

Дәріс 13. Дауыс пен естуге байланысты техникалық және психологиялық аспектілер.

- ✓ Кіріспе
- ✓ Сөйлеуді тану жүйелерінің жіктелінуі
- ✓ Сөйлеуді тану тәсілдері
- ✓ Сөздік енгізу, шығару құрылғылары

Адамның қабылдау жүйесі. Есту. Дауыс. Дыбыссыз дауыс. Өзара әрекеттесуді жобалауда дыбысты қолдану. Дыбыс арқылы дауыстық әрекеттесудің артықшылықтары мен кемшіліктері. Дыбыс арқылы дауыссыз өзара әрекеттесудің артықшылықтары мен кемшіліктері. Дыбыспен байланысты техникалық мәселелер. Дыбыстық толқындар. Машина жасаған дыбыс. Сөйлеуді тану.



Кіріспе

Сөйлеуді тану - сөйлеу сигналын сандық ақпаратқа түрлендірудің автоматты процесі (мысалы, мәтіндік деректер).

Кері әдіс-сөйлеу синтезі.

Сөйлеуді тануға арналған алғашқы құрылғы 1952 жылы пайда болды, ол адам айтқан сандарды тани алды.

1962 жылы Нью-Йорктегі компьютерлік технологиялар жәрмеңкесінде IBM Shoebox құрылғысы ұсынылды.



1963 жылы АҚШ-та "Сперри" корпорациясының инженерлері әзірлеген "Септрон" деп аталатын талшықты-оптикалық сақтау құрылғысы бар миниатюралық тану құрылғылары (Sceptron, бірақ ['septra:n] "к" жоқ) адам-оператор айтқан белгілі бір сөз тіркестеріне сәйкес белгілі бір әрекеттер тізбегін орындайтын болды.



"Септрондар" дауыстық теруді автоматтандыру нөмірлерді және дауыстап оқылған мәтінді телетайппен автоматты түрде жазу үшін тіркелген (сымды) байланыс саласында қолдануға жарамды болды, ол әскери салада (әскери техниканың күрделі үлгілерін дауыстық басқару үшін), авиацияда (ұшқыш пен экипаж мүшелеріне жауап беретін "ақылды авионика" жасау үшін), автоматтандырылған басқару жүйелерінде және т. б. қолданылуы мүмкін болды.



1983 жылы "Апач" тікұшақтарына арналған "ақылды авионика" интерактивті кешені таныстырылды, ол командалар мен ұшқыштың сұраныстарын танып, оларды борттық жабдыққа басқару сигналдарына түрлендіреді және оларға қойылған міндеттерді іске асыру мүмкіндігіне қатысты бір буынды дауыспен жауап береді.



Сөйлеуді танудың коммерциялық бағдарламалары тоқсаныншы жылдардың басында пайда болды. Әдетте оларды қолдың жарақатына байланысты көп мәтін тере алмайтын адамдар пайдаланады. Бұл бағдарламалар (мысалы, Dragon NaturallySpeaking (ағылш.,орыс., VoiceNavigator (ағыл.,орыс.) пайдаланушының дауысын мәтінге аударады. Мұндай бағдарламалар үшін аударма сенімділігі өте жоғары емес, бірақ жылдар өте келе ол біртіндеп жақсарады.



Мобильді құрылғылардың есептеу қуатын арттыру олар үшін сөйлеуді тану функциясы бар бағдарламалар құруға мүмкіндік берді. Осындай бағдарламалардың ішінде Microsoft Voice Command қосымшасын атап өткен жөн, ол сізге көптеген қосымшалармен дауыс беру арқылы жұмыс істеуге мүмкіндік береді. Мысалы, ойнатқышта музыка ойнатуды қосуға немесе жаңа құжат жасауға болады.



Сөйлеуді тану технологиясының келесі қадамы үнсіз қол жеткізу интерфейстерін (silent speech interfaces - үнсіз сөйлеу интерфейстері, SSI) дамыту деп санауға болады. Сөйлеуді өңдеудің бұл жүйелері артикуляцияның ерте сатысында сөйлеу сигналдарын алуға және өңдеуге негізделген. Сөйлеуді танудың дамуының бұл кезеңі қазіргі тану жүйелерінің екі маңызды кемшіліктерінен туындайды: шуылға шамадан тыс сезімталдық, сондай-ақ тану жүйесіне жүгінген кезде нақты және нақты сөйлеу қажеттілігі. SSI-ге негізделген тәсіл - өңделген акустикалық сигналдарға қосымша ретінде шуылға ұшырамайтын жаңа сенсорларды қолдану.



