УДК 004.4.277

Вінічук І.М.

ПЕРСПЕКТИВИ ЗАСТОСУВАННЯ ТА ОЦІНЮВАННЯ ЯКОСТІ ІНТЕРАКТИВНИХ ГОЛОСОВИХ СИСТЕМ У СУЧАСНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЯХ

Національна академія керівних кадрів культури і мистецтва, e-mail: irbinzyk@ukr.net

Розглянуті проблеми визначення параметрів якості каналів передачі мовної інформації в сучасних інформаційних системах. Оцінені можливості практичного застосування інтерфейсу користувача голосового типу.

Вступ

Уведення інформації оператором в електронно-обчислювальну машину (ЕОМ) традиційним способом - з клавіатури є неприйнятним при певних ситуаціях. По-перше, коли введення з клавіатури заважає виконанню оператором інших функцій, а по-друге небезпечністю відволікання уваги оператора, якщо неможливо відстежувати набір тексту на пристрої відображення. В цих випадках використання голосу для вводу текстової інформації в інформаційних системах стає актуальним.

Пристрій, що міг тільки розпізнавати вимовлені людиною цифри створено на початку п’ятдесятих років минулого сторіччя. Програми по розпізнаванню мови з'явилися на початку дев'яностих років минулого сторіччя (наприклад, Dragon NaturallySpeaking, Voice Navigator). Ці програми перетворюють голос користувача в текст. На даний час ситуація швидко змінюється на краще завдяки збільшенню обчислювальних потужностей сучасних ЕОМ та наявності віброакустичних перетворювачів голосу на нових принципах.

Аналіз досліджень і публікацій

На даний час практичного застосування набув програмний продукт Microsoft Voice Command, який дозволяє працювати з багатьма програмами Microsoft за допомогою голосу. Є приклади практичних реалізацій, що дозволяють керування побутовими пристроями, транспортними засобами та ін.

Наступним щаблем розвитку інтерактивних голосових систем стали так звані інтелектуальні мовні рішення (IVR), що сприяють автоматичному синтезу й розпізнаванню людської мови в телефонії. Переваги, що доводять доцільність подібних рішень - зниження навантаження на операторів контакт-центрів і секретарів, скорочення витрат на оплату праці й підвищення продуктивності систем обслуговування. Спілкування з голосовим порталом є більш природним для людини. При цьому системи розпізнавання є незалежними від дикторів, тобто розпізнають голос будь-якої людини. Основною перевагою голосових систем є сприятливий інтерфейс користувача - він позбувається від необхідності користуватися складним голосових меню.

Розвиток інтерфейсів безмовного доступу Silent Speech Interfaces (SSI) можна вважати наступним кроком технологій розпізнавання мови. Ці системи обробки мови базуються на одержанні й обробці мовних сигналів на ранній стадії артикуляції. Даний етап розвитку розпізнавання мови викликаний двома істотними недоліками сучасних систем розпізнавання: надмірна чутливість до акустичних шумів, а також необхідність чіткої і ясної мови при звертанні до системи розпізнавання. Підхід, заснований на SSI, полягає в тім, щоб використовувати нові перетворювачі, на які мінімально впливають зовнішні акустичні завади. На даний час практично застосовуються голосове керування, голосовий пошук та голосовий набір номеру.

Голосове керування - це спосіб взаємодії із пристроєм за допомогою голосу. Широкому поширенню голосового керування заважає недостатня обчислювальна потужність електронних обчислювальних засобів і проблема наявності сторонніх (зовнішніх) акустичних завад.

Голосовий пошук - система розпізнавання мови, що дозволяє здійснювати переклад мовного запиту користувача в текстовий вид, що потім передається в стандартну систему пошуку по базі даних. Голосовий пошук реалізується в наступних напрямках:

- алфавітний довідник, пошук компанії по назві або категорії, пошук людини за списком;
- пошук інформації, такий як новини, фінанси, пробки, напрямок руху, погода або інформація з кінотеатрів (при цьому часто використовується керування багаторівневим голосовим меню);

- пошук в мережі інтернет;
- вибір опцій зі списку служб мобільного сервісу та ін.

Інша категорія додатків які можна вважати частиною голосового пошуку - "голосовий набір номеру" - пошук контакту в каталозі.

Перетворення «голосових» файлів у текст для подальшого текстового пошуку по них існують, наприклад, у таких програмах, які дозволяють перетворювати голосову пошту у текст для більш легкого пошуку й перегляду, а так само дозволяють подальше пересилання голосової пошти у вигляді електронної пошти або sms-повідомлення. Так само доступні послуги, що надають можливість залишати голосові замітки по телефоні й перетворювати їх у текст. Одна з головних цілей такого обслуговування полягає в тому, щоб зробити голосовий вміст, що легко архівується і зручним для пошуку. Цей спосіб відкриває значні перспективи в забезпеченні швидкого доступу до інформації, особливо на мобільних пристроях. Він надає деякі з переваг письмової мови, зберігаючи переваги розмовної мови.

