

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА КОМП'ЮТЕРНА ТЕХНІКА

УДК 681.3.06

Т. В. Грищук, канд. техн. наук

АМПЛІТУДНИЙ МЕТОД ВИЗНАЧЕННЯ ТОНАЛЬНИХ ТА СОНОРНИХ ЗВУКІВ У МОВНОМУ СИГНАЛІ

Розглянуто метод визначення тональних та сонорних ділянок в сигналі, що характеризуються наявністю квазіперіодів. Ця інформація може використовуватись для підвищення ефективності процесу розпізнавання мови. Введено поняття квазіперіоду та відстані між квазіперіодами. Надано рекомендації з вибору експериментальних параметрів методу.

Вступ та постановка задачі дослідження

Наявність мови є однією з найсуттєвіших відмінностей людини від інших представників тваринного світу. Впродовж всієї своєї історії людство намагається перенести цю властивість на інші види. Ну хто з нас не намагався в дитинстві повчити домашню папужку розмовляти і мріяв, що настане той час, коли відбудеться природній діалог з улюбленцем.

З розвитком науково-технічного прогресу людство отримало нових «улюбленців» — комп'ютери. Для багатьох комп'ютер став не просто засобом роботи, але й чи не найулюбленішим засобом проведення вільного часу. З початку 50-х років минулого століття багато науковців намагаються навчити комп'ютер розуміти людську мову і надавати відповідь. Голосові інтерфейси є природними і зручними, але для їх масового поширення потрібно розв'язати ще багато задач для підвищення швидкодії та надійності роботи таких систем.

Мовний сигнал є складним за своєю природою. Інформація, яку він несе, має складний ієрархічний характер. На перший погляд, це є звукова хвиля, тобто в часовому просторі ознак мовний сигнал представляється зміною амплітуди в часі. Окрім цього мовний сигнал є зміною частотних характеристик. Прослуховуючи голосове повідомлення, ми отримуємо семантичну інформацію — ми розрізняємо різні звуки, слова та цілі речення з їх змістом. Безперечно всі ці дані необхідні для розпізнавання вхідної фрази.

Голосові інтерфейси мають чітке обмеження на вхідну множину команд користувача. Тобто заздалегідь обумовлюється набір фраз, які система буде розуміти. Маючи обмежений набір фраз, зовсім не обов'язково отримати з вхідного мовного сигналу всі можливі часові та частотні ознаки. За цим принципом будуються багаторівневі системи розпізнавання мови (СРМ). В таких системах є декілька рівнів розпізнавання, які обриють з усієї множини кандидатів певну (меншу) кількість для подальшого аналізу. На рис. 1 показано схему роботи такої системи. Так, окремі слова можуть бути описані в різних алфавітах (як послідовність звуків, фонем, складів тощо).

На сьогодні можна виділити два основних напрямки побудови систем розпізнавання мови: еталонний і фонемно-орієнтований [1]. Еталонний метод ґрунтується на порівнянні деяких характеристик мови (енергетичних, спектральних тощо). В якості еталонів в більшості випадків використовують цілі слова. Фонемно-орієнтований метод ґрунтується на виділенні фонем з потоку мови. Фонема — це одиниця мови. Як слово складається з літер, так мова складається з фонем.

В цій статті розглядається фонемно-орієнтований підхід до побудови системи розпізнавання мови.

Метою статті є підвищення ефективності процесу розпізнавання мови на первинному рівні аналізу мовного сигналу, що дозволить застосовувати алгоритми розпізнавання на малопотужних пристроях з голосовими інтерфейсами.

Ефективність процесу розпізнавання є комплексною величиною швидкості та достовірності розпізнавання. В статті пропонується метод, який на відміну від існуючих, на основі амплітудної інформації про мовний сигнал дозволяє скоротити розмір вхідного сигналу на 30...50 % в залежності від фонемного складу вхідного сигналу без втрати корисної інформації про вхідний сигнал.

Постановка задачі

Нехай на вхід системи розпізнавання подається послідовність амплітуд мовного сигналу $A(t)$, де $t \in 1..T$.

Потрібно знайти множину $P = \{P_i\}$, де $P_i = [P_{ib}, P_{ie}]$ — квазіперіодичні інтервали, що відповідають фонемам мови.

