

«Сейфуллин оқулары – 16: Жаңа формациядағы жастар ғылыми – Қазақстанның болашағы» атты халықаралық ғылыми-теориялық конференциясының материалдары = Материалы Международной научно-теоретической конференции «Сейфуллинские чтения – 16: Молодежная наука новой формации – будущее Казахстана». - 2020. - Т.І, Ч.3 - С.186-189

АҚПАРАТТЫҚ ЖҮЙЕЛЕРДЕ СӨЙЛЕУДІ ТАНЫП БІЛУ НЕГІЗДЕРІ

Қуандық К.Н.

Ақпараттық қоғамның қазіргі заманында, сөйлеуді тануды зерттеу және талдау, дауыстық басқару секілді сөйлеу технологияларын пайдалану күн сайын дамуда. Пайдаланып жүрген барлық мобильді құрылғылар, дербес компьютерлер, интернет-браузерлер, ақылды үй жүйелері, тіпті тұрмыстық техника құралдары да сөйлеу технологияларымен жұмыс істейді. Себебі, табиғи, тілдік қарым-қатынас адамдар үшін ең ыңғайлы және үйреншікті болып табылады. Сондықтан ғалымдар мен инженерлер көп жылдар бойы адам мен машина арасындағы тілдік қарым-қатынас мәселесін шешуге тырысуда.

Дауыс арқылы қажетті ақпараттарды іздеуге, тауарларды сатып алуға, шет тілін үйреніп, көптеген басқа құрылғыларды оңайлықпен басқаруға болады. Бірақ бұл технологиялар көп жағдайда дамыған елдердің тілдерінде, яғни ағылшын, неміс, жапон, орыс, қытай және т.б. тілдерде жұмыс істейді. Қазақстан нарығында бұл бағыттағы технологиялар дамудың бастапқы кезеңінде, яғни, жаппай пайдалану байқалмайды, бұл дикторға тәуелділіктің салдары болуы мүмкін, үздіксіз сөйлеу үшін айырып тану дәлдігінің жеткіліксіздігі және әртүрлі кедергілердің болуына жоғары сезімталдығы болуы мүмкін.

Ақпаратты кез келген тілде енгізу мәселесі келесі факторлармен күрделене түседі: тіл құрылымының айырмашылығымен, айтылу ерекшелігімен, шуылмен және кедергілермен, жалғаулармен, екпіндермен және т. б.

Бүгінгі таңда қолданыстағы сөйлеуді тану жүйелері лексикалық элементтерді тану үшін қажетті барлық қолжетімді және тіпті артық ақпаратты жинауға негізделеді. Google Inc. секілді ірі компаниялардың сөздерді тану жүйелері жүздеген және тіпті мыңдаған дикторлардың сөйлеу мәнерлерінің кең базасын пайдаланады, бұл оларға белгісіз дикторлардың көптеген сөздерін сенімді тануға қол жеткізуге мүмкіндік береді. Кейбір зерттеушілер[1,2] осылай өзгерістерге ұшырайтын сигналдың сапасына негізделген сөйлеу үлгісін тану міндеті тану үшін жеткілікті болады деп санайды, алайда қазіргі уақытта қалыпты сөйлеудің шағын хабарламаларын тану кезінде де, әр түрлі нақты сигналдарды алғаннан кейін лингвистикалық символдарға тікелей трансформацияны жүзеге асыру қалаған нәтиже болып табылады, бірақ әзірге мүмкін емес .

Сөйлеуді тану - бейнелерді танудың күрделі, кезеңдік міндеті. Осы міндетті шешу барысында сөйлеу деректері талданады және берілген

иерархияға сәйкес жіктеледі. Жіктелген бейнелер әр түрлі құрылымдық элементтер, белгілі бір ұзақтықтағы (фонема, буындар, сөздер) тілдік деректер кесінділері болуы мүмкін. Кіріс сигналы туралы априорлық ақпаратты көп болжаған сайын, соғұрлым сапалы түрде оны өңдеу және тануды жүзеге асыруға болады.

Жалпы жағдайда сөйлеуді сәйкестендірудің әрбір жеке міндеті кіріс сигналынан адам сөзін білдіретін акустикалық тербелістерді бөліп, жіктеуге және тиісті түрде әрекет етуге негізделеді.

Дауыстық басқару сөзді тану технологиясына негізделген: жүйе микрофон арқылы ауа тербелісі туралы ақпаратты алады, алынған деректерді жүйеде жазылған командалармен салыстырады және сәйкес келген жағдайда, ұйғарылған әрекетті орындайды. Неғұрлым лексикалық бірліктерді дәлдікпен тануға болатын болса, жүйенің команданы қатесіз тану мүмкіндігі соғұрлым жоғары болады.

Сөйлеуді тану үшін акустикалық сигнал анықтаушы және цифрлаушы құрылғылардың және машинамен өңдеудің көмегімен тіркеледі және сандық нысанға айналады. Дискретизация нәтижесінде үздіксіз (аналогтық) сигнал сандардың кезектілігіне аударылады. Сөйлеу сигналдарын цифрлық өңдеудің ең танымал әдістері: Фурье базисіндегі жиіліктік талдау, вейвлет талдау, кепстральды талдау, субполқынды талдау.