В цьому сенсі стає актуальною акустична експертиза якості подання нформації у мовному форматі.

Постановка задачі

Оцінка якості мовних повідомлень базується на положеннях теорії розбірливості мови. Така оцінка може бути зроблена завдяки визначення параметру «розбірливість» (правильна артикуляція). Всі методи виміру розбірливості мови діляться на суб'єктивні, тобто вироблені за участю дикторів і аудиторів, і об'єктивні (рис.1), у яких замість дикторів і аудиторів використовуються спеціальні технічні пристрої - штучний голос, штучний рот і штучне вухо. Коректне співставлення цих методів є проблемою оцінки розбірливості мови. Особливо актуальною ця проблема стає при створенні інтерактивних голосових систем та систем синтезу мови. Таким чином постає задача визначення критеріїв, за якими можна порівнювати різні методи.

Фізичні основи каналу передачі інформації у мовному форматі

Згідно [7,6] в залежності від фізичної природи виникнення, способів реєстрації, технічні канали акустичної (мовної) інформації підрозділяють на повітряний, вібраційний, електроакустичний, оптико-електронний і параметричний.

Так, у вібраційних каналах здійснюється перетворення мовного сигналу в механічні вібрації, і навпаки. Прикладом вібраційного каналу є передача мовного сигналу за межі даного приміщення, обумовлена вібрацією стін, вікон, металевих радіаторів і т.п.

Радіоканали (радіотелефонний, мобільна зв'язок) містять системи перетворення акустичних сигналів в електромагнітні хвилі, і навпаки.

Основними типами каналів електроакустичного тракту є параметричний канал і оптико-електронний канал. Є приклади формування параметричного каналу базовані на зміні параметрів електронних схем, обумовлених впливом акустичних хвиль. У якості «носіїв» таких електронних схем можуть бути, наприклад, гетеродинні приймачі, телевізори й т.п. Інший варіант реалізації - високочастотне опромінення мініатюрних мікрофонів. Оптико-електронний канал може будуватися з використанням інфрачервоних лазерів, разом із прийомною апаратурою, що фіксує відбиття лазерного променя від шибок, дзеркал і т.д. («лазерні мікрофони»).

Рис.1. Об'єктивні методи оцінки якості мовних повідомлень

Найчастіше буває важко визначити конкретний тип каналу через його змішаний характер. Так, наприклад, в учбовому процесі мовний сигнал викладача сприймається мікрофоном, фільтрується, підсилюється й випромінюється спеціальний випромінювачем, до якого приєднані диктофони або спеціальні мініатюрні передавачі. Таким чином, мовна інформація може передаватися по радіоканалу, оптичному каналу, телефонній лінії, мережі змінного струму, допоміжним технічним засобам і системам.

У повітряних технічних каналах інформації середовищем розповсюдження акустичних сигналів є повітря, і для його перетворення в сигнал фізичної природи використовуються мікрофони різних модифікацій. Якщо система обробки звуку реалізована програмно-апаратними засобами (наприклад, на комп'ютері), то такий канал можна віднести й до класу цифрових.

Визначення розбірливості мови

Згідно [4], основними критеріями якості каналів мовної комунікації є:

- розбірливість;
- гучність;
- натуральність
- артикуляція.

Розбірливість (зрозумілість) мови є найважливішим параметром, що характеризує канали передачі мови.

На відміну від розбірливості мови, гучність не є самодостатнім параметром – її застосовують разом з розбірливістю. Нею позначають бажаний (комфортний) рівень прийнятих сигналів. З особистої практики кожної людини відомо, що надто низький рівень гучності приводить до зниження розбірливості. Розбірливість знижується й при надто високому рівні гучності мови.

Ще одним параметром, що характеризує канали мовної комунікації, є натуральність мови, що розуміється тобто здатність системи відтворювати не тільки зміст переданої мови, але і її тембр, індивідуальні особливості мови диктора. Цей параметр не настільки важливий, як розбірливість. Виключенням є ті випадки, коли вартим є завдання високоякісного відтворення мови диктора (або співу). У виключно технічних системах натуральність мови є другорядною, якщо тільки не маємо завдання визначення особистості диктора.