Періодичність, як характеристика звуків

Звуки, що приймають участь в формуванні мови, мають дві основні класифікації: по артикуляційним та по акустичним ознакам [2, 3].

Класифікація звуків по артикуляційним ознакам є надзвичайно важливою для методів розпізнавання мови, що ґрунтуються на моделюванні носоглотки, але для вирішення задач розбиття на фонемі більш цікаво розглядати акустичні відмінності звуків. За акустичними ознаками звуки розділяються на:

1. Тональні звуки — утворюються голосом при практично повній відсутності шумів. До таких звуків належать голосні звуки.
2. Сонорні — чия якість визначається характером звучання голосу, який грає головну роль в їх утворенні, а шум приймає участь мінімально.
3. Шумні — їх якість визначається характером шуму.

Потрібно зазначити, що голосні і сонорні звуки складаються з ділянок затухання імпульсів від основних коливань голосових зв'язок. В літературі їх називають доменами [1]. Домени однієї фонемі тільки на перший погляд є ідентичними. В реальності форма домен може видозмінюватись, тому сигнал на таких проміжках не є цілковито періодичним в класичному значенні цього терміну. Саме тому в даній статті пропонується називати домени квазіперіодами. Зміни в квазіперіодах не відчутні на слух та на погляд, але вони ускладнюють процес виділення їх з сигналу. Тому розв'язання задачі побудови надійного методу знаходження квазіперіодів в сигналі є актуальною.

На рис. 2 показано приклад звукового сигналу, що містить квазіперіодичну ділянку, що відповідає голосному звуку «О».

Розбиваючи мовний сигнал на квазіперіоди, отримуємо ще один рівень розпізнавання. В загальній ієрархії цей рівень знаходиться нижче, ніж рівень розпізнавання фонем.

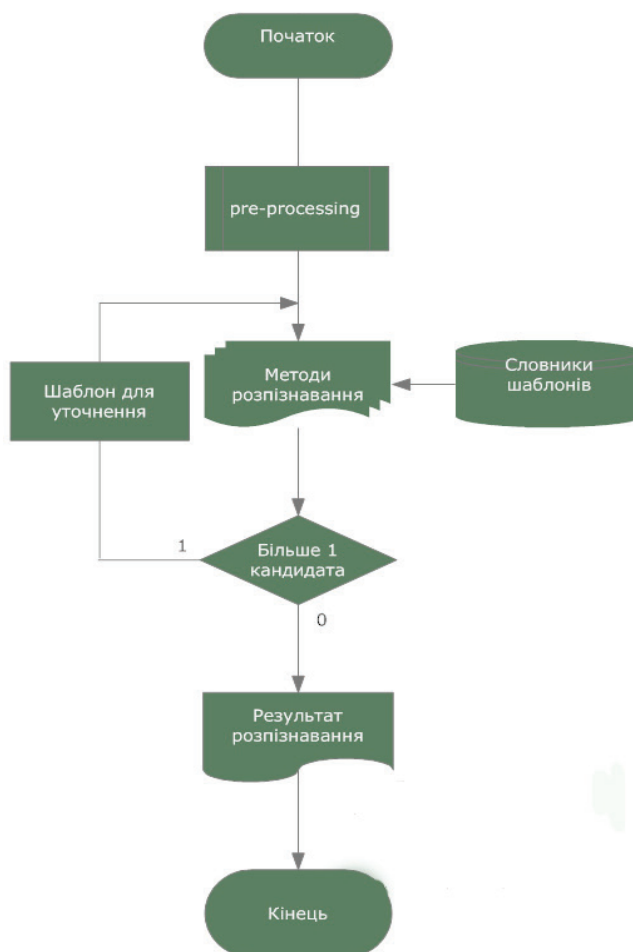


Рис. 1. Схема роботи багаторівневої CPM

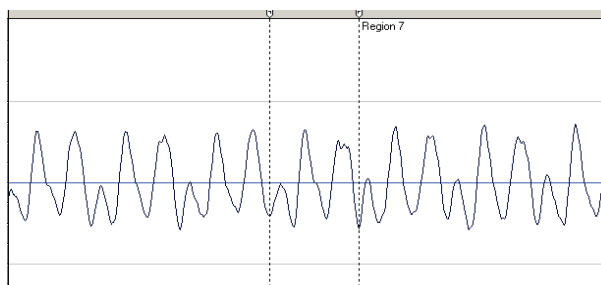


Рис. 2. Приклад тонального звуку

Загальний алгоритм виділення квазіперіодів

Спочатку дамо визначення квазіперіодичного сигналу.

