üçün analoq sıxma alqoritmi, Şimali Amerika və Yaponiya rəqəmsal rabitə sistemlərində istifadə olunur. O, Avropada istifadə edilən A-qanuna oxşardır. x signalı üçün μ -qanun belə yazılır:

$$F(x) = \text{sgn}(x) \frac{\ln(1 + \mu |x|)}{\ln(1 + \mu)}, -1 \leq x \leq 1,$$

burada $\mu = 255$ (8 bit).

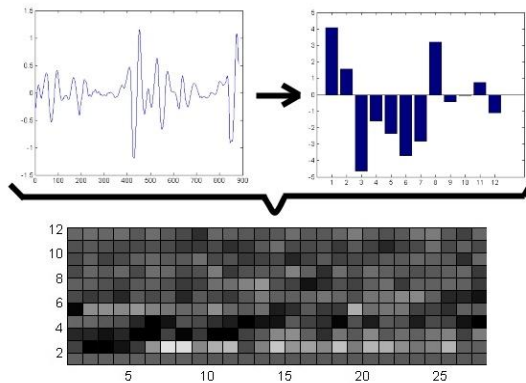
VoIP (Voice over IP) IP-телефония

IP üzərindən səs – istənilən IP-şəbəkə (Internet, LAN) üzərindən səsin ötürülməsi və qəbulu texnologiyası.

Nitqin avtomatik tanınması

Acoustic model Акустическая модель

Akustik model – nitq signalında linqvistik informasiyanın kodlaşdırılmasının ehtimal davranışını təsvir edən modeldir. Kontekstdən asılı olmayan (hər bir fonem üçün ən böyük ehtimallı əlamət vektorları) və kontekstdən asılı (senondan kontekstə görə qurulmuş) modellər mövcuddur. LVCSR sistemləri fonlara uyğun gələn akustik bloklardan istifadə edirlər.



Daha geniş yayılmış yanaşmada kontekstdən asılı fonların təsviri üçün kəsilməz vəziyyətli gizli Markov modellərindən (CDHMM) istifadə edirlər. Aşağıdakı akustik modellər daha geniş yayılıb:

- *Şablon modeli (Template model, Шаблонная модель).* Akustik model olaraq tanınanacaq struktur elementinin hər hansı şəkildə saxlanmış nümunəsi çıxış edir (sözlər, komandalar). Belə modeldə tanınmanın variativliyi eyni elementin müxtəlif tələffüz variantlarının yadda saxlanması yolu ilə təmin edilir (diktörlər çoxluğu eyni bir komandanı dəfələrlə təkrar edirlər). Əsasən, sözlərin vahid tam kimi tanınması üçün istifadə edilir (komanda sistemləri).
- *Vəziyyətlər modeli (State models, Модель состояний).* Hər bir söz ehtimal qaydaları əsasında sözün verilmiş hissəsində

eşidilməsi mümkün olan səslər toplusunu göstərən vəziyyətlər ardıcılığı kimi modelləşdirilir. Bu yanaşma daha böyük sistemlərdə istifadə edilir.

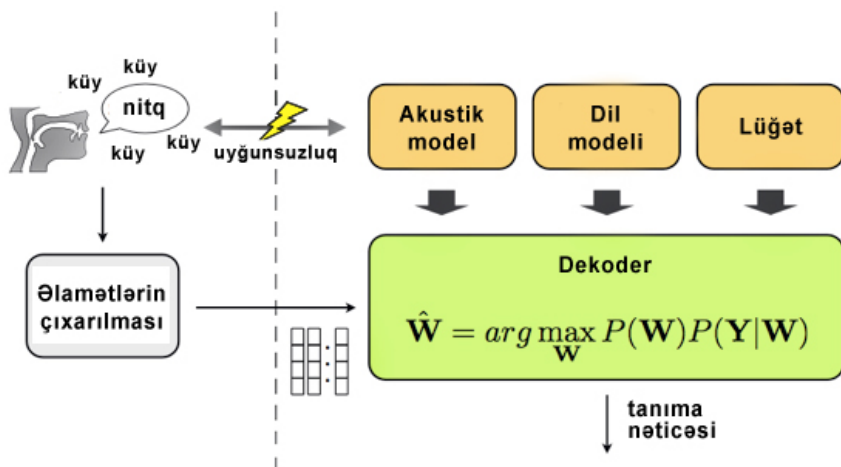
ASR Accuracy Точность APP

Nitqin tanınması sisteminin dəqiqliyi – 1-WER kimi müəyyən olunur. *Bax. Word Accuracy.*

Automatic Speech Recognition (ASR)

Автоматическое распознавание речи (APP)

Nitqin avtomatik tanınması – kompüterin nitq signalını sözlər ardıcılığına çevirməsi prosesidir. *Nitqin mətnə çevrilməsi* də adlanır.



Nitqin avtomatik tanınması sistemlərini aşağıdakı əlamətlərə görə təsnif etmək olar:

- lüğətin həcmi (sözlərin məhdud çoxluğu; böyük lüğətlər);
- diktordan asılılıq (diktordan asılı və diktordan asılı olmayan sistemlər);
- nitqin tipi (fəsiləsiz (təbii) və ya bölgülü (aramla));
- təyinatı (diktə sistemləri; əmr sistemləri);

- struktur vahidi (frazalar, sözlər; fonemlər, difonlar, allofonlar).

Backing-off method Метод возврата

Qayıtma metodu – müəyyən n -qram təlim korpusunda olmadıqda və ya onun rastgəlmə tezliyi çox kiçik olduqda onun yerinə $(n-1)$ -qramın ehtimalı istifadə edilir.

Bayes classifier Байесовский классификатор

Bayes klassifikatoru – ehtimal paylanmalarının məlum olduğunu fərz etsək, Bayes klassifikatoru

$$\bar{U} = \arg \max_{U \in U^*} P(U|X)$$

verilmiş X əlamətlər vektoru üçün (müşahidələr ardıcılığı) ən ehtimallı U sözlər ardıcılığını müəyyən edir. Bayes klassifikatoru səhv ehtimalının minimal olması mənasında optimaldır. HMM-ə Bayes klassifikatorunun xüsusi halı kimi baxmaq olar.

Confidence score Оценка доверия

İnam səviyyəsi – hipotezlə (məsələn, tanınan söz, identifikasiya edilən şəxs və s.) əlaqələndirilən aposterior ehtimal. Nitqin tanınması sistemlərində sözlərin inam səviyyələrinin cəmi düzgün sözlərin qiymətləndirilmiş sayına bərabərdir. Adətən, inam qiymətləri NCE metrikasının hesablanması yolu ilə müəyyən edilir.

Connected speech Слитная речь

Bitişik nitq – təbii tələffüz edilən cümlələrdir. Bitişik nitqin tanınması daha mürəkkəbdir, çünki tələffüz edilən sözlərdə dəqiq sərhəd yoxdur.

Context dependent phoneme model Контекстно-зависимая модель фонем

Kontekstdən asılı fonem modeli – koartikulyasiya hadisəsinə görə fonemə qonşu fonemlər güclü təsir edir və fonetik kontekstdən asılı olaraq müxtəlif səslər yaranır. Koartikulyasiya ona görə olur ki, artikulyasiya orqanları bir vəziyyətdən digərinə avtomatik keçir.

bilmir. Koartikulyasiya hadisəsini daha yaxşı nəzərə almaq üçün kontekstdən-asılı vahidlər istifadə edilir, məsələn, bifonlar və trifonlar. Bifonem HMM-də fonem özünün sol və ya sağ konteksti ilə modelləşdirilir. Trifonem HMM-də fonem özünün sol və sağ konteksti ilə modelləşdirilir.

Daha detallı kontekstdən asılı vahidlər istifadə edildikcə vahidlərin sayı artır. Trifonların sayının lüğətdəki sözlərin sayından çox olması mümkündür. Bu yenidən öyrənmə problemi yaradır. Parametrləri etibarlı qiymətləndirmək üçün öyrənmə üçün böyük miqdarda verilənlər tələb edilir. Kontekst modellərinin fonem və ya vəziyyət səviyyələrində birləşdirilməsi istifadə edilə bilər. Yeni və ya naməlum mühitlər üçün yenidən öyrənmə verilənlərin toplanması baxımından çətin başa gəlir. Bu halda parametrlərin adaptasiyası kimi yanaşmalardan istifadə edilir.

Context independent phoneme model

Контекстно-независимая модель фонем

Kontekstdən asılı olmayan fonem modeli – fonem qonşu fonemlərdən fonetik asılı olmaması fərz olunaraq modelləşdirilir. Məsələn, monofonem HMM bir fonemi modelləşdirir. Bu kontekstdən (fonetik) o mənada asılı olmayan vahiddir ki, onu qonşu fonetik kontekstdən fərqləndirmir.

Continuous Density HMM (CDHMM)

Скрытые марковские модели с непрерывной плотностью

Kəsilməz sıxlıqlı gizli Markov modeli – daha çox Qarışıq Gauss modeli (ing. Gaussian mixtures model) kimi tanınır.

Deep Neural Network (DNN) Глубокая нейронная сеть

Dərin neyron şəbəkələri – akustik modellərin öyrədilməsi üçün istifadə edilir. Qarışıq Gauss modelindən daha yaxşı nəticə verir (tanıma keyfiyyətini 10-20 % artırmağa imkan verir). Qarışıq Gauss modeli üçün xarakterik olan bir çox nöqsandan azaddır və daha yaxşı ümumiləşdirmə qabiliyyətinə malikdir. Bundan başqa, neyron

şəbəkələrdə akustik modellər küyə daha dayanıqlıdır və yüksək sürətə malikdir.

Dərin neyron şəbəkələri insanların anlayışları bir neçə səviyyədə təsvir etmələri ideyasından çıxış edir – insanlar aşağı səviyyələrdə sadə əlamətləri və onların üzərində yüksək səviyyə abstraksiyalarını qururlar. İnformasiyanı belə daha abstrakt yolla təsvir etməklə daha asan ümumiləşdirmə aparmaq olar.

Bizim sifəti və ya səsi tanımağımız bizə ani görünə bilər, həqiqətdə isə burada bir neçə əməl mərhələsi vardır. Əməl iyerarxiyasında beynin müxtəlif hissələri iştirak edir, sensor verilənləri bir hissədən digərinə ötürülür, bu zaman çevirmələr aparılır və digər informasiya ilə birləşdirilir (əlamətlərin birləşdirilməsi daha böyük obrazı tanımağa imkan verir).

Mütəxəssislər, praktiki məsələlər üçün neyronlarda 8-10 layın kifayət edəcəyini düşünürlər. Bəziləri bunun neyrologiyaya əsaslandığını güman edir. 0.1 san ərzində neyron (insan beyində) 10 dəfə “yanır-sönür” – bu təxminən, ümumi qavrayış məsələlərinin (nitq və vizual tanıma) çoxu üçün tələb olunan zamana bərabərdir.

Keyword spotting Поиск ключевых слов

Açar sözlərin axtarışı – nitqdə açar sözlərin identifikasiyasıdır; nitqin avtomatik tanınmasının bir sahəsidir. Səs yazıları bazasında və səs axınlarında fraqmentin avtomatik axtarışının effektiv üsuludur. Sorğu, axtarış və idarəetmə sistemlərində istifadəçinin danışığında açar sözləri tanımaq zəruridir.

Language model Лингвистическая (языковая) модель

Dil modeli – danışiq dilindəki qanunauyğunluqları əhatə edir və nitq tanınması sistemində sözlər ardıcılığının ehtimallarını qiymətləndirirməklə sözlərin axtarışını məhdudlaşdırmaq üçün istifadə edilir. Bu model sonra hansı sözün gələ biləcəyini müəyyən edir və ehtimal edilməyən sözləri atmaqla axtarış prosesini əhəmiyyətli dərəcədə məhdudlaşdırmağa kömək edir. N-qram modeli ən populyar modellərdən biridir, n-qram ardıcılıqlarının

tezliklərini qiymətləndirmək yolu ilə dilin sintaktik və və semantik məhdudiyyətlərini əhatə etməyə çalışır.

Dil modelinin yaradılması konkret dildən çox asılıdır. Məsələn, ingilis dili üçün statistik model (n-gram) istifadə edilməsi kifayətdir. Yüksək flektivliyə malik dillər (eyni sözün bir çox forması olan dillər) üçün təkcə statistikadan istifadə edilməklə yaradılan dil modelləri lazımi nəticəni vermir – sözlər arasındakı statistik əlaqələri qiymətləndirmək üçün etibarlı, olduqca böyük həcmdə verilənlər lazımdır. Hibrid dil modellərində dilin qaydaları, nitq hissələri və söz formaları haqqında informasiya və statistik model istifadə edilir.

