

СУЧАСНІ МЕТОДИ АНАЛІЗУ МОВНОГО СИГНАЛУ

Розглянуто сучасні методи аналізу мовного сигналу в інформаційних технологіях передачі інформації. Наведено основні методи кодування мовного сигналу, а також методи синтезу мовного сигналу.

The modern methods of the analysis in information technologies of transfer of the information are considered. The basic methods of coding of a speech signal, and also methods of synthesis of a speech signal are given.

Вступ

Ця стаття присвячена огляду сучасних методів моделювання мовних сигналів, результати використання яких наведені у значній кількості наукових публікацій [1—4]. Методи обробки мовних сигналів активно розвивалися протягом останніх десятиріч стосовно широкого кола завдань — від техніки зв'язку до читаючих автоматів. Для синтезу мови у більшості випадків використовують два класи синтезаторів: синтезатори, що відтворюють передатну функцію голосового тракту, і синтезатори — аналоги фізичного каналу голосового тракту. Серед методів моделювання реалізації мовного сигналу розрізняють: спектральний аналіз на коротких інтервалах часу та синтез мови; гомоморфний аналіз та синтез мовних сигналів; синтез методом лінійного прогнозу; моделювання мови з використанням передатної функції; хвильовий метод кодування; метод параметричного уявлення; синтез за правилами [2].

Постановка проблеми

Методи аналізу та синтезу мовних сигналів можна розглядати за допомогою моделі, у якій мовний сигнал є відгуком системи з повільно змінними параметрами на періодичне чи шумове збуджувальне коливання. Взагалі механізм, що формує мову (голосовий тракт) — це акустична трубка, що збуджується відповідним джерелом при утворенні бажаного звуку. Для дзвінких звуків джерелу збудження відповідає квазіперіодична послідовність імпульсів, що становлять потік повітря, який протікає крізь голосові зв'язки, що коливаються. Фрикативні звуки утворюються при виштовхуванні повітря крізь звуження голосового тракту. При цьому виникає турбулентний потік, котрий є джерелом шуму, що збуджує голосовий тракт, який наведено на рисунку.

Існують різні методи синтезу мови. Вибір того або іншого методу визначається різними обмеженнями. Розглянемо ті чотири види обмежень, які впливають на вибір методу синтезу.

Можливості синтезованої мови залежать від того, в якій області вона буде застосовуватися. Коли необхідно вимовляти обмежене число фраз (і їх вимовлення лінійно не змінюється), необхідний мовний матеріал просто записується на плівку. З іншого боку, якщо завдання полягає в стимулюванні пізнавального процесу при читанні вголос, використовується абсолютно інший ряд методик.

Голосовий апарат людини

Усі системи синтезу мови мають утворювати на виході якусь мовну хвилю, але це не довільний сигнал. Щоб сформувати мовну хвилю певної якості, сигнал проходить шлях від джерела в мовному тракті, який порушує дію артикуляторних органів, які діють ніби фільтри, що змінюються в часі.

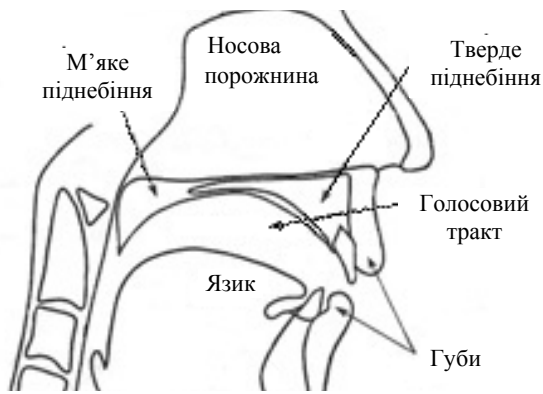


Схема голосового апарату людини

морфологічні і синтаксичні структури, фонотак-тичні обмеження.

Артикуляторні органи також накладають обмеження на швидкість зміни сигналу. Вони також мають функцію згладжування: гладкого зчеплення окремих базових фонетичних одиниць у складний мовний потік.

