

NUTQNI TANIB OLISHNI OCHIQ MANBALI DASTURLARI TAHLILI

**t.f.d., prof. Mamatov Narzullo Solidjonovich,
Jalelov Quwanish Moyatdinovich**

“Toshkent irrigatsiya va qishloq xo’jaligini mexanizatsiyalash muhandislari”

Milliy tadqiqot universiteti

kjalelov007@gmail.com

Annotatsiya: Mazkur maqolada ochiq manbali nutqni tanib olish texnologiyalari tahlili keltirilib, unda ularning foydalanish imkoniyatlari baholangan va ushbu tadqiqot doirasidagi tahlillar nutqni aniqlash tizimlarining kelajagini shakllantirishda ochiq manbali tizimlar ta'siri yoritilgan.

Kalit so‘zlar: TTS texnologiyasi, nutq sintezi, tizim, model, dasturlash tili, matn.

Matndan nutqqa yoki nutq sintezi - bu so‘zlarni taniydigan va inson nutqini shakllantiradigan matndan sun’iy ravishda yaratilgan inson sifat tovushli nutqdir.

Dastlabki Text-to-Speech tizimi 1968 yilda Yaponiyadagi Elektrotexnika laboratoriyasida Noriko Umeda va boshqalar tomonidan dunyoga taqdim etilgan. 1961 yilda fizik Jon Larri Kelli va uning hamkasbi Lui Gerstman IBM 704 kompyuteridan nutqni sintezlash uchun foydalanishdi, bu Bell Laboratoriyalari tarixidagi eng mashhur voqeadir.

TTS ilovalari yoki Text-to-Speech ilovalari yozma matnni og‘zaki so‘zlarga aylantirish uchun nutq sintezidan foydalanadigan dasturiy ilovalardir. TTS ilovalari ko‘plab afzalliklarga ega [1]. Masalan, TTS ilovalari avtomobil haydash, mashq qilish yoki vizual e’tiborni talab qiladigan boshqa harakatlar bilan shug‘ullanayotganda yoki maqolalar, kitoblar va boshqa yozma materiallarni o‘qishda foydalanish mumkin. Bundan tashqari, TTS ilovalari yangi tilni o‘rganayotgan yoki talaffuzini yaxshilashga harakat qilayotganlar uchun juda foydali hisoblanadi. Ular matnni bir necha tilda o‘qish uchun moslashtirilishi mumkin, bu esa til o‘rganuvchilarga nutq qobiliyatlarini mashq qilishni osonlashtiradi. Shuningdek, foydalanuvchilar TTS ilovasining tezligini, ohangini va talaffuzini shaxsiy ehtiyojlariga moslashtirishi mumkin. TTS ilovalari asosan inson va mashina o‘rtasidagi aloqani simulyatsiya qilishda masalan, dasturlarni ovozli boshqarishda foydalaniladi. Operatsion tizimlarning katta qismi xususiy mahsulotlardir, ya’ni foydalanuvchi yoki potentsial ishlab chiquvchi ularning manba kodiga kirish huquqiga ega emas. Bu nutqni tanib olish tizimlarini ochiq kodli loyihalarga integratsiya qilish qobiliyatiga salbiy ta'sir ko'rsatadi. Ochiq manbali nutqni tanib olish tizimlarini ijobiy va salbiy tomonlarini tavsiflovchi markazlashtirilgan ma'lumotlar manbai ham mavjud emas. Natijada, muammoni hal qilishda nutqni optimal tanib olish tizimini tanlash muammosi yuzaga keladi [2].

Tadqiqotning maqsadi tijorat yoki tadqiqot faoliyati uchun tizimni tanlashda xarajatlarni kamaytirish uchun ochiq manbali nutqni tanib olish tizimlaridan foydalanish bo'yicha hamda matnni nutqga o'tkazishda tavsiyalarni ishlab chiqishdan iborat.

CMU Sphinx, HTK, iAtros, Julius, Kaldi va RWTH ASR tanlangan tizimlar tanib olishning aniqligi va tezligi, foydalanish qulayligi va ichki tuzilishi kabi ko'rsatkichlar bo'yicha taqqoslandi.

CMU Sphinx. Ushbu tizimning kuchli tomoni shundaki, bepul kirishda ushbu tillarning til va akustik modellari mavjud. Qo'llab-quvvatlanadigan tillar orasida standart ingliz tilidan tashqari, rus, qozoq va boshqa bir qator tillar ham mavjud. CMU Sphinx BSD litsenziyasi ostida tarqatiladi, bu uning tijorat loyihalariga integratsiyalashuviga imkon beradi. Ushbu tizim tijorat loyihalarida qo'llanilishi mumkin, chunki u Kaldi-ning ko'pgina afzalliklariga ega, garchi u tanib olish aniqligini biroz yomonroq ta'minlasa ham, shuningdek, ushbu tizim asosida uchinchi tomon ilovalarini yaratish uchun ishlatilishi mumkin bo'lgan API-ni taqdim etadi.