Language recognition

Автоматическое распознавание языка

Dilin tanınması – kompüterin nitq signalında hansı dildə danışılmasını müəyyən etməsi prosesidir.

Large vocabulary

Böyük lüğət – lüğətin həcmi müvafiq tətbiqi sistemin konkret tələbləri ilə müəyyən edilir. Adətən, lüğətlərin həcmi aşağıdakı siniflərə bölünür:

- Kiçik lüğət – onlarla söz;
- Orta lüğət – yüzlərlə söz;
- Böyük lüğət – minlərlə söz;
- Çox böyük lüğət – on minlərlə söz.

Lattice Решетка (сетка)

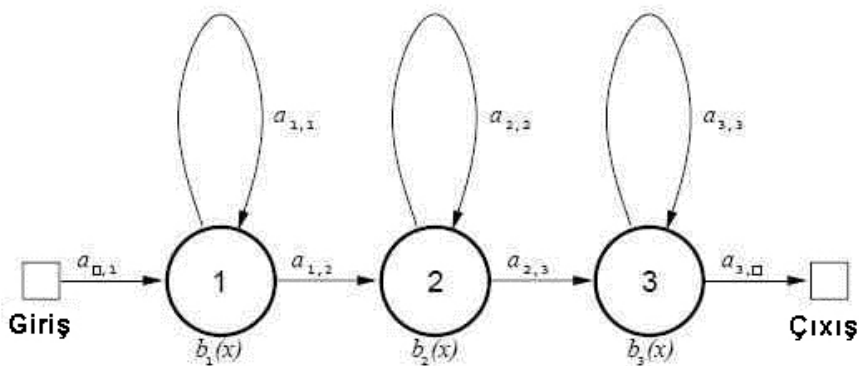
Söz qəfəsi – çəkili asiklik qrafıdır, burada söz nişanları qrafın tillərinə və ya təpələrinə mənimsədilir. Akustik və linqvistik modellərdə hər bir til ilə çəkilər, hər bir təpə ilə zaman vəziyyəti əlaqələndirilir.

Left-to-right HMM CMM слева-направо

Soldan-sağa yönəlmiş GMM – nitqin diqqətəlayiq (fərqləndirici) xüsusiyyətlərindən biri onun dinamikliyidir. Məsələn, hətta fonem

kimi kiçik seqmentdə də səslər tədricən dəyişir. Fonemin başlanğıcı əvvəlki fonemlərdən asılıdır, fonemin orta hissəsi ümumi halda stabildir və fonemin sonuna sonrakı fonemlər təsir edir. Əlamət vektorları haqqında zaman məlumatları tanıma prosesində vacib rol oynayır. Bayes klassifikatorunda nitqin dinamik xarakteristikalarını nəzərə almaq məqsədilə fonemin təsviri üçün, adətən, 3 vəziyyətdən ibarət, soldan-sağa yönəlmiş HMM istifadə edilir.

Şəkildə belə HMM-ə nümunə göstərilir, burada a_{ij} qiyməti i -ci vəziyyətdən j -cu vəziyyətə keçid ehtimalını $b_i(x)$ – i -ci vəziyyətdə x əlamət vektorunun müşahidə edilməsi ehtimalını göstərir. HMM-də hər bir vəziyyət fonemdə səsin paylanması modelləşdirir.



Şəkildə fonem 3 ardıcıl paylanmadan ibarətdir. Söz HMM-ni fonem HMM-lərini konkatensasiya etməklə qurmaq olar. HMM cümləsini söz HMM-lərini birləşdirməklə almaq olar. HMM-in törətdiyi əlamət vektorlarının ehtimalı vəziyyətlər arasında keçid ehtimalları və baxılan vəziyyətlərdə əlamət vektorlarının müşahidə edilməsi ehtimallarını istifadə etməklə hesablanır.

Lexicon or pronunciation dictionary (vocabulary)

Лексикон или словарь произношений

Leksikon və ya tələffüz lüğəti – tələffüzləri ilə birgə sözlərin siyahısı. Nitqin tanınması üçün sistemə məlum olan bütün sözlər leksikona daxildir, hər bir sözün uyğun ehtimallarla bir və ya bir neçə tələffüzü olur.

LVCSR (Large-Vocabulary Continuous Speech Recognition)

Распознавание слитной речи с большим словарем

Böyük lüğətlə bitişik nitqin tanınması (böyük lüğət 20 min və daha çox sözdən ibarətdir). Tanıma lüğətinin həcmi əməl tələblərinə təsir edir.

MAP decoding MAP декодирование

MAP dekodlaşdırma – dekodlaşdırma prosedurudur (nitqin tanınması), verilmiş X nitq signalı və M modeli üçün W söz transkripsiyasının $\Pr(W|X, M)$ aposterior ehtimalını maksimallaşdırmağa cəhd edir.

Maximum A Posteriori (MAP) Estimation

Максимальная апостериорная оценка

Maksimal aposterior qiymətləndirməsi – model parametrlərinin $\Pr(M|X, W)$ aposterior ehtimalını maksimallaşdırmağa cəhd edən öyrənmə prosedurudur (deməli, parametrlərə təsadüfi kəmiyyətlər kimi baxılır), burada X – nitq signalı, W – söz transkripsiyası və M modelin parametrləridir).

Maximum Likelihood Estimation (MLE)

Оценка максимального правдоподобия

Maksimal həqiqətəoxşarlıq qiymətləndirməsi – öyrənmə prosedurudur (modelin parametrlərinin qiymətləndirilməsi). Verilmiş model üçün təlim verilənlərinin $f(X|W, M)$ həqiqətəoxşarlıqını maksimallaşdırmağa cəhd edir (X – nitq signalı, W – söz transkripsiyası və M – modeldir).

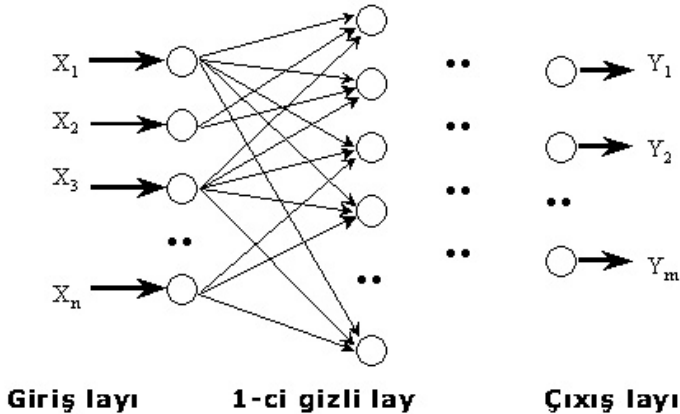
Maximum Mutual Information Estimation (MMIE)

Максимальная оценка взаимной информации

Qarşılıqlı informasiyanın maksimal qiyməti – söz transkripsiyasının $\Pr(W|X, M)$ aposterior ehtimalını maksimallaşdıran diskriminativ təlim prosedurudur (X – nitq signalı, W – söz transkripsiyası və M – modeldir). Bu təlim metodunu *maksimal şərti həqiqətəoxşarlıq qiymətləndirməsi* də adlandırırlar.

Multilayer perceptron Многослойный персептрон

Çoxlaylı perseptron – süni neyron şəbəkələrinin bir sinfidir. Bu bir istiqamətli şəbəkə giriş verilənlərini müəyyən çıxış verilənlərinə əks etdirir. O, qeyri-xətti aktivləşmə funksiyasına (adətən, siqmoid) malikdir və üç və daha artıq laydan ibarətdir.



N-gram model Модель N-грамм

N-gram modeli – tərtibi N-1 olan Markov zəncirinə əsaslanan ehtimallı linqvistik model

Normalized cross entropy (NCE)

Нормированная кросс-энтропия

Normallaşdırılmış kross-entropiya – normallaşdırılmış qarşılıqlı entropiya; bu metrika inam səviyyəsini (ing. confidence score) qiymətləndirmək üçün istifadə edilir.

OOV word (Out Of Vocabulary word)

Не входящее в словарь слово

Lüğətdə olmayan söz – lüğətdə olmayan hər bir söz birdən çox tanıma səhvinə səbəb olur (adətən, 1,5 və 2 arasında səhv). Lüğətdə olmayan sözlərdən gələn səhvlərin səviyyəsini azaltmağın aşkar yolu lüğətin həcmi artırmaqdır.

Percent of correct words Процент соответствия слов

Düzgün sözlərin faizi – düzgün tanınmış etalon sözlərin faizi. Daxiletmə səhvləri nəzərə alınmadıqda tanıma sistemlərinin qiymətləndirilməsi üçün istifadə edilə bilər. O, $\%WACC + \%Ins$ kimi müəyyən edilir, burada $\%Ins$ – daxil edilmiş sözlərin sayının etalon sözlərin sayına nisbətinin 100 mislidir, $WACC$ – sözün tanınması dəqiqliyidir.

Perplexity Показатель связности (Перплексия)

Dolaşıqlıq – dil modelinin relevantlığı çox zaman test toplusunun dolaşıqlığı ilə ölçülür. Entropiya anlayışına əsaslanır, onu $\text{pow}(\text{Prob}(\text{text}|\text{language-model}), -1/n)$ kimi müəyyən edirlər, burada n – test toplusundakı sözlərin sayıdır. Dolaşıqlıq həm modeləşdirilən dildən, həm də modelin özündən asılıdır. Modelin nə dərəcədə yaxşı olmasını və dilin nə dərəcədə mürəkkəb olmasını birlikdə qiymətləndirir.

Signal smoothing Сглаживание сигнала

Signalın hamarlanması – nöqtədəki qiymətin nöqtə ətrafındakı qiymətlərin ədədi ortası ilə əvəz edilməsidir.

Silent Speech Interfaces (SSI)

Интерфейс Безмолвного Доступа

Səssiz nitq interfeysləri – bu sistemlər artikulyasiyanın erkən mərhələsində alınmış nitq signallarının emalına əsaslanır. Bu texnologiya müasir nitq tanıma sistemlərinin iki əhəmiyyətli nöqsanı ilə əlaqəlidir: küyə həddindən artıq həssaslıq və nitq tanıma sistemlərinə müraciət zamanı dəqiq və aydın nitq tələbi. SSI yanaşması küyün təsirinə məruz qalmayan yeni sensorlardan emal edilən akustik signallara əlavə kimi istifadə edilməsindən ibarətdir.

Siri (Speech Interpretation and Recognition Interface)

Интерфейс для интерпретации и распознавания речи

Siri iOS üçün yaradılmış şəxsi köməkçi və sual-cavab sistemidir. Suallara cavab vermək və məsləhətlər vermək üçün Siri nitqin

tanınması texnologiyasından istifadə edir. Siri-də nitqin tanınması Nuance Communications şirkəti tərəfindən yaradılmış səs texnologiyalarına əsaslanır. Siri – SRI International-ın beynəlxalq süni intellekt mərkəzində yaradılmışdır (SRI (Stanford Research Institute) 1970-ci illərə kimi Stenford Universitetinin tərkibində idi, 1977-ci ildən SRI International adlanır və müstəqil elmi-tədqiqat qurumudur).

Spontaneous speech Спонтанная речь

Spontan nitq – hazırlıqsız şifahi nitq.

SUR (Speech Understanding Research)

Программа по разработке системы распознавания речи

1971-1976-cı illərdə DARPA (Defense Advanced Research Projects Agency) tərəfindən maliyyələşdirilən nitqin tanınması sistemlərinin yaradılması üzrə tədqiqat layihəsi. Layihə çərçivəsində Karnegi Mellon Universitetində yaradılmış “Harpy” sistemi 1011 sözü tanıyırdı.

Token Токен

Token – mətnin linqvistik əhəmiyyətli minimal elementi (sözlər, durğu işarələri, pul vahidlərinin işarələri, ədədlər, hərf-rəqəm birləşmələri, tarixlər, telefon nömrələri, IP-ünvan, fayl adı və s.)

Tokenization Токенизация

Tokenizasiya – giriş mətninin sözlərə, rəqəm birləşmələrinə, düsturlara və s. bölünməsi.

Topic Detection Обнаружение темы

Mövzunun aşkarlanması – böyük lüğətdə nitqin tanınması zamanı danışq mövzusunun müəyyən edilməsi modulu da istifadə edilir. Bu danışq mövzusunun asılı olaraq lüğət və dil modellərini avtomatik dəyişməyə (seçməyə) imkan verir. Mövzunun müəyyən edilməsi modulu Data Mining texnologiyaları istifadə edilməklə işlənir.

Voicing Озвучивание

Səslənmə – səslənmə səviyyəsi signalın nə dərəcədə periodik olmasını ölçür (periodiklik, harmoniklik və ya *HNR (Harmonics-to-Noise Ratio)* kimi də məlumdur). Praktikada signalın periodiklik dərəcəsini avtokorrelyasiya funksiyasının maksimumunun nisbi hündürlüyündən almaq olar.