Бұдан әрі түрлі сөйлеуді тану жүйелерін жіктеуге болатын негізгі факторлар келтіріледі.

Сөздік көлемі. Сөйлеуді тану жүйелерінің қателіктері тану жүйесінің сөздіктеріндегі сөздердің санына тікелей байланысты. Бірнеше ондаған сөздерден тұратын сөздік айтарлықтай дәлдікпен тани алады. Ал, жүздеген және мыңдаған сөздерден тұратын сөздіктердің тану дәлдігі елеулі қателіктерге әкеліп соқтырады, сөздер көп болған сайын, олардың бір-бірімен ұқсастықтары да көп болады[3].

Дикторға тәуелділік. Дикторға тәуелді жүйе бір қолданушының пайдалануына арналған, яғни, ол сөйлеудің жеке ерекшеліктеріне сәйкес конфигурацияланады, ал дикторға тәуелсіз жүйе кез-келген қолданушымен жұмыс істеуге арналған және айтылымның жеке ерекшеліктерін ескермейді. Дикторға тәуелсіздік - бұл қиын мақсат, өйткені жаттықтыру кезінде жүйе оқып жатқан диктордың параметрлеріне бейімделеді. Дикторға тәуелсіз жүйелердегі қателіктер дикторға тәуелді жүйелерге қарағанда айтарлықтай жоғары болады[3].

Сөйлеу түрлері. Әдетте енгізу үшін пайдаланушы бір сөзді, болмаса сөз тіркесін қолдануы немесе үздіксіз сөйлеу кезінде маркер сөздерді табуы керек. Үздіксіз сөйлеуді тану әлдеқайда қиын, өйткені жекелеген сөздердің шекаралары нақты анықталмаған және олардың айтылуы кейбір айтылған дыбыстардың бұлдырланып, жұтылуы арқылы едәуір бұрмаланып кетеді[3].

Қолдану саласы. Жүйенің мақсаты тану жүргізілетін абстракцияның талап етілетін деңгейін анықтайды. Сөйлеуді тану қолданылатын жүйелердің бірнеше түрі бар: маркер сөздерді іздеу жүйесі, сөйлеу-мәтін жүйесі және сөйлеу-сөйлеу жүйесі. Маркер сөзді немесе сөздер тізбегін іздеу жүйесінде

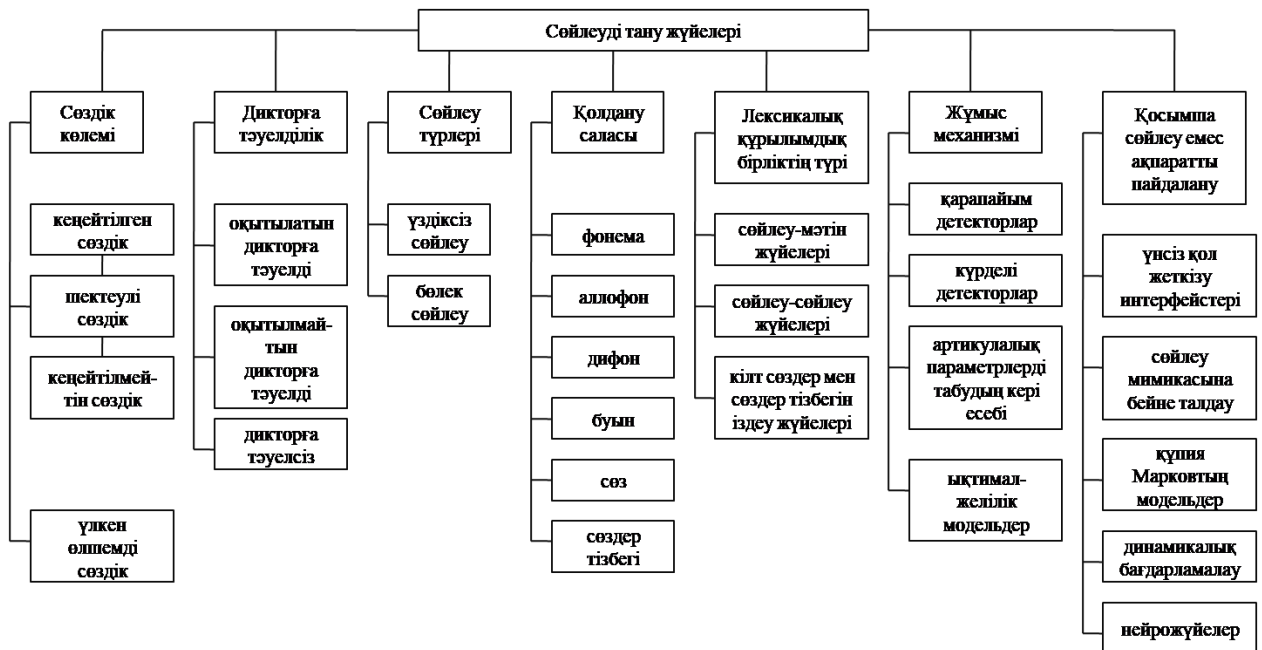
маркерді тану бірыңғай бейнені тану ретінде жүреді. Сөйлеу-мәтін және сөйлеу-сөйлеу (синхронды аударма жүйесі) қазіргі кезде танылатын паттерндерді ғана емес, бұрын танылған паттерндерді де ескеретін айырудың жоғары дәлдігін талап етеді. Осылайша, жүйе тек кейбір жекелеген сөздер мен сөз тіркестерін ғана емес, сонымен қатар олар айтылған контекстті ескеруі керек.

