До питання виділення фонем у мовному сигналі за допомогою ефекту стоячої хвилі

І.А. Безвербний

Інститут кібернетики імені В.М. Глушкова НАН України, 03187, м. Київ, проспект Академіка Глушкова, 40, ihorbezverbnyi@gmail.com

I.A. Bezverbnyi

ON THE ISSUE OF PHONEME DETECTION IN THE SPEECH SIGNAL BY THE STANDING WAVE EFFECT

Abstract. *An important scientific question is to simplify the computations needed to recognize a speech signal, first of all by decomposing a process called a decomposition. There are many different methods of decomposition and empirical approaches may have different levels of complexity and scope. A quite commonly used class of decomposition methods is spectral analysis, one of which is the Fourier transform. Fourier transform is the decomposition of a process into harmonic functions with fixed frequency and amplitude, which is a significant disadvantage, since the language process is always non-stationary. These methods are very common in the field of time process analysis, but their versatility implies a disadvantage, such as redundancy in calculations and as a result of high hardware requirements. There are methods today that allow for simplicity. These analytical methods are so-called adaptive transformations, for which the basis is determined directly by the input itself, that is, the linguistic signal. This conversion is a Hilbert – Huang transform. But unlike Fourier transforms and wavelet transforms, in process of Hilbert – Huang transformations, decomposition into empirical modes or intrinsic modes (IMFs) occurs, which do not ask. This approach greatly simplifies the analysis of the time process and avoids the problems of increasing the error because of a large number of computations, resulting in reduced time and freeing up the hardware resources required to use such an algorithm. Therefore, the task of this work is to develop a new adaptive method, based on the algorithm for the allocation of phonemes in the speech signal developed earlier. A feature of the new method is to reduce the number of computations by using a standing wave effect to identify individual phonemes in recognition. Thus, the article presents an adaptive method of segmentation of speech signals using the standing wave effect, which is constructed using the Hilbert – Huang transform principles and the method of empirical mode decomposition.*

Key words: *Hilbert – Huang transformations, Intrinsic Mode Functions, IMF, empirical mode, decomposition, numerical and analytical procedure, sliding spectral analysis.*

Анотація. *Завданням цієї роботи є вироблення нового адаптивного методу, спираючись на розроблений раніше алгоритм виділення фонем у мовному сигналі. Особливістю нового методу є скорочення кількості обчислень завдяки використанню ефекту стоячої хвилі.*

Ключові слова: *перетворення Гільберта – Хуанга, мода внутрішніх коливань, емпірична мода, декомпозиція, чисельно-аналітичний метод, ковзний спектральний аналіз.*

Аннотация. *Задача этой работы – создание нового адаптивного метода на основе ранее разработанного алгоритма выделения фонем в речевом сигнале. Особенность нового метода в сокращении количества вычислений благодаря использованию эффекта стоячей волны.*

Ключевые слова: *преобразование Гильберта – Хуанга, мода внутренних колебаний, эмпирическая мода, декомпозиция, численно-аналитический метод, скользящий спектральный анализ.*

Вступ. Важливим науковим питанням є спрощення обчислень, необхідних при розпізнаванні мовного сигналу. Справа в тому, що мовний сигнал в певному наближенні є результатом композиції тональних сигналів. Відбиваючись від стінки завитки вуха, сигнал утворює стоячу хвилю у тих фрагментах вибірки, де спостерігається голосна фонема. Цей ефект доцільно використати для скорочення обчислень при розпізнаванні мовних фонем. В процесі реалізації розпізнавання необхідно здійснити декомпозицію. Існує багато різноманітних методів декомпозиції. Ці методи засновуються на математичних та емпіричних підходах, можуть мати різний рівень складності та області застосування, є дуже поширеними в галузі аналізу часових процесів, однак їх універсальність передбачає такий недолік як надлишковість обчислень у першу чергу з метою уникнення наслідків нестаціонарності й як наслідок високі вимоги до апаратури.

Тим часом на сьогодні є методи, які дозволяють спростити обчислення, 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 [1]. Воно дозволяє

знаходити миттєвий спектр нелінійних нестаціонарних послідовностей. Але на відміну від перетворення Фур'є і вейвлет-перетворення в процесі перетворення Гільберта – Хуанга відбувається розкладання на емпіричні моди або внутрішні коливання (Intrinsic Mode Functions, IMF), 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. Як результат скорочується час і звільнюються апаратні ресурси, необхідні для використання такого алгоритму.

I. Сутність проблеми

Дослідження людського вуха в 19 ст. італійським дослідником Альфонсом Корті дозволило з'ясувати, що звуковний сигнал ідентифікується у так званому кортієвому органі, що в сучасній анатомії називається вушна завитка. Цей орган має форму згорнутих у двовимірну спіраль двох трубок, з'єднаних у центрі (рис. 1).