Word Accuracy (WACC) Точность распознавания слов

Sözün tanınması dəqiqliyi – nitqin tanınmasını qiymətləndirmək üçün istifadə edilən göstəricidir. Dəqiqlik faizi $WACC = 100 - WER$ kimi müəyyən edilir, burada *WER* – sözlərin tanınma səhvlərinin səviyyəsidir. Qeyd etmək lazımdır ki, *WACC* qiyməti mənfi ola bilər.

Word Error Rate (WER)

Уровень ошибок распознавания слов

Sözlərin tanınma səhvlərinin səviyyəsi – nitqin tanınmasını qiymətləndirmək üçün geniş istifadə edilən metrikadır. Bu göstərici üç növ səhv nəzərə alınmaqla sözlərin tanınma səhvlərinin orta sayını müəyyən edir: əvəzlənmə (etalon söz başqa sözlə əvəzlənib), daxiletmə (etalon sözün olmaması fərziyyəsi) və silmə (transkriptor etalon sözü buraxıb). Sözlərin tanınma səhvlərinin səviyyəsi bu səhvlərin cəminin etalon sözlərin sayına nisbəti kimi müəyyən olunur. Bu tərifə görə *WER* qiyməti 100 %-dən çox ola bilər. *WER* əsasında müxtəlif ASR platformaları müqayisə etmək çətindir. Linqvistik modelin verdiyi töhfəni aydınlaşdırmaq da çətindir.

WordNet – ingilis dili üçün leksik verilənlər bazasıdır. Tezaurus/semantik şəbəkədir, 1980-ci illərin əvvəlində Princeton Universitetində yaradılmışdır.

Lüğət əsas nitq hissələri: isim, fel, sifət və zərf üçün dörd şəbəkədən ibarətdir. WordNet-də əsas lüğət vahidi oxşar mənalı sözləri birləşdirən sinonimlər çoxluğu (‘‘synset’’ – sinset). Hər bir sinset tərif və sözlərin kontekstdə işlənməsi misalları ilə verilir. WordNet-in digər dillər üçün də variantları var, məsələn, EuroWordNet, BalkanNet.

Nitqin avtomatik sintezi

Concatenative synthesis

Конкатенативный синтез/Компилятивный синтез

Konkatenativ sintez – yazılmış səs seqmentlərinin birləşdirilməsinə (“yapışdırılmasına”) əsaslanır. Seqmentlərin uzunluğu 10 ms-ədən 1 sən qədər olur. Ümumiyyətlə, konkatenativ sintez səslənməsinə görə daha təbii sintetik nitq verir. Buna baxmayaraq, nitqdə təbii dəyişmələr ilə nitq siqnallarının seqmentlərə avtomatik bölünməsi üsullarının təbiəti arasındakı fərqlər bəzən çıxışda eşidilən qüsurlara səbəb olur.

Konkatenativ sintezin üç növü var:

- Unit selection synthesis – Vahidlərin seçilməsi ilə sintez
- Diphone synthesis – Difon sintezi
- Domain-specific synthesis – Konkret sahələr üçün sintez

Diphone synthesis Дифонный синтез

Difon sintezi – dildə meydana çıxan bütün difonların daxil olduğu minimal nitq verilənləri bazasından istifadə edir. Difonların sayı dilin phonotactics-indən asılıdır: məsələn, ispan dilində təxminən 800, alman dilində isə təxminən 2500 difon var. Difon sintezində hər bir difonun yalnız bir nümunəsi nitq verilənləri bazasında olur. Sintez zamanı cümlə bu minimal vahidlərdən xətti proqnoz kodlaşdırması, PSOLA və ya MBROLA kimi üsulların köməyi ilə yığılır. Elmi tədqiqatlarda istifadə edilir, çünki bir sıra pulsuz əlyetər proqram məhsulları var.

Festival – çoxdilli nitq sintezi sistemidir, nitq sintezi metodlarının işlənməsi və tədqiqi üçün də mühit təqdim edir. Edinburq Universiteti Centre for Speech Technology Research (CSTR) əməkdaşı Alan Blek tərəfindən yaradılıb. Karnegi Mellon Universiteti və digər tərəfdaşlar da əhəmiyyətli töhfələr vermişdir. C++-da yazılıb, BSD lisenziyasına oxşar pulsuz proqram təminatı lisenziyası ilə yayılır.

Formant synthesis Формантный синтез

Formant sintezi – sintez zamanı insan nitqi nümunələrindən istifadə etmir. Süni nitq dalğalarını yaratmaq üçün əsas tonun tezliyi, səslənmə və küy səviyyəsi kimi parametrləri zamana görə dəyişirlər. Kaskadlı paralel formant sintezatoru Denniz Kalt tərəfindən yaradılmışdı.

Formant sintezi texnologiyası əsasında sistemlərin çoxu süni, robot səsinə bənzər səs yaradırlar. Lakin nitq sintezi sisteminin məqsədi həmişə maksimal təbiilik olmur və formant sintezi konkatenativ sintez qarşısında üstünlüklərə malikdir. Formant sintezi sistemləri emosiyanı və intonasiyanı daha yaxşı verə bilər.

HMM synthesis Синтез на основе СММ

HMM-sintez – bu sistemdə tezlik spektri (nitq traktı), əsas tezlik (səs mənbəyi) və nitqin müddəti (prosodiya) gizli Markov modeli ilə modelləşdirilir. Nitq siqnalları maksimal həqiqətəoxşarlıq kriteriyası əsasında generasiya edilir.

Hazırda Unit Selection (səs elementlərinin verilənlər bazalarından seçilməsi) və HMM-sintez nitqin sintezində iki əsas istiqamətdir. Unit Selection-da nitq tembrinin təbiiliyi keyfiyyətlidir, sintez edilmiş səs donor-diktorun səsinin tembr çalarını saxlayır. Lakin intonasiya tərtibatı bir qədər qeyri-təbii, yeknəsək ola bilər.

HMM-sintezdə intonasiya tərtibatı, yüksəklik və tembr diktorun təbii səsndən sanki “klonlaşdırılır”, lakin sintez edilən səs bir qədər mexaniki və “robotabənzər” səslənir.

Java Speech API Markup Language (JSML) – səsi dəstəkləyən veb-brauzerlər və e-poçt sistemlərində nitqin sintezi üçün nişanlama kodlarının və sözlərinin toplusudur. JSML sənədin strukturunu təsvir edən teqlərin və mətnin tələffüzünü, aksent, yüksəklik və sürəti idarə edən diktor teqlərini təsvir edir. SABLE funksiyaları üçün JSML və digər nişanlama dili Spoken Text Markup Language (STML) birləşdirilib.

JAWS (Job Access With Speech) – Microsoft Windows üçün kompüter ekran oxuyucusudur, görmə əngəlli insanlara mətnin nitqə çevrilməsi yolu ilə və ya yenilənə bilən Breyl displeyi ilə ekranı oxumağa imkan verir.

MBROLA – nitq sintezi üçün xüsusi alqoritmdir, onun əsasında bir çox proqram təminatı yaradılmışdır, onlar pulsuz yayılır, lakin açıq kodlu deyil, binar kodludur.

MBROLA difon sintezi texnologiyası əsasında qurulsa da, nitqin keyfiyyəti digər difon sintezatorlarından yüksək olur. Bu qismən onunla izah olunur ki, MBROLA-da istifadə edilən difonların ilkin emalı metodu onların çıxışda hamar yapışdırılmasını təmin edir. Layihənin rəsmi saytında xeyli sayda dil üçün difon bazaları əlçatandır.

MBROLA mətnin nitqə çevrilməsi üçün nəzərdə tutulmuş tam proqram təminatı deyil; əvvəlcə mətn fonemlərə çevrilməlidir, sonra sintez edilən nitqin ümumi intonasiyası haqqında prosodik informasiya xüsusi MBROLA formatında əlavə edilməlidir, daha sonra bu verilənlər ayrıca proqram təminatı tərəfindən emal edilir və xüsusi formata salınmış informasiya səs çıxışına verilir.

MIT Oxygen Project

MIT Oxygen layihəsi – Massaçusets Texnologiya İnstitutu, Kompüter Elmləri Laboratoriyasının 35 illiyi münasibətilə başlatılmış layihə (bu laboratoriya 1963-cü ildə yaradılıb, hazırda Computer Science and Artificial Intelligence Laboratory – CSAIL kimi tanınır). Nitq texnologiyaları gələcək hesablama texnologiyalarının “oksigeni” olaraq inkişaf etdirilirdi. Layihə çərçivəsində aşağıdakı sistemlər yaradılmışdır:

- Galaxy – nitq texnologiyalarının inteqrasiyası üçün arxitektura.
- SUMMIT – nitqin tanınması sistemi
- TINA – təbii dilin emalı sistemi
- Dialog manager – sorğunun semantik kontentini analiz edərək cavabın semantik kontentini formalaşdırır.

- GENESIS – semantik kontenti təbii dildə mətnə çevirən sistem
- ENVOICE – nitq sintezi sistemi.

SABLE – Sun Microsystems, AT&T, Bell Labs və Edinburq Universiteti arasında birgə layihə ("SABLE" sözünün ilk hərfləri). Nitqin sintezi üçün mətnlərin annotasiya edilməsində istifadə edilən XML-nişənləmə dilidir. Səsləndirmə proqramlarında səsləndirilən yazılmış sözləri, rəqəmləri və cümlələri idarə edən nişənləri müəyyən edir.

SAPI (Speech Application Programming Interface) – Windows tətbiqi proqramlarında nitqin tanınması və nitqin sintezi texnologiyalarının istifadəsi üçün Microsoft tərəfindən yaradılmış tətbiqi proqramlaşdırma interfeysidir. İki əsas SAPI versiyası vardır (**SAPI 4** və **SAPI 5**), onlar bir-birindən fərqlidir (uyuşmazdır).

Windows 2000 əməliyyat sistemində SAPI 4, Windows XP və bütün sonrakı versiyalarda SAPI 5 vardı. Bir kompüterə hər iki SAPI versiyasını yükləmək olar; bu bütün mövcud (köhnə və yeni) nitq mühərriklərindən istifadə etməyə imkan verir.

Speech synthesis Синтез речи

Nitqin sintezi – insan nitqinin süni səsləndirilməsi.

Speech Synthesis Markup Language (SSML) – nitqin sintezi proqramları üçün XML əsasında nişənləmə dilidir. W3C işçi qrupu tərəfindən tövsiyə olunub. SSML teqlərin köməyi ilə pauzanı vermək, əsas ton tezliyini azaltmaq və yüksəltmək, sözün tələffüzünü dəyişmək və s. olar. SSML çox zaman interaktiv telefon sistemləri üçün VoiceXML ssenarilərinə daxil edilir. Audio kitabların yaradılması üçün də istifadə edilə bilər. SSML Sun Microsystems tərəfindən yaradılmış Java Synthesis Markup Language (JSML) nişənləmə dilinə əsaslanır.

Speech synthesizer Синтезатор речи

Nitq sintezatoru – nitqin sintezi üçün istifadə edilən kompüter sistemi. Proqram və aparat təminatı ilə reallaşdırıla bilər. Sintez edilən nitq verilənlər bazasında saxlanan nitq nümunələrinin birləşdirilməsi yolu ilə yaradıla bilər. Sistemlər saxlanan nitq vahidlərinin həcminə görə fərqlənirlər; fonları və difonları saxlayan sistem çıxışda geniş seçim diapazonu dəstəkləyir, lakin aydın başa düşülməsi ləngiyə bilər. Konkret istifadə sahələri üçün bütöv sözlərin və ya frazaların saxlanması çıxışda daha keyfiyyətli nəticə almağa imkan verir. Bundan başqa, çıxışda tamamilə «sintetik» səs yaratmaq üçün sintezatora səs traktı modeli və insan səsinin digər xarakteristikaları da daxil edilə bilər. Nitq sintezatorunun keyfiyyətini insan səsinə oxşarlığı və aydın başa düşülməsi qabiliyyətinə görə qiymətləndirirlər. Bir çox kompüter əməliyyat sistemlərinə 1990-cı illərdən başlayaraq nitq sintezatorları daxil edilib.

