

ВИКОРИСТАННЯ ВЕЙВЛЕТ-ПЕРЕТВОТЕННЯ ДЛЯ ПРЕДСТАВЛЕННЯ МОВНИХ СИГНАЛІВ У ЗАДАЧІ РОЗПІЗНАВАННЯ МОВИ

Вступ

У сучасному урбанізованому і комп'ютеризованому світі однією з найбільш прибуткових комерційних діяльностей є інформаційні технології і суміжні з ними галузі, зокрема мовні технології. Це навіть у деякій мірі стає показником рівня розвитку науки в країні. Бурхливий розвиток інформаційного і телекомунікаційного ринків, а разом з ним і апаратно-програмної бази, дозволяє братися за реалізацію проектів, що є алегоричними ознаками цивілізації третього тисячоріччя.

У літературі та кінематографі фантастичного жанру одним з найяскравіших прикладів техніки майбутнього довго служили системи розпізнавання мови. І тільки в останні десятиліття високий рівень розвитку технологій в області створення високоефективних процесорів цифрової обробки сигналів уможливило рішення задачі розпізнавання мови в реальному масштабі часу. Однак основним стимулом розвитку таких систем є практичні застосування, що стали можливими в сучасному інформаційному суспільстві, що прагне до всеосяжної інформатизації й інтелектуалізації всіх життєвих процесів.

Розглядаючи можливості застосування систем розпізнавання мови можна виділити наступні пріоритетні напрямки:

- застосування мовного введення команд при неможливості використання альтернативних методів керування;
- використання системи розпізнавання мови як мовний інтерфейс у мініатюрних комп'ютерних системах і засобах комунікації;
- використання розпізнавання мови для спрощення операцій, що вимагають складного шляху через різні ієрархії інтерфейсу;
- створення сучасних телекомунікаційних довідкових служб, а також підтримка банківських операцій з використанням процедури аутентифікації клієнта ;
- розробка на основі сучасних технологій пошуку в інформаційних мережах мовної інформації по заданих ключових словах або проблема-тиці.

Постановка задачі

Задача розпізнавання мови полягає в адекватному відображенні символами тієї або іншої мови вимовленої звукової послідовності. Задачу

© Л.О. Фаніна, А.І. Папченко, 2004

розпізнавання мови вважають зосередженням усіх задач штучного інтелекту [1].

У загальному випадку задачу можна представити як неоднорідний багатостадійний процес перетворення вхідного мовного сигналу, що представляє собою коливання тиску повітряного середовища, формування векторів ознак, вироблення гіпотез і відшукування найбільш оптимальної гіпотези, відповідно до обраної логіки прийняття рішень.

Схематично процес розпізнавання представлений на рис.1.



Рис. 1 – Процес обробки мовного сигналу в системах розпізнавання мови

Більшість параметричних систем розпізнавання мови включають два основних етапи. Найбільш значимим фактором, що визначає якість розпізнавання, є ефективне параметричне представлення мовного сигналу, тобто послідовний процес оцифровки, фільтрації, сегментації, зважування.

Оптимальний вибір методів обробки мовного сигналу повинен відповідати наступним вимогам:

Аналоговий мовний сигнал від деякого джерела необхідно перетворювати в цифровий з максимальним співвідношенням сигнал/шум велика акустична мінливість мовного сигналу та широкий різновид зовнішніх умов вимагають використання різних алгоритмів фільтрації для усунення деяких небажаних явищ у сигналі.

З цілком зрозумілих причин часові відліки мовного сигналу неможливо використовувати як параметри для розпізнавання мови. Дві часові реалізації одного і того ж виспіву можуть значно відрізнятися як за формою, так і по амплітуді. (Темп мови, тривалість вимови окремих слів і звуків навіть для одного диктора варіюється в дуже широких межах. На рис.2 фонограми слова “мама”, вимовлені тим самим диктором з інтервалом у 5с.)

Таким чином, можливі значні розбіжності між окремими ділянками збереженого еталона і відповідного йому вхідного акустичного сигналу за рахунок часової неузгодженості. У цьому зв'язку звичайно застосовують різні методи спектрального аналізу для одержання адекватного набору параметрів, що відбивають внутрішні характеристики мовного сигналу.