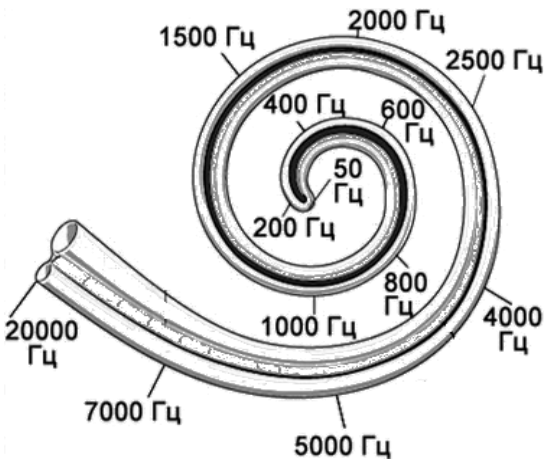


РИС. 1. Виконане Георгом Бекеші графічне зображення завитки людського вуха [3]

Подальші дослідження дозволили висунути гіпотезу, згідно якої сприйняття звуку відбувається за рахунок резонансних коливань волосних клітин, що покривають поверхню базальної мембрани вушної завитки. На початку двадцятого століття Георгу Бекеші вдалося вибудувати механічну модель завитки людського вуха. В результаті спостережень Бекеші помітив, що коливання певних частот знаходять резонансний відгук на відповідних ділянках завитки. Подальші дослідження, зокрема, групи співробітників Гарвардського університету під керівництвом Чжина Куна, підтвердили правильність припущень Бекеші про завитку, як орган перетворення звукового сигналу в нервовий імпульс, який у свою чергу розпізнається відповідними відділами півкуль головного мозку. Важливим результатом цих досліджень є визначення швидкості передачі таких імпульсів – це 200–300 імпульсів на секунду. Тобто саме в середині трубок завитки відбуваються резонансні коливання, які виникають як результат проходження через них у двох протилежних напрямках звукового коливання. Таким чином відбувається накладання двох тотожних хвиль, що поширюються в протилежних напрямках. Це дає підстави розглядати процес виникнення стоячої хвилі, яка утворюється у відповідних ділянках завитки. Саме в тих ділянках завитки, де відбувається накладання коливань, відбувається і резонанс. Завданням запропонованого далі алгоритму є відтворення процесу виникнення стоячої хвилі, і, як подальший результат, частотний аналіз. Далі за частотним портретом ідентифікується фонема. Для автоматизації і спрощення обчислень використовується перетворення Гільберта – Хуанга, яке дозволяє підготувати вхідний часовий масив для цифрового моделювання стоячої хвилі у завитці.

II. Ступінь розробки

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]. В роботі [3] розглянуто адаптивний алгоритм виділення фонем в мовному сигналі, який побудовано з використанням принципів перетворення Гільберта – Хуанга. Використовуючи розроблені там положення, створено вдосконалений метод, що спирається на виявлення ефекту стоячої хвилі у фрагментах загальної вибірки.

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 б суттєво спростити кількість обчислень.

III. Мета

Мета даної роботи – розроблення нового адаптивного методу фонетичного аналізу, що враховував би особливості процесу мовлення на прикладі української фонетики.

IV. Ідея алгоритму

Ідея алгоритму базується на тому, що передача фонем у 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, що несуть інформацію про відповідну фонему, і є предметом дослідження.

Перше ніж сигнал надійде як вхідний аргумент алгоритму, він фільтрується низькочастотним фільтром з метою вилучення високих частот, які є у переважній більшості шумовими перешкодами.

Наступним етапом потрібно розділити низькі та високі частоти на два часових ряди. Найзручніше, з точки зору кількості обчислювальних операцій, отримати ці часові ряди як першу і другу моди перетворення Гільберта – Хуанґа. Отримані моди потрібно піддати фільтрації рекурсивним фільтром рухомого середнього для згладжування нерівностей, які можуть накопичуватися як помилки протягом реалізації алгоритма. Експериментально виявлено, що для аналізу розглянутих у процесі створення алгоритму мовних сигналів найбільш прийнятним для довжини фільтру є число 64 або близьке до нього.

Наступний етап – програмна імітація стоячої хвилі, яка утворюється у вусі людини під час сприйняття звукового сигналу і є важливим елементом розпізнавання людиною мовних фонем. Сприйняття звукового сигналу обумовлюється індивідуальними анатомічними особливостями людини, зокрема, довжиною завитки, де утворюється стояча хвиля від звукового сигналу та якості сигналу. Для моделювання цієї індивідуальної характеристики використовується динамічна довжина вибірки. Для регулювання довжини чергової вибірки необхідно провести експериментальну кількість децимацій залежно від якості відтворення конкретної фонеми. В процесі роботи алгоритму аналізується динамічна ділянка загальної осцилограми звукового сигналу експериментально визначеної довжини, яка достатня для розпізнавання конкретної фонеми. Для подальшого визначення характеристик наявної у сигналі фонеми було розроблено такий алгоритм моделювання стоячої хвилі у завитці.

