УДК 004.95

НУТҚНИ АВТОМАТИК ТАНИБ ОЛИШ МУАММОЛАРИ ВА ЕЧИМЛАРИ

Маматов Н.С., Абдуллаев Ш.Ш., Юлдошев Ю.Ш.

Сўнгги йилларда нутқни автоматик таниб олишда сезиларли ривожланишлар кузатилиб, кўплар ютуқларга эришмоқда. Бироқ, нутқни автоматик таниб олишнинг мавжуд воситалари инсон имкониятларига нисбатан анча чегаралангандир. Масалан, инсон узоқдаги нутқни, сифатсиз узатилган ёки шовқинга эга ҳамда акцентли нутқни қийинчиликсиз таниб олади. Бундан ташқари инсон кўплаб овозлар орасидан маълум суҳандон ва унинг нутқини ажратиб олиш ва спонтан нутқни таниб олиш имкониятига ҳам эга. Инсоннинг бундай имкониятларидан айниқса, сўнглари нутқни таниб олишнинг замонавий тизимларини яратишда кўплаб муаммоларни келтириб чиқаради. Нутқни автоматик таниб олишнинг мавжуд тизимлари эса фақатгина ажратилган буйруқлар ёки сонларни таниб олишда инсондан устун бўлиб келмоқда.

Нутқни таниб олишнинг мавжуд тизимлари нейрон тармоқлар структураларини мукаммаллаштириш, улардаги турли даражадали тескари алоқаларни таъминлаш ва ўқитишни янги усулларини ишлаб чиқиш билан боғлиқ бўлиб, унда нутқ таркибининг маъноли қисмини қолдирувчи семантика соҳасидаги ишланмаларни қўллаш ҳамда уларни имкониятларидан фойдаланиш мақсадга мувофиқ бўлади.

Мазкур мақолада нутқни автоматик таниб олишниҳозирги кунда кенг қўлланилаётган иловаларининг қисқача тавсифлари, ўзига хос хусусиятлари, ривожланишининг асосий босқичлари келтирилган бўлиб, унда ушбу масалани ечиш учун Марков занжирлари асосида нутқни таниб олиш усули кўриб чиқилган. Бундан ташқари, нутқни таниб олиш масаласини ечишга қаратилган ёндашувларнинг қиёсий таҳлили ҳамда коартикуляция муаммосини ечишда контекстга боғлиқ трифонлар ва бифонларни моделлаштириш йўлларниҳамдамоделларни адаптация қилиш, белгиларни нормаллаштиришнинг суҳандонга боғлиқ бўлмаган, алоқа канали ва аддитив ҳалақитларга инвариантлик таъминланган тизимларни ишлаб чиқишдаги роли кўрсатилган. Нутқни автоматик таниб олиш самарадор тизимларини ишлаб чиқиш усули сифатида чуқур нейрон тармоқлар ва рекуррент нейрон тармоқлари келтирилган. Кўпқатламли нейрон тармоқларнинг биологик тизимлар билан ўхшашлиги асослаб берилган. Хулосада нутқни таниб олишнинг замонавий тизимлари муаммо ва камчиликлари баён этилиб, ривожлантириш учун тавсиялар келтирилган.

***Таянч иборалар*:** Нутқни автоматик таниб олиш, динамик дастурлаш, Марков модели, моделлар адаптацияси, белгиларни нормаллаштириш, ҳолатларни боғлаш, чуқур нейрон тармоқлар, рекуррент нейрон тармоқлар.

В последние годы наблюдается существенное развитие в автоматическом распознавании речи и достигаются больших успехов. Но возможности существующих средств автоматического распознавания речи намного ограничены чем человеческие возможности. Например, человек без никаких затруднений легко распознаёт удалённую, некачественную или шумную, а также акцентную речь. Кроме того, человек имеет возможность распознавания спонтанной речи и определённого голоса диктора среди множества голосов. Такие человеческие возможности, особенно последние, создают много проблем при создании современных систем распознавания речи. А существующие системы автоматического распознавания речи являются преимущественным чем человек при распознавании выделенных команд и чисел.

Существующие системы распознавания речи связаны с усовершенствованием нейро-сетевых структур, обеспечением различных обратных связей и разработкой новых методов обучения, где целесообразно применение разработок из области семантики, в которых можно выделить смысловую часть состава речи и использование их возможностей.

В статье приведены короткий обзор в настоящее время широко применяемых приложений автоматического распознавания речи, их особенности, основные этапы их развития. А также рассмотрен метод решения задачи распознавания речи на основе Марковских цепей, сравнительный анализ подходов и методов моделирования контексосвязанных трифонов и бифонов при решении проблемы коартикуляция и адаптация моделей, показаны роли разработки систем, обеспечивающих инвариантность аддитивных шумов и каналов связи, нормализация признаков дикторонезависимых систем. Для разработки эффективной системы автоматического распознавания речи предложен метод глубоких и рекуррентных нейронных сетей. Отмечено сходство глубоких (многослойных) нейронных сетей с биологическими системами. В заключении описаны проблемы и недостатки современных систем распознавания речи и дано предложение их развития.

***Ключевые слова*:** автоматическое распознавание речи, метод динамического программирования, марковская модель, адаптация моделей, нормализация признаков, связывание состояний, глубокие нейронные сети, рекуррентные нейронные сети.

