

УДК 004 (004.652.2)( 51-77)

**Маматов Н.С., Юлдошев Ю.Ш., Тўрақулов О.Х.****Нутқни автоматик таниб олиш талаблари ва ёндашувлари**

**Аннотация.** Нутқни автоматик таниб олиш микрофондан олинган акустик сигнални сўзлар кетма-кетлигига ўтказиш жараёнидир. Ҳосил бўлган сўзлар кетма-кетлиги берилган нутқли мулоҳаза маъносини тушунишда фойдаланилади. Мазкур ишда нутқни таниб олиш тизимларига қўйиладиган асосий талаблар ва ишлаб чиқиш ёндашувлари таҳлил қилинган.

**Калим сўзлар:** тизим, ажралган ва ёпишган нутқ, луғат, халақит, яширин марков модели, сунъий нейрон тармоғи, таниб олиш, динамик дастурлаш, мезон.

**Кириш.** Ҳозирги кунда сунъий интеллект соҳасининг нутқни автоматик таниб олиш йўналиши жадал ривожланиб бормоқда. Сўнги йилларда мазкур йўналишда кўплаб ютуқларга эришилди. Олинган натижалар инсон ва машина орасидаги матнли мулоқот ўрнига тезкорлик ва табиийликни таъминловчи овозли мулоқотдан фойдаланиш имконини бермоқда. Айни пайтда инглиз, рус, испан, француз, япон, хитой ва араб тилларидаги нутқ сигналларни юқори аниқликда таниб олувчи тизимлар ишлаб чиқилган. Бу тилларнинг танланиши нутқ технологияларини ривожланишининг иқтисодий ва сиёсий жиҳатлари билан боғлиқ. Бошқа тилларга эса етарли даражада эътибор қаратилмаяпти. Жумладан ўзбек тилига ҳам.

Мавжуд ва ишлаб чиқиладиган тизимлар самардорлигини баҳолаш тизимларга қўйиладиган талаблар ва бошқа кўплаб кўрсаткичлар орқали амалга оширилади. Куйида нутқни таниб олиш тизимларига қўйиладиган талаблар ва уларни ишлаб чиқиш ёндашувлари таҳлили келтирилган.

**Нутқни таниб олиш тизимларига қўйиладиган талаблар.** Нутқни таниб олиш масаласи нутқни узатиш каналининг хоссаси, луғат ўлчами, нутқнинг ўзгарувчанлиги, атрофдаги халақит даражаси, нутқ тури (ажралган ёки ёпишган) каби кўплаб параметрлар билан характерланади [15].

Ажралган нутқни таниб олишда суҳандон сўзлар орасида қисқа танаффус қилиб нутқни киритишни секинлаштиради. Бу эса табиийликни сусайтиради. Ёпишган сўзни таниб олишда сўзлар орасидаги танаффус талаб этилмайди. Табиий нутқ матн ёки сунъий сигналлардан фарқли равишда элементларни (фонема, сўз, жумла) оддий ва бир қийматли ажратиш имконини бермайди. Мазкур элементлар аниқ бир физик чеграга эга бўлмай, улар эшитувчининг тушуниши асосида ажратилади [23]. Агар тингловчидан нотаниш сўзни фонемалар орқали ёзиш сўралса, у жумла ва сўзларни ажратишда кўплаб хатоликларга йўл қўяди. Ҳаттоки инсон ҳам нутқни граматик, лексик ва маънога оид билимларсиз ажрата олмайди. Сўзлар чегарасини таниб олиш жараёнида аниқлаш мушкул бўлади ва бу чегара лингвистик ҳамда акустик мезонлар асосида кирувчи нутқ оқими билан энг макбул ҳолда оптимал сўзлар кетма-кетлигини аниқлаш орқали аниқланади.

Нутқни таниб олиш масаласининг мураккаблиги кўплаб омиллар таъсир қилувчи асосий параметрларининг ўзгарувчанлиги билан боғлиқ. Булардан бири нутқни ўрганишининг асосий компонентаси бўлиб, айнан бир суҳандон томонидан айтилган ва айнан бир сўзни кўп хилда ифодаланшидир. Турли суҳандонларни индивидуал нутқ ускуналарининг фарқи билан боғлиқ бўлган ўзгариш муҳимроқ ҳисобланади. Бунда суҳандоннинг жинси, ёши, диалекти, эмоцияси ва жисмоний ҳолати ҳам алоҳида таъсир кўсади. Бундан ташқари акустик томони яъни