Найчастіше повний опис мовного сигналу тільки в термінах основних параметрів неможливо. Спектральні ознаки, що містять в собі значиму частотну інформацію про сигнал, не здатні врахувати динаміку мовного сигналу. З цією метою використовують додаткові так названі дельта-параметри, які є похідними за часом різних порядків від основних параметрів [2].

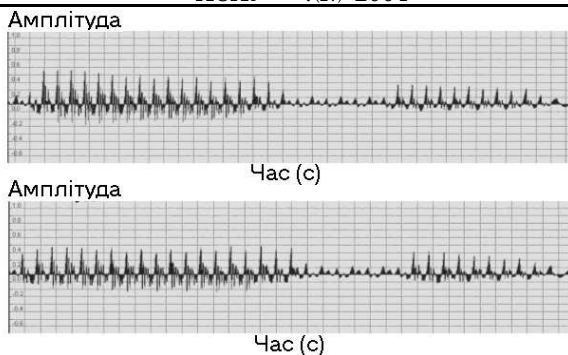


Рис. 2 – Фонограма слова “мама”

Таким чином, недосконалість сучасних систем розпізнавання мови, як складового елемента мовного інтерфейсу в значній мірі зв’язано з низькою ефективністю математичних методів моделювання сигналу. Для рішення цих проблем приходиться знаходити нові методи й алгоритми відновлення вимовленої звукової послідовності.

Обмеження перетворення Фур’є щодо задачі розпізнавання мови

Більшість методів моделювання й обробки мовних сигналів у задачі розпізнавання мови (спектральний аналіз, кепстральне представлення) спираються на перетворення Фур’є (ПФ), коефіцієнти Фур’є або кепстральні коефіцієнти, що одержано у результаті перетворення піддаються досить простій фізичній інтерпретації. Однак тому що в якості основних базисних функцій використовуються синуси, косинуси та їхні комплексні експоненти перетворення Фур’є має ряд обмежень щодо задачі розпізнавання мови:

- втрата інформації в часі існування частотних компонентів сигналу;
- вихідний сигнал замінюється на періодичний, з періодом рівним тривалості досліджуваного зразка;
- ПФ не дає інформацію про динаміку зміни спектрального складу сигналу;
- ПФ погано працює при зміні параметрів процесу згодом (нестационарності), оскільки дає усереднені коефіцієнти для всього досліджуваного сигналу;
- ПФ вимагає знання сигналу не тільки в минулому, але й у майбутньому, що є неможливим для систем реального часу;
- погана пристосованість до сигналів з локальними особливостями.

У загальному випадку при довільній залежності $f(t)$ пряме ПФ означає перехід від часового представлення сигналу до його частотного представлення відповідно до узагальненого вираження:

$$F(w) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt. \quad (1)$$

У деяких випадках фізична інтерпретація за допомогою цієї формули буває досить складною. Так, щоб одержати спектральну інформацію на обраній частоті, необхідно мати і минулу, і майбутню інформацію, що може бути застосовано лише до стаціонарних сигналів, у яких частота не міняється згодом. До того ж формула (1) не враховує, що частота може еволюціонувати згодом.

Зворотнє перетворення забезпечує відновлення сигналу:

$$f(t) = \frac{1}{2\pi} \int F(w) e^{j\omega t} dw. \quad (2)$$

Це перетворення унітарне:

$$\int |f(t)|^2 dt = \frac{1}{2\pi} \int |F(w)|^2 dw. \quad (3)$$

Формула (3) є рівністю Парсеваля і вказує на збереження енергії при переході від часової області до частотного.

Таким чином, частотно-часова смуга виявляється добре локалізованою по частоті і практично необмеженою по осі часу. Більш того, як помітно з формули (1), сигнал $f(t)$ повинен досить швидко спадати на нескінченності в минулому й у майбутньому.

Спроби подолання цих труднощів і поліпшення розуміння локальних властивостей сигналу на теперішній час реалізують за допомогою так названого короткочасного або віконного перетворення Фур'є. Сигнал $f(t)$ аналізується лише усередині деякого вікна і виявляється локалізованим у часі. Але при короткочасному перетворенні Фур'є вікно має фіксований розмір, що не залежить від масштабу, що розглядається [4].