In recent years, there has been a significant development in automatic speech recognition and great success has been achieved. But the capabilities of existing automatic speech recognition tools are much more limited than human capabilities. For example, a person without any difficulties easily recognizes remote, substandard or noisy, as well as accent speech. In addition, a person has the ability to recognize spontaneous speech and a certain voice announcer among many voices. Such human capabilities, especially the latter, create many problems when creating modern speech recognition systems. And existing systems of automatic speech recognition are predominant than a person in recognizing selected commands and numbers.

Existing speech recognition systems are associated with the improvement of neural network structures, the provision of various feedbacks and the development of new teaching methods, where it is advisable to use developments from the field of semantics in which the semantic part of speech composition and the use of their capabilities can be distinguished.

The article provides a brief overview of currently widely used automatic speech recognition applications, their features, the main stages of their development. We also considered a method for solving the speech recognition problem based on Markov chains, a comparative analysis of approaches and methods for modeling context-related triphons and biphons in solving the problem of co-articulation and adaptation of models; For the development of an effective automatic speech recognition system, a method of deep and recurrent neural networks has been proposed. The similarity of deep (multilayer) neural networks with biological systems is noted. In conclusion, the problems and shortcomings of modern speech recognition systems are described and a proposal for their development is given.

***Keywords*:** automatic speech recognition, dynamic programming method, Markov model, model adaptation, feature normalization, linking states, deep neural networks, recurrent neural networks.

I. КИРИШ

Айни пайтда нутқни автоматик таниб олиш сунъий интеллект соҳасидаги жадал ривожланаётган йўналишлардан бири ҳисобланади. Ҳозирги кунда ушбу йўналишга оид кўплаб тизимлар ишлаб чиқилган. Бундай тизимларга ногиронларга ёрдам бериш, кириш-чиқишни назорат қилиш, турли техник воситаларни бошқариш каби тизимларини мисол қилиб келтириш мумкин. Нутқ билан ишловчи тизимларнинг ўзига хос хусусияти фойдаланувчининг тизимдан овози орқали фойдаланиш имконияти бўлиб, бунда бошқа бир шахснинг ёрдами талаб этилмайди.

Нутқни таниб олишнингдастлабки тизимларидасаноқли буйруқларни таниб олиш имкони бўлган. Замонавий тизимлар эса катта ҳажмдаги

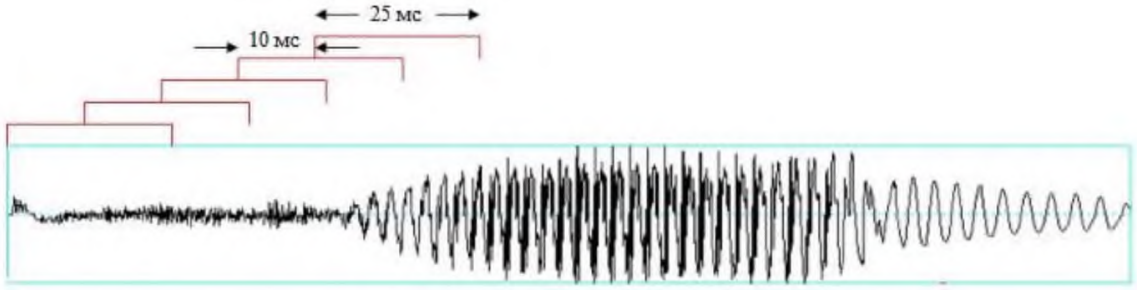
буйруқларни таниб олиш имкониятига эга. Масалан, суст эшитувчилар учун мўлжалланганреал вақтда субтитрларни ҳосил қилишнинг автоматик тизимичекланган луғатдан фойдаланилади [1]. Яна бир чекланган луғатдан ва структуралашган гаплардан фойдаланувчи тизим АҚШнинг кўплаб тиббиёт марказларидашифокорлар фаолиятини ҳужжатлаштиришда қўлланилмоқда [2]. Чекланган луғат ва структураланган гаплардан фойдаланувчи тизимларни яратиш нисбатан осон ҳисобланади. Ҳозирда кўплаб йирик компаниялар ўзларининг нутқни автоматик таниб олиш тизимларига эга. Масалан, Яндекс компанияси томонидан яратилган «Алиса» тизими виртуал овозли ёрдамчи ҳисобланади. Мазкур тизим оғзаки ва ёзма табиий нутқни таниш ҳамда саволларга жавоб бериш орқали жонли суҳбатни интерпритация қила олади. Ушбу тизим берилган саволларга жавобларни экранга чиқариш ва уларни эшиттириб ўқиш орқали тақдим этади. “Алиса” тизимининингинтерфейси рус тилида бўлиб, у биринчи марта 2017-йил 10-октябрда ишга туширилган. Тизим C++ дастурлаш тилида яратилган бўлиб, Android, iOS, Windows операцион тизимларида ишлайди. Айни пайтда тизимни мукаммаллаштириш ишлари олиб борилмоқда. Ушбу тизим Яндекс компаниясинингSpeechKit платформаси ёрдамида нутқни таниб олади ва синтез қилади. Нутқни таниб олишнинг яна бир замонавий тизими бу Google Assistantдир. У Google компанияси томонидан ишлаб чиқилган ақлли шахсий ассистентдир. Ушбу тизимининг интерфейси кўп тилли бўлиб, у биринчи марта 2016 йилда Google I/O презентациясида намойиш этилган. У ҳам “Алиса” тизими каби C++ дастурлаш тилида яратилган бўлиб, Android, iOS ва Wear OSақлли соатлар учун мўлжалланган операцион тизимларида ишлайди. Ҳозирги кунда кенг қўлланиладиган нутқни таниб олиш тизимларидан бири бу Siri(Speech Interpretation and Recognition Interface) тизимидир. У булутли шахсий ёрдамчи ва савол-жавобли тизим бўлиб, Apple компаниясининг iOS, watchOS, macOS, ва tvOS операцион тизимлари таркибига киритилган. Дастлаб Siri App Store да iOS тизими учун илова сифатида ишлаб ишлаб чиқарилган. Ҳозирда ушбу илова Apple компанияси операцион тизимлари учун таркибий қисм сифатида қўлланилмоқда. Ушбу тизим Objective-C тилида яратилган бўлиб, у 21 хил тил ва унинг турли шакллари билан ишлай олади. Ҳозирги кундакенг қўлланилаётган тизимлар сифатида Amazon компаниясининг ALEXA тизимини, IBM, Microsoft компанияларинг ҳамда Хитойнинг Baidu компаниясининг нутқни автоматик таниб олиш тизимларини келтириш мумкин.