микрофоннинг алмашиши ва жойлашувининг ўзгариши ҳамда хонанинг акустик ҳолати ҳам жиддий салбий таъсир кўрсатади. Луғатнинг катталашиши ҳам таниб олишга жиддий таъсир кўрсатади, яъни бунда ўхшаш акустик бир қийматли бўлмаган сўзларнинг акустик гуруҳи ҳосил бўлади. Бу эса луғатнинг ўсишига боғлиқ экспоненциал ўсишига олиб келади. Таниб олинмиши зарур бўлган луғат ўлчамини бир нечта тоифаларни мавжуд. Бир ва ўнлаб сўзларга эга бўлган луғат кичик луғат деб аталади [3]. Кичик ўлчамли луғатдан фойдаланувчи масала ва иловалар мавжуд. Масалан, рақамларни кетма-кет аниқлаш (телефон рақамлари), техник объектларни буйруқлар асосида бошқариш (автомобил, самолёт) [20], роботларни масофадан бошқариш тизимлари [12,22], қурилмаларни бошқариш тизимлари (медицина) [32] ва бошқалар. Ўртача луғатлар одатда юзлаб сўзларни ўз ичига олади. Бундай луғатларга асосланган кўплаб мулоқот ёки савол-жавоб тизимлари мажуд [7,34]. Катта луғатлар минглаб сўзлардан иборат бўлади [3]. Бундай луғатлар чекланган предмет соҳаларида ўзини тизимлари ёки автоматлашган ёрдам тизимларида кўп ишлатилади. Ўта юқори ўлчамли луғатлар йуз минглаб сўзлардан иборат бўлиб, улар ихтиёрий матнларни стенографиясида тадбиқ этилиши мумкин [27].

Диалогли тизимда ишлашда ёки нутқ орқали матнларни киритишда фойдаланувчи тизимдан жавобни тез олишни хошлайди. Шунинг учун нутқни таниб олиш тизими реал вақтда кутишсиз жавобни таъминлашни талаб этилади. Шундай тимсолларни аниқлаш масалалари мавжудки, уларда реакция вақти муҳим рол ўйнайди. Масалан, архив нутқ ёзувларини матнга ўтказиш [29]. Ҳозирги кунда бундай тизимлар кўплаб ишлаб чиқилган.

Нутқни автоматик таниб олишнинг замонавий тизимларига қўйиладиган ўта муҳим талаблар куйидагилардан иборат.

- ёпишган нутқни юқори аниқликда таниб олиш;
- суҳандонга боғлиқ бўлмаган;
- кўп сўзларни таниб олиш имконияти;
- юқори тезлик.

Нутқни таниб олиш тизимларига ўхшаш мураккаб тизимларга кўп мезонли баҳолаш ва асосланган макбул модел ҳамда унинг параметрларини танлаш шу куннинг долзарб масалаларидан биридир [21,28].

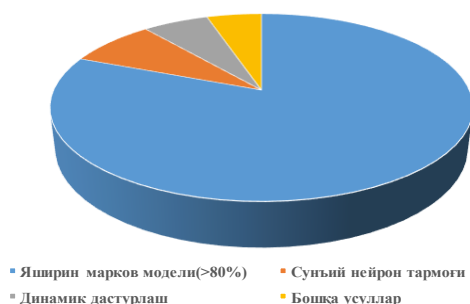
Ишлаб чиқилган нутқни таниб олиш тизимлари самардорлигини баҳолаш учун нутқнинг қайта ишлашнинг ҳар бир поғонасида кўплаб мезонлар тадбиқ этилади. Бу мезонларнинг таниб олиш аниқлиги ва тизимнинг реакция вақти(жавоб) интеграл мезон ҳисобланади. Идеал автоматик тизим бехато натижани тақдим этиши шарт. Бироқ айни пайтдаги мавжуд компьютер тизимлари бундай кўрсаткичга эришгани йўқ, инсон даражасида нутқни таниб олишга интилиш шарт.

**Нутқни автоматик таниб олиш ёндашувлари.** Нутқни таниб олиш масаласи қайта ишланган нутқ

сигналига катта эҳтимоллик (ўхшашлик) билан мос келувчи оптимал сўзлар моделлари кетма-кетлигини аниқлашдан иборат.