HTK. Bu tizim faqat ingliz tilini qo'llab-quvvatlaydi. Tizimning manba kodini tarqatish imkonini beruvchi HTK litsenziyasi ostida tarqatiladi.

iAtros. Tizmidan foydalanish qulayligi nuqtai nazaridan, iAtros tadqiqot davomida tekshirilgan barcha tizimlardan past. Ushbu tizim standart tillar ingliz va ispan tillarida qo'llab-quvvatlanadi. Ushbu tizimning kamchiligi faqat Linux oilasining operatsion tizimlarida ishlaydi. GPLv3 litsenziyasi ostida tarqatiladi, bu esa ushbu tizimni tijorat loyihalariga manba kodini oshkor qilmasdan integratsiya qilish imkonini bermaydi, bu o'z navbatida uni tijorat maqsadlarida foydalanish uchun yaroqsiz qiladi.

Julius. Bu tizim C tilida amalga oshirilgan. Tizim konsol interfeysi va uchinchi tomon ilovalariga integratsiya qilish uchun API taqdim etadi. Julius ingliz va yapon tillarini qo'llab-quvvatlaydi. BSD-ga o'xshash litsenziya ostida tarqatiladi. Julius tizimi ta'lim faoliyati uchun ham tavsiya etilishi mumkin, chunki u HTK ning barcha afzalliklariga ega, shuningdek, yapon tili kabi ekzotik tilni tanib olish imkoniyatini beradi.

Kaldi. Bu tizim akustik signal xususiyatlarining hajmini kamaytirish va shunga mos ravishda tizim ish faoliyatini oshirish uchun ko'plab algoritmlardan foydalanishga imkon beradi. Kaldi C++ dasturlash tilida yozilgan bo'lib, tizim tezligiga ijobiy ta'sir ko'rsatadi va modulli tuzilishga ega bo'lib, tizimni qayta tiklash, yangi funksiyalarni qo'shish va mavjud xatolarni tuzatishni osonlashtiradi. Kaldi faqat konsol interfeysini taqdim etadi, bu esa uchinchi tomon ilovalariga integratsiyani murakkablashtiradi. Ushbu tizim faqat ingliz tilini qo'llab-quvvatlaydi, butunlay bepul Apache litsenziyasi ostida tarqatiladi, ya'ni uning kodini oshkor qilmasdan tijorat mahsulotiga birlashtirilishi mumkin.

RWTH ASR. Ushbu tizim, iAtros kabi, akustik modellashtirish bosqichida Gauss aralashmasi modellaridan foydalanishi mumkin. O'ziga xos xususiyati - kirish signalining akustik xususiyatlarini olishda tovush xususiyatidan foydalanish qobiliyati. Bu tizim C++ tilida amalga oshirilgan va modulli arxitekturaga ega. RWTH ASR litsenziyasi ostida tarqatiladi, unga ko'ra tizim kodi faqat notijorat maqsadlarda foydalanish uchun taqdim etiladi, bu esa ushbu tizimni tijorat loyihalariga integratsiya qilish uchun yaroqsiz qiladi. Ushbu tizim tanib olishning aniqligi muhim bo'lgan, ammo vaqt muhim bo'lmagan muammolarni hal qilish uchun foydalanilishi mumkin. Shuni ham ta'kidlash joizki, litsenziya bilan belgilangan cheklovlar tufayli u har qanday tijorat faoliyati uchun mutlaqo yaroqsiz [4].

Aniqlik nuqtai nazaridan, tizimlar eng keng tarqalgan ko'rsatkichlar [3] yordamida taqqoslandi: so'zni aniqlash tezligi (WRR), so'z xatosi darajasi (WER), ular quyidagi formulalar yordamida hisoblanadi:

$$WER=(S+I+D)/T, \quad WRR=1-WER$$

bu erda S - so'zlarni almashtirish operatsiyalari soni, I - so'zlarni kiritish bo'yicha operatsiyalar soni, D - asl iborani olish uchun tan olingan iboradan so'zlarni olib tashlash bo'yicha operatsiyalar soni va T - asl nusxadagi so'zlar soni ibora va foiz sifatida o'lchanadi.

Tanib olish tezligi nuqtai nazaridan, taqqoslash Real Time Factor yordamida amalga oshirildi - tanib olish vaqtining tan olingan signalning davomiyligiga nisbati ko'rsatkichi, shuningdek Speed Factor (SF) deb nomlanadi. Ushbu ko'rsatkichni quyidagi formula yordamida hisoblash mumkin:

$$SF=T_{sav}/T$$

bu erda T_{sav} - signalni aniqlash vaqti, T - uning davomiyligi va real vaqtning fraktsiyalarida o'lchanadi.

Tahlil natijalari quyidagi jadvalda keltirilgan.