Юқорида санаб ўтилган тизимлар нутқни ўзига хос хусусиятлари асосида яратилган бўлиб, қуйида ушбу хусусиятлар баён этилади.

II. АСОСИЙ ҚИСМ

Нутқни автоматик таниб олишнинг ўзига хос хусусиятлари. Акустик сигналнинг таянч белгиларини шакллантиришда нисбатан юкори даражадаги белгиларни куриш учун Фурье спектридан асос сифатида фойдаланилади. Бунинг сабаби инсон эшитиш тизимининг асосий элементларидан ҳисобланган қулоқ ичидани чиғаноксимон орган тузилиши ва эшитиш тизимининг тонотопик ташкил этилиши ҳисобланади. Тонотопок ташкил этилиш деганда акустик сигналнинг спектрал компонентларининг эшитиш органи бўйлаб миянинг эшитиш бурмаларигача таркалиши назарда тутилади. Бунда бир-бирига қўшни бўлган частоталар қўшни нейрон каналлари бўйлаб таркалади.

Кўпинча спектни олиш учун “ойнали” таҳлилқўлланилади. Ойнали таҳлил 15-25 мс вақт оралиғидаги силлиқловчи ойнада ёки шу вақт оралиғида рекурсив фильтрларни қўллаш орқали амалга оширилади. Ойна узунлигини 15-25 мс олиниши нутк сигналининг ўзгариш хусусиятлариёки нутқнинг 8-12 Гц оралиғида жойлашувчи силлабик частотаси билан боғлиқ. Одатда таҳлил килиш ойнаси 10 мс кадам билан 100 Гцда вектор-белгиларни олиш частотасини таъминлаш оркали силжитилади (1-расм). Аксарият тизимлар 16кГцли дискретлаш частотасидаги 400 та қийматдан фойдаланади.



1-расм. Нутк сигналида ойналарни жойлаштириш

Нутк сигналини таниб олишда кенг қўлланиладиган белгилардан бири бу мел-частотали кепстрал коэффициент (mel-frequency cepstral coefficients, MFCC)дир [4]. Бунда ҳар бир канаб бўйича энергия қиймати логарифмланади. Мел шкаласи психоакустик тажрибалардан олинган псевдологарифмик частоталар шкаласини ўзида акс эттиради. Унинг аҳимияти инсон эшитиш тизимининг ишлашини тасаввур килишга ёрдам бериш билан бирга белгилар вектори ўлчамини сезиларли кискартиришдан иборат. Логарифмлаш (1) формула оркали амалга оширилади ва у сигналларни таркалиши учун характерли бўлган амплитудавий сикилишни моделлаштиради. Бундан ташкари тескари косинус алмаштириши оркали кепстр ўтади ва факат биринчи 12 та компонентани колдиради:

$$M(f) = 1125\ln(1 + f/700) \quad (1)$$

бу ердаf − частота қиймати.

Мел-частотали кепстрал коэффициентларни аниқлаш бир неча босқичларда амалга оширилади (2-расм).

2-расм. MFCC алгоритми қадамлари.

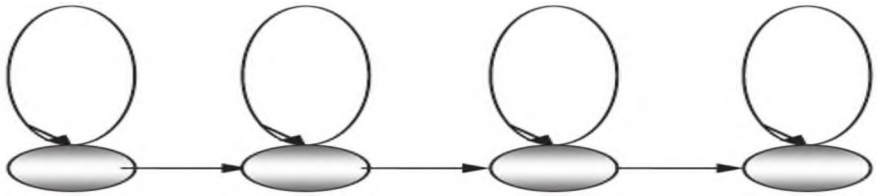
Спектрал таҳлилда 15-25 мс узунликдаги ойнадан олинган MFCC белгилар физиологик ҳисобланмайди. Чунки инсон эшитиш тизими сезиларли узун бўлакларни таниб олишга мослашган бўлади. Лекин нисбатан юқори даражадаги нейронлар сигналнинг ушбу кичик бўлакларни умумлаштириб фонема даражасидаги элементларни ажратишни амалга оширади. Қуйида ушбу муаммони нутқни таниб олишнингзамонавий тизимларида ҳал килинишини кўриб чикилади.

Яширин марков модели ёрдамида нутқни таниб олиш

Яширин марков модели динамик дастурлаш усулидаги камчиликларни бартараф этди. Бунинг учунҳолатларни баён этувчи луғатдаги сўзларни фрагментлар орқали ифодалаш ва белгилар фазосида эҳтимоллик зичлик функциялари ёрдамида ҳолатларни баён этиш киритилди.