Квазіперіодичний сигнал — це сигнал, форма якого «приблизно» повторюється через кожні $T \pm \delta$ відліків.

Коротко опишемо загальний алгоритм методу.

Крок 1. Знаходимо всі існуючі екстремуми, що утворюють контур сигналу.

Крок 2. Знаходимо перший квазіперіод.

Крок 3. Знаходимо наступний квазіперіод, відстань від якого до попереднього не перевищує заданого значення. Якщо відстань задовільна, то повторюємо Крок 3, в іншому випадку переходимо до Кроку 2.

Пошук закінчується, коли проведено аналіз всіх екстремумів вхідного сигналу.

Далі детальніше розглянемо кожний крок цього алгоритму.

Процедура знаходження локальних екстремумів

На підготовчому етапі необхідно перетворити вхідний сигнал $A(t)$ в вектор локальних екстремумів $E(t)$, що буде описувати загальну форму сигналу. Для цього введемо визначення локального екстремуму:

Істотним локальним максимумом/мінімумом в ε -околі будемо вважати точку, яка має найбільше/найменше значення на відрізку звукового файлу $\pm\varepsilon$ від даної точки. Локальні максимуми/мінімуми, що не задовольняють даній умові, будемо вважати неістотними.

Розмір ε -околу визначається експериментально.

На рис. 3 проілюстровано вищевказане визначення.

Таким чином, кожний наявний в звуковому сигналу квазіперіод буде характеризуватись такими параметрами:

- довжина L (в семплах);
- послідовність векторів, що з'єднують ε -істотні екстремуми.

Квазіперіод позначимо таким чином: $P = [P_b, P_e]$.

Для того, щоб знайти перший квазіперіод знадобиться ще один параметр, значення якого визначається експериментальним шляхом, — мінімальна довжина квазіперіода. Маючи вхідну послідовність екстремумів на початковому етапі, нам потрібно виділити з них послідовність векторів, що будуть утворювати початковий квазіперіод, який далі буде уточнюватись. Зрозуміло, що початковий і кінцевий екстремуми квазіперіода є однаковими за видом (максимуми або мінімуми) за близькими за значеннями.

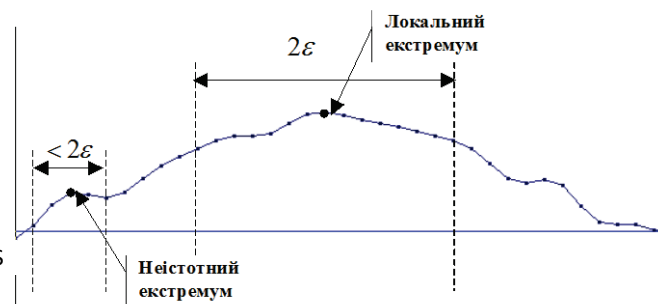


Рис. 3. Процедура знаходження істотних екстремумів

Знаходження кандидатів на порівняння

Для обраного квазіперіода кандидатами на порівняння будуть відрізки звукового сигналу, що відрізняються за довжиною не більше ніж на задану величину від довжини початкового квазіперіода (приблизно 10%). Окрім цього, послідовність векторів початкового квазіперіода і кандидата повинні починатись і закінчуватись векторами одного типу. Тобто, якщо послідовність векторів початкового квазіперіода починається з вектора, початок якого — максимум, а кінець — мінімум, то і у всіх кандидатів на порівняння перший вектор повинен з'єднувати максимум з мінімумом.

Процедура порівняння двох квазіперіодів

В якості вхідних параметрів процедури порівняння маємо:

- квазіперіод $P = [P_b, P_e]$ довжиною L , для якого необхідно знайти наступний квазіперіод;

— N кандидатів на порівняння, довжини яких знаходяться в інтервалі $[0,9L; 1,1L]$.

На рис. 4 показано фрагмент звукового сигналу, який ми будемо використовувати для ілюстрації. В цьому прикладі визначено, що фрагмент сигналу від маркера 1 до маркера 2 є квазіперіодом. Тепер необхідно визначити, де закінчується наступний квазіперіод. В межах допустимого відхилення знайдено 2 кандидата на порівняння:

- квазіперіод 2—3;
- квазіперіод 2—4.

