УДК 681.58

СЕГМЕНТАЦІЯ МОВНОГО СИГНАЛУ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

Дюжаєв Л.П., к.т.н., доцент; Соколов Д.Ю., студент
Національний технічний університет України
«Київський політехнічний інститут», м. Київ,Україна

Вступ

У системах автоматичного розпізнавання мови важливим завданням є сегментація мови відповідно до фонетичної транскрипції мови. У процесі розпізнавання необхідно спочатку сегментувати мовний сигнал на характерні елементи (сегменти), визначити тип сегмента, а потім проводити порівняння за різними ознаками.

У дослідних системах і на етапі попередньої розробки можливе використання ручної сегментації. Однак, даний підхід вимагає значних затрат сил і часу. Крім того, практично неможливо точно відтворити результати ручної сегментації внаслідок суб'єктивності людського слухового і зорового сприйняття. Подібних проблем не виникає при автоматичній сегментації, яка все ж не безпомилкова, але дає відтворювані результати.

Існує два основних типи алгоритмів сегментації мови. До першого типу відносяться алгоритми, які проводять сегментацію мови за умови, що відома послідовність фонем даної фрази. Інший тип алгоритмів не використовує апріорної інформації про фразу, і при цьому межі сегментів визначаються по мірі зміни акустичних характеристик сигналу. Також існує один тип алгоритмів, які приймають рішення, як на основі апріорної інформації, так і на основі зміни акустичних характеристик.

При автоматичній сегментації бажано використовувати тільки загальні характеристики мовного сигналу, оскільки зазвичай на цьому етапі немає конкретної інформації про зміст мовного висловлювання.

Постановка задачі

У даній статті наводиться алгоритм сегментації мовного каналу з використанням кратномасштабного аналізу та вейвлет-перетворення, що дозволяє виділити границі сегментів фонем, але має точність, яка залежить від конкретних параметрів для конкретного диктора. Запропоновані у статті модифікації алгоритму кратномасштабного аналізу дозволяють знаходити мітки границь межфонемних переходів у звукових сигналів різних дикторів.

Теоретичній підхід до сегментації

Мовний сигнал складається з квазістаціонарних ділянок, відповідних дзвінким і шиплячим фонемам, що чергуються з ділянками з порівняно швидкими змінами спектральних характеристик сигналу (межфонемні переходи, внутріслівні переходи мова-пауза). В межах стаціонарних ділянок значну роль для аналізу мови відіграють спектральні особливості сигналу, що визначаються передатковою характеристикою мовного тракту, яка змінюється в процесі артикуляції. Можна сказати, що мовний сигнал характеризується нелінійними флуктуаціями різних масштабів. Тому досить ефективним для аналізу мовного сигналу є кратномасштабний аналіз та вейвлет-перетворення.

Вейвлети мають істотні переваги в порівнянні з перетворенням Фур'є, тому що вейвлет-перетворення дозволяє судити не тільки про частотний спектр сигналу, але також про те, в який момент часу з'явилася та чи інша гармоніка. З їх допомогою можна легко аналізувати переривчасті сигнали, або сигнали з гострими сплесками. Крім того, вейвлети дозволяють аналізувати дані згідно з масштабом на одному із заданих рівнів (дрібному або великому). Основна область застосування вейвлет-перетворень — аналіз та обробка сигналів і функцій, нестаціонарних у часі або неоднорідних в просторі, коли результати аналізу повинні містити не тільки загальну частотну характеристику сигналу але й відомості про певні локальні координати, на яких себе виявляють ті чи інші групи частотних складових або на яких відбуваються швидкі зміни частотних складових сигналу.

Якщо в якості структурних одиниць мови розглядати фонеми, то задача сегментації зводиться до виявлення межфонемних переходів. В рамках традиційних підходів, вирішення цього завдання є вельми проблематичним. Однак вейвлет перетворення дозволяє вирішити цю проблему принаймні для фонем, що відповідають порівняно протяжним квазістаціонарним ділянкам мовного сигналу. Справа в тому, що на межфонемних переходах сигнал зазнає значних змін відразу на багатьох масштабах дослідження, і, відповідно, характеризується зростанням вейвлет-коефіцієнтів для багатьох рівнів деталізації, в той час як на стаціонарних ділянках фонем вейвлет-коефіцієнти виявляються згрупованими поблизу певних масштабів. Таким чином, пошук межфонемних границь може бути зведений до відшукання моментів збільшення вейвлет-коефіцієнтів на значну кількість рівнів масштабування. При цьому суттєвим є вибір вейвлетного базису, який повинен дозволяти описувати стаціонарний мовний сигнал з порівняно малим числом ненульових коефіцієнтів. Можливе використання декількох вейвлетних базисів для пошуку межфонемних переходів у кожному з них з наступним об'єднанням результатів.