TD-PSOLA (Time-Domain – Pitch Synchronous Overlap and Add) Технология перекрытия и синхронизации с частотой основного тона во временной области

Zaman oblastında əsas ton tezliyi ilə sinxron örtmə və toplama – parametrik modellər istifadə etmədən, bilavasitə signalın özünün prosodik xarakteristikalarının (əsas tonun tezliyi, enerji və nitq signalının müddətinin) modifikasiya edilməsi metodu. TD-PSOLA metodunun mahiyyəti aşağıdakından ibarətdir: ilkin allofon əsas tonunun periodları üzrə nişanlarla sinxronlaşdırılmış müvəqqəti pəncərələr ardıcılığına vurulur:

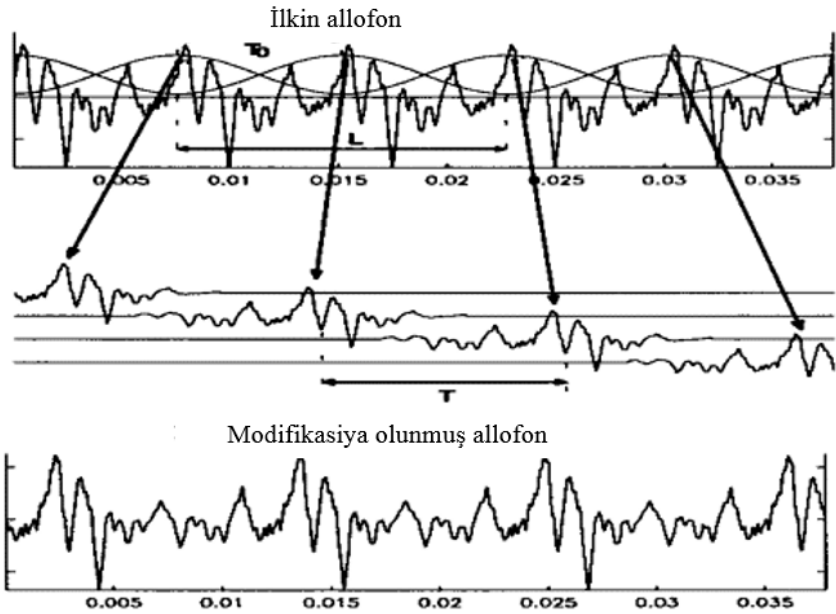
$$x_m(n) = v_m(n - t_m)x(n),$$

burada $x(n)$ – ilkin kəsilməz rəqəmsallaşdırılmış nitq signalı, $x_m(n)$ – alınmış akustik hissələr ardıcılığı, $v(n)$ – pəncərə funksiyası, t_m – ilkin allofonun əsas tonunun periodları üzrə nişanlama işarələridir. Alınmış akustik hissələr ardıcılığı bir-birinə nəzərən sürüşdürülür və sonra cəmlənir, nəticədə modifikasiya olunmuş allofon alınır:

$$\tilde{x}(n) = \sum_{m=-\infty}^{\infty} k_m(n)x_m(n - (t'_m - t_m)),$$

burada $k(n)$ – sintez edilmiş nitqdə ilkinlə müqayisədə enerji dəyişməsinə kompensasiya edən normallaşdırma funksiyasıdır,

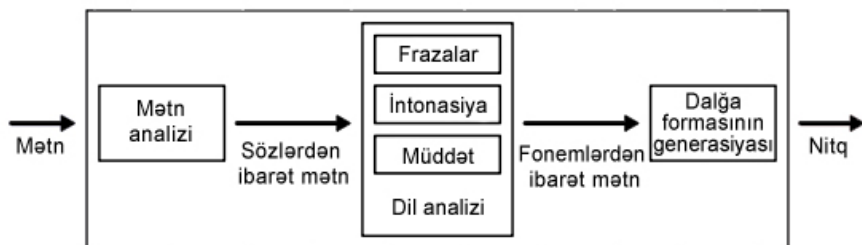
enerjinin dəyişməsi pəncərə funksiyasının və siqnalın zamana görə dəyişməsi alqoritmlərinin təsiri nəticəsində meydana çıxır. Belə pəncərə çevirməsi siqnalın spektral qurşayanını təxminən saxlayır və əsas tonun periodunu dəyişməyə imkan verir. Şəkində bu metodun sxemi göstərilib.



Pəncərə funksiyasının növü və onun ölçüləri sintez edilən siqnalın ilkin siqnala maksimal dəqiq uyğun olmasını təmin etmək üçün diqqətlə seçilir.

Text-To-Speech (TTS) Текст-В-речь

Mətdən-nitq sintezi sistemi adi mətni nitqə çevirir; digər sistemlər fonetik transkripsiya kimi simvolik linqvistik təsvirləri nitqə çevirir. Mətnin səsə çevrilməsi sistemi (və ya "mühərrik") iki: düz və tərs hissədən ibarətdir. Düz hissənin iki əsas vəzifəsi var. Birincisi, rəqəmlər və ixtisarlər kimi simvollar olan giriş mətni mətn ekvivalentinə çevrilir.



Bu prosesi çox zaman *mətnin normallaşdırılması*, ilkin emal və ya *tokenizasiya* adlanır. Sonra hər bir sözə fonetik transkripsiya təyin edilir və mətn fraza, cümlə kimi prosodik vahidlərə bölünərək nişanlanır.

Sözlərə fonetik transkripsiyanın təyin edilməsi prosesi mətnin fonemə (*text-to-phoneme*) və ya qrafemin fonemə (*grapheme-to-phoneme*) çevrilməsi adlanır. Fonetik transkripsiya və prosodik informasiya birlikdə simvolik dil təsvirini verir. Tərs məsələyə çox zaman sintezator adı verilir və simvolik dil təsvirini səsə çevirir. Bəzi sistemlərdə bu hissəyə hədəf prosodiyanın (əsas tonun konturu, fonemlərin müddəti) hesablanması da daxildir.

Vurğuların düzgün qoyuluşu, ixtisarlara, ədədlərin, xüsusi işarələrin düzgün oxunması, səslənmənin yüksək dərəcədə təbiiliyi və intonasiya axıcılığı əsas problemlərdəndir.

VoiceXML (Voice Extensible Markup Language, VXML) – insanla kompüter arasında interaktiv media və səsli dialoqu təsvir etmək üçün rəqəmsal sənəd standartıdır. Bank sistemləri və müştərilərə avtomatlaşdırılmış xidmət portalları kimi səsi dəstəkləyən tətbiqi proqramların yaradılmasında istifadə olunur. VoiceXML tətbiqi proqramları veb-brauzerin veb-serverdən aldığı HTML (Hypertext Markup Language) sənədini emal etməsinə oxşar işləyirlər. VoiceXML sənəd formatı XML-ə (Extensible Markup Language) əsaslanır. W3C (World Wide Web Consortium) tərəfindən yaradılmış standartdır.

Səsə görə şəxsin tanınması

ALIZE – səsə görə biometrik autentifikasiya sistemlərinin işlənməsi üçün C++ açıq kodlu platforma (LGPL lisenziyası ilə yayılır). Avignon Universiteti (Fransa) kompüter elmləri laboratoriyasında yaradılmışdır.

Background Model Фоновая модель

Fon modeli – böyük sayda səs nümunələri əsasında öyrədilmiş model; ümumi, diktordan asılı olmayan xarakteristikaları təsvir etmək üçün istifadə edilir.

Cepstral Mean Normalisation (CMN)

Метод нормализации кепстрального среднего

Kepstral ortanın normallaşdırılması – kepsral ortanın çıxılması metodu (CMS) kimi də məlumdur, kepsral xarakteristikalardan onların orta qiymətini (riyazi gözləməsini) çıxmaqla signalda «bükülmə» təhriflərini aradan qaldırmağa cəhd edilir:

$$\bar{c}_t = c_t - \frac{1}{T} \sum_{i=1}^M c_i.$$

Beləliklə, istənilən stasionar spektral təhrif, məsələn, mikrofonun tezlik xarakteristikası yox edilir, bu onunla əlaqədardır ki, bükülmə oblastındakı təsirlər kepsral oblastında additivdir. Bundan başqa, eksperimental sübut olunub ki, CMN metodu səsə həssaslığı da azaldır. Nəzəri olaraq bunu onunla izah etmək olar ki, danışan şəxs barədə informasiya təkcə spektrin sabit komponentində toplanmaya bilər, modulyasiya spektrinin aşağı tezliklərində də paylana bilər.

Cepstral Mean Subtraction (CMS)

Метод вычитания кепстрального среднего

Kepstral ortanın çıxılması – diktorun verifikasiya/identifikasiya sistemində əlamətlərin çıxarılması mərhələsində nitq signallarının ötürülməsi kanalının bükülmə təhrifini kompensasiya etmək üçün istifadə olunur. Kepsral ortanın normallaşdırılması (Cepstral Mean Normalization, CMN) metodu kimi də məlumdur.

Cohort models Когортные модели

Koqort modelləri – test ardıcılığının axtarılan diktör deyil, başqa birisi tərəfindən tələffüz edilməsi fərziyyəsinin həqiqətəoxşarlığının qiymətləndirilməsi üsuludur, “bütün dünyanın” modelləşdirilməsi yerinə, mövcud modellərdən “koqort çoxluq” adlandırılan kiçik altçoxluq seçilir.

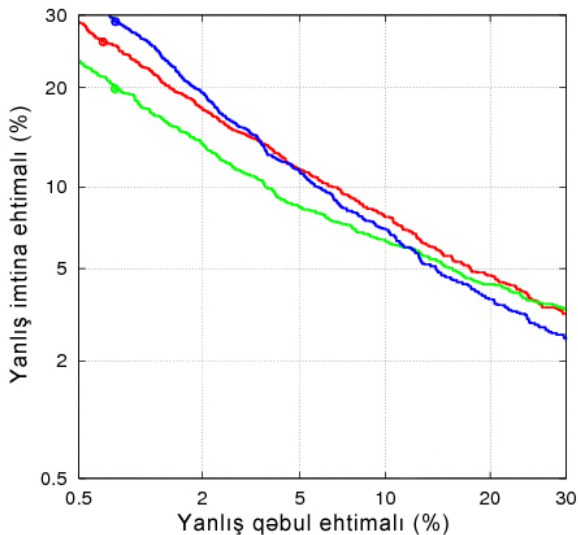
Criminalistic phonoscopy Криминалистическая фоноскопия

Kriminalistik fonoskopiya (yunanca phono – səs və latınca scopo – seyr etmək, düşünmək) – səsə görə şəxsin müəyyən edilməsidir (identifikasiyasıdır); səs yazısı, maqnit və digər daşıyıcılarda səs informasiyası tədqiq edilir. Fonoskopik ekspertiza rüşvət, ələalma, şantaj, müxtəlif növ dələduzluq və telefon terrorizmi üzrə cinayət işlərinin araşdırılması zamanı istifadə olunur.

DET curve (Detection Error Trade-off)

Параметрическая кривая ошибок

Tanıma səhvlərinin balansı – FRR-in FAR-dan asılılıq qrafikidir.



Detection Cost Function (DCF)

Функция стоимости определения

Tanınmanın dəyər funksiyası – dikturun tanınması sistemlərinin qiymətləndirilməsi parametridir:

$$DCF = C_{FR} P_{tar} R_{FR} + C_{FA} P_{imp} R_{FA},$$

burada P_{tar} və P_{imp} – həqiqi şəxsin və "yalançının" cəhdlərinin aprior ehtimallarıdır, R_{FR} və R_{FA} – uyğun olaraq yanlış imtina və yanlış qəbul ehtimallarının alınmış qiymətləridir, C_{FA} – yanlış qəbulun və C_{FR} – yanlış imtinanın qiymətidir. Qərarların qəbul olunması sərhədi, dəyər funksiyasının qiymətini minimallaşdırmaq üçün optimallaşdırılır.

Dynamic Time Warping (DTW)

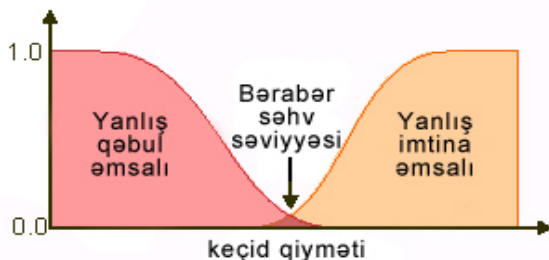
Динамическое искажение времени

Zamanın dinamik təhrifi alqoritmi – nümunələrin müqayisəsinə əsaslanan tanımadır. O dinamik proqramlaşdırma prinsipinə əsaslanır və seqmentlərin düzləşdirilməsinin qlobal yolunun qiymətləndirilməsi məsələsini həll edir.

Equal Error Rate və ya ERRor Rate (EER və ya ERR)

Равный уровень ошибок

Bərabər səhv səviyyəsi – səhvlərin tezliyidir. FAR və FRR öz aralarında əlaqəlidir – onlardan birinin qiyməti azalırsa, digərinin qiyməti artır.



Əgər hansısa bir keçid həddində (ing. *threshold*) $FAR=FRR$ olarsa, onda bu qiymət ERR kimi götürülür.