У задачах розпізнавання мовних сигналів особливу роль грає локалізація і класифікація особливих характеристик сигналу (піки й розриви, різкі перепади рівня спектра), частотно-часовий аналіз. Перетворення Фур'є не дозволяє в точній мірі відбити локальні особливості сигналів, тому що його базисною функцією є синусоїда, що робить застосування методів ПФ в аналізі мовних сигналів небажаним.

Основні поняття вейвлет-перетворення

Альтернативою можуть стати вейвлети, основним застосуванням яких є аналіз і обробка нестаціонарних (у часі) або неоднорідних (у просторі) сигналів.

Вейвлети - це математичні функції, що дозволяють аналізувати різні частотні компоненти даних. Вейвлет-перетворення дозволяє судити не тільки про частотний спектр сигналу, але також про те, у який момент часу з'явилася та або інша гармоніка. З їхньою допомогою можна

легко аналізувати переривчасті сигнали, або сигнали з гострими сплесками. Крім того, вейвлети дозволяють аналізувати дані відповідно до масштабу. Унікальні властивості вейлетів дозволяють сконструювати базис, у якому представлення даних буде виражатися лише декількома ненульовими коефіцієнтами. Ця властивість робить вейвлети дуже ефективними для аналізу даних, цифрової обробки зображень та обробки складних акустичних сигналів.

Вейвлет-перетворення у великому ступені дозволяє подолати перераховані недоліки перетворення Фур'є, оскільки базисні функції мають властивість часової локалізації, тобто мають кінцеву енергію (норму):

$$E_f = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty. \quad (4)$$

У даній роботі для представлення мовних сигналів пропонується використання вейвлетного базису, що має ряд переваг:

- локалізація в часовій і частотній області, що дозволяє робити ефективний частотно-часовий аналіз нестационарних сигналів, прикладом яких є мова;
- можливість масштабного перетворення й зсувів (стиск-розтягання досліджуваного сигналу за допомогою хвильових функцій різної періодичності);
- математичний апарат, який було розроблено для локалізації й класифікації особливих точок сигналу;
- відмінно відображають динаміку зміни сигналу уздовж „осі масштабів”, локалізацію різномасштабних деталей (спектр мови людини радикально змінюється з часом, а характер цих змін - дуже важлива інформація).

Аналіз мовних сигналів може стати класичним прикладом застосування вейвлет-перетворення, загальний принцип побудови, базису якого складається у використанні масштабного перетворення й зсувів. У такий спосіб за допомогою дискретних масштабних перетворень і зсувів можна описати всі частоти і покрити усю вісь, використовуючи один базисний вейвлет.

Для практичного застосування дуже важливо знати ознаки, якими повинна володіти функція, щоб бути вейвлетом:

Локалізація. Вейвлет перетворення у відмінності від ПФ використовує локалізовану базисну функцію. Вейвлет повинний бути локалізований і в часовому просторі і по частоті.

Нульове середнє:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0. \quad (5)$$

Часто для практичного застосування виявляється необхідним, щоб не тільки нульовий, але й усі перші m моментів дорівнювали нулю:

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0. \quad (6)$$

Такий вейвлет називається вейвлетом m -го порядку. Вейвлети, що володіють великим числом нульових моментів, дозволяють аналізувати дрібномасштабні флуктуації й особливості високого порядку.

Обмеженість:

$$\int |\psi(t)|^2 dt < \infty. \quad (7)$$

Оцінка гарної локалізації й обмеженості може бути записана у вигляді:

$$|\psi(t)| < (1 + |t|^n)^{-1} \quad \text{или} \quad |\hat{\psi}(w)| < (1 + |k - w_0|^n)^{-1}, \quad (8)$$

де w_0 - домінуюча частота вейвлета, число n повинне бути як можливо великим.

Автоподібність базису. Характерною ознакою базису вейвлет-перетворення є його самоподоба. Усі вейвлети даного сімейства $\psi_{ab}(t)$ мають те ж число осциляцій, що й базисний вейвлет $\psi(t)$, оскільки отримані з нього за допомогою масштабних перетворень і зсувів.