Сегментація з використанням кратномасштабного аналізу

Розкладання на вейвлети мовного сигналу довжиною N відліків являє собою суму:

$$f(t)=\sum_{k=0}^{\frac{N}{2^n}-1} s_{n,k}\varphi_{n,k}+\sum_{j=1}^{n}\sum_{k=0}^{\frac{N}{2^j}-1} d_{j,k}\psi_{j,k},$$

де $s_{n,k}$ — коефіцієнти апроксимації, $d_{j,k}$ — деталюючі коефіцієнти, $\varphi_{n,k}$ та $\psi_{j,k}$ — масштабовані та зсунуті версії скейлінг-функції (масштабної функції) φ та «материнського вейвлета» ψ, n — рівень деталізації.

Маштабування та зсування φ та ψ знаходиться наступним чином:

$$\varphi_{j,k}=2^{\frac{j}{2}}\varphi\left(2^j t-k\right),\ \psi_{j,k}=2^{\frac{j}{2}}\psi\left(2^j t-k\right).$$

Опис алгоритма

Данний алгоритм сегментації базується на кратномасштабному аналізі сигналу, а саме:

1. Мовний сигнал, оцифрований з частотою дискретизації 22050 Гц, розбивається на вікна розміром 512 відліків, що перекриваються з половинним перекриванням вікна.
2. Сигнал розкладається по U рівням ($U=6$, використовувалося кратномасштабне вейвлет-перетворення в базисі Добеші, що дорівнює 8)
3. Для кожного *j*-го рівня будується числова послідовність $\left\{e_{i,j}\right\}_{i=1}^{N/256}$:

$$e_{i,j}=10\lg\sum_{k=0}^{n_{j-1}} d_{j,i+k}^2,$$

де i — номер вікна, $n_j=\dfrac{n}{2^j}$ — розмір вікна на *j*-ом рівні, n - розмір вікна у вихідному сигналі (в нашому випадку $n=512$).

4. Використовуючи наступне співвідношення визначають передбачувані межі між вікнами з номерами i та $i+1$:

$$\left|e_{i+1,j}-e_{i,j}\right|\geq\eta,$$

де $\eta=3{,}5$ — визначається експериментально.

5. Знаходиться загальна кількість передбачуваних границь для всіх рівнів:

$$T=(T_1,T_2,...,T_{i,}...,T_N)$$

6. Обираючи пороговий коефіцієнт $g_{пор}$, що змінюється в межах (0; 1), отримуємо нерівність для пошуку межфонемного переходу: $\dfrac{T}{U}\geq g_{пор}$

7. Обчислюємо координату границі межфонемного переходу, усереднюючи сформований з нерівності вище масив знайдених границь.

Результат роботи алгоритму приведений на рис.1.