Expectation-Maximization (EM) Максимизация ожидания

Riyazi gözləmənin maksimallaşdırılması – ehtimal modeli bəzi gizli dəyişənlərdən asılı olduqda modelin parametrlərinin maksimal həqiqətəoxşarlıq qiymətləndirilmələrinin tapılması üçün riyazi statistikada istifadə edilən alqoritmdir. Alqoritm hər iterasiyası iki addımdan ibarətdir. E-addımda (ing. expectation) həqiqətəoxşarlıq funksiyasının gözlənen qiyməti hesablanır, bu halda gizli dəyişənlərə müşahidə edilən dəyişənlər kimi baxılır. M-addımda (ing. maximization) maksimal həqiqətəoxşarlığın qiyməti hesablanır, Beləliklə E-addımda hesablanmış gözlənen həqiqətəoxşarlıq artır. Sonra bu qiymət növbəti iterasiyada E-addım üçün istifadə olunur. Alqoritm yığılana qədər yerinə yetirilir.

False Accept Rate (FAR) Ошибка ложного принятия

Yanlış qəbul əmsalı – biometrik sistemin şəxsi başqa şəxs kimi tanıması ehtimalıdır (riyazi statistikada "2-ci növ səhv" adlanır).

False Reject Rate (FRR) Ошибка ложного отказа

Yanlış imtina əmsalı – biometrik sistemin şəxsi tanınamaması ehtimalıdır (riyazi statistikada "1-ci növ səhv" adlanır).

Forensic Speaker Recognition

Криминалистическое распознавание диктора

Səsin kriminalistik tanınması – məhkəmə ekspertizası üçün səs yazısına əsasən danışan şəxsin kimliyinin müəyyən edilməsi.

Gaussian Mixture Model (GMM)

Смешанные гауссовы модели

Qarışıq Gauss modeli – model qaussianların çəkili cəmindən ibarətdir:

$$p(x|\lambda) = \sum_{i=1}^M w_i p_i(x),$$

burada λ – diktor modeli, M – modelin komponentlərinin sayı, w_i – komponentlərin çəkildirid, $\sum_{i=1}^M w_i = 1$.

Hər bir komponentin ehtimalının sıxlıq funksiyası aşağıdakı kimi verilir:

$$p_i(x) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right),$$

burada D – əlamətlər fəzasının ölçüsü, μ_i – riyazi gözləmə vektoru, Σ – kovariasiya matrisidir. Əksər hallarda diaqonal kovariasiya matrisinə baxılır. Diktor modelinin qurulması üçün riyazi gözləmə vektorlarını, kovariasiya matrislərini və komponentlərin çəkirlərini müəyyən etmək lazımdır.

Google Voice Search Голосовой поиск Google

Google Voice Search – istifadəçilərə axtarış məlumatını mobil telefon və ya kompüterə deməklə Google Search-dən istifadə etməyə imkan verən tətbiqi proqramdır (2014-cü ildə 5 dil dəstəklənirdi). Proqram Iphone üçün yaradılmışdı, bulud texnologiyalarına əsaslanır. 2010-cu ildə Google Voice Search Androidlə işləyən telefonlara və 2011-ci ildə Chrome brauzerinə əlavə edilmişdir.

Hidden Markov Model (HMM) Скрытая модель Маркова

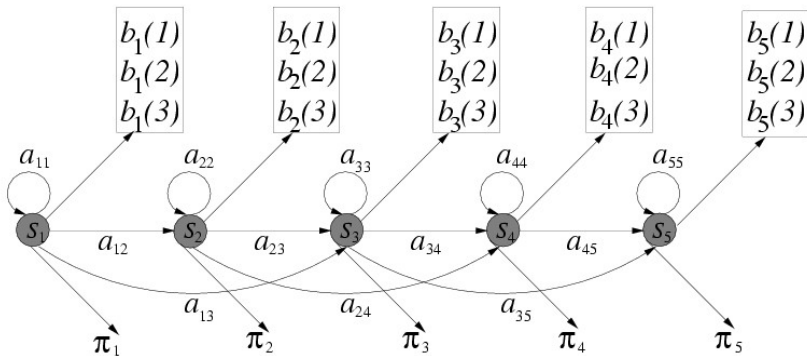
Gizli Markov Modeli (GMM) – model N vəziyyətdən ibarətdir, onların hər birində sistem hər hansı parametrlin M qiymətindən birini qəbul edə bilər. Vəziyyətlərin arasındakı keçid ehtimalları $A = \{a_{ij}\}$ matrisi ilə verilir, burada a_{ij} – i -ci vəziyyətdən j -cu vəziyyətə keçid ehtimalıdır. Parametrlin M qiymətindən hər birinin N vəziyyətdən birində olması ehtimalları $B = \{b_j(k)\}$ vektoru ilə verilir, burada $b_j(k)$ – j -cu vəziyyətdə k -cı qiymətin düşməsi ehtimalıdır. İlkin vəziyyətlərin ehtimalı da $\pi = \{\pi_i\}$ vektoru ilə verilir, burada π_i – başlanğıc anda sistemin i -ci vəziyyətdə olması ehtimalıdır. Beləliklə, gizli Markov modeli $\lambda = \{A, B, \pi\}$ üçlüyü ilə təyin olunur.

Gizli Markov modellərində üç əsas məsələyə baxılır:

- Verilmiş müşahidələr ardıcılığının baxılan modelə generasiya edilməsi ehtimalını hesablamaq.
- Verilmiş müşahidələr ardıcılığı əsasında gizli vəziyyətlərin ən ehtimalı ardıcılığını hesablamaq
- Verilmiş müşahidələr ardıcılığı əsasında modelin parametrlərini elə seçməli ki, bu modelə görə müşahidələrin ehtimalı maksimum olsun.

HMM-də ehtimalların qiymətləndirilməsi alqoritmləri:

- Viterbi alqritmi (Viterbi algorithm);
- Baum-Veş alqritmi (Baum-Welch algorithm);
- Düz-tərs gediş alqritmləri (Forward-backward procedure).



Hnorm (Handset normalization) Н-норма

Hnorm – bu yanaşmada telefonda asılı parametrlər hər bir diktor modeli bədənyyətli çoxluğu ilə müqayisə edilməklə qiymətləndirilir. Alınmış qiymətlər uyğunluq qiymətlərinin normalaşdırılması üçün istifadə edilir.

Impostor Злоумышленник

Bədənyyətli – özünü başqa bir şəxs kimi qələmə verən şəxs.

i-vector (identity vector) i-вектор

i-vektor – həm dikturun, həm də kanalın dəyişkənliyini özündə saxlayan ümumi dəyişkənlik matrisini (ing. *total variability matrix*) idarə edən faktordur. Bu zaman diktordan və kanaldan asılı M supervektoru aşağıdakı kimi göstərilir:

$$M = m + Tw$$

burada T – ümumi dəyişkənlik matrisidir, m – riyazi gözləmə vektoru, w – i-vektordur. i-vektorların klassifikasiyasına altfəzanın öyrədilməsi, xətti diskriminant analizi, sinifdaxili kovariasiyanın normallaşdırılması (ing. *Within-Class Covariance Normalization, WCCN*) və kosinus nüvəsi funksiyasından istifadə etməklə klassifikasiya mərhələləri daxildir.

Joint Factor Analysis (JFA) Совместный факторный анализ

Birgə faktor analizi – nitqin paylanması riyazi gözləmələrin sabit uzunluqlu supervektorları baxımından modelləşdirməyə imkan verir, arzu edilməyən sessiyalararası dəyişikliyin aşkar modelləşdirməsi mexanizmini təmin edir.

Diktordan asılı supervektor aşağıdakı kimi müəyyən edilir:

$$M = m + Vy + Ux + Dz,$$

burada m – diktordan və kanaldan asılı olmayan supervektordur, V və D – diktör altfəzasını (uyğun olaraq, məxsusi səslər matrisi və diaqonal qalıq), U (məxsusi kanallar matrisi) – sessiya altfəzasını müəyyən edir. x , y və z vektorları öz altfəzalarında diktordan və sessiyadan asılı faktorlardır, hər biri ixtiyari $N(0, 1)$ normal paylanmalı dəyişəndir. JFA ilə PLDA arasında əsas fərq ondadır ki, JFA metodu GMM riyazi gözləmə supervektoru fəzasında, PLDA isə birbaşa akustik əlamətlər fəzasında işləyir.

Linear Discriminant Analysis (LDA)

Линейный дискриминантный анализ

Xətti diskriminant analizi – obrazların tanınması zamanı ölçünün azaldılması üçün geniş istifadə olunan metoddur. Bu yanaşmanın ideyası siniflər arasındakı dispersiyanın (ing. *between-class variance*) maksimallaşdırılması və siniflər daxilində (ing. *intra-*

class variance) dispersiyanın minimallaşdırılmasından ibarətdir. Metod səsə görə şəxsin tanınması üçün də tətbiq edilir.

Maximum A Posteriori (MAP)

Максимум апостериорной вероятности

Aposterior ehtimalın maksimumu – təsnifat səhvi eyni olduqda orta risk funksionalının minimallaşdırması baxımından optimal olan klassifikator obyektinə ən çox ehtimal edilən sinfə aid edir.

Maximum Likelihood Linear Regression (MLLR)

Линейная регрессия максимальной правдоподобности

Maksimal həqiqətəoxşarlıq xətti regressiyası – optimallaşdırma alqoritmlərindən biridir. Məqsəd fonemlərin bütün modellərinin vəziyyətlərinin vektor kvantlaması metodu ilə müəyyən sayda regressiya siniflərinə ayırmaqdır.

NIST SRE (Speaker Recognition Evaluation) Оценка систем распознавания личности по голосу Национального института стандартов и технологий

NIST SRE – səsə görə şəxsi tanıma sistemlərini qiymətləndirmək məqsədilə ABŞ Milli Standartlar və Texnologiya İnstitutu (National Institute of Standards and Technology) tərəfindən 1996-cı ildən təşkil edilən müsabiqə (1996-2006-cı illərdə hər il və 2008, 2010, 2012-ci illərdə keçirilmişdir).

Nuisance Attribute Projection (NAP)

Метод проекции атрибутов искажений

Təhrif atributlarının proyeksiyası metodu – klassifikatorlarda təhrif atributlarının filtrasıyası üçün istifadə edilən metoddur. Təhrif atributları müşahidələrin nəticələrinə təsir göstərən və konkret şəxslə əlaqədar olmayan xüsusiyyətlərdir. Məsələn, müxtəlif növ mikrofonlar eyni şəxsin nitqində müxtəlif dəyişikliklər yaradır. NAP metodunun məqsədi belə dəyişiklikləri aradan qaldırmaqdan ibarətdir. Fərz olunur ki, kanal dəyişiklikləri kiçik ölçülü altfəzalarda

yerləşir və bu dəyişiklikləri yox etmək üçün xüsusi ortoqonal proyeksiya matrisi qurulur.

Pattern recognition Распознавание образов

Obrazların tanınması – obyektlərin bir neçə sinif və ya kateqoriya üzrə klassifikasiyasını öyrənir. Obyektlərə *obraz* deyilir.

Obrazların tanınması məsələsi intellektual sistemlərin əksəriyyətində əsas məsələlərdən biridir. Nitqin, səsə görə şəxsin tanınmasına obrazların tanınması məsələləri kimi baxmaq olar.

Klassifikasiya presedentlərə əsaslanır və presedentlər əsasında qərar qəbul edilməsi geniş yayılmış yanaşmadır.

Presedent – klassifikasiyası məlum olan obrazdır və klassifikasiya məsələlərinin həllində nümunə kimi qəbul edilir.

Fərz edək ki, bütün obyektlər sonlu sayda siniflərə bölünüb. Hər bir sinif üçün sonlu sayda obyekt – presedent məlumdur. Obrazların tanınması yeni obyektə siniflərdən birinə aid etmək prosesidir.

Phonogram Фонограмма

Fonoqram – danışığın və ya onun tədqiq olunması tələb edilən fraqmentinin audioyazısıdır.

Principal Component Analysis (PCA)

Метод главных компонент

Əsas komponentlər metodu – minimal miqdarda informasiya itirməklə verilənlərin ölçüsünün azaldılmasının əsas metodlarından biridir. K. Pirson tərəfindən 1901-ci ildə təklif edilmişdir. Metodun mahiyyəti koordinat oxları ilkin verilənlərin maksimal dispersiyası istiqamətində yönəlmiş yeni ortoqonal bazisə keçməkdən ibarətdir. Yeni bazisin birinci oxu üzrə dispersiya maksimaldır, ikinci ox dispersiyanı birinci oxa ortoqonallıq şərtində maksimallaşdırır və s. Əsas komponentlərin hesablanması ilkin verilənlərin kovariasiya matrisinin məxsusi qiymətlərinin və məxsusi vektorlarının hesablanmasına gətirilir.