Сөйлеуді тану жүйелерінің жіктелінуі:

- ✓ сөздік өлшемі бойынша (шектеулі сөздер жиынтығы, үлкен сөздік);
- ✓ дикторға байланысты (дикторға тәуелді және дикторға тәуелді емес жүйелер);
- ✓ сөйлеу түрі бойынша (біріктірілген немесе бөлек сөйлеу);
- ✓ мақсаты бойынша (диктант жүйелері, командалық жүйелер);
- ✓ қолданылатын алгоритм бойынша (нейрондық желілер, жасырын Марков модельдері, динамикалық бағдарламалау);
- ✓ құрылымдық бірліктің түрі бойынша (фразалар, сөздер, фонемалар, дифондар, аллофондар);
- ✓ құрылымдық бірліктерді бөлу принципі бойынша (шаблон бойынша тану, лексикалық элементтерді таңдау).



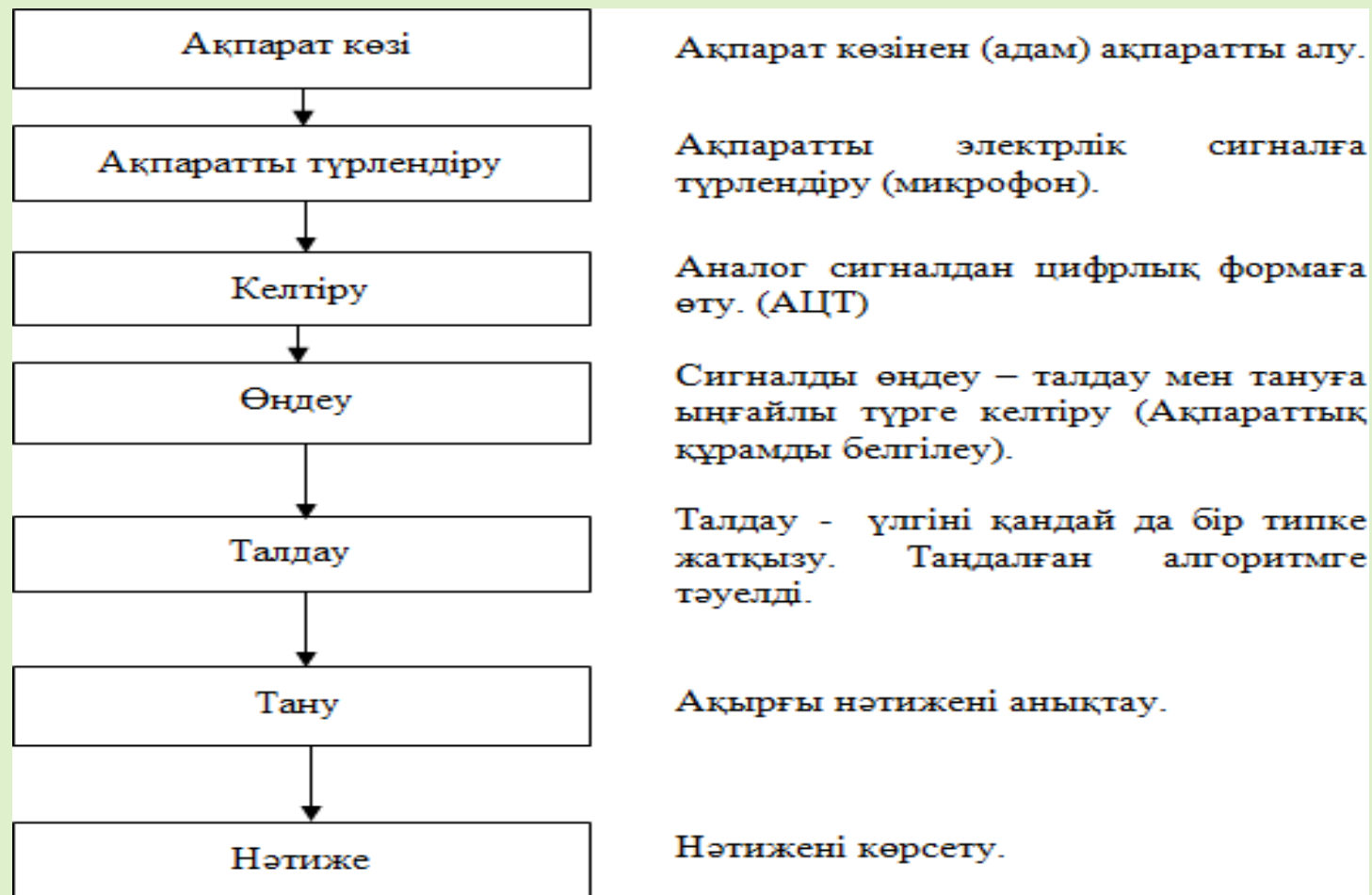
Табиғи тілде сөйлеуді автоматты тану жасанды интеллект дамуының ең өзекті бағыты болып табылады.