Марков занжирлари учун ўкитиш ва таниб олишнинг математик аппаратнитурли тасодифий жараёнларда ўтган асрнинг 60-йиллар охирида ривожланди ва 70-йиллар бошида эса нутк сигнали учун қўлланилди [5]. Марков занжирини нутк сигнали учун қўллаганда занжир дискрет вактдаги ҳолатлараро бир томонлама ўтишли жараён сифатида каралиб, кейинги ҳолатга ўтиш эҳтимолиги фақат жорий ҳолатга боғлиқ бўлади ва олдинги вақт моментларида жараён қандай ҳолатларда бўлганига боғлиқ бўлмайди (3-расм) [6].

Бу талаб ҳолатларни яшаш вақтининг нотўғри гистограммасига олиб келади ва шунинг учун ундан замонавий тизимлар воз кечди. Бу каби кейинги ҳолатга ўтиш эҳтимоли жорий ҳолатга боғлиқ бўлган моделларни бир жинсли бўлмаган марков моделлари ёки ярим марков моделлари деб аталади.



3-расм. Тўрт ҳолатли марков занжири.

Шундай килиб, марков модели маълум бир товуш ёки сўз учун бир ёки бир нечта ўтишлар эҳтимолликлари ва белгилар фазосида эҳтимолли зичлик функциялари аникланган кетма-кет ҳолатларни ўзида акс эттиради. Эҳтимоллик зичлик функциялари дискрет ёки узлуксиз ҳолатда берилиши мумкин. Узлуксиз ҳолатда улар одатда диагонал ковариация матрицали гаусс функциялари йиғиндиси орқали аппроксимация қилинади.

Ковариация матрицасининг диагоналлиги эса ўкитиш параметрлари сонини камайтиришга ва баъзи алгоритмларни соддалаштиришга ва баъзи муаммоларни аналитик ечишгаимкон беради. Масалан, ковариацион матрица тўлиқ бўлганда фақат сонли усуллардан фойдаланишга тўғри келади.

Ўқитиш нутқли маълумотлар омбори (ёзиб олинган нутқ сигналлари ва уларга мос келувчи матнлар) асосида амалга оширилади. Омборда тажрибали лингвист мутахассислар томонидан маълумотларнинг маълум кисми бўлакларга ва фонемаларга ажратилганбўлади ва улардан марков моделини куришда фойдаланилади. Дастур ҳар бир фонема учун белгилар векторларини алоҳида тўпламларга жамлайди. Натижада гаусс функциялари тўплами аппроксиямацияси оркали эҳтимоллик зичлик функцияларини қуриш қийинчилик туғдирмайди. Бундан ташқари дастур бўлаклар давомийлигини таҳлил килиб гистограммаларини куради. Ҳар бир фонема учун давомийлик гистограммаларини билган ҳолда жорий ҳолатда маълум вакт бўлгандан кейин берилган фонемага мос ҳолатдан чикиш эҳтимоллиги ҳисобланади.

Ҳолатлар параметрлари дастлабки баҳолашдан ўтиб, ҳолатлар занжирида маълумотлар омборидаги белгилар векторлари кетма-кетликларни ҳосил бўлиш эҳтимолликларини максималлаштириш орқали параметрларни кайта баҳолаш учун Баума-Уэлш [6] ёки Витерби [6] алгоритмлари қўлланилади.

Навбатдагибоскич нутқ маълумотлар омборини бўлакларга ажратилмаган кисмидан фойдаланади. Ҳосил килинган ҳолатлар кетма-кетлиги ўзига олдиндан маълум бўлаклар учун яхши натижа беради, лекин бошка нуткни таниб олиш учун яхши натижа бермайди. Шунинг учун бу кадамда улкан нутқ базаси ёрдамида “мажбурий текислаш” усулидан фойдаланилади. Бироқ бу каби ёндашув ҳам динамик дастурлашга нисбатан таниб олиш сифатини сезиларли яхшилашга имкон бермайди [7].

Қаралаётган фонеманинг айтилишига кўшни фонемалар таъсир

кўрсатади, бу эффект “коартикуляция” эффекти деб аталади. Бундан ташқари, “коартикуляция” эффекти ва бир фонема учун турлича белгилар векторларининг шакллантирилиши ҳам натижани яхшилашга йўл қўймайди. Шундай қилиб, фонемани аниқроқ ифодалашда ўзидан олдинги ва кейинги товушлар билан бирикмаларини алоҳида акустик объект сифатида қараш ва мос ҳолатлар кетма-кетлигини қуриш талаб этилади. Бундай объектлар “трифон” деб аталади ва улар учта кетма-кет фонемаларни ўзаро боғлайди. “Бифонлар” эса фонеманинг ўзидан олдинги ёки кейинги фонема билан бирикмасини ифодалайди. Бифонлар асосан нутқ бўлагининг боши ёки охирини ифодалашда ёки трифонлар ҳолатларини қуриш учун маълумот етарли бўлмаганда қўлланилади. Контекстни инобатга олмайдиган фонемаларни монофонлар деб аталади. Одатда трифонларни ифодалашда уч ҳолатдан фойдаланилади. Четки ҳолатлар фонеманинг қўшни фонемалар таъсирига учраган қисми ва марказда энг кам таъсирга учраган қисми жойлашган бўлади. Ҳолатлар сони контекстга боғлиқлик даражасига тенг миқдорда бўлиши шарт эмас. Масалан, “пентафонлар”ни учта ёки монофонларни бир нечта ҳолат орқали моделлаштириш мумкин [8].

