КОМП'ЮТЕРНІ ЗАСОБИ, МЕРЕЖІ ТА СИСТЕМИ

M.V. Semotyuk, I.A. Bezverbnyi

ADAPTIVE ALGORITHM FOR ALLOCATION PHONEMES IN SPEECH SIGNAL

Today the speech recognition problems tasks adaptive methods are widely used. This approach considerably simplifies the analysis and helps to avoid the issues concerning the increase of errors from a large number of calculations. One of these methods is described in the article.
Key words: empirical mode, phonemes analysis.

На сегодня задачи распознавания речевых сигналов с использованием адаптивных методов находят широкое применение. Такой подход значительно упрощает анализ и позволяет избежать проблемы нарастания ошибки в результате большого количества вичислений. В статье рассматривается один из таких методов.
Ключевые слова: эмпирическая мода, анализ фонем.

На сьогодні задачі розпізнавання мовних сигналів з використанням адаптивних методів знаходять широке застосування. Такий підхід значно спрощує аналіз і дозволяє уникнути проблем з наростанням похибки від великого числа обрахунків. В статті розглядається один з таких методів.
Ключові слова: емпірична мода, аналіз фонем.

УДК 004.934.2

М.В. СЕМОТЮК, І.А. БЕЗВЕРБНИЙ

АДАПТИВНИЙ АЛГОРИТМ ВИДІЛЕННЯ ФОНЕМ У МОВНОМУ СИГНАЛІ

Вступ. Реальні мовні процеси, з якими доводиться мати справу в задачах розпізнавання звуку, є процесами досить високої складності, такими що містять значну кількість компонент. Аналіз звукової осцилограми мовного сигналу відображує взаємодію великої кількості механічних процесів, що відбуваються під час мовлення. Тому аналіз осцилограми мовного сигналу постає досить складним завданням, так як його складові компоненти взаємодіють і таким чином маскують і спотворюють закономірності, притаманні такому сигналу.

Ступінь розробки. Важливим науковим питанням є спрощення обчислень, необхідних при розпізнаванні мовного сигналу. В першу чергу це досягається шляхом розкладання процесу на компоненти. Такий процес називається декомпозицією. Існує багато різноманітних методів декомпозиції. Ці методи засновуються на математичних та емпіричних підходах, можуть мати різний рівень складності та області застосування. Методи, що дозволяють суттєво спростити обчислення при застосуванні цих аналітичних методів, – це так звані адаптивні перетворення, для яких базис визначається безпосередньо самими вхідними даними, тобто мовним сигналом. Таким перетворенням є перетворення Гільберта – Хуанга. Воно дозволяє знаходити миттєвий спектр нелінійних нестаціонарних послідовностей. В процесі перетворення Гільберта – Хуанга відбувається розкладання на емпіричні моди або внутрішні коливання, які не задаються аналітично і визначаються

виключно самою послідовністю. Такий підхід значно спрощує аналіз часового процесу і дозволяє уникнути проблем з наростанням похибки від великої кількості обчислень [1, 2].

Сутність проблеми. Методологія аналізу коливальних процесів, що знаходиться в основі алгоритму Хуанга, може бути використана для побудови алгоритмів, що для аналізу мовного сигналу оперують не аналітичними залежностями, а емпіричними модами, отриманими як результат обробки вхідної часової послідовності. На сьогодні такі задачі вирішуються переважно аналітичними методами з елементами масштабування компонент вхідного сигналу. Тим часом важливо мати адаптивний алгоритм фонемного аналізу, розроблений з використанням емпіричних мод, що дозволило би суттєво спростити кількість обчислень.

Тому мета цієї роботи – це розроблення нового адаптивного методу фонетичного аналізу, що враховував би особливості процесу мовлення на прикладі української фонетики.

Ідея алгоритму базується на тому, що передача фонем в середньостатистичному мовному сигналі повторюється (в середньому 20–40 разів). Для відділення кожної наступної фонеми в мовному сигналі зростає амплітуда, а наприкінці фонеми амплітуда знижується.