Щоб визначити, який з N кандидатів є правильним, нам потрібно почергово порівняти початковий квазіперіод з усіма кандидатами. Для цього будемо використовувати процедуру порівняння квазіперіодів.

Нехай маємо два квазіперіода з довжинами L_1 і L_2 , де $L_1 < L_2$.

Встановлюємо параметр h — максимальний зсув.

В циклі потрібно зміщати квазіперіод довжиною L_1 вздовж квазіперіода довжиною L_2 в межах кроку в межах $\pm h$. При кожному зміщенні знаходимо площу, що характеризує відстань між двома квазіперіодами. Детальніше процес знаходження відстані опишемо пізніше.

В результаті отримуємо множину площ, з яких нам необхідно обрати мінімальну. Це і буде відстань між двома квазіперіодами.

Таким чином, використовуючи наведену процедуру порівняння, знаходимо відстані початкового квазіперіода до всіх кандидатів. Для нашого прикладу знаходимо дві такі відстані.

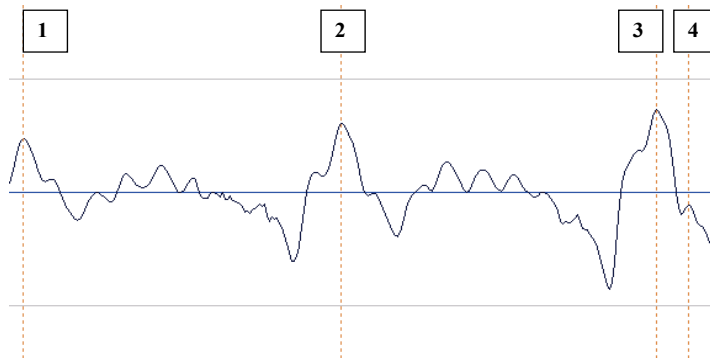


Рис. 4. Знаходження наступного квазіперіода

Вибір найкращого кандидата

Отримано дві відстані і потрібно вирішити, який з двох кандидатів є найкращим. Проведені автором експериментальні дослідження показали, що найкращим є кандидат, що забезпечує подальше продовження квазіперіодичної ділянки.

Для цього пропонується використовувати процедуру прогнозу, яка полягає у такому. Найкращим кандидатом є той кандидат, який забезпечує найменшу відстань між початковим квазіперіодом і ділянкою, яка йде безпосередньо після цього кандидата.

Ілюстрація вищесказаного подана на рис. 5.

Таким чином, отримуємо відстані $DistFuture$ для кожного кандидата. Найкращим вважається кандидат, який дає мінімальне значення $DistFuture$.

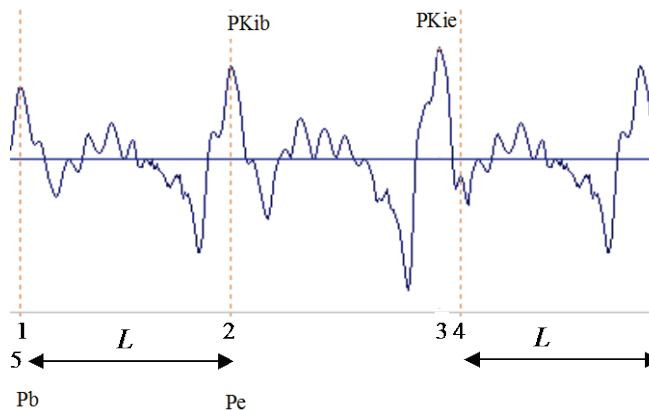


Рис. 5. Обчислення відстані в процедурі прогнозу

Відстань між двома квазіперіодами

Зазначимо, що відстанню між двома квазіперіодами слід вважати площу між ними. Нагадаємо, що квазіперіод — це послідовність векторів. Площа між квазіперіодами є сумою площ між відповідними векторами (див. рис. 6).