Бұл бағыттағы нәтижелер адам мен компьютер арасындағы тиімді қарым-қатынас орнату жабдықтарын жасау мәселесін шешуге мүмкіндік береді.

Сөйлеуді автоматты тану – компьютердің акустикалық сөйлеу сигналын мәтін немесе сәйкес басқарушы командаға түрлендіру процесі. Бұдан күрделірек есеп сөйлеуді автоматты түсіну болып табылады, ол сөйлеуді тануға қоса танылған мәтіннің семантикалық талдауын қамтиды. Мұндай жүйелер әлі де зерттелу үстінде.



Кез-келген сөйлеуді тану жүйесіндегі сигналды өңдеу есебін келесі түрде көрсетуге болады:



Сөйлеуді тану жүйелерін келесі белгілері бойынша классификациялары.

Дикторға тәуелді жүйелер жеке пайдаланушыға жасалады. Бұл жүйелер құрылымы бойынша қарапайымырақ, арзанырақ бірақ дикторлар тобына бейімдеуге немесе белгісіз диктормен жұмыс жасауға жарамсыз.

Дикторға тәуелсіз жүйелер кез-келген диктормен жұмыс істеу үшін жасалады. Бұл жүйелер әлі де қиындықпен дамып келеді, қымбатырақ болып табылады және дикторға тәуелді жүйелермен салыстырғанда тану сенімділігі төменірек болып келеді.

Дикторға бейімделуші жүйелер нақты бір дикторға немесе дикторлар тобына баптау жасауға мүмкіндік береді. Мұндай жүйелер танудың жеткілікті тиімділігін қамтамасыз етеді, бірақ оны диктор немесе дикторлар тобына баптау уақыты ұзақ болуы мүмкін.



Сөйлеуді тану жүйесінің көлемі оның күрделілігімен тікелей байланысты және тану дәлдігі сипаттамасына елеулі әсер етеді.

Сөздік көлемі қолданбалы жүйенің нақты талаптарына сәйкес анықталады. Кейбір қолданбалы программалар тек бірнеше сөзді талап етеді (мысалы сандар), ал келесілері өте үлкен сөздіктерді қажет етеді (мысалы, мәтінді автоматты теру жүйелері).

Әдетті сөздік көлемінің келесі градациялары қарастырылады:

кішкентай сөздік – ондаған сөздер;

орташа сөздік – жүздеген сөздер;

үлкен сөздік – мыңдаған сөздер;

өте үлкен сөздік – он мыңдаған сөздер.



Сөйлеу ағынының сипатына байланысты сөйлеуді тану жүйелері жеке сөздерді тану және үздіксіз сөйлеуді тану жүйелері болып бөлінеді.

Жеке сөздерді тану жүйелері жеке сөздерді бөлек бөлек кідіріс жасап айту арқылы тануға арналған.

Бұл әдіс танудың ең қарапайым түрі болып табылады, себебі мұнда сөздің бастапқы және соңғы нүктелерін табу оңай, және көршілес сөздер бір-біріне әсер етпейді.

Бұл жоғары сапалы тануға қол жеткізуді қамтамасыз етеді.



Үздіксіз сөйлеуді тану жүйелері сөздер бірге, кідіріссіз табиғи түрде айтылатын сөйлеуді тануға арналған. Үздіксіз сөйлеуді өңдеу қиынырақ, оған себеп ең алдымен сөздердің бастапқы және соңғы нүктелерін белгілеудің қиындығы. Сонымен қатар «коартикуляция» мәселесі бар. Коартикуляция дегеніміз көршілес дыбыстардың өзара әсер етуі. Сәйкесінше, сөздің басы мен соңындағы дыбыстарға көршілес сөздердің басы мен соңындағы дыбыстар әсер етеді. Үздіксіз сөйлеуді тануға сөйлеу темпі де әсер етеді. Оның үстіне үздіксіз сөйлеуі тану жүйелерінің сөздік көлемі үлкен болуы шарт (он мыңдаған сөздер), себебі олар негізінен мәтінді автоматты теруге арналған.



Қазіргі жүйелерде сөйлеуді тануда екі түрлі тәсіл қолданылады:

- Дыбыстық белгілерді тану.
- Лексикалық элементтерді тану.