Структура мови

Ряд можливих звукових поєднань визначається природою тієї або іншої мовної структури. При дослідженні мовного сигналу було виявлено, що одиниці і структури, використовувні лінгвістами для опису і пояснення мови, можуть також використовуватися для характеристики і побудови мовної хвилі. Таким чином, при побудові вихідної мовної хвилі використовуються основні фонологічні закони, правила наголосу,

Математичний апарат дослідження

При дослідженні характеристик голосового тракту й основної моделі утворення мовних сигналів обґрунтовано, що голосовий тракт може описуватися за допомогою раціональної передатної функції. Теоретично розроблено декілька методів моделювання мовного сигналу на основі передатної функції, що мають полюси та нулі. Методи, що вимагають проведення операцій з великими матрицями, і способи з ітеративною оптимізацією звичайно виявляються малоприйнятними для аналізу мови, коли ставляться високі вимоги до швидкодії та об'єму пам'яті й часто вимагається обробка сигналів у дійсному масштабі часу. Тому найбільш прийнятними виявляються методи, засновані на критерії найменших квадратів і сформульовані так, щоб вони зводилися до розв'язання лінійних рівнянь. Звичайно ця умова приводить до методів, коли спочатку визначаються полюси, а потім нулі, а не одночасно одне й друге. Оцінити полюси незалежно від нулів можна за допомогою автоковаріаційного варіанта методу лінійного прогнозу.

Щоб описати імпульсну характеристику голосового тракту $u(n)$ моделлю, що містить полюси та нулі, розглянемо послідовність $u(n)$

$$u(n) = \sum_{k=0}^q b_k \delta(n-k) + \sum_{k=1}^p a_k u(n-k). \quad (1)$$

При $n > q$, де q — загальна кількість нулів, рівняння набуде вигляду

$$u(n) = \sum_{k=1}^p a_k u(n-k). \quad (1a)$$

Таким чином, при будь-якому $n > q$ послідовність $u(n)$ можна лінійно передбачити за її p попередніми відліками. Отже, навіть за наявності в моделі нулів полюси можна оцінити, визначивши в рівнянні коефіцієнти a_k , що мінімізують помилку прогнозу. Але оскільки рівняння справедливе лише при $n > q$, у даному випадку необхідно виконувати умови, щоб значення $u(n)$ не виходили за ці межі, навіть якщо після інверсної фільтрації $u(n)$ відома при всіх n . Припустимо, що $e(n)$ означає помилку прогнозу

$$e(n) = u(n) - \sum_{k=1}^p a_k u(n-k), \quad n > k. \quad (2)$$

Виберемо коефіцієнти a_k так, щоб сумарна середня квадратична похибка прогнозу

$$E_T = \sum_{n=q+1}^{\infty} e^2(n), \quad (3)$$

була мінімальною.

У результаті після ряду перетворень отримаємо систему рівнянь для коефіцієнтів оператора прогнозу a_k у вигляді

$$\sum_{k=1}^p a_k \varphi_{ik} = \varphi_{i0}, \quad i = 1, 2, \dots, p, \quad (4)$$

$$\varphi_{ik} = \sum_{n=q+1}^{\infty} u(n-i)u(n-k). \quad (4a)$$

Матриця коефіцієнтів Φ_{ik} симетрична, але не є теплицевою. Після визнання коефіцієнтів a_k можна використати метод запропонований Шенксом, і знайти коефіцієнти b_k у рівнянні через мінімізацію середньої квадратичної похибки між $u(n)$ та імпульсною характеристикою шуканого фільтра з полюсами та нулями. У результаті отримуємо систему лінійних рівнянь

$$\sum_{k=0}^q u_k R_{ik} = V_i, \quad i = 0, 1, \dots, q, \quad (5)$$

де
$$R_{ik} = \sum_{n=0}^{\infty} w(n-i)w(n-k), \quad (5a)$$

причому послідовність $w(n)$ — імпульсна характеристика фільтра, який має лише полюси, тобто

$$w(n) = \sum_{k=1}^p a_k w(n-k) + \delta(n) \quad (6)$$

і

$$V_i = \sum_{n=0}^{\infty} u(n)w(n-i). \quad (7)$$

Отже, теоретично розроблено метод моделювання мовного сигналу на основі передатної функції, що мають нулі та полюси.

Технологія

Можливості успішно моделювати і створювати пристрої для синтезу мови значною мірою визначаються техніко-технологічними характеристиками пристрою. Мовна наука зробила великий крок уперед завдяки появі різних технологій, зокрема: рентгенографії, кінематографії, теорії фільтрів і спектрів, а головним чином — цифровим комп'ютерам. З приходом інтегральних мережевих технологій з постійно зростаючими можливостями стала можливою побудова потужних, компактних, недорогих пристроїв, що діють в реальному часі. Цей факт, разом із ґрунтовними знаннями алгоритмів синтезу мови, стимулював подальший розвиток систем синтезу мови і перехід їх в практичне життя, де вони широко застосовуються.