Відстань обчислюється за такою формулою:

$$Dist = \sum_i S_i$$

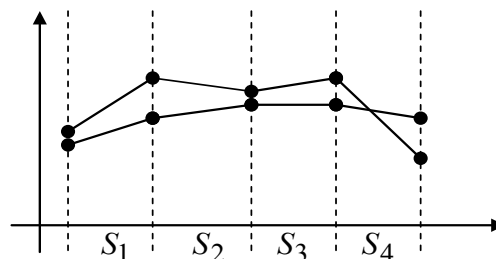


Рис. 6. Порівняння двох квазіперіодів

Для знаходження площі між двома векторами потрібно розглядати два випадки.

1 випадок (рис. 7).

Якщо $(y_1^b - y_1^a)(y_2^b - y_2^a) < 0$;

$$h = |x_2 - x_1|;$$

$$S = \frac{1}{2} h \left[\frac{|y_1^b - y_1^a|^2 + |y_2^b - y_2^a|^2}{|y_1^b - y_1^a| + |y_2^b - y_2^a|} \right].$$

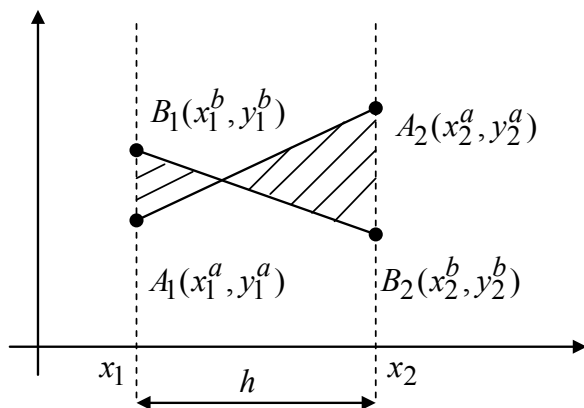


Рис. 7. Вектори мають точку перетину

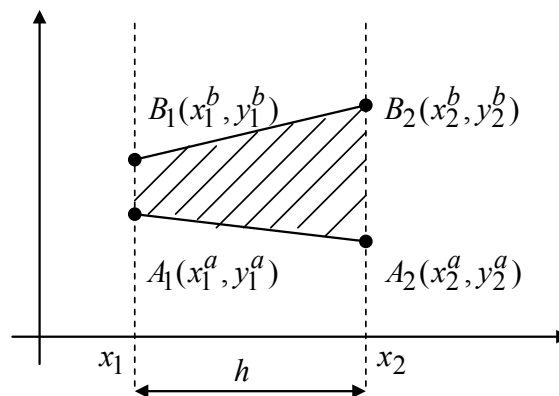


Рис. 8. Вектори не мають точки перетину

2 випадок (рис. 8).

Якщо $(y_1^b - y_1^a)(y_2^b - y_2^a) \geq 0$,

$$S = \frac{1}{2} h (|y_1^b - y_1^a| + |y_2^b - y_2^a|).$$

Висновки

Розроблений метод визначення тональних і сонорних звуків у звуковому сигналі. Проведені експериментальні дослідження підтвердили адекватність розробленого методу. Подальші дослідження можуть полягати у визначенні можливості використання такого підходу для розв'язання задачі зменшення розмірності словників СРМ в процесі розпізнавання. Також запропонований підхід можна використовувати для стиснення звукових сигналів при їх передаванні. На відміну від існуючих, запропонований підхід ґрунтується лише на амплітудному аналізі вхідного мовного сигналу та дозволяє скоротити вхідний сигнал для пересилання на вищі рівні розпізнавання від 30...50 % в залежності від фонемної структури вхідного сигналу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Проблемы создания многоуровневой системы распознавания речи [Электронный ресурс] / С. Н. Курочкин, А. Г. Бродин // Автоматизация и управление в машиностроении — 1997. — № 2. — Режим доступа до журн. : http://magazine.stankin.ru/arch/n_02/automation/art05.html.
2. Рабинер Л. Р. Цифровая обработка речевых сигналов / Л. Рабинер, Р. Шафер — М. : Радио и связь, 1981. — 496 с.
3. Фант Г. Акустическая теория речеобразования / Г. Фант. — М. : Наука, 1964. — 284 с.

Рекомендована кафедрою комп'ютерних систем управління

Стаття надійшла до редакції 25.02.11
Рекомендована до друку 1.12.11

Гришук Тетяна Вікторівна — доцент кафедри комп'ютерних систем управління.

Вінницький національний технічний університет, Вінниця