Дунёнинг етакчи олимлари обзорли таҳлили. Ҳозирги кунда деярли барча нутқни автоматик таниб олиш тизимлари яширин марков модели, сунъий нейрон тўрлари, динамик дастурлаш каби асосий ёндашувларга асосланган эканлигини кўрсатди [24-26, 33].

Узоқ вақтлар мобайнида динамик дастурлаш асосидаги ёндашув доминират бўлиб келди. У нутқни таниб олиш нутқ фрагментини олдиндан яратилган эталон сўз билан таққослаш орқали амалга оширилган. Динамик дастурлаш масаласи нутқни икки қисми учун ночизикли мосликни топишга келтирилади. Мазкур мосликни куришда Р.Беллмон фундаментал ишларига асосланган динамик дастурлаш алгоритмларидан фойдаланган [1]. Динамик дастурлаш ни нутқни таниб олишга қўллашга оид дастлабки ишларидан бири украинлик олим Т.К.Винцюкга тегишлидир [2].



1.1-расм. Нутқни автоматик таниб олишнинг асосий ёндашувлари

Динамик дастурлаш усули ёрдамида ёпишган нутқни таниб олишни бир нечта ёндашувлари мавжуд. Булар икки поғонали динамик дастурлаш алгоритми, поғоналарни куриш ва битта ўтувчи усуллардир [30]. Мазкур алгоритмлар бир хил боғланган тамойиллардан фойдаланади ва жорий этиш мураккаблиги билан фарқланади. Мазкур усуллардан ташқари динамик дастурлаш га сирпанувчи ойна ва размитих тўплалар назарияси тадқиқ этиш орқали нутқни таниб олиш усули ҳам мавжуд [14].

Динамик дастурлаш га асосланган барча усуллар суҳандонга боғлиқ бўлиб, бу унинг асосий камчилиги ҳисобланади. Бундан ташқари динамик дастурлаш асосида яратилган тизимдан фойдаланишда ҳар бир фойдаланувчи ўзини эталонини яратиш талаб этилади, яъни луғатдаги барча сўзлар талаффуз қилиниши керак бўлади. Таниб олиш сифатини ошириш учун фойдаланувчидан луғатдаги барча сўзларни бир неча марта такрорлаши талаб этилади. Шунинг учун бундай ёндашув кичик луғатли иловаларда қўлланилмоқда. Масалан дастурларни овозли бошқаришда.

Кўплаб иловаларда нутқни таниб олиш сунъий нейрон тармоғи орқали амалга оширилади. Бунда аниқ бир ёзув танланганда шуғулланган ва тўғри танланган тармоқ юқори наижани таъминлайди. Амалиётда эса бир ёки бир неча қатламли нейрон тармоқлари қўлланилади [4,10]. Бунда тармоқ мураккаблиги яширин қатламдаги нейронлар сони асосида аниқланади. Бунда кириш ва чиқиш қатламдаги нейронлар сони фиксирланган бўлсада ва бу масала шартга боғлиқ. Нейрон тармоғи киришга нутқ сигнали белгилари вектори бериладиган, чиқишлар эса олинган луғатга боғлиқ бўлган ёндашув кенг тарқалган. Бунда кўпинча чиқишлар сони луғатдаги сўзлар сонига тенг қилиб олинади. Нейрон тўри бир нечта суҳандон

нутқини ўрганишга қодир бўлганлиги учун, у суҳандонга боғлиқ бўлмаган тизимларни яратишда қўлланилади. Нейрон тўрларини ёпишган нутқга тадқиқ этиш кўплаб муаммоларни юзага келтиради. Масалан, нутқ сигналларининг давомийлиги олдиндан аниқ бўлмайди. Бунга ма равишда белгилар векторларисони ҳам. Бундан ташқари айтилган сўзлар сони ва тартиби ҳам. Бу эса ўқитиш ва яратишни мураккаблаштиради.

Нутқни таниб олишни айрим тизимларида нейрон тармоғи яширин марков модели билан бирга қўлланилади [11]. Бунда нейрон тармоғи белгилар векторларини қайта ишлаш босқичида ёки таниб олиш гепотизалар матндаги ишлов беришда қўлланилади. сунъий нейрон тармоғи юқори потенциалга эга бўлишига қарамай, ҳозирда нутқни таниб олиш соҳасида кенг тадқиқ этилмайди. Чунки нейрон тармоғини ўқитиш ўта мураккаб ва катта ҳажмдаги ҳисоблашни талаб қилади.