1-jadval

TTS tizimlari tahlili

Tizim	WER%	WRR	SF
CMU Sphinx (pocketsphinx/sphinx4)	21,4/22,7	78,6/77,3	0,5/1
HTK	19,8	80,2	1,4
iAtros	16,1	83,9	2,1
Julius	23,1	76,9	1,3
Kaldi	6,5	93,5	0,6
RWTH ASR	15,5	84,5	3,8

Mazkur maqolada TTS tizimlarini tahlil qilish natijasida quyidagi natijalar xulosa sifatida qabul qilindi:

TTS ilovalari juda ko'p afzalliklarni taqdim etadi, jumladan, matnni bir nechta tillarda o'qish qobiliyati va og'zaki matnning tezligi, ohangi va talaffuzini sozlash qobiliyati;

TTS ilovalari ko'rishda nuqsoni yoki o'qishda qiynaladigan odamlar, til o'rganuvchilar uchun foydalidir.

Foydalanilgan adabiyotlar ro'yxati:

1. Jha M. Improved unsupervised speech recognition system using MLLR speaker adaptation and confidence measurement / M. Jha et al. // V Jornadas en Tecnologias del Habla (VJTH'2008) – 2008. – P. 255-258.

2. Sundermeyer M. The rwth 2010 quaero asr evaluation system for english, french, and german / M. Sundermeyer et al. // Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP) – 2011. – P. 2212-2215.

3. Гусев М.Н. Система распознавания речи: основные модели и алгоритмы / М.Н. Гусев, В.М. Дегтярев. – СПб.: Знак, 2013. – 128 с.

4. Карпов А.А. Методология оценивания работы систем автоматического распознавания речи / А.А. Карпов, И.С. Кипяткова // Известия высших учебных заведений. Приборостроение. – 2012. – Т. 55. – №. 11. – С. 38-43.

TASVIR SHOVQINLARI TAHLILI

t.f.d., prof. Mamatov Narzullo Solidjonovich

“Toshkent irrigatsiya va qishloq xo’jaligini mexanizatsiyalash muhandislari”

Milliy tadqiqot universiteti

Jalelova Malika Moyatdin qizi

Berdaq nomidagi Qoraqalpoq davlat universiteti

jalelova97@mail.ru

Annotatsiya: Hozirgi vaqtda turli xil tasvirga olish qurilmalaridan olingan raqamli tasvirlar soni keskin oshib bormoqda. Ayrim hollarda ushbu tasvirlar tasvirga olish, qayta ishlash, siqish, saqlash, uzatish va ko’paytirish jarayonida turli shovqinlardan ta’sirlanib, sifatsiz holatga kelishi mumkin. Shovqin tasvirdagi ma’lumotlarning sifatini buzadigan tasodifiy o’zgarishlarni ifodalaydi va tasvirdan muhim ma’lumotlarni olishga to’sqinlik qiladi. Bu o’z navbatida, tasvirda uchraydigan asosiy shovqin turlarini o’rganish tasvirlash sohasida dolzarb ekanligini ko’rsatadi. Mazkur tadqiqot ishi Gauss, tuz va qalampir, uniform, Puasson, spekl, rayleigh va risian kabi shovqin turlarini o’rganishga bag’ishlangan.

Kalit so’zlar: tasvir, shovqin, Gauss, tuz va qalampir, Puasson, spekl, uniform, piksel, ehtimollik zichligi.

Tasvirlash sohasida raqamli tasvirlarning yuqori sifatli vizual ko’rinishga ega bo’lishi o’ta muhim. Biroq, raqamli kameralar, skanerlar yoki boshqa tasvirga olish uskunalari tomonidan olingan tasvirlar har doim ham yuqori sifatga ega bo’lmasligi mumkin. Tasvir sifatini sezilarli darajada pasaytiruvchi keng tarqalgan muammolardan biri bu shovqindir.

Shovqin- tasvir tafsilotlarini yashiradigan piksel qiymatlarining istalmagan va tasodifiy o’zgarishlarini ifodalaydi. Ushbu o’zgarishlar uskunadagi sensor cheklovlari, atrof-muhit sharoitlari yoki tasvirlash qurilmasining elektron komponentlari kabi ko’plab omillar sabab yuzaga kelishi mumkin. Shuning uchun mazkur tadqiqot ishida tasvirda uchraydigan asosiy shovqin turlarini o’rganish maqsad qilingan.

Shovqin turlari

Gauss shovqini deyarli barcha turdagi raqamli tasvirlarda uchraydigan va butun tasvir bo’ylab taqsimlanadigan shovqin turi hisoblanadi:

$$p_G(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

bu yerda $p_G(z)$ – ma’lum bir shovqin qiymati Z paydo bo’lish ehtimolini tavsiflaydi, z – intensivlik sifatida berilgan, μ – z ni o’rtacha qiymati va σ – standart og’ish.