Трифонлар учун ҳолатларни қуриш зарурияти фонетик бирликлар сонининг кескин ошишига олиб келади ва жуда катта маълумотлар омбори ҳам уларнинг статистикасини баҳолашга етарли ҳисобланмайди. Масалан, Wall Street Journal Pronunciation Lexicon маълумотлар базаси учун параметрлар сони жуда кўпайиб кетиши [9] ишда кўрсатиб берилган. Мазкур муаммо ҳолатларни боғлаш усулиорқали ҳал этилади [9]. Бунда эҳтимолли зичлик функцияси нисбатан кучли қамраб олган ҳолатлар бирлаштирилади ёки ўзаро боғланади. Жараён қуйидан, яъни морофондан бошлаб амалга оширилади. Бунда дастлаб монофонларни эҳтимолли зичлик функцияси энг кам қоплагантрифонларга ажратишдан бошланади ва янги трифонларни ўқитиш учун маълумотлар етарли бўлмаганда жараён тўхтатилади. Шу орқали зичлик функцияси морофонларни кам кесишувчи ва ўқитилиши қулай бўлган йирик трифонларга ажратади.

Ушбу усул орқали ишлаб чиқилган таниб олиш тизимлари динамик дастурлаш усулига асосланган тизимларга нисбатан самарадор эканлигини амалда кўрсатди, аммо суҳандон ёки алоқа канали алмашганда таниб олиш сифати кескин тушиб кетади. Шунинг учун тизимни янги сухандонга адаптациялаш масаласини ҳал қилишга йўналтирилган кўплаб ишлар амалга оширилди ва белгиларни нормаллаштириш ҳамда моделни адаптация қилиш ёндашувлари ишлаб чиқилди. Маълумотлар омборини ташкил этувчи векторларни ўртача характеристикаларига яқинлаштиш мақсадида кирувчи нутқ сигнали ёки унинг белгилар векторларини созлаш белгиларни нормаллаштириш деб аталади. Шу мақсадда ўртача кепстрни айириш ва овоз тракти узунлиги бўйича нормаллаштириш амалга оширилади [10].

Адаптациялашда Байес адаптацияси [11] ёки апестериор эҳтимолликни максималлаштириш ва ўхшашликни максималлаштирувчи чизиқли регрессия қўлланилади [10]. Бундан ташқари адаптация хос суҳандонлардан фойдаланиш орқали ҳам ҳал қилиниши мумкин [12]. Ҳалақитли шароитларга моделни адаптациялашда эса Тейлор вектор қаторлари қўлланилади [3].

Яширин марков модели ёрдамидатуташ нутқни таниб олиш учун қўшимча сифатида тил моделиданфойдаланилади. Тил модели нутқни автоматик таниб олиш сифатини оширишга хизмат қилади ва сўзларни эҳтимолий кетма-кетлигини аниқлаш орқали таниб олишда нисбатан юқори аниқликка эришилади. Тил модели мумкин бўлган сўзлар кетма-кетлигининг энг юқори ва энг паст эҳтимолликка эга эканлигини аниқлаш имконини беради. Ўзбек тилидаги нутқни таниб олишда тил модели нисбатан пастроқ самара бериши мумкин. Чунки сўзларни тасодифий кетма-кетликда келиши, сўзларнинг турли шакллари мавжудлиги, синтактик жиҳатлари ҳамда гап охирида овоз баландлигининг пасайиши каби омиллар таниб олиш самарадорлигига салбий таъсир кўрсатади. Шунга қарамай катта маълумотлар базаси асосида қурилган тил модели нутқ таниб олиш аниқлигини ошириш имконини беради.

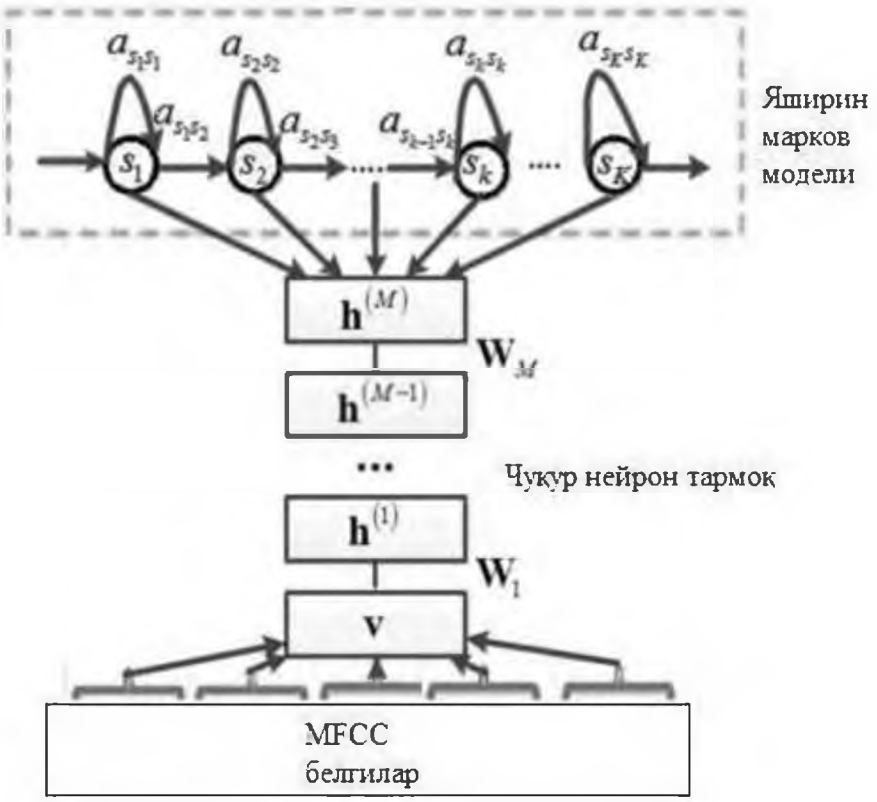
Чуқур нейрон тармоқлар.