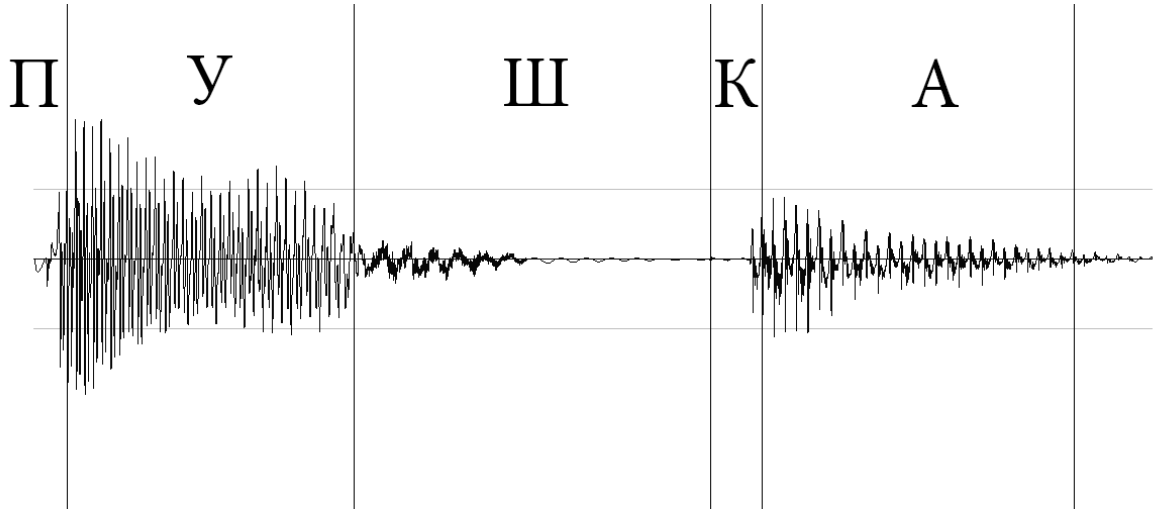

Рис.1. Сегментація мовного сигналу

Визначення оптимального порогового коефіцієнта

Проведемо дослідження результатів роботи описаного алгоритму при різному $g_{пор}$ для визначення його оптимального значення. Параметри запису мовного сигналу: 22050Гц, 16 біт.

Результати експерименту наведені в таблиці 1.

З результатів експерименту видно, що із збільшенням порогового коефіцієнту зменшується чутливість алгоритму до змін мовного сигналу. Так, при значеннях 0.2-0.4 помітне виділення зайвих сегментів для голосних. При цьому добре розділяються голосові звуки, що стоять поряд з -оа-, -ія-, розділяється поєднання -кс-. При великих значеннях $g_{пор}$ кількість зайвих сегментів мале, але перестають поділятися голосові звуки і поєднання "к" з шиплячими.

Табл.1. Результати фонемної сегментації в залежності від $g_{пор}$

$g_{пор}$ / слово	0,2	0,4	0,6	0,8
акація	[а|а|к|а|ц|ц|і|я|я]	[а|а|к|а|ц|і|я]	[а|а|к|а|ц|ц|ія]	[а|к|а|ц|ія]
факс	[ф|а|а|к|с]	[ф|а|к|с]	[ф|а|к|с]	[ф|а|кс]
пушка	[п|у|у|ш|ка|а]	[п|у|ш|к|а|а]	[п|у|ш|ка|а]	[п|у|ш|ка|а]
коала	[к|о|а|а|а|л|а]	[к|оа|л|а]	[к|оал|а]	[к|оала]
рак	[р|а|а|к|к]	[р|а|а|к|к]	[ра|а|к|к]	[ра|а|к]
миля	[м|ми|и|л|я|я]	[м|ми|и|л|я]	[м|миля|я]	[м|миля|я]

До того ж якість сегментації дуже сильно залежить від фонемного складу мовного сигналу. Наприклад, для слова "факс" найкращий результату досягається при значенні порогового коефіцієнта $g_{пор} = 0,4$,

для слова "акація" — 0,8. Таким чином, для оптимальної роботи представленого алгоритму необхідно змінювати $g_{пор}$, тобто зробити його адаптивним.

Модифікація алгоритму

З метою удосконалення алгоритму пропонується наступна модифікація:

1. Проводити підрахунок кількості необхідних міток сегментування *N*:

$$N = S - 1,$$

де *S* — кількість фонем в транскрипції слова.

2. Проводити зміну параметрів η і $g_{пор}$ в діапазонах [2.5; 4] та [0.4; 1] відповідно для формування вектора *T*, що задає унікальні мітки границь.

3. Перевіряти виконання наступної умови у векторі T:

$$T_i - T_{i+1} < 512,$$

де T_i — значення границі *i*-ї позиції вектора, T_{i+1}— значення границі в наступній границі вектора.

У разі виконання наведеної вище умови, границю T_i прирівнювати до $T_i = \frac{T_i - T_{i+1}}{2}$, а значення T_{i+1} видаляти з вектору.

4. Якщо довжина вектора *T* більше *N*, то формується вектор *P*, що описує швидкість зміни потужності спектру на границях

$$P_i = \left|F_{\max}(j) - F_{\max}(j-1)\right| + \left|F_{\max}(j) - F_{\max}(j+1)\right|,$$

де $j = \frac{T_i}{256}$ — номер вікна довжиною в 256 відліків звукового сигналу для границі T_i, $F_{\max}(j)$ — максимальне значення спектра звукового сигналу в *j*-му вікні.