Айни пайтда кўплаб тизимлар нутқни таниб олишда машҳур математик аппарат яширин марков моделидан фойдаланади [6,8]. Яширин марков модели етарли даражада математик структурага эга бўлганлиги учун, нафақат нутқда балки турли соҳаларда тасодифий жараёнларни тадқиқ этишда назарий асос бўлиб хизмат қилмоқда. Яширин марков модели нутқни таниб олиш масаласини ечиш билан бирга халақит ва сакрашларни бартараф этиш орқали сигнал сифатини оширади, нутқ сигнали манбасини моделлаштириш, суҳбат структурасини оптималлаштириш ва бошқаларда кенг тадқиқ этилмоқда.

Яширин марков модели асосида суҳандонга боғлиқ бўлмаган ёпишган нутқни таниб олишнинг етарли даражадаги самарадор усули ишлаб чиқилган.

Нутқни автоматик таниб олишнинг Support Vector Machines [31], нутқни вейвлет таҳлили [3] ва инсон эшитишини моделлаштирувчи каби тизимларда ҳам мавжуд. Бироқ мазкур технологиялар замонавий нутқни таниб олиш тизимларида кенг тадқиқ этишда ўз ўрнини топмаган.

#### Фойдаланилган адабиётлар

1. Беллман, Р. Динамическое программирование / Р. Беллман; М.: ИЛ, 1960, 400 с.
2. Винцюк, Т. К. Распознавание слов устной речи методами динамического программирования/ Т. К. Винцюк // М.: Кибернетика, 1968. – №1. –С. 15-22.
3. Галунов, В.И. Состояние исследований в области речевых технологий и задачи, выдвигаемые государственными заказчиками / В.И. Галунов, и др. // Доклад на секции по автоматическому распознаванию и синтезу речи РАН. М., 2002.
4. Галушкин, А.И. Теория нейронных сетей / А.И. Галушкин; М.:ИПРЖР, 2000, 416 с.
5. Геппенер, В.В. Вейвлет-преобразование в задачах цифровой обработки сигналов: Учебное пособие / В.В. Геппенер, Д.А. Черниченко, С.А. Экало // СПб.: Изд-во СПбГЭТУ, 2002. 78 с.
6. Желинек, Ф. Распознавание непрерывной речи статистическими методами / Ф. Желинек // ТИИЭР 64, № 4, 1976, с. 131-160.
7. Ли, И.В. Проектирование систем речевого диалога / И.В. Ли, А.Л. Ронжин // Труды СПИИРАН. Вып. 3, т. 1. — СПб.: Наука, 2006, С. 320-338.
8. Марков, А.А. Об одном применении статистического метода / А.А. Марков // Известия АН, сер.6, X, №4, 1916, 239 с.
9. Моттль, В.В. Скрытые Марковские модели в структурном анализе сигналов / В.В. Моттль, И.Б. Мучник; М.: Физматлит, 1999, 351 с.