Ця властивість використовується для автоматизації дослідження фонемних складових звукового сигналу. Однак величина зростання і занепаду амплітуди в кожній новій фонемній моді різна. Власне пропорції зростання і занепаду амплітуди звукового сигналу в межах аналізу фонем несуть інформацію про відповідну фонему і є предметом дослідження. Спершу сигнал фільтрується низькочастотним фільтром з метою вилучення високих частот, які є в переважній більшості шумовими перешкодами. Цей важливий етап показано на рис. 1.

Всі мінімальні екстремуми відфільтрованого сигналу встановлюють рівними нулю шляхом віднімання апроксимованої кривої, що проведена по точках мінімальних екстремумів, подібно як це робиться для визначення емпіричних модових функцій перетворення Гільберта – Хуанґа (рис. 2).

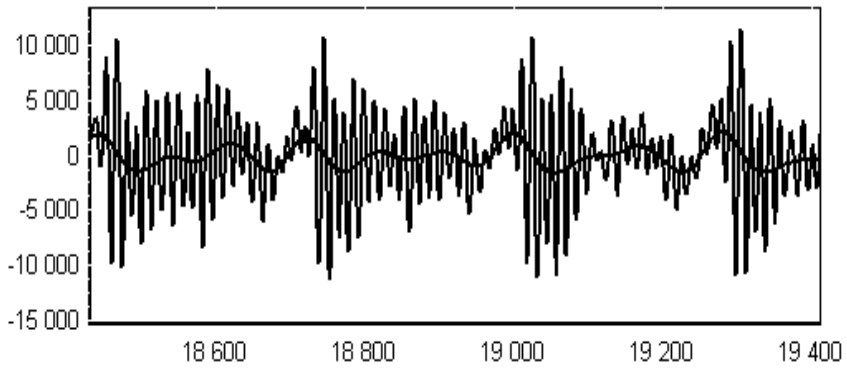

РИС. 1. Приклад фільтрації звукового сигналу нерекурсивним фільтром 128 порядку

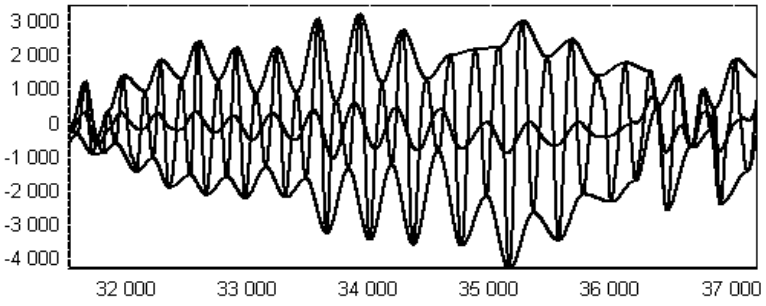

РИС. 2. Перша мода перетворення Гільберта – Хуанґа і моди екстремумів

Іншими словами якщо ϕ_i – мода фільтрованого сигналу, а ξ_i, ψ_i – апроксимовані криві, що огинають ϕ_i по максимальних і мінімальних екстремумах. Тоді

$$\alpha_i = \phi_i - \psi_i \qquad (1)$$

– експериментальна мода з вхідними значеннями решти вибірок мовного сигналу в натуральну величину, яка отримує умовну назву мода α-фонем або мода вхідних значень (рис. 3).