Значення розбірливості мови можна виміряти або розрахувати. При розрахунку й вимірюванні розбірливості мови необхідно мати у своєму розпорядженні критерії якості передачі мови взагалі, і міру розбірливості мови, зокрема. Наступний важливий момент - вибір способу оцінювання розбірливості мови. Так, розбірливість мови можна виміряти експериментально, а можна розрахувати теоретично (спрогнозувати розбірливість мови). Внаслідок наявності різних методів розрахунку й вимірювання, важливо вміти обґрунтовано вибирати раціональний метод - такий метод, що досить точний, з одного боку, і разом з тим, досить простий і дешевий - з іншого.

Визначення розбірливості мови можна робити як за участю дикторів і аудиторів (суб'єктивні методи вимірів), так і без їхньої участі (об'єктивні методи вимірів). Для розрахунків і вимірювання розбірливості мови необхідно мати у своєму розпорядженні математичну модель, що теоретично обґрунтовує особливості мовного тракту суб’єкта-джерела мовного сигналу, зміни в каналі передачі, а також особливості слухового тракту аудитора.

Так, в гучному приміщенні напружується голос і слух тим сильніше, ніж сильніше був навколишній шум, що відповідає ситуації малого співвідношення сигнал-шум.

Ситуація більших відносин сигнал-шум типовий для «каналів» передачі мовної інформації, таких як приміщення й відкритий простір, оснащені системами звукопідсилення або без таких, різних ліній зв'язку - телефонних, стільникових і т. ін. У таких системах рівень сигналу в точці прийому звичайно перевищує рівень завад. Завадами в такому разі є навколишній шум, а також шум, переданий разом з мовним сигналом по каналу передачі. При озвучуванні приміщень істотним може бути вплив завад у вигляді реверберації й навіть явища акустичного відлуння в дуже великих приміщеннях. В разі застосування систем звукопідсилення можливі завади у вигляді нелінійних спотворень, що виникають через перевантаження підсилювачів. У

загальному випадку можливо мати одночасний вплив всіх перерахованих факторів, що спотворюють мову й приводять до зменшення її розбірливості.

Ситуація малих відношень сигнал-шум типовий для систем захисту мовної інформації, оскільки завдання таких систем саме й полягають в навмисному та істотному послабленні сигналу в точці прийому. Говорячи про рівень захисту інформації, мають на увазі той або інший ступінь втрат інформації. Очевидно, при найвищому рівні захисту, в місці прийому взагалі повинні відсутні ознаки мови. Частіше, однак, задовольняються приховуванням, у тім або іншому ступені, змісту мовного повідомлення.

Суб'єктивний і об'єктивний підходи до вимірів розбірливості мови

На даний час для акустичної експертизи мовної комунікації застосовуються різноманітні підходи. Існують декілька «визнаних» версій формантного методу оцінки розбірливості мови, який зводиться, по суті, до двох стандартів AI и S11 [1] (рис. 1). Існують також формантні методи, які носять імена авторів: метод Н.Б. Покровського, метод Ю.С. Бикова и метод М.А. Сапожкова.

Розглядаючи суб'єктивні методи, часом розрізняють «чисто» суб'єктивні методи, за яких в оцінюванні розбірливості мови бере участь цілісна пара диктор-аудитор, і «об'єктивізовані» методи, де розбірливість мови оцінюють різні групи дикторів і аудиторів, з наступним усередненням отриманих оцінок [4, с.10].

Прикладом «чисто» суб'єктивного методу є випробування радіостанцій по рекомендаціях Міжнародного консультативного комітету з радіозв'язку [5]. При випробуваннях радіостанція працює в нормальним режимі, на передавальній стороні радіоканалу диктор читає текст, а на прийомній стороні радіоканалу аудитор виставляє оцінку розбірливості мови по п'ятибальній шкалі:

- нерозбірливо;
- періодично розбірливо;
- важкорозбірливо;
- розбірливо;
- повністю розбірливо.

Очевидний недолік «чисто» суб'єктивних підходів - неминучий вплив на результати вимірів, у якому мірою розбірливості є відношення числа правильно прийнятих по випробуваному каналі елементів. Таким чином, вплив суб'єктивних особливостей окремих операторів усереднюється, у результаті чого артикуляційні виміри дають статистично стійкі, а тому досить об'єктивні результати. По закінченні вимірів розбірливості, тобто після того, як визначений відсоток правильно розпізнаних складів, виникає питання про оцінку класу якості випробуваного об'єкта (тобто відповідає об'єкт пропонованим до нього вимогам чи ні). Якщо випробуваним об'єктом є апаратура зв'язку, то застосовують норми, погоджені між різними відомствами, що виробляють і експлуатують таку апаратуру. По суті, наявність саме останньої процедури - класифікації об'єкта, дозволяє говорити не просто про визначення розбірливості мови, а про акустичну експертизу [2].