Для побудови базису вейвлет-перетворення використовується одна функція, що іменується материнським вейвлетом - $\varphi(t)$. На рис.3 представлені приклади часто використовуваних вейвлетів: а) WAVE; б) мексиканський капелюх; в) Морле; г) Пауля; д) LMB; е) Добеши. Праворуч показані вейвлети в залежності від часу $\psi(t)$, ліворуч – їхні образи Фур’є $\hat{\psi}(w)$ [3].

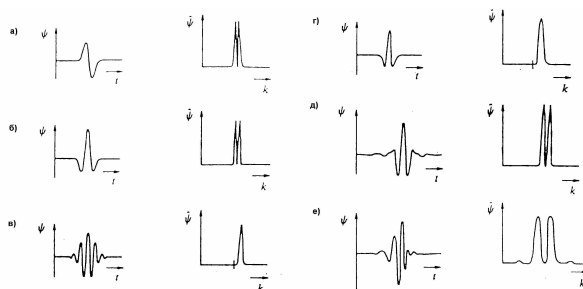


Рис. 3 – Приклади найбільш часто використовуваних вейвлетів

Слід зазначити, що вейвлет-перетворення не завжди може підміняти Фур’є-аналіз. Багато теорем вейвлет-аналізу доводяться за допомогою

розкладання Фур'є. Таким чином, ці два типи аналізу є скоріше доповнучими одне одного, ніж такими, що змагаються у застосуваннях.

Застосування вейвлет-перетворення до моделювання мовних сигналів

Мовний сигнал $s(t)$ відповідно вейвлет-перетворенню прийме вид:

$$s(t) = c_0 \varphi(t) + \sum_{j=1}^N \sum_{k=0}^{2^j-1} c_{jk} \psi_{jk}(t), \quad (9)$$

де $\psi(t)$ – функція, що деталізує, з нульовим значенням інтеграла, визначає особливості сигналу і породжує коефіцієнти, що деталізують $\varphi(t)$ – масштабуюча функція з одиничним значенням інтеграла, визначає грубе наближення сигналу і породжує коефіцієнти апроксимації. Функції $\varphi(t)$ властиві лише ортогональним вейвлетам. Функція $\psi(t)$ задається базисною функцією $\psi_0(t)$, що як і $\psi(t)$ визначає тип вейвлета. Базисна функція $\psi_0(t)$ як і функція $\psi(t)$ повинні забезпечувати виконання двох операцій:

зсув по часовій осі $t - \psi_0(t - b)$ при $b \in R$;

масштабування $-a^{-\frac{1}{2}} \psi_0\left(\frac{t}{a}\right)$ при $a > 0$ і $a \in R^+ - \{0\}$.

Параметр a задає ширину цього пакета, а b – його положення на часовій осі. У такий спосіб для заданих a і b функція і є вейвлет:

$$\psi(t) = a^{-\frac{1}{2}} \psi_0\left(\frac{t-b}{a}\right). \quad (10)$$

У такий спосіб за допомогою дискретних масштабних перетворень $(1/2^j)$ і зсувів $(k/2^j)$ можна описати всі частоти і покрити усю вісь, маючи один базисний вейвлет $\psi_0(t)$ (рис.4) [5].

На відміну від перетворення Фур'є, у якому координата x заміняється на одну частотну змінну p , у теорії вейвлет-перетворення x заміняється на дві змінні і b . У визначеному змісті b є аналогом координати x , а параметр a – аналогом зворотної частоти $\frac{1}{p}$, тобто $c(a, b)$ містить інформацію про просторові (або часові) і частотні властивості сигналу одночасно. Це й дозволяє вивчити сигнал більш детально, ніж за допомогою Фур'є-аналізу.

Основна ідея представлення нестационарних сигналів за допомогою вейвлетів полягає в розкладанні досліджуваних сигналів на дві складові – апроксимуючу й уточнюючу з наступним уточненням (рис. 5), таке представлення можливе як у частотній, так і в часовій області.

Вейвлет-перетворення може використовуватися як для представлення сигналів, так і для їхньої фільтрації. Для фільтрації і видалення шумів звичайно, використовують загальновідомий прийом – видалення високочастотних складових із спектра сигналу.