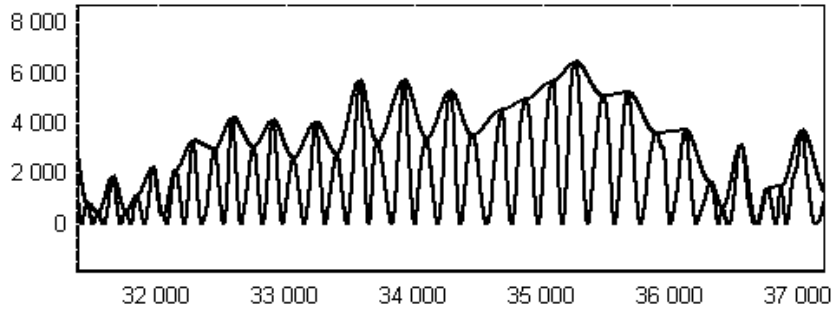

РИС. 3. Мода α-фонем та моди максимальних екстремумів

Наступний етап – це приведення максимальних екстремумів до значення одиниці з паралельним зменшенням значення решти точок у числовий проміжок $0 < w < 1$. Для цього отриману моду α-фонем піддають аналогічній процедурі віднаходження апроксимованої моди, що містить точки максимальних екстремумів або інакше моди максимальних екстремумів ϑ_i. Значення моди α-фонем діляться на відповідні значення моди максимальних екстремумів:

$$\beta_i = \frac{\alpha_i}{\vartheta_i}. \qquad (2)$$

В результаті отримується емпірична мода β_i, властивістю якої є однакова (одинарна) висота максимальних екстремумів у точках визначення фонем, решта

максимальних екстремумів операції ділення виявляються суттєво меншими за одиницю, а мінімальні екстремуми такої функції є рівними нулеві або близькими до нуля.

Цю емпіричну модову функцію назвемо модою одинарних фонем (рис. 4). Таким чином автоматизовано розділ мовного сигналу на фонеми, після чого відбувається порівняльний аналіз кожної фонеми за допомогою бази даних, створеної шляхом експериментальних досліджень.

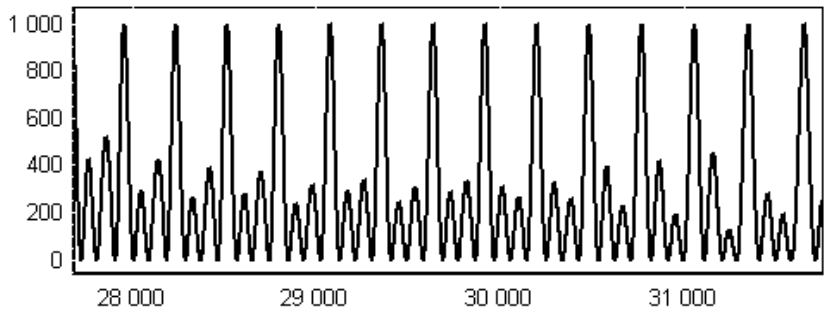

РИС. 4. Мода одинарних фонем

Робота алгоритму наводиться на прикладі мовного сигналу протягом 1,12 с, частота дискретизації 44100 кГц, частота квантування 16 біт. Робота алгоритму починається з фільтрації сигналу нерекурсивним фільтром 128 порядку:

$$b_i = \frac{a_i + ... + a_{i+128}}{128}. \quad (3)$$

Другий етап алгоритму – це віднаходження гладкої огинаючої за точками мінімумів з подальшим визначенням різниці вхідного масиву та отриманої огинаючої.

Цей етап передбачає такі дії.

1. Визначаються екстремуми відфільтрованої модової функції ϕ_i. Питання визначення екстремумів у дискретних рядах може розглядатися лише в практичній площині.

2. Між максимальними і мінімальними екстремумами проводяться апроксимаційні криві. Апроксимацію між точками було вирішено впроваджувати косинусоїдальними функціями (рис. 5).

Кожна точка між двома екстремумами отримує значення:

$$y_i = \frac{dy}{2}\cos\left(\pi + \pi\frac{i}{dx} + 1\right) + y_0. \quad (4)$$

Це давало кращі результати порівняно з прямолінійною, синусоїдальною апроксимацією.

3. Відбувається покрокове віднімання апроксимованої кривої, що проведена через точки мінімальних екстремумів, ψ_i відфільтрованого масиву ϕ_i. Резуль-

татом операції віднімання є емпірична мода α_i, мінімальні екстремуми якої завжди дорівнюють нулеві за виключенням аномальних шумових мінімальних екстремумів.