Нутқни таниб олиш технологиялари янги технология бўлишига қарамай тижорий нуқтаи назардан жуда истиқболли ҳисобланади. Динамик дастурлаш усули ёки марков модели асосида қурилган нутқни таниб олиш тизимлари кутилган натижаларни таъминламагани учун оммавий тижорий иловалар яратишга имкон бермади. Бироқ, улар тор предмет соҳаларда маълум ютуқларга эришди.

Кўплаб тадқиқотчилар сунъий нейрон тармоқлари нутқни таниб олиш масаласи характерига мос деб ҳисоблашади. Сунъий нейрон тармоқлардан фойдаланишга уринишлар етарлича эрта бошланган бўлиб, кўплаб истиқболли ғоялар таклиф этилган [13]. Мисол сифатида 9 та кетма-кет келган мел-спектр векторларидан ташкил топган супервектор шаклидаги кўп муддатли белгилардан фойдаланиш, чиқиш ва кириш қатламлар орасидаги рекуррент боғланишлар ёрдамида контекстга боғлиқликни инобатга олиш ғояларини келтириш мумкин. Таклиф этилган ғоялар асосида яратилган тизим натижалари гаусс аралашмаси асосида яратилган тизим натижаларидан кўра пастроқ бўлди [14]. Бунинг асосий сабабларидан бири сифатида ўша даврдаги компьютерларнинг қуввати бир нечта қатламли ва чиқишларида мавжуд трифонлар учун бир неча минг нейронлари бўлган нейрон тармоқларни жуда катта маълумотлар базаси билан ўқитиш учун етарли бўлмаганлиги келтириш мумкин. Нейрон тармоқ ёки ихтиёрий сондаги яширин қатламли перцептрон универсал аппроксиматор ҳисобланади [14], яъни ҳатто битта яширин қатламли нейрон тармоқ ҳам белгилар фазосидаги ҳар қандай текисликни аппроксимация қилиш имконини беради. Кўп

қатламли нейрон тармоқлардан фойдаланиш ва тармоқни ўқитишда хатоларни тескари тарқатиш алгоритмидан фойдаланиш нутқни таниб олишда юкори натижалар бера бошлаганини алохида таъкидлаб ўтиш жоиз [15,16]. Ушбу ёндашув классик ёндашувларга нисбатан яккол афзалликларни кўрсатди. Масалан, 309 соатли нутк асосида ўкитилган кўп катламли нейрон тармок 2000 соатли нутк асосида ўкитилган гаусс аралашмали тизимдан яхширок натижа бериши кузатилди.

Кўп қатламли нейрон тармоғи бир нечта тилдаги нутқларни таниб олиш имкониятига эга бўлиб, бир тил учун ўкитилган нейрон тармок фақат кирувчи катламни бошка тилларнинг нутқли маълумотлари билан ўкитиш орқали бир нечта тилдаги нутқни таниб олиш вазифасини ҳам бажара олиши тажрибаларда аникланди [16]. Х.Германский нейрон тармоқлар чикишидаги трифонлар апостериор эҳтимолликларини марков модели киришидаги белгилар вектори сифатида фойдаланиб ўз натижаларини эълон килди [17]. Ушбу ишда гибрид модел таклиф этилган бўлиб, унинг кўриниши 4-расмда келтирилган бўлиб, бунда $a_{s_1,s_1},\dots,a_{s_k,s_k}$ - ҳолатлараро ўтиш матрицаси, $s_1,\dots,s_k$ - ҳолатлар матрицаси матрицаси яширин марков моделининг асосий параметрлари, $h^{(1)},\dots,h^{(M)}$ – нейрон тўри қатламлари, $W_1,\dots,W_M$ – вазн коэффициентлари матрицаси.



4-расм. Чукур нейрон тармок ва яширин марков модели гибриди.

Кўплаб мутахассислар рекуррент нейрон тармоқлар ёрдамида фонемаларни аниқлаш мумкин деб ҳисоблашади. Рекуррент нейрон тармоқлар айланма жараёнга йўналган нейронларни умумлаштиради. Бу нейрон тармоғини хотира билан таъминлайди ва шу орқали нейрон тармоқ статик объектларни таниб олишдан ташқари жараёнларни ҳам таъниб олиш имконига эга бўлади. Рекуррент нейрон тармоқларининг ушбу хоссаси чуқур нейрон тармоқларга нисбатан афзаллигини кўрсатади. Бу каби тармоқлар фонема ёки бошқа акустик объектни аниқлаш масаласини ҳал қилишда динамик дастурлаш усули ёки марков модели ёндашувларидан воз кечишга имкон беради. Рекуррент нейрон тармоқлар бўйича тадқиқотлар етарлича олдин бошланган [18], лекин компьтерларнинг қувват етишмовчилиги бу каби нейрон тармоқларнинг ўша пайтдаги бошқа усулларга нисбатан устун томонлари кўрсатиш имконини бермаган.