10. Осовский, С. Нейронные сети для обработки информации / С. Осовский, пер. с польского И. Рудинского. -М.: Финансы и статистика, 2004, 344 с.
11. Петровский, А.А. Методы построения устройств распознавания речи на базе гибрида нейронная сеть/скрытая Марковская модель / А.А. Петровский // Нейрокомпьютеры: разработка, применение, 2002, № 12, с. 26-36.
12. Потапова, Р.К. Речевое управление роботом / Р.К. Потапова //М.:КомКнига, 2005, 328 с.
13. Рабинер, Л. Цифровая обработка речевых сигналов / Л. Рабинер, Р. Шафер– М.: Радио и связь, 1987.
14. Ронжин, А. Метод распознавания слитной речи на основе анализа сигнала в скользящем окне и теории размытых множеств / А. Ронжин и др. // Научно-теоретический журнал «Искусственный интеллект», №4. – Донецк, Украина, 2002, С. 256-263.
15. Ронжин, А.Л. Речевой и многомодальный интерфейсы / А.Л. Ронжин, А.А.Карпов, И.В. Ли; - М.: Наука, 2006 - (Информатика: неограниченные возможности и возможные ограничения), 173 с.
16. <http://www.sakrament.com/viewprod.php?TopId=30&ProdId=24>
17. <http://htk.eng.cam.ac.uk/>
18. <http://www.newvoice.ru/>
19. <http://www.nuance.com>
20. <http://speechpro.com/production/?id=471&fid=44>
21. Соколов, Б.В. Концептуальные основы оценивания и анализа качества моделей и полимодельных комплексов / Б.В. Соколов, Р.М. Юсупов // Теория и системы управления. – 2004. –№ 6 –С. 5–16.
22. Станкевич, Л.А. Интеллектуальные роботы и системы управления /Л.А. Станкевич // Нейрокомпьютеры: разработка и применение, № 8-9, 2005.
23. Ушакова, Т.Н. Проблема внутренней речи в психологии и психофизиологии. Психологические и психофизиологические исследования речи / Т.Н. Ушакова – М.: Наука, 1985. – С. 13-26.
24. Cox, R.V. Speech and Language Processing for Next-Millennium Communications Services / R.V. Cox, et al // Proceedings of the IEEE, Vol. 88, No. 8, 2000, pp. 1314-1337.
25. Furui, S. 50 years of progress in speech and speaker recognition / S. Furui // Proceedings of SPECOM'2005, Patras, Greece, 2005, pp. 3-9.
26. Haton, J.-P. Automatic speech recognition: Past, Present and Future / J.-P. Haton // Proceedings of SPECOM'2004, St. Petersburg: "Anatoliya", 2004, pp. 3-7.
27. Hori, T. An extremely-large-vocabulary approach to named entity extraction from speech / T. Hori, A. Nakamura // Proceedings of ICASSP'2006, Toulouse, France, 2006.
28. Potryasaev, S. Quality and Quantity Estimation and Analysis of Multimodal Systems for Human-Computer Interaction / S. Potryasaev, B. Sokolov, Yusupov // Proceedings of SPECOM'2006, St. Petersburg: "Anatoliya", 2006, pp. 158-167.
29. Psutka, J. Large Vocabulary ASR for Spontaneous Czech in the MALACH Project / J. Psutka, et al // Proceedings of Eurospeech'2003, Geneva, Switzerland, 2003, pp. 1821-1824.
30. Rabiner, L. Fundamentals of Speech Recognition / L. Rabiner, B. Juang – New Jersey: Prentice-Hall, Englewood Cliffs, USA, 1993.
31. Surendran, D. Dialog Act Tagging with Support Vector Machines and Hidden Markov Models / D. Surendran, G. Levow // Proceedings of Interspeech'2006, Pittsburgh, PA, USA, 2006, pp. 1950-1953.
32. Timofeev, A.V. Development of man-machine interfaces and virtual reality means for integrated medical systems / A.V. Timofeev, et al. // Proceedings of SPECOM'2006, St. Petersburg: "Anatoliya", 2006, pp. 175-178.
33. Trentin, E. A survey of hybrid ann/hmm models for automatic speech recognition / E. Trentin, M. Gori // Neurocomputing, vol. 37, no. 1-4, 2001, pp. 91-126.
34. Turunen, M. Evaluation of a Spoken Dialogue System with Usability Tests and Long-term Pilot Studies: Similarities and Differences / M. Turunen, J. Hakulinen, A. Kainulainen // Proceedings of Interspeech'2006, Pittsburgh, USA, 2006, pp. 1057-1060.

**Маматов Нарзилло Солиджонович** – техника фанлари доктори, Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети хузуридаги Ахборот-коммуникация технологиялари илмий-инновацион маркази катта илмий ходими.  
Тел.: (+99897) 403-56-22  
Факс: (0371) 237 62 48

E-mail: [m\\_narzullo@mail.ru](mailto:m_narzullo@mail.ru)

**Юлдошев Юсуф Шералиевич** – Муҳаммад ал-Хоразмий номидаги ахборот технологиялари университети хузуридаги ахборот – коммуникация технологиялари илмий – инновацион маркази таянч  
+998909615739  
E-mail: [yusuf\\_yuldoshev@mail.ru](mailto:yusuf_yuldoshev@mail.ru)

**Тўракулов Отабек Холмирзаевич**

Жиззах политехника институти таянч докторанти  
+998974553700  
E-mail: [o\\_turakulov@mail.ru](mailto:o_turakulov@mail.ru)

**N.S.Mamatov, Yu.Sh.Yuldoshev, O.X.To'raqulov.**  
**Approaches and requirements to the development of speech recognition systems**

Automatic speech recognition is the process of transferring an acoustic signal from a microphone to a sequence of words. The sequence of words used is used to understand the meaning of speech. In this study, the main requirements and approaches to the development of speech recognition systems were analyzed.

**Keywords:** system, split and stuck speech, dictionary, noise, hidden Markov model, artificial neural network, identification, dynamic programming, criterion.