Probabilistic Linear Discriminant Analysis (PLDA)

Вероятностный линейный дискриминантный анализ

Ehtimallı xətti diskriminant analizi – xətti diskriminant analizinin (linear discriminant analysis, LDA) ehtimal üçün genişlənməsidir. Diktorun tanınmasında PLDA spesifik sinif (məsələn, bir diktor) üçün müşahidələrin dəyişkənliyini sinfin bütün müşahidələri üçün ortaq olan identifikasiya dəyişəninə və hər bir müşahidədən asılı olan kanal dəyişəninə ayırır. Kanal dəyişəni hər bir müşahidəyə kanal küyünün səbəb olduğu dəyişkənliyi izah etmək üçün istifadə edilir. PLDA modelində D_x ölçülü i -vektor aşağıdakı şəkildə təsvir edilə bilər:

$$x_{ij} = \mu + Fh_i + Gw_{ij} + \varepsilon_{ij}.$$

Bu model iki hissədən ibarətdir: $\mu + Fh_i$ yalnız diktordan asılı komponent və küy komponenti $Gw_{ij} + \varepsilon_{ij}$. Küy komponenti hər sessiyada fərqlənir və diktorun tələffüzündə küyü xarakterizə edir. ε_{ij} – riyazi gözləməsi 0 və kovariasiya matrisi Σ olan qaussiandır.

RASTA (RelAtive SpecTrAl) – təhrif edilmiş nitq seqmentinə tətbiq edilən qısamüddətli çevirməni və nitq spektrinə tətbiq edilən loqarifmik çevirməni özündə birləşdirən kanalın kompensasiyası metodudur. Hynek Hermansky və həmmüəllifləri tərəfindən 1991-ci ildə təklif edilmişdir (Berkli Universiteti International Computer Science Institute, ICSI).

Score Normalization Нормализация значений соответствия

Uyğunluq qiymətinin normallaşdırılması – Səsə görə şəxsin verifikasiyası sistemlərinin dəqiqliyini artırmaq üçün uyğunluq qiymətləri üzərində aparılan çevirmə.

Speaker Diarization Обнаружение диктора

Diktorun diarizasiyası – audio axının danışanlara aid seqmentlərə bölünməsidir, diktor seqmentlənməsi və ya klasterizasiya da adlanır. Diarizasiya danışan şəxslərdən kimin hansı zaman anında danışmasının müəyyən edilməsi prosesidir. Nitqin danışan şəxslərə

aid seqmentlərə bölünməsi nitqin avtomatik tanınması və ya şəxsin səsə görə identifikasiyası zamanı istifadə edilir. Əsas problemlərdən biri odur ki, diktorların sayı əvvəlcədən məlum deyil və avtomatik müəyyən etmək lazımdır.

Speaker Identification Идентификация диктора

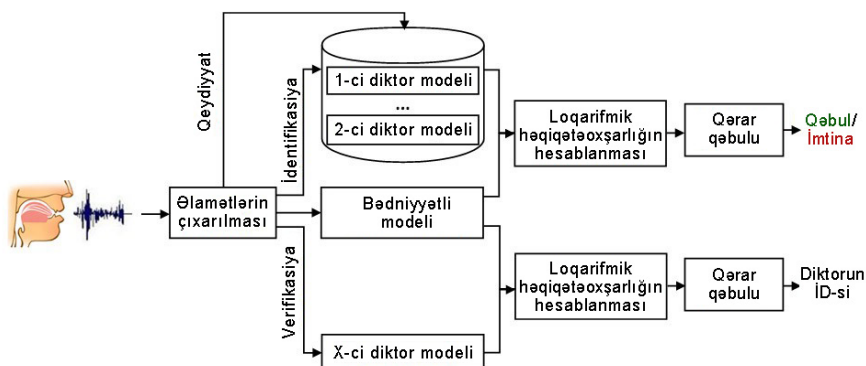
Diktorun identifikasiyası – səs nümunəsini bazada saxlanan nümunələrlə müqayisə etməklə danışan şəxsin müəyyən edilməsi prosesi. İdentifikasiyanın nəticəsi olaraq səs nümunəsi sistemdə qeydiyyatda alınmış və axtarılan nümunəyə ən yaxın olması ehtimal edilən şəxsin (və ya bir neçə şəxsin) adı bildirilir.

Speaker Model Модель диктора

Diktor modeli – bir diktora aid səs nümunələri əsasında öyrədilmiş model; həmin şəxsin xarakteristikalarını təsvir etmək üçün istifadə edilir.

Speaker Recognition Распознавание диктора

Diktorun tanınması – nitq signalı əsasında danışanın kimliyinin avtomatik müəyyən edilməsi prosesidir. Bəzən səsin tanınması (ing. voice recognition) kimi də işlədilir.



Speaker Segmentation and clustering Сегментация дикторов и кластеризация

Важ. Speaker diarization

Speaker Verification Верификация диктора

Diktorun verifikasiyası – diktorun özünü təqdim etdiyi şəxs olduğunun müəyyən edilməsi prosesidir. Diktorun səsi yalnız bir şablonla – özünü təqdim etdiyi şəxslə müqayisə edilir.

Səsə görə şəxsin verifikasiyası binar klassifikasiya məsələsidir. Formal olaraq, məsələ iki hipotezin yoxlanmasından ibarətdir:

H_0 – Y frazasını S adlı şəxs tələffüz edib;

H_1 – Y frazasını S adlı şəxs tələffüz etməyib.

İki hipotezdən birinin seçilməsi üçün həqiqətəoxşarlığın nisbətində baxılır. Bu zaman qərar qəbuletmə proseduru aşağıdakı kimidir:

$$\frac{P(Y|H_0)}{P(Y|H_1)} \begin{cases} \geq \theta \rightarrow H_0 \text{ qəbul edilir} \\ < \theta \rightarrow H_1 \text{ qəbul edilir} \end{cases}$$

burada $P(Y|H)$ – H hipotezinin doğru olması şərtində Y nitq seqmenti əsasında qiymətləndirilmiş ehtimal sıxlığı funksiyasıdır, θ – keçid qiymətidir. H hipotezi S adlı diktoru əlamətlər fəzasında xarakterizə edən diktor modeli ilə müəyyən edilə bilər.

Supervector Супервектор

Supervektor – Qarışıq Gauss modelinin riyazi gözləmə vektorlarından ibarət vektordur. JFA metodunda supervektor iki supervektorun – s diktor supervektoru və c kanal supervektorunun cəmi şəklində göstərilir:

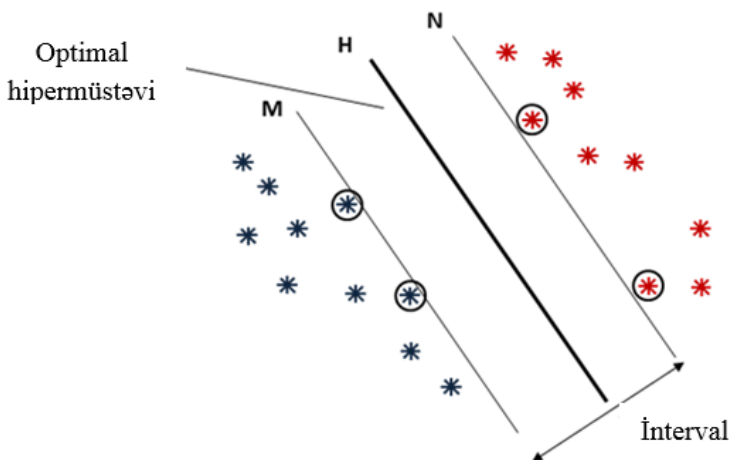
$$M = s + c,$$

burada s və c normal paylanmaya malikdir.

Support Vector Machines (SVM) Метод опорных векторов

Dayaq vektorları metodu – xətti ayrılan iki sinif üçün təlim (nümunələr) çoxluğunun ən yaxın nöqtələrindən maksimum məsafədə keçən ayrıcı hiperüstəvinin tapılmasından ibarətdir.

Verilənlər xətt ayrılan olmadıqda giriş nöqtələrini Hilbert fəzasına inikas etdirən qeyri-xətti çevirmələrdən istifadə etmək olar.



Text Dependent Speaker Recognition

Текстозависимое распознавание диктора

Diktorun məndən asılı tanınması – diktordan açar sözlərin və ya cümlələrin tələffüz edilməsini tələb edən verifikasiya növü; eyni mətn həm öyrənmə (təlim), həm də tanıma üçün istifadə edilir.

Text Independent Speaker Recognition

Текстнезависимое распознавание диктора

Diktorun məndən asılı olmadan tanınması – konkret mətnin tələffüz edilməsindən asılı olmadan diktorun verifikasiyası.

T-norm (Test normalization) Т-норма

Tnorm – uyğunluq qiymətlərində sessiyalararası dəyişməni kompensasiya edir. Hər bir diktor üçün yalançı və doğru uyğunluq qiymətləri arasındakı üst-üstə düşməni azaltmağa cəhd edir.

Universal Background Model (UBM)

Универсальная фоновая модель

Universal fon modeli və ya "bədniyyətli" modeli – nıtq bazasında istifadə edilən bütün diktorların bəzi ortalama xarakteristikalarını təsvir edir. Universal fon modeli böyük sayda (məsələn, 2048) gaussianla təsvir olunur.

Modelin öyrədilməsi diktorlar çoxluğunun böyük həcmdə səs yazıları əsasında həyata keçirilir. Bu halda H_1 -in həqiqətəoxşarlığı belə hesablanır:

$$p(S | H_1) = p(X | \lambda_{UBM}),$$

burada λ_{UBM} – universal fon modeli, H_0 – “X frazasını S tələffüz etmişdir”, H_1 – “X frazasını S tələffüz etməmişdir” hipotezləridir.

Vector Quantization (VQ) Векторное квантование

Vektor kvantlaması – çoxparametrlı signal parametrlərin hər biri üzrə kvantlanır və nəticə vahid verilənlər bloku kimi saxlanır və emal edilir. Vektorların nizamlanmış çoxluğu *kod kitabı* adlanır. Vektor kvantlaması rəqəmsal mobil rabitə sistemlərində nitqin kodlaşdırılması üçün geniş istifadə edilir.

Vektor kvantlamasına vektorun mümkün qiymətlər fəzasının sonlu sayda oblasrlara bölünməsi kimi baxmaq və klassifikasiya metodu kimi də istifadə etmək olar.

Within-Class Covariance Normalization (WCNN)

Внутриклассовая ковариационная нормализация

Sinifdaxili kovariasiyanın normallaşdırılması – i -vektor əsasında sistemlərin kompensasiyası metodudur. Onun ideyası təlim mərhələsində FAR və FRR səhvlərini minimum etmək üçün axtarılan diktorun və "yalançıların" xətti ayrılmasından ibarətdir.

Z-norm (Zero normalization) Z-норма

Z-norma – uyğunluq qiymətlərində diktorlararası dəyişməni kompensasiya edir. Ümumi, diktordan asılı olmayan keçid qiymətindən istifadə etməyə imkan verir.

Nitq bazaları

ELDA (Evaluations and Language resources Distribution Agency) Агентство по оценке и распространению языковых ресурсов

Qiymətləndirmə və Dil Resurslarının Yayımı Agentliyi
<http://www.elda.org/>

ELRA (European Language Resource Association) Европейская ассоциация языковых ресурсов

Avropa Dil Resursları Assosiasiyası – ingilis dilindən fərqli dillərdə və ingilis dilinin daşıyıcısı olmayan insanların səsi toplanmış bir neçə verilənlər bazasına malikdir. Səsə görə şəxsin identifikasiyası və verifikasiyası üçün italyanca danışan 671 insandan toplanmış səs yazıları bazası istifadə edilə bilər. Diktorlar telefonla ofis və ya ev şəraitində mətni oxuyurlar və rəqəmləri söyləyirlər. Səs bazası fasilə bir neçə gün olmaqla bir neçə sessiya ərzində toplanmışdır.

<http://www.elra.info/>

LDC (Lingustic Data Consortium) Консорциум лингвистических данных

Linqvistik Verilənlər Konsorsiumu <http://www ldc.upenn.edu/>

NIST speech databases Речевой базы данных NIST

NIST səs nümunələri bazası – ilk olaraq Switchboard I və Switchboard II bazalarından alınmışdır. Verilənlər LDC konsorsiumu tərəfindən toplanmışdı. Bazada kollec tələbələrinin 5-dəqiqəlik dialoqları olan 2728 səs yazısı var. Tələbələr müxtəlif telefonlardan gündə bir dəfə zəng edirdilər, zəng etdikləri şəxsə isə həmişə eyni telefon olurdu.