Щодо вейвлет-перетворення фільтрацію й очищення від шумів можна здійснювати шляхом обмеження рівня коефіцієнтів, що деталізують. Тому що короткочасні особливості сигналу, до яких можна віднести

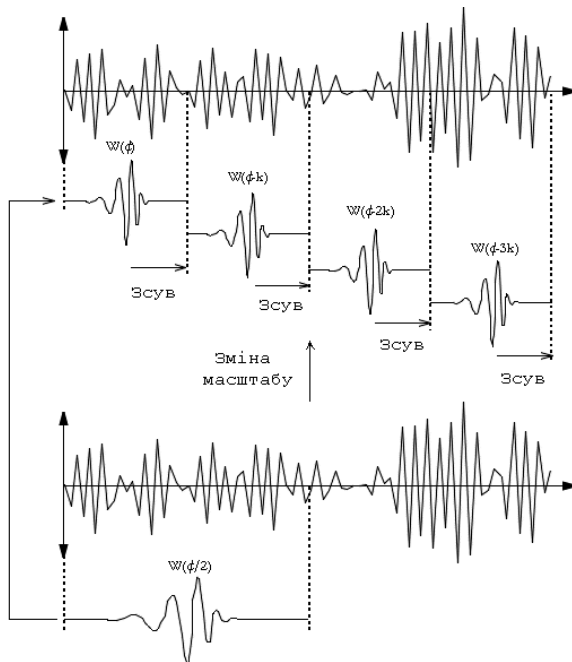


Рис. 4 – Представлення сигналу за допомогою масштабування і зсуву базисного вейвлета

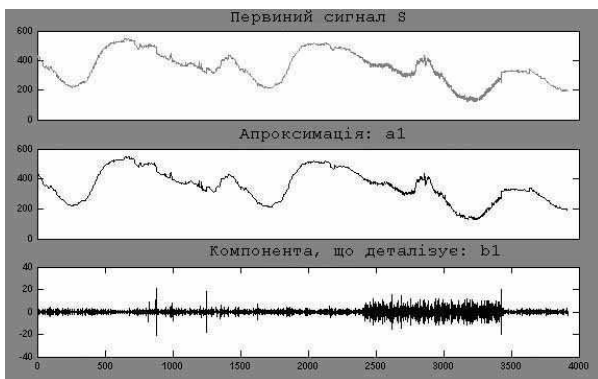


Рис. 5 – Представлення сигналу за допомогою масштабування і зсуву базисного вейвлета

шуми, створюють коефіцієнти, що деталізують, з високим змістом шумових складових, що мають великі випадкові викиди значення сигналу, рівень шумів можна зменшити, задавши деякий поріг, зрівнявши коефіцієнти по їх рівню. Необхідно підкреслити, що рівень обмеження можна задавати для кожного коефіцієнта окремо, що дозволяє синтезувати адаптивні до змін сигналу системи очищення сигналів від шумів. Для видалення шумів може використовуватися штрафний поріг, що визначається правилом вибору коефіцієнтів вейвлетного розкладання з використанням “штрафного” методу Бірге-Массарта [5].

На основі викладеного теоретичного матеріалу і практичних досліджень можна зробити висновок про доцільність представлення мовних сигналів за допомогою вейвлет-перетворення, як на етапі попередньої обробки, так і при формуванні векторів ознак для подальшого процесу розпізнавання.

Література

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К.: Наукова думка, 1987.
2. Е.И. Бовбель, И.Э. Хейдеров. Статистические методы распознавания речи: скрытые Марковские модели // Зарубежная радиоэлектроника. Успехи современной радиоэлектроники. 1998 3, С.45-65.
3. Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения // Успехи физических наук. - Ноябрь 1996. – Том 166. - 11. – С.1145-1170.
4. Дремин И.М., Иванов О.В., Нечитайло В.А. Вейвлеты и их использование // Успехи физических наук. - Май 2001. – Том 171. - 5. – С.465-501
5. Касап А.М., Фаніна Л.О. Використання вейвлет-перетворення для фільтрації мовних сигналів // Proceedings of 7-th INTERNATIONAL MODELLING SCHOOL of AMSE-UAPL. 12-17 September 2002, Alushta – Ukraine.