Методи синтезу

Різні підходи до аналізу мови можуть бути згруповані за сферами їх застосування, за складністю їх втілення. Синтезатори ділять на два типи: з обмеженим і необмеженим словником. У пристроях з обмеженим словником мова зберігається у вигляді слів і пропозицій, які виводяться в певній послідовності при синтезі мовного повідомлення. Мовні одиниці, використовувані в синтезаторах подібного типу, вимовляються диктором заздалегідь, а потім перетворюються в цифрову форму за допомогою різних методів кодування, що дає змогу компресувати мовну інформацію і зберігати її в пам'яті синтезатора. Існує кілька методів запису і компонування мови.

Хвильовий метод кодування

Найпростіший шлях — записати матеріал на магнітну плівку і при необхідності відтворювати. Цей спосіб забезпечує високу якість синтезованої мови, оскільки дає змогу відтворити форму природного мовного сигналу. Проте цей шлях синтезу не надає можливості реалізувати побудову нової фрази, оскільки не передбачає звернення до різних осередків пам'яті і виклику з пам'яті потрібних слів. Залежно від використовуваної технології цей спосіб може затримувати доступі і мати обмеження, пов'язані з можливостями запису. Ніяких знань про будову мовного тракту і структуру мови не потрібні. Єдине серйозне обмеження в даному випадку має об'єм пам'яті для запису сигналу.

Існують способи кодування мовного сигналу в цифровій формі, що дають змогу у кілька разів ущільнювати інформацію: проста модуляція даних, імпульсно-кодова модуляція, адаптивна дельтова модуляція, адаптивне кодування. Дані способи можуть зменшити швидкість передачі даних

від 50 кбіт/сек (типовий варіант) до 10 кбіт/сек, тоді як основні характеристики мовного сигналу зберігаються. Природно, складність операцій кодування і декодування збільшується зі зниженням числа біт за секунду. Такі системи ефективні, коли словник повідомлень невеликий і фіксований. У разі ж, коли потрібно з'єднати довге повідомлення, згенерувати високоякісну мову важко, оскільки значення параметрів мовної хвилі не можна змінити, а вони можуть не підійти в новому контексті. В усіх системах синтезу мови встановлюється деякий компроміс між якістю мови і гнучкістю системи.

Збільшення гнучкості неминує веде до ускладнення обчислень.

Параметричне уявлення

Задля подальшого зменшення необхідної пам'яті для зберігання і забезпечення необхідної гнучкості мовного сигналу було розроблено кілька способів, які абстрагуються від мовної хвилі як такої, а подають її у вигляді набору параметрів. Ці параметри відображають найбільш характерну інформацію або в часовій, або в частотній області. Наприклад, мовна хвиля може бути сформована складанням окремих гармонік заданої висоти і заданими спектральними виступами на даній частоті. Альтернативний шлях полягає в тому, щоб форму мовного тракту описати в термінах акустики і штучно створити набір резонансів. Цей метод синтезу економічніше хвилевого, оскільки вимагає значно меншого об'єму пам'яті, але при цьому він вимагає більше обчислень, щоб відтворити початковий мовний сигнал.

Даний спосіб дає можливість використовувати ті параметри, які відповідають за якість мови (значення формант, ширина смуг, частота основного тону, амплітуда сигналу). Це надає змогу синтезувати сигнали, так що переходи на межах абсолютно не помітні. Зміни таких параметрів як частота основного тону впродовж всього повідомлення дають можливість істотно змінювати інтонацію і часові характеристики повідомлення. Найбільш популярними сьогодні методами кодування в пристроях, що використовують параметричне зображення сигналів, є метод, заснований на формантних резонансах, і метод лінійного прогнозу (LPC — *linear predictive coding*). Для синтезу використовуються одиниці мови різної довжини: параграфи, пропозиції, фрази, слова, склади, півсклади, дифони. Чим менше одиниця синтезу, тим менша їх кількість потрібна для синтезу.

При цьому потрібно більше обчислень, і виникають труднощі коартикуляції на стиках. Переваги цього методу: гнучкість, малий об'єм пам'яті для зберігання початкового матеріалу, збереження індивідуальних характеристик диктора. Потрібна відповідна цифрова техніка і знання моделей мовоутворення, при цьому лінгвістична структура мови не використовується.