NTIMIT database База данных NTIMIT

NTIMIT verilənlər bazası – DARPA (Müdafiə elm-tədqiqat işlərinin perspektiv planlaşdırılması) İdarəsinin kəsilməz akustik-

fonetik nitq fondu səs nümunələri verilənlər bazalarından biridir. Verilənlər bazasında 630 insan qeydə alınıb, cümlələri akustik kabinədə oxuyurlar. NYNEX (NTIMIT) verilənlər bazası süni səs köməyi ilə yaradılıb, bu səs telefon xətti ilə böyük məsafəyə verilir və o biri uca yazılırdı.

Oregon İnstitutunun səsin tanınması üçün verilənlər bazasında 100 obyektin səs yazıları var. Obyektlər iki il müddətində müxtəlif küy səviyyəsində müxtəlif telefon istifadə edərək və müxtəlif mühitdə 12 dəfə zəng edirdi. Verilənlər bazasını məndən asılı etməmək üçün obyektlər müxtəlif mövzularda danışırıdılar.

POLYCOST – bu verilənlər bazası Avropa layihəsi üçün toplanmışdı. Yazıların əksəriyyəti ingilis dilində danışan, lakin doğma dili ingilis dili olmayan şəxslərin nitqidir, yazılarda onların doğma dillərində kiçik nitq fraqmentləri də var. Verilənlər bazasında 13 Avropa ölkəsinin nümayəndəsi qeydə alınıb. Yazılar ISDN telefon xətlərində toplanıb. Bu sistemin köməyi ilə danışanın doğma dilinin səsin tanınmasına təsiri test edilə bilər.

PolyVar – həm fransız dili daşıyıcılarının, həm də fransız dili ikinci dil olan şəxslərin, əsasən isveçrəlilərin, səs yazıları toplanmış səs bazasıdır. Səs bazasında alman və fransız dillərində 160 saatlıq səs yazısı var. Səs bazası yaradılarkən 31 insan 2-dən 10-a qədər, 41 insan isə 10-dan artıq zəng vurmuşdur.

TIMIT (Texas Instruments Massachusetts Institute of Technology) geniş yayılmış nitq verilənləri bazalarından biridir. Demək olar ki, ideal şəraitdə identifikasiya etməyə imkan verir. TIMIT məlumat bazasına ABŞ dialektlərinin 10 müxtəlif regionundan olan 630 diktordan (70 %-i kişilər və 30 %-i qadınlar) alınmış 6300 səs yazısı daxildir. Hər bir dikturun on danışığının (tələffüzünün) hər biri üçün təxminən 30 saniyəlik nitq yazısı var. Səs bazası frazalar üzrə təşkil edilmişdir: tələffüzlər ayrıca fayla yazılmış və onlara uyğun fonetik nişanlama faylları verilmişdir.

YOHO – səsə görə şəxsin verifikasiyası üzrə YOHO səs bazası sabit mətnlə səsə tanınması üzrə eksperimentlər üçün nəzərdə tutulub. Bu bazada yüksək keyfiyyətli telefonla deyilən frazaları tələffüz edən 128 obyektin səsi yazılıb. Səs bazası dörd sessiya ərzində toplanmış, sessiyadan sonra verifikasiya aparılmış, sessiyalar arasında fasilə bir neçə gün çəkmişdi.

Nitq siqnallarının emalı üçün proqram təminatı

Adobe Audition (CoolEditPro) – nitq siqnallarının analizi və emalı alətidir, bir sıra funksional imkanlara malikdir.

<http://www.adobe.com/special/products/audition/syntrillium.html>

CMU Sphinx – Karnegi Mellon Universitetində yaradılmış nitqin tanınması sistemlərini bildirən ümumi termindir. Sphinx diktordan asılı olmayan, kəsilməz (təbii danışiq) nitqin tanınması sistemidir. HMM akustik modelinə və N-qram statik linqvistik modeldən istifadə edir.

CMU Sphinx müxtəlif nitq emalı məsələlərini əhatə edən bir neçə paketdən ibarətdir:

- Pocketsphinx – C-də yazılmış nitq tanıma sistemi;
- Sphinxbase – Pocketsphinx-in tələb etdiyi dəstək kitabxanası;
- Sphinx4 – Java-da yazılmış modifikasiya edilə bilən tanıma sistemi;
- Sphinxtrain – akustik modeli öyrətmə alətləri.

GoldWave – peşəkar rəqəmsal audio redaktordur. Səs yazıları üzərində sadə əməliyyatlardan səsin ən mürəkkəb emal üsullarına kimi bir çox funksiyanı yerinə yetirir. Öyrənilməsi və istifadəsi sadədir.

<http://www.goldwave.com/>

HTK (Hidden Markov Model Toolkit) – Kembric universitetində yaradılmışdır, gizli Markov modellərindən istifadə etdikdə tez-tez müraciət edilən proqram paketlərindən biridir. Paketi bir çox tədqiqat sahəsində istifadə etmək olar, lakin nitqin analizi üçün çox sayda köməkçi funksiyanın, məsələn, akustik və dil modellərinin yaradılması funksiyalarının mövcudluğu onu nitqin tanınması üzrə proqram vasitələrinin yaradılması üçün xüsusilə faydalı edir.

MATLAB – yüksək səviyyəli proqramlaşdırma dili və alqoritmlərin yaradılması, verilənlərin analizi və vizuallaşdırılması, ədədi hesablamalar üçün interaktiv mühitdir. MATLAB-ın baza kitabxanasında və Signal Processing, Filter Design və Communications əlavə paketlərində siqnalların rəqəmsal eməli ilə əlaqədar müxtəlif hesablamaları asan və sürətlə yerinə yetirməyə imkan verən çox sayda funksiyalar var. MATLAB-da səs fayllarının yazılması və ya oxunması, sonoqram və spektroqramların çıxarılması, seqmentasiya, formantların analizi, əsas ton periodunun alınması, filtrasiya və s. üçün çoxsaylı funksiyalar vardır.

MATLAB mühitində maşın təlimi üsullarına daxildir: klassifikasiya üsulları (dayaq vektorları metodu (SVM), *k*-ən yaxın qonşular metodu, sadə Bayes, diskriminant analizi, loqistik reqressiya və neyron şəbəkələri) və klasterizasiya (*k*-ortalılar metodu, Qarışıq Gauss modeli (GMM), gizli Markov modelləri (HMM), qeyri-səlis *c*-ortalılar klasterizasiya metodu və s.)

NLTK – dil alətləri toplusudur. Alət toplusuna kompüter linqvistikası və təbii dilin eməli sahəsində təhsil və elmi araşdırmaları dəstəkləmək üçün proqram modulları, verilənlər toplusu və dərslilər daxildir. NLTK alətləri təbii dilin eməli (Natural Language Processing, NLP) və empirik linqvistika, süni intellekt, informasiya axtarışı və maşın təlimi kimi sahələrdə təhsil alan tələbələr və ya elmi araşdırmalar aparan tədqiqatçılar üçün ideal vəsaitdir.

<http://www.nltk.org>

PRAAT (doing phonetics by computer) – Amsterdam Universiteti Fonetik Tədqiqatlar İnstitutunda hazırlanmış (1992-ci il, professorlar Paul Boersma və David Weenink) kompüter proqramıdır, pulsuz paylaşılır (holland dilində danışmaq deməkdir). Praat nitq siqnallarını analiz, sintez və eməli etməyə imkan verir. Proqramın köməyi ilə tonun yüksəkliyini, nitqin format və intensivliyini analiz/sintez etmək, onun xarakterini və tonunu dəyişmək olar. WAV, AIFF, FLAC və digər formatlarda

audiofaylları emal edə bilər. Həm 8-, həm də 16-bitlik formatları dəstəkləyir.

Funksional imkanları:

- səsənin analizi (statistik, spektral, intensivlik, yüksəklik, formant və s.);
- nitqin sintezi (artikulyasiya, fonetik və akustik xarakteristikalar üzrə);
- səs seqmentlərini redaktə etmək imkanı;
- spektroqramı çap etmək imkanı;
- annotasiya (spektroqram, tonal kontur, formant konturu, güc);
- fonetik eksperimentlər.

Proqrama riyazi və fonetik simvollar da daxildir və Encapsulated PostScript faylları generasiya edir, bu onu dissertasiya və məqalələrin yazılmasında faydalı edir. Bundan başqa, proqramda sadə proqramlaşdırılan ssenari dili də var və digər proqramlarla yaxşı işləyir.

<http://www.fon.hum.uva.nl/praat/>

SpeechAnalyzer nitqin analizi üçün kompüter sistemidir və səs materialının fonetik və fonoloji analizi üçün tətbiq edilir. Proqram səs signalını spektroqram, dalğa formasının qrafiki təsviri, tonun tezliyinin dəyişməsi diaqramları, intensivliyin, müddətin və formant komponentinin diaqramları şəklində təqdim edərək vizuallaşdırır.

<http://www-01.sil.org/computing/sa/>

WinCeCil – səs yazısına baxmaq və spektroqramlar tərtib etmək üçün proqramdır. Səs yazısının müddəti 3 saniyə təşkil edir. Təkmilləşdirilmiş versiyası *Speech Analyzer* adlanır.

www-01.sil.org/computing/speechtools/softdev2/Cecil2/CECdownloads2.htm

Ədəbiyyat

1. Huang X., Acero A., Hon H. Spoken language processing: a guide to theory, algorithm, and system development. Prentice Hall, 2001.
2. Denes P.B., Pinson E.N. The speech chain: the physics and biology of spoken language. New York: Anchor Books, 1998.
3. Rose P. Forensic speaker identification. New York: Taylor & Francis, 2002.
4. Rabiner L.R., Juang B.N. Fundamentals of speech recognition. New Jersey: Prentice Hall, 1995.
5. Quatieri T.F. Discrete-time speech signal processing: principles and practice. New Jersey: Prentice Hall, 2001.
6. Vaseghi S.V. Advanced digital signal processing and noise reduction. Chichester: John Wiley & Sons, 2006.
7. Benesty J., Sondhi M., Huang Y. Springer handbook of speech processing. Berlin: Springer Verlag, 2008.
8. Hess W.J. Pitch determination of speech signals. Berlin: Springer-Verlag, 1983.
9. Сергиенко А.Б. Цифровая обработка сигналов. – СПб Питер, 2002.

Veb-saytlar

Signalların rəqəmsal emalı

1. Wikipedia <http://en.wikipedia.org/>
2. А.В. Фролов, Г.В. Фролов, Синтез и распознавание речи. Современные решения. <http://www.frolov-lib.ru/books/hi/>
3. Грибунин В.Г., Глоссарий по цифровой обработке сигналов - www.autex.spb.ru/
4. Audio signalların emalı üçün C++ kitabxanası <http://www.music.helsinki.fi/research/spkit/documentation/SPKit.html>.
5. Nitq texnologiyalarına həsr olunmuş portal <http://speech-soft.ru>
6. Signalların rəqəmsal emalı və nitqin kodlaşdırılması – müəhazirələr və materiallar <http://aprodeus.narod.ru/teaching.htm>

7. *Signalların rəqəmsal emalına həsr olunmuş forum*
<http://www.hydrogenaudio.org>
8. *Signalların rəqəmsal emalının müxtəlif aspektlərinə həsr olunmuş portal* <http://www.dsprelated.com/>
9. *SuperCollider – Real zamanda audio sintezi üçün proqramlaşdırma dili* <http://www.audiosynth.com/>
10. *Edinburq Universiteti Nitq Texnologiyaları Tədqiqat Mərkəzi* <http://www.cstr.ed.ac.uk/>
11. *Karnegi Mellon Universiteti, Nitq texnologiyaları qrupu*
<http://www.speech.cs.cmu.edu/>

Nitqin sintezi

1. *Açıq kodlu (Java) nitq sintezatoru*
<http://freetts.sourceforge.net/docs/index.php>
2. *Java speech API (Sun Microsystems) – nitqin sintezi və tanınması* <http://java.sun.com/products/java-media/speech>.
3. *Açıq kodlu eSpeak layihəsi – ingilis və digər dillər üçün nitq sintezatorları* <http://espeak.sourceforge.net/>
4. *MARY nitq sintezi sistemi (Java) – ingilis və alman dilləri*
<http://mary.dfki.de/>
5. *Festival nitq sintezi sistemi*
<http://www.cstr.ed.ac.uk/projects/festival/>
6. *Karnegi Mellon Universitetinin nitq sintezi sistemi*
<http://www.festvox.org/>
7. *Naqoya Texnologiya İnstitutunun HMM-əsasında nitq sintezi sistemi (HTS)* <http://hts.sp.nitech.ac.jp/>
8. *MBROLA layihəsi (Faculté Polytechnique de Mons)*
<http://tcts.fpms.ac.be/synthesis/mbrola.html>
9. *LPC sintezatoru və analizatoru*
http://soundlab.cs.princeton.edu/software/rt_lpc/.
10. *GSM vokoderlər və standartlar*
http://www.vocal.com/data_sheets/gsmvocoders.html
11. *İngilis dilinin leksik bazası Wordnet*
<http://wordnet.princeton.edu/>