Рекуррент нейрон тармоқларининг яна бир афзаллиги кичик ўлчамли векторлар билан ишлай олиши ҳисобланади. Контекстга боғлиқлик, яъни фонемаларнинг бир-бирига таъсири чуқур нейрон тармоқларда жуда катта ўлчамдаги супер-векторлар билан ишлашни талаб этади. Масалан, 25 мс ли бўлакларда 10 мс қадам билан олинган 300 мс лик сигнални белгилар вектори кўринишида ифодалаш учун тахминан 30 та белгилар вектори зарур бўлади. Шундай қилиб, натижавий супер-вектор ўлчами 300 дан 1000 гача бўлиши мумкин. Бундай ўлчамли векторлар билан ишлаш ноқулайликлар келтириб чиқаради ва бу ноқулайликни олдини олишда рекуррент алоқали нейрон тармоқлардан фойдаланиш қўл келиши олинган натижаларда ўз аксини топди [19].

Нутқни автоматик таниб олиш сифатни баҳолаш.

Нутқни автоматик таниб олиш тизимлари сифатни баҳолаш тўғри таниб олинган сўзлар фоизи (WRR — Word Recognition Rate) ёки нотўғри таниб олинган сўзлар фоизи (WER — Word Error Rate) билан баҳоланади. Баъзи ҳолларда тизимни баҳолашда гап ёки жумлани таниб олишдаги хатолик кўрсатгичидан ҳам фойдаланилади (SER — Sentence Error Rate). Мазкур баҳолаш диалогли тизимларда муҳим аҳамият касб этади, чунки айтилган сўзлар кетма-кетлигини таниб олиш масаласидан фарқли равишда гипотезани тўғрилаш имкони мавжуд эмас. Сўнгги вақтларда нутқ таниб олишнингтурли тизимлари таққослашда асосий кўрсаткич сифатида WER(унинг абсолют ёки нисбий қиймати) фойдаланилмоқда. Нутқ технологиялари ривожланиши WER кўрсаткич қийматини тобора нолга яқинлаштирмоқда. WER ни аниқлаш усули Левенштейн масофасини ҳисоблашга асосланган динамик дастурлаш алгоритми йўли орқали иккита матнли сатрларни таққослашдан иборат. Бунда биринчи сатр таниб олиш натижаси ва иккинчи сатр аслида айтилган сўзлардан ташкил топади [1]. Левенштейн масофаси биринчи сатрдаги маълумотни иккинчисига айлантиришдаги матнли маълумотларни таҳрирлаш суммаси ёки минимал сони, яъни сўзни алмаштириш(S),

ўчириш(D) ва сўзни қўйиш(I) каби амалларнинг минимал сонини ўзида акс эттиради [2].

$$WER = \frac{S+D+I}{T} * 100\% \quad (2)$$

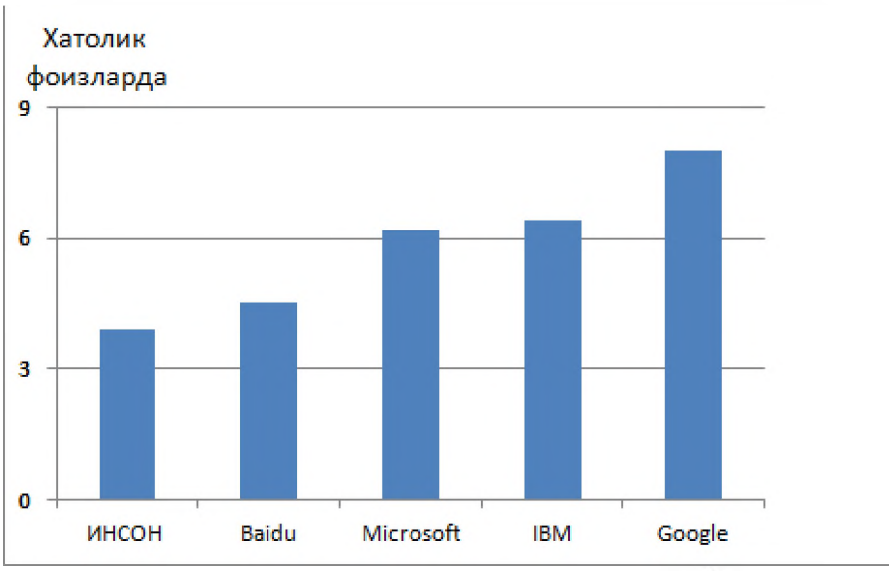
бу ерда T -таниб олинаётган жумладаги сўзлар сони.

Бундан ташқари нутқни таниб олиш сифатини баҳолашда тўғри таниб олинган сўзлар фоизи кўрсаткичидан (WCR — Word Correctly Recognized) ҳам фойдаланилади. Бунда тизим томонидан хатолик билан қўйилган сўзлар ҳисобга олинмайди:

$$WCR = \frac{H}{T} * 100\%, \quad H = N - D - S, \quad (3)$$

бу ерда H-тўғри таниб олинган сўзлар сони, N-диктор томонидан талаффуз қилинган сўзлар сони.

5-расмда инсон ва айрим тизимларнинг нутқни автоматик таниб олиш сифатни баҳолаш натижалари келтирилган.