Методи синтезу

Описані вище методи синтезу орієнтовані на такі мовні одиниці, як слова, заздалегідь введені в пристрій з голосу диктора. Цей принцип лежить в основі функціонування синтезаторів з обмеженим словником. У синтезаторах з необмеженим словником елементами мови є фонемні або склади, тому в них застосовується метод синтезу за правилами, а не просте компонування. Такий метод перспективний, оскільки забезпечує роботу з будь-яким необхідним словником, проте якість мови значно нижча ніж при використанні методу компонування.

При синтезі мови за правилами також використовуються хвильовий і параметричний методи кодування, але вже на рівні складів.

Параметричний метод потребує компромісу між якістю мови і можливістю змінювати параметри. Дослідники виявили, що для синтезу мови високої якості необхідно мати кілька різних вимов одиниці синтезу (наприклад, складу), що веде до збільшення словника початкових одиниць без будь-яких відомостей про контекстну ситуацію, що визначає той чи інший вибір. Тому процес синтезу набуває більш абстрактного характеру і переходить від параметричного зображення до розробки набору правил, за якими обчислюються необхідні параметри на основі ввідного фонетичного опису. Він сам по собі містить мало інформації. Це звичайно імена фонетичних сегментів (наприклад, голосні і приголосні) із знаками наголосу, позначеннями тону і часових характеристик.

Таким чином, метод синтезу за правилами використовує малоінформаційний опис на вході (менше 100 біт/сек). Цей метод дає повну свободу моделювання параметрів, але необхідно підкреслити, що правила моделювання недосконалі. Синтезована мова гірша натуральної, проте вона задовольняє тести на розбірливість і зрозумілість. На рівні пропозиції правила надають необхідний ступінь свободи для створення плавного мовного потоку.

Оцінка синтезованої мови

З погляду зрозумілості, розбірливості синтезована мова досить якісна. Був проведений тест, де одна досліджувана група прослуховувала синтезовану мову з письмовим варіантом перед очима, а інша — без. З'ясувалося, що результати прослуховування мало відрізняються один від одного. Проте синтезованій мові не вистачає жвавості і природності, тому сприймати її впродовж тривалого часу важко. Дослідження показали, що фрикативні і назальні звуки вимагають подальшого поліпшення якості.

Висновки

У даній праці були представлені основні ідеї та методи цифрової обробки мовних сигналів. Більш детально розглядалися методи та системи, які опираються на основну модель формування мови. Переважна частина досліджень стосувалася того, що головною проблемою в обробці мовних

сигналів є інверсна фільтрація сигналу на компоненти, що окремо являють собою характеристики збуджувального сигналу та голосового тракту.

Усі розглянуті методи добре узгоджуються із сучасним станом цифрової техніки і дають змогу застосовувати при створенні апаратури великі інтегральні схеми, мікропроцесори і т. д. Оскільки подібні пристрої дуже зручні для створення систем спектрального аналізу з високою роздільною здатністю, вони можуть стати базою для недорогих мовних систем, в яких необхідний такий аналіз. Отже, можна очікувати продовження досліджень і прискореного розвитку всіх методів і систем.

У сфері обробки мовних сигналів залишається ще багато невирішених проблем. При вивченні можливостей застосування розглянутих методів слід враховувати характер умов, на яких вони засновані. Однак при подальшому розвитку алгоритмів обробки мови, мабуть, дедалі більше буде враховуватися непостійність характеристик систем в часі, більше застосовуватимуться методи аналізу таких систем, як оптимальне управління і моделювання систем зі змінними параметрами.

ЛІТЕРАТУРА

1. *Аналогово-цифровой* процессор для вычисления акустических параметров речевого сигнала. Пакет программного обеспечения первичного анализа речевых сигналов для ПИЭМ / Л. С. Чудновский, В. М. Агеев, Н. В. Ко-ролев и др. // Автоматическое распознавание и синтез речевых сигналов. — К., 1989. — С. 62—64.
2. *Карпов О. Н.* Методы анализа и распознавания речевых сигналов. — Днепропетровск, 2001. — С. 25—36.
3. *Steward J. Q.* An Electrical Analogue of the Vocal Organs, *Nature*, 110, 311—312 (1922).
4. *Golden R.M.* Digital Computer Simulation of a Sampled-Data Voice Excited Vocoder, *J. Acoust. Soc. Amer.*, 35, 1358—1366 (1963).
5. *Применение* цифровой обработки сигналов // Под ред. Э. Оппенгейма // Цифровая обработка речевых сигналов. — М., 1980. — С. 185—192.