Nitqin tanınması

1. *CMU Speech Software*
<http://www.speech.cs.cmu.edu/hephaestus.html>
2. *CMU Sphinx layihəsi* <http://cmusphinx.sourceforge.net/>
3. *Nitqin tanınması üçün səs korusu (GPL lisenziyası)*
<http://www.voxforge.org/>
4. *Julius layihəsi* http://julius.sourceforge.jp/en_index.php
5. *HTK (Hidden Markov Model Toolkit)* <http://htk.eng.cam.ac.uk/>
6. *Voice Code layihəsi*
<http://voicecode.iit.nrc.ca/VoiceCode/public/ywiki.cgi>
7. *ISIP (Institute for Signal and Information Processing)*
<http://www.isip.piconepress.com/projects/speech/>
8. *MIT Oxygen layihəsi* <http://oxygen.csail.mit.edu/Speech.html>

Səsə görə şəxsin tanınması

1. *MIT Lincoln Laboratory, Human Language Technology Group*
<http://www.ll.mit.edu/mission/cybersec/HLT/HLT.html>
2. *SRI International Speech Technology and Research (STAR) Laboratory* <http://www.speech.sri.com/>
3. *Idiap Research Institute, Biometric Person Recognition (Biometrics) Research Group* <http://www.idiap.ch/scientific-research/research-groups/biometric-person-recognition>
4. *University of Eastern Finland, Speech and Image Processing Unit (SIPU)* <http://cs.uef.fi/pages/tkinnu/webpage/>
5. *Brno University of Technology, Speech Processing Group*
<http://speech.fit.vutbr.cz/>
6. *Центр речевых технологий (Санкт-Петербург)*
<http://www.speechpro.ru/>

Adlar göstəricisi

1/f noise 30

μ -law 62

A

A/D converter 24

Accent 16

Acoustic model 63

Acoustic signal 5

Acoustic theory of speech
production 10

Acoustical spectrum 52

Acoustics 5

ADC 24

Additive noise 30

Additive white Gaussian
noise 30

Adobe Audition 99

ALIZE 82

Allophone 16

Alveolar tubercle 10

AMDF algorithm 40

Amplitude 5

Amplitude method for pitch
detection 40

Amplitude spectrum 40

Amplitude-frequency
characteristic 52

Analog audio signals 57

Analog signal 24

Analog-to-digital converter 24

Anti-aliasing filter 30

Articulation 10

ASR 64

ASR Accuracy 64

Audio codec 57

Aural perception 16

Autocorrelation 40

Autocorrelation method for
pitch detection 41

Automatic Speech
Recognition 64

A-law 57

B

Background Model 82

Backing-off method 65

Band pass filter 31

Band stop filter 31

Bandwidth filter 31

Bark scale 5

Bayes classifier 65

Bessel filter 31

Bigram 16

Biquad filter 32

Bit rate 24

Blackman window 41

Blackman-Harris window 41

BPF 31

Butterworth filter 32

C

CDHMM 66

CELP 57

Cepstral analysis 52

Cepstral coefficients 52

Cepstral	Mean	DAC	25
Normalisation	82	DAM	58
Cepstral Mean Subtraction	82	DCC	52
Cepstrum	42	DCF	84
Cepstrum method for pitch detection	42	DCT	44
Chebyshev filter	33	DDCC	53
CMN	82	Decibel	6
CMS	82	Decimation	25
CMU Sphinx	99	Decoder	59
Code	Excited	Deep Neural Network	66
Prediction	57	Delta MFCC	47
Codec	58	Delta-Cepstral Coefficients	52
Cohort models	83	Delta-delta MFCC	48
Comb filter	33	Delta-Delta-Cepstral Coefficients	53
Companding	58	DET curve	83
Compression	58	Detection Cost Function	84
Concatenative synthesis	75	Detection Error Trade-off	83
Confidence score	65	DFT	44
Connected speech	65	Diagnostic	Acceptability
Context dependent phoneme model	65	Measure	58
Context	independent	Diagnostic Rhyme Test	59
phoneme model	66	Digital filters	33
Continuous Density HMM	66	Digital signal	25
Continuous	Wavelet	Digital signal processing	25
Transform	43	Digital signal processor	25
CoolEditPro	99	Digital-to-analog converter	25
Cosinus window	43	Diphone	16
Criminalistic phonoscopy	83	Diphone synthesis	75
Critical band	5	Discrete cosine transform	44
CWT	43	Discrete Fourier transform	44
		Discrete signal	26
		Discretization step	26
		Discriminativ models	87
		Double delta	48
D			
D/A converter	25		

Double Delta Cepstral
Coefficients 53
Downsampling 26
DNN 66
DRT 59
DSP 25
DTW 84
Dynamic range 26
Dynamic Time Warping 84

E

Expectation-Maximization 85
Equal Error Rate 84
ELDA 96
ELRA 96
EM 85
enCOder 57
Energy 16
Epiglottis 10
ERR 84
ERror Rate 84
European Language Resource
Association 92
Evaluations and Language
resources Distribution
Agency 92

F

False Accept Rate 85
FAR 85
Fast Fourier Transform 44
Festival 75
FFT 41
Filter bank 33
Filter design 33

Filtration 34
Finite Impulse Response 34
FIR 34
Forensic Speaker
Recognition 85
False Reject Rate 85
Formant 17
Formant extraction
algorithm 18
Formant synthesis 76
Fourier transform 45
Frames 27
Frequency spectrum 45
Frequency-domain Features 53
Fricative sound 18
FRR 85
Fundamental frequency 18

G

Gaussian Mixture Model 85
Gaussian window 45
Generative models 88
Glottal stop 11
Glottis 11
GMM 85
GoldWave 99
Google Voice Search 86
Gullet 11

H

H.261 59
H.264 59
H.323 59
Hamming window 45
Handset normalization 87

Hanning window 45
Harmonics 19
Hidden Markov Model 86
High-level features 53
High-pass filter 34
HMM 86
HMM synthesis 76
Hnorm 87
HTK 99

I

identity vector 88
IDFT 46
IIR 35
IIR filter 32
Impostor 87
Impulse 34
Impulse characteristic 35
Infinite impulse response 35
Intonation 19
Inverse Discrete Fourier Transform 46
i-vector 88

J

Java Speech API Markup Language 76
JAWS 77
JFA 88
Joint Factor Analysis 88
JSML 76

K

Keyword spotting 67

L

Language model 67
Language recognition 68
Large vocabulary 68
Large-Vocabulary Continuous Speech 63
Larynx 11
Lattice 68
LDA 88
LDC 96
Left-to-right HMM 68
Lexicon 69
Levinson-Durbin algorithm 53
Linear digital filter 35
Linear Discriminant Analysis 88
Linear Prediction Cepstral Coefficients 54
Linear Prediction Coefficients 54
Linguistic Data Consortium 92
Lips 11
Low-pass filter 35
LP 54
LPC 49
LPCC 54
LVCSR 70

M

Maximum A Posteriori 89
Maximum A Posteriori Estimation 70
Maximum Likelihood Estimation 70

Maximum Likelihood Linear
 Regression 89
 Maximum Mutual
 Information Estimation 70
 MAP 89
 MAP decoding 77
 MATLAB 100
 MBROLA 77
 Mean Opinion Score 60
 Mel 55
 Mel Scale 6
 Mel Frequency Cepstrum
 Coefficients 55
 MFCC 50
 MFCC FB-20 filter bank 35
 MIT Oxygen Project 77
 MLE 89
 MLLR 89
 MMIE 70
 Modified Rhyme Test 60
 MOS 60
 MRT 60
 Multilayer perceptron 71
 Multiplicative noise 36

N

NAP 89
 Nasal cavity 11
 NCE 65
 N-gram 19
 N-gram model 71
 NIST speech databases 96
 NIST SRE 89
 NLTK 100
 Noise 36

Normalized cross entropy 71
 Nose 11
 NTIMIT database 96
 Nuisance Attribute
 Projection 89
 Nyquist frequency 27

O

OOV word 71
 Oral cavity 12
 Oregon Institute 97
 Oscillogram 46
 Out Of Vocabulary word 71
 Overtone 19

P

Palate 12
 Pattern recognition 90
 PCA 90
 PCM 60
 Peak signal-to-noise ratio 37
 peak SNR 37
 Pentaphone 21
 Percent of correct words 72
 Perceptual audio coding 60
 Perceptual Linear Prediction 55
 Perplexity 72
 Pharynx 12
 Phase 46
 Phone 7, 20
 Phoneme 20
 Phonetics 20
 Phonogram 90
 Pitch 7
 PLDA 91

Plosive sound 20
 PLP 55
 POLYCOST 97
 PolyVar 97
 PRAAT 100
 Pre-emphasis 37
 Principal Component
 Analysis 90
 Probabilistic Linear
 Discriminant Analysis 91
 Pronunciation 20
 Pronunciation dictionary 69
 Prosodic features 20, 56
 Prosody 21
 PSNR 37
 Psychoacoustics 12
 Pulse Code Modulation 60

Q

Quantization 27
 Quantizing level 27
 Quinphone 21

R

RASTA 91
 Rectangular window 47
 Resampling 28
 Reverberation 7

S

SABLE 78
 SAMPA 21
 Sample 28
 Sampling 28
 Sampling frequency 28

Sampling rate 28
 Sampling resolution 29
 SAPI 78
 Score Normalization 91
 Senon 21
 Short-term spectral features 56
 Short-Time Fourier
 Transform 47
 Short-time magnitude 56
 Signal 8
 Signal smoothing 72
 Signal-to-Noise-Ratio 37
 Silent Speech Interfaces 72
 Siri 72
 SNR 37
 Soft palate 12
 Sound 8
 Sound intensity 8
 Sound spectrum 47
 Source-filter model 13
 Speaker Diarization 91
 Speaker Identification 92
 Speaker Model 92
 Speaker Recognition 92
 Speaker Recognition
 Evaluation 81
 Speaker Segmentation and
 clustering 92
 Speaker Verification 93
 Spectrogram 47
 Spectrum analysis 48
 Spectrum inversion method 37
 Spectrum method for pitch
 detection 48
 Spectrum reversion method 38

Speech Application
 Programming Interface 78
 Speech codec 61
 Speech frequency range 29
 Speech intelligibility 61
 Speech Interpretation and
 Recognition Interface 72
 Speech naturalness 61
 Speech signal 29
 Speech sounds 8
 Speech synthesis 78
 Speech Synthesis Markup
 Language 78
 Speech synthesizer 79
 Speech tempo 21
 Speech Understanding
 Research 73
 SpeechAnalyzer 101
 Spontaneous speech 73
 SSI 72
 SSML 71
 STFT 47
 Stress 21
 Subband Coding 61
 Supervector 93
 Support Vector Machines 93
 SUR 73
 SVM 93
 Syllable 22

T

TD-PSOLA 79
 Teeth 13
 Test normalization 94

Text Dependent Speaker
 Recognition 94
 Text Independent Speaker
 Recognition 94
 Text-To-Speech 80
 Timbre 22
 TIMIT 97
 T-norm 94
 Token 73
 Tokenization 73
 Tone 22
 Tongue 13
 Topic Detection 73
 Trachea 13
 Trigram 22
 Triphone 22
 TTS 80

U

UBM 94
 U-law 62
 Unigram 22
 Universal Background
 Model94
 Unvoiced sound 23
 Uvula 14

V

VAD 56
 Vector Quantization 95
 VXML 81
 VQ 95
 Vocal chords 14
 Vocal source 14
 Vocal source features 56

Vocal tract 15
Vocoder 61
Voice Activity Detector 56
Voice codec 61
Voice Extensible Markup
Language 81
Voiced sound 23
VoiceXML 81
Voicing 74
VoIP 62
Volume 8
Volume level 9

W

WACC 74
Wavelet Transform 49
WCNN 95
WER 74
White noise 38
Wiener filter 38
WinCeCil 101
Window function 49
Within-Class Covariance
Normalization 95
Word Accuracy 74
Word Error Rate 74
WordNet 74
WT 49

Y

YIN algorithm 50
YOHO 98

Z

Zero normalization 95

Zero-Crossing Rate 56
Z-norm 95
Z-transform 5

Texniki redaktorlar: Anar Səmidov

Zülfiyyə Hənifəyeva

Korrektor: Günay Muradova

Kompüter tərtibatı: Xətai Əlizadə

Çapa imzalanmışdır: 07.12.2015. Çap vərəqi 60x84, 1/16

Sifariş №72, sayı 100 nüsxə