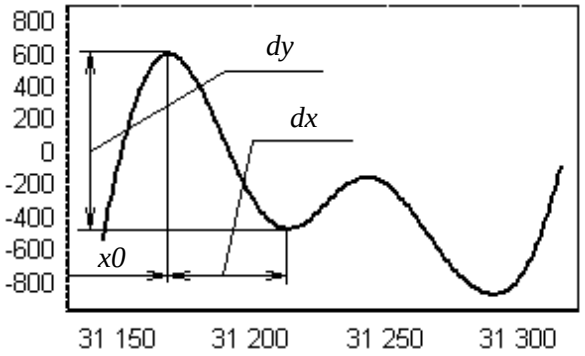

РИС. 5. Апросимація косинусоїдальними функціями

Третій етап алгоритму – це віднаходження верхньої огинаючої для отриманої на попередньому етапі емпіричної моди α_i. Цей етап іноді варто повторити двічі, результатом чого є емпірична мода ϑ_i.

Четвертий етап алгоритму – це ділення моди α_i на моду ϑ_i, в результаті маємо емпіричну моду β_i, яка є модою одинарних фонем. Як видно з рис. 4 фонема «о» має одну пелюстку, яка більше відрізняється амплітудою за інших. Вона має стандартну або близьку до неї величину і умовно називається основним максимумом, тоді як подавлені в результаті ділення (2) пелюстки умовно називаються локальними максимумами. Основний максимум слугуватиме обмеженням фонеми. Для його виявлення виставляється індивідуальний поріг, в середньому 85–90 % стандартної величини основного максимуму.

П’ятий етап алгоритму – це сегментація і визначення подібностей у межах сегментованих фонем. Експериментально виявлено, що відстані між екстремумами і значення амплітуди екстремумів у межах кожного такого сегмента фонеми на ділянках стаціонарного відтворення звуку дуже подібні. Надалі алгоритм розпізнавання фонем розглядається на прикладі осцилограми звукового сигналу у вигляді слова «мімо». На рис. 6 показано точки, які відповідають максимальним екстремумам, по осі абсцис відкладено одинарні відстані між границями фонем, а по осі ординат відкладено відстані між двома сусідніми екстремумами в межах однієї фонеми. Тобто точки максимумів в межах однієї фонеми на рис. 6 вистроєні у вертикальному порядку одна над одною. Фактично цей графік відображає спектр довжин хвиль для кожної окремої фонеми. Перший, другий і третій (і так далі, якщо такі існують в межах фонеми) екстремуми кожної наступної фонеми на цьому графіку фактично утворюють послідовні криві, які відтворю-

ють характер вимовляння фонеми конкретним мовцем. Аналіз цих кривих дозволяє визначати характер вимовляння фонеми.

РИС.6. Послідовності відстаней між локальними екстремумами

На рис. 6 простежуються ділянки стаціонарності, які відповідають голосним звукам «і» та «о». Для такої ділянки може бути знайдений шаблон у вигляді числового масиву, що складається з усереднених значень відстаней між кожним з екстремумів у межах виділеної фонеми. В такий спосіб фонема може бути розпізнана з наперед заданою точністю. Для цього створюється база даних, яка містить відповідний набір відстаней між локальними екстремумами в межах визначеної фонеми.

Висновки. Таким чином в статті представлено адаптивний метод сегментації мовних сигналів, що побудований з використанням принципів перетворення Гільберта – Хуанга і методу емпіричної модової декомпозиції. Середовище моделювання було створене на базі програмного продукту Delphi7 фірми Borland.

1. Давыдов А.Г., Лобанов Б.М. Использование периодичности речевого сигнала при фонемной сегментации речи. Доклады БГУИР. 2006 апрель–июнь, № 2 (14). http://ssrlab.by/wp-content/uploads/2006/12_100229_1_57873.pdf
2. Huang N. E. Introduction to the Hilbert Huang transform and its related mathematical problems, http://www.worldscientific.com/doi/suppl/10.1142/5862/suppl_file/5862_chap1.pdf

Одержано 16.11.2017