Лексикалық құрылымдық бірліктің түрі. Сөйлеуді талдау кезінде талдаудың базалық бірлігі ретінде жеке сөздер мен сөз тіркестері, буындар, сондай-ақ фонемалар, аллофондар, дифондар сияқты элементтер таңдалуы мүмкін. Лексикалық құрылымдық бірлік түріне тұтастай алғанда жүйенің қиындығы да, тану сапасы да, қолданылатын сөздіктің өлшемі де байланысты.

Жұмыс механизмі. Қазіргі заманғы жүйелерде танушы жүйелердің жұмыс істеу механизміне әртүрлі тәсілдер кеңінен қолданылады. Ықтималдық-желілік тәсіл сөйлеу сигналы белгілі бір бөліктерге бөлінеді (кадрлар, немесе фонетикалық белгі бойынша), содан кейін танылатын сөздіктің қандай элементіне осы бөлік және барлық кіріс сигналы қатысы бар деген ықтималдық баға жүреді. Дыбыс синтезінің кері есебін шешуге негізделген тәсіл кіріс сигналы бойынша сөйлеу жолының артикуляторларының қозғалыс сипаты анықталынады және арнайы сөздік бойынша айтылған фонемамен анықталады.

Қосымша сөйлеу емес ақпаратты пайдалану. Соңғы уақыттарда акустикалық емес параметрлерді қолданатын сөйлеуді тану жүйелері, мысалы: еріннің, тілдің, бет бұлшық еттерінің қозғалысы (камерамен бекітілетін), ультрадыбыс, бас сүйегінің тербелістері, сондай-ақ дауыстық байламдар мен көмейдің белсенділігін бекітетін электромиография танымал бола бастады. Мұндай әдістердің пайда болуының негізгі себебі - тануға жарамды ақпарат көлемін ұлғайту ниеті. Мұндай ларингофондар алғаш рет танк әскерлерінің байланыс жүйелерінде қолданылды, онда көлік құралдарының ішіндегі дыбыстық шу деңгейі өте жоғары болды және дыбыстық тербелістерді тікелей жазумен қатар қосымша ақпарат көздері қажет болды. Осындай тәсілдерді қолдана отырып, сөйлеуді тану жүйелері жоғары дәлдікке қол жеткізе алады, сонымен қатар оған шу мен шу әсерін ескере отырып, акустикалық жолға қойылған көптеген шектеулерді алып тастай алады.

Ұсынылған факторларға сәйкес сөйлеуді танудың қолданыстағы жүйелерін неғұрлым толық сипаттайтын классификациясы жасалды (сурет 1).



Сурет 1. Сөйлеуді тану жүйелерінің классификациясы

Сөйлеуді танып білу жүйелерінің құрылымына әсер ететін негізгі факторларды талдау қазіргі уақытта әмбебап, сенімді және қарапайымдылыққа ие жүйенің жоқ екенін көрсетті. 1-суретте ұсынылған сөйлеуді тану жүйелерінің классификациясы жаңа алгоритмдерді, сөйлеуді танудың әдістері мен жүйелерін жасау кезінде осы бағыттағы зерттеулердің көлемін тарылтады.

Сөйлеуді танып білу жүйелері қазіргі уақытта қарқынды дамып келе жатқан технологиялардың бірі болып саналады. Компьютерлік жүйелердің даму шамасына қарай, тікелей компьютермен жұмыс істеу кезінде адам тілін пайдалану мүмкін болса, жүйелерді пайдалану әлдеқайда кеңейеді. Атап айтқанда, нақты уақытта машинаны қарапайым дауыспен басқару, сондай-ақ қарапайым адам сөзі түрінде ақпаратты енгізу және шығару мүмкін болады.

Әдебиеттер тізімі

1. Brumberg J.S., Nieto-Castanon A, Kennedy P.R., Guenther F.H. (2010). Brain-computer interfaces for speech communication. *Speech Communication* 52:367-379.
2. Jorgensen C, Dusan S. (2010). Speech interfaces based upon surface electromyography. *Speech Communication*, 52: 354-366.
3. Хабрахабр. Распознавание речи [Электрондық ресурс] – Жаңартылған күні: 15.07.2009, URL: <http://habrahabr.ru/post/64572>

Ғылыми жетекші э.ғ.к. Айдынов З.П