5-расм. Инсон ва айрим тизимларнинг WER кўрсаткичлари

III. ХУЛОСА

Сўнгги йилларда нутқни автоматик таниб олишда сезиларли ривожланишларнинг кузатилиши ва кўплар ютуқларга эришганлига қарамай нутқни автоматик таниб олиш инсон имкониятларига нисбатан анча чегараланган. Масалан, инсон узоқлашган нутқни, сифатсиз алоқа канали орқали узатилган ёки акс садоли, акцентли нутқни қийинчиликсиз таниб олади ёки кўплаб овозлар орасидан маълум суҳандон нутқини ажратиб олиш ва спонтан нутқни таниб олиш имкониятига эга. Санаб инсон имкониятларидан айниқса, сўнгги иккитаси нутқни таниб олишнингзамонавий тизимларида катта муаммоларни келтириб чиқаради. Нутқни автоматик таниб олиш тизимлари фақатгина ажратилган буйруқлар

ёки сонларни таниб олишда инсондан устун келмоқда. Нутқни таниб олишнинг мавжуд тизимларини ривожлантириш нейрон тармоқлар структураларини мукаммаллаштириш, уларда турли даражадаги тескари алоқаларни таъминлаш ва ўқитишнинг янги усулларини ишлаб чиқиш билан боғлиқ. Бундан ташқари нутқ таркибининг маъноли қисмини қолдирувчи семантика соҳасидаги ишланмаларни қўллаш ҳамда уларни имкониятларидан фойдаланиш мақсадга мувофиқ бўлади.Агар ушбу ишлар амалга оширилса, у ҳолда нутқни автоматик таниб олишда Марков модели, динамик дастурлаш ва Витерби алгоритмларига эҳтиёж қолмаслиги мумкин.

АДАБИЁТЛАР

[1] Levin K., Ponomareva I., Bulusheva A., Chernykh G., Medennikov I., Merkin N., Prudnikov A., Tomashenko N. Automated closed captioning for Russian live broadcasting // Proceedings of the Annual Conference of the International Speech Communication Association INTERSPEECH. Singapore, 2014. P. 1438–1442.

[2] Terry K. Instant patient records and all you have to do is talk // Medical Economics. 1999. V. 76. N 19. P. 101–102, 107–108, 111–112.

[3] Huang X., Acero A., Hon H.-W. Spoken Language Processing. Prentice Hall, 2001. 1008 p.

[4] The HTK book [Электронный ресурс]. Cambridge University Engineering Department. Режимдоступа: http://speech.ee.ntu.edu.tw/homework/DSP_HW2-1/htkbook.pdf, свободный. Яз. англ. (дата обращения 22.10.2015).

[5] Baker J.K. The dragon system – an overview // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1975. V. ASSP 23. N 1. P. 24–29.

[6] Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // Proceedings of the IEEE. 1989. V. 77. N 2. P. 257–286. doi: 10.1109/5.18626

[7] Ramesh P., Wilpon J.G. Modeling state durations in hidden Markov models for automatic speech recognition // IEEE Transactions on Acoustics, Speech, and Signal Processing (ICASSP-92). San Francisco, USA, 1992. V. 1. P. 381–384.

[8] Shafran I., Ostendorf M. Use of higher level linguistic structure in acoustic modeling for speech recognition // Proc. IEEE Int. Conf. on Acoustic Signal and Speech Processing. Istanbul, Turkey, 2000. V. 2. P. 1021–1024.

[9] Digalakis V., Murveit H. Genones: optimizing the degree of mixture tying in a large vocabulary hidden Markov model-based speech recogniz-

er // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP). Adelaide, South Australia, 1994. V. 1. P. 537–540.

[10] Hain T., Woodland P.C., Niesler T.R., Whittacker E.W.D. 1998 HTK system for transcription of conversational telephone speech // Proc. Int. Conf. on Acoustics, Speech and Signal Processing.1999,V. 1. P. 57–60.

[11] Leggetter C.J., Woodland P.C. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models // Computer Speech and Language. 1995. V. 9. N 2. P. 171–185. doi: 10.1006/csla.1995.0010

[12] Kuhn R., Junqua J.-C., Nguen P., Niedzielski N. Rapid speaker adaptation in eigenvoice space // IEEE Transactions on Speech and Audio Processing. 2000. V. 8. N 6. P. 695–706. doi: 10.1109/89.876308

[13] Bourlard H., Hermansky H., Morgan N. Towards increasing speech recognition error rates // Speech Communication. 1996. V. 18. N 3. P. 205–231. doi: 10.1016/0167-6393(96)00003-9

[14] Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators // Neural Networks. 1989. V. 2. N 5. P. 359–366. doi: 10.1016/0893-6080(89)90020-8

[15] Hinton G., Deng L., Yu D., Dahl G., Mohamed A.-R., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T., Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups // IEEE Signal Processing Magazine. 2012. V. 29. N 6. P. 82–97. doi: 10.1109/MSP.2012.2205597

[16] Hermansky H., Ellis D., Sharma S. Tandem connectionist feature extraction for conventional HMM systems // Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP). Istanbul, Turkey, 2000. V. 3. P. 1635–1638.

[17] Robinson A.J. An application of recurrent nets to phone probability estimation // IEEE Transactions on Neural Networks. 1994. V. 5. N 2. P. 298–305. doi: 10.1109/72.279192

[18] Robinson T., Hochberg M., Renals S. The use of recurrent neural networks in continuous speech recognition / In: Automatic Speech and Speaker Recognition. Advanced Topics / Eds. C.H. Lee, F.K. Soong, K. Paliwal. Kluwer Academic Publishers, 1996. 518 p. doi: 10.1007/978-1-4613-1367-0

[19] Triefenbach F., Demuynck K., Martens J.-P. Large vocabulary continuous speech recognition with reservoir-based acoustic models // IEEE Signal Processing Letters. 2014. V. 21. N. 3. P. 311–315. doi: 10.1109/LSP.2014.2302080