Алғашқы тәсіл сөйлеу фрагменттерін алдын ала жазылып алынған эталон бойынша тануға саяды.

Бұл әдіс алдын ала жазылған сөйлеу командаларын орындау үшін арналған қарапайым жүйелерде кеңінен қолданылады.

Екінші тәсіл күрделірек. Оны жүзеге асыру кезінде жеке фонем, аллофон секілді лексикалық элементтер бөлініп алынады да буын және морфемдерге біріктіріледі.



Қазірде сөйлеуді тану жүйесін құру кезінде негізгі ұмтылыс келесі бағыттарда жасалады:

- Үздіксіз сөйлеу — пайдаланушыға табиғи (үздіксіз) сөз арасында кідіріс жасамай сөйлеуге мүмкіндік беру.
- Дикторға тәуелсіздік — жүйенің сөйлеу сигналын қайталау арқылы компьютерді дербес баптаусыз сөздерді тану мүмкіндігі.
- Үлкен көлемді сөздіктер — сөйлеуді тану жүйесінің қуаты мен тиімділігін арттыру мақсатында жалпы және арнайы пәндік облыстардан сөздердің үлкен санын тану мүмкіндігі.



Сөздік енгізу – шығару құрылғылары жылдам дамушы мультимедиа құралдарына жатады.

***Сөздік енгізу құрылғылары**, бұл – мысалы, адам айтатын әріптер мен сөздерді тануға және оларды кодтауға мүмкіндік беретін күрделі программалық жасауы бар түрлі микрофондық акустикалық жүйелер, «дыбыстық маустар».*

***Сөздік шығару құрылғылары** – бұл компьютерге қосылған дауыс динамиктер немесе дыбыстық колонкалар арқылы ойнатылатын, цифрлық кодтарды әріптер мен сөздерге айналдыруды орындайтын түрлі дыбыс синтезаторлары.*



SILK (SPEECH, IMAGE, LANGUAGE, KNOWLEDGE) – РОЯЗ (РЕЧЬ, ОБРАЗ, ЯЗЫК, ЗНАНИЯ) ИНТЕРФЕЙСИ

- Интерфейс наиболее приближен к обычной человеческой форме общения.
- Компьютер находит для себя команды, анализируя человеческую речь и находя в ней ключевые фразы.
- Результаты выполнения команд он также преобразует в понятную человеку форму.



SILK- ИНТЕРФЕЙС ДЛЯ ОБЩЕНИЯ ЧЕЛОВЕКА С МАШИНОЙ ИСПОЛЬЗУЕТ:

- речевую технологию;
- биометрическую технологию (мимический интерфейс);
- семантический (общественный) интерфейс.





РЕЧЕВАЯ ТЕХНОЛОГИЯ



- Это простейшая реализация SILK- интерфейса появилась в середине 90-х годов после появления недорогих звуковых карт и широкого распространения технологий распознавания речи.
- При этой технологии команды подаются голосом путем произнесения специальных стандартных слов (команд), которые должны выговариваться четко, в одном темпе с обязательными паузами между словами.
- Требуется индивидуальная предварительная настройка компьютерной системы на конкретного пользователя.

БИОМЕТРИЧЕСКАЯ ТЕХНОЛОГИЯ («МИМИЧЕСКИЙ ИНТЕРФЕЙС»)



- Возникла в конце 90-х годов и в настоящее время находится в стадии разработки.
- Для управления компьютером используется выражение лица, направление взгляда, размер зрачка и другие признаки человека.
- Для идентификации пользователя используется рисунок радужной оболочки его глаз, отпечатки пальцев и другая уникальная информация, которая считывается с цифровой камеры, а затем с помощью программы распознавания образов из изображения выделяются команды.



СЕМАНТИЧЕСКИЙ (ОБЩЕСТВЕННЫЙ) ИНТЕРФЕЙС

- Возник еще в конце 70-х годов XX века, с развитием искусственного интеллекта.
- Может не являться самостоятельным видом интерфейса, так как он включает в себя и интерфейс командной строки, и графический, и речевой, и мимический интерфейсы.





АНТИЧЕСКИЙ (ЕСТВЕННЫЙ) ИНТЕРФЕЙС



- Основной его особенностью является отсутствие команд при общении с компьютером. Запрос формируется на естественном языке, в виде связанного текста и образов.
- По сути - это **моделирование общения человека с компьютером.**
- В настоящее время используется для военных целей. Такой интерфейс крайне необходим в обстановке ведения воздушного боя.