Таким чином, величина P_i характеризує швидкість зміни потужності спектру на границі T_i. На справжніх границях величина P_i має набагато більше значення ніж на несправжніх. Це пояснюється тим що на межфонемних переходах спектр буде відрізнятися, а під час вимови окремої фонеми буде практично ідентичний на сусідніх вікнах.

5. До тих пір, поки розмір вектора *T* є більшим значення *N*, з вектора *T* видаляють границі, відповідно яким елемент вектора *P* є мінімальним.

Висновок

У статті запропоновані модифікації алгоритму кратномасштабного аналізу з адаптивним вибором порогового коефіцієнту міжфонемних переходів $g_{пор}$, що дозволяють знаходити мітки границь межфонемних переходів у звукових сигналів різних дикторів. При використанні методу кратномасштабного аналізу на записах цілих слів було визначено, що можлива поява в результатах міток границь, не відповідних жодної з позицій транскрипції (табл.1.). Тому що цей метод є залежним від диктора — для кожного диктора і окремих випадків необхідні свої параметри η і $g_{пор}$. Внесення описаних змін у алгоритм дозволяє

незалежно від заданого диктора отримувати оптимальну сегментацію мовного сигналу у системах автоматичного розпізнавання мови.

Література

1. Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных. / Информационные процессы. —2004 г. — т. 4. — № 2. — С. 202-220.

2. Дремин И.М., Иванов О.В., Нечитайло В.А. Вейвлеты и их использование. / Успехи физических наук. — 2001 г. — т. 171. — №5. — С. 465-500.

3. Lyudovyk T. Unit Selection Speech Synthesis Using Phonetic-Prosodic Description of Speech Databases / Lyudovyk T., Sazhok M. // Proceedings of the International Conference "Speech and Computer" (SPECOM'2004). — St.-Petersburg (Russia).—2004. — P. 594-599.

Дюжаєв Л.П., Соколов Д.Ю., ***Сегментація мовного сигналу з використанням вейвлет-перетворення****. У даній статті наводиться алгоритм сегментації мовного каналу з використанням кратномасштабного аналізу та вейвлет-перетворення, що показує як в межах стаціонарних ділянок значну роль для аналізу мови відіграють спектральні особливості сигналу, що визначаються передатною характеристикою мовного тракту, яка змінюється в процесі артикуляції. Тобто мовний сигнал характеризується нелінійними флуктуаціями різних масштабів, які є унікальними для різних дикторів. Запропоновані у статті модифікації алгоритму кратномасштабного аналізу дозволяє знаходити мітки границь міжфонемних переходів у звукових сигналах різних дикторів.*

Ключові слова: *вейвлет-перетворення, деталізація, апроксимація, сегментація.*

Дюжаев Л.П., Соколов Д.Ю., ***Сегментация речевого сигнала с использованием вейвлет-преобразования.*** *В данной статье приводится алгоритм сегментации речевого канала с использованием кратномасштабного анализа и вейвлет-преобразования, который показывает как в пределах стационарных участков значительную роль для анализа языка играют спектральные особенности сигнала, которые определяются передаточной характеристикой речевого тракта и изменяется в процессе артикуляции. Т.е. речевой сигнал характеризуется нелинейными флуктуациями различных масштабов, которые являются уникальными для различных дикторов. Предложенные в статье модификации алгоритма кратномасштабного анализа позволяет находить метки границ межфонемних переходов звуковых сигналов различных дикторов.*

Ключевые слова: *вейвлет-преобразование, детализация, аппроксимация, сегментация.*

Dyuzhayev L.P., Sokolov D.Y., ***Voice signal segmentation using wavelet transforms****. This article provides a voice channel segmentation algorithm using the multiresolution analysis and wavelet transform, which features both within the stationary sites significant role for the analysis of language play signal spectral characteristics that are determined by the transfer function of the vocal tract, and is changed in the process of articulation. I.e., speech is characterized by nonlinear fluctuations of different size, which are unique to the various speakers. Proposed in this paper algorithm modification for multiresolution analysis allows you to find borders labels of interphoneme transitions for the sounds of different speakers.*

Keywords: *wavelet transform, detailing, approximation, segmentation.*