Процедура артикуляційних випробувань стандартизована - прикладом може служити Державний стандарт Російської Федерації ГОСТ 50840-95 [3]. Один з найпростіших методів такого роду, названий у роботі [6] «об'єктивним».

Даний метод точніше й швидше тонального, для його проведення не потрібні оператори (диктори й аудитори). Нарешті, даний метод принципово дозволяє повністю автоматизувати процедуру вимірів на базі сучасних ЕОМ.

Як і тональний метод, «об'єктивний» метод є непрямим, тобто розбірливість мови оцінюється не шляхом підрахунку правильно розпізнаних мовних одиниць, а шляхом проведення спеціального вимірювального експерименту зі звуковими сигналами у вигляді тону й смугового шуму, у ході якого виміряються рівні відчуттів у декількох смугах частот. Далі ці рівні відчуттів за спеціальною методикою перераховуються в розбірливість мови.

Вочевидь, об'єктивні методи вимірів розбірливості мови досить перспективні в силу можливості значного прискорення й здешевлення процедури вимірів. Так, наприклад, час вимірів у сучасних апаратно-програмних вимірювальних системах становить одиниці хвилин і навіть секунд.

Для реалізації об'єктивних методів визначення розбірливості мови, а також для створення математичної моделі каналу передачі мови необхідно мати у дослідити властивості мовного й слухового трактів, оскільки вони є складовими частинами каналу передачі мовної інформації.

Спектральні властивості звуків мови

Спектральні розходження між звуками мови є головними (хоча й не єдиними) критеріями, що їх визначають. Спектри голосних звуків являють собою (у першому наближенні) періодичну послідовність спектральних піків. Відстань між цими піками визначається частотою основного тону. Виражені сплески рівня огинаючої спектральних піків іменують «формантами» .

Спектри голосних звуків або повністю суцільні, тобто зовсім не містять дискретних компонентів, або суцільні в окремих смугах частот. Ці спектри також містять локальні сплески. Деякі з них є формантами, деякі - ні [6].

Щоб вирішити, які сплески рівня спектра є формантами, варто врахувати, що фізична природа формант - явище резонансу в порожнинах глотки й носоглотки . В окремих звуках можна помітити до 6 спектральних підйомів. До формант відносять тільки ті, які обумовлені явищем резонансу в мовному апараті людини. Частина формант (одна-дві в російській мові) забезпечує розбірливість мови, інші форманти забезпечують індивідуальність голосу диктора, що може бути використане в завданнях розпізнавання голосу (ідентифікації).

Висновки

Програми по розпізнаванню мови є дуже перспективними засобами формування інтерфейсу користувача засобів інформаційних систем завдяки розвантаженню його рук для інших функцій, зменшують навантаження на психофізичний апарат, природності і простоті процесу взаємодії машина-людина. Це може сприяти виходу на якісно новий рівень функціонування соціокультурної сфери сучасного суспільства. Практичне застосування доступних програмових засобів цього типу таких, як додаток Microsoft Voice Command сприятиме напрацюванню нових форми роботи в різних галузях. Інтерфейси безмовного доступу на базі одержання й обробки мовних сигналів на ранній стадії артикуляції з використанням акусто-електричних перетворювачів захищених від зовнішніх акустичних завад є перспективною технологією розпізнавання мови. При визначенні розбірливості (зрозумілості) мови, доцільно користуватися вимірювальними методами. Метод артикуляції незважаючи на універсальність і порівну простоту є громіздким, тривалим і має високу вартість процедури визначення.

Список використаних джерел

1. ANSI S3.5-97, American National Standard Methods for calculation of the Speech Intelligibility Index - American National Standards Institute. - New York, 1997. - 35 p.

2. Дидковский, В. С. Акустическая экспертиза каналов речевой информации / В.С. Дидковский, М.В. Дидковская, А.Н. Продеус. – К., 2008. – 420 с.

3. Калинцев, Ю. К. Разборчивость речи в цифрових вокодерах / Ю.К. Калинцев. – М.: Радио и связь, 1991. – 219 с.

4. Покровский, Н. Б. Расчет и измерение разборчивости речи / Н.Б. Покровский. – М.: Связьиздат, 1962. - 390 с.

5. Сапожков, М. А. Вокодерная связь / М.А. Сапожков, В.Г. Михалков. - М.: Радио и связь, 1983.- 246 с.

6. Хорев, А. А. Классификация и характеристика технических каналов утечки информации, обрабатываемой ТСПИ и передаваемой по каналам связи (http//st.ess.ru//publications/articles/ /tsp/tspi.htm)

Хорев, А. А. Технические каналы утечки акустической (речевой) информации (http//st.ess.ru//publications/articles/tsp/tspi.htm).