1. Визначається n – довжина масиву Q, отриманого в результаті децимації.
2. Задаються порожні масиви Z та R вхідного та вихідного звукових сигналів.
3. Задається значення довжини масиву F $m = 0$.
4. Запускається цикл 1 за лічильником $i \in \left[0 : \frac{n}{2}\right]$.
5. В середині циклу обнуляється попереднє значення сумарної вибірки $\sigma_i = 0$ стоячої хвилі; обнуляються допоміжні параметри $q = 0$, $h = 0$.
6. Запускається внутрішній цикл 2 по лічильнику $j \in [0 : n]$.
7. Якщо $i < j + 1$, тоді відбувається формування масивів Z та R;
   a) якщо $i + q \le j$, тоді
      i. $Z_{j-q} = Q_j$, $R_{j-q} = 0$;
   b) якщо ні
      ii. $dec(h)$, $inc(m)$;
      iii. $Z_{j-q} = Q_{i+h}$, $R_{j-q} = Q_j - Z_{j-q}$.
8. Якщо ні $inc(q)$.
9. Кінець циклу 2.
10. Кінець циклу 1.

Розпізнавання фонеми відбувається в момент утворення стоячої хвилі певної конфігурації.

На рис. 2 та 3 показано результат вияв-

лення стоячої хвилі на різних ділянках вибірки. Утворення стоячої хвилі від часової вибірки динамічної довжини, відбувається в результаті накладання цієї вибірки самої на себе у зворотньому напрямку.

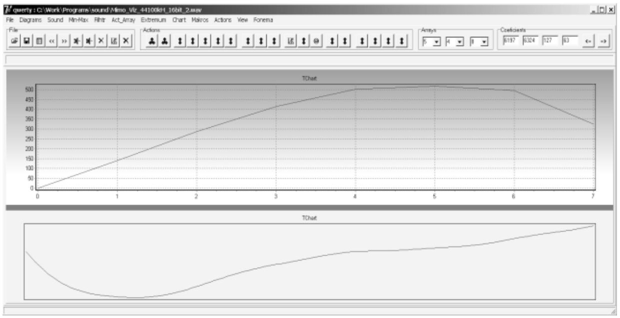


РИС. 2. Стояча хвиля на ділянці на кроці, що передує розпізнаванню фонеми

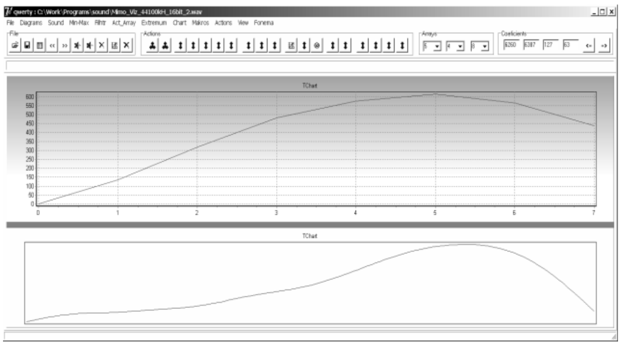
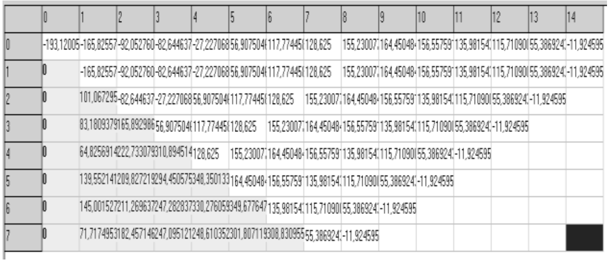


РИС. 3. Стояча хвиля на ділянці на кроці, що дозволяє розпізнавання фонеми

Таким чином розпізнавання фонеми відбувається в межах певного слайду, який послідовно з певним експериментально визначеним кроком переміщується по всій довжині вибірки (рис. 4). З фонем, які було розпізнано в результаті роботи алгоритму, формується індивідуальна база даних. Також індивідуальна характеристика мовця передбачає експериментальне встановлення необхідної довжини вхідної вибірки, з якої формується пересувний слайд із стоячою хвилею. Довжина слайду може змінюватися в процесі розпізнавання залежно від конкретної фонеми. Тому процес розпізнавання відбувається у циклі, який забезпечує визначення всіх можливих фонем.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	-193,12005	-165,82557	-92,052760	-82,644637	-27,227068	56,907504	117,77445	128,625	155,23007	164,45048	156,55759	135,98154	115,71090	55,386924	-11,924595
1	0	-165,82557	-92,052760	-82,644637	-27,227068	56,907504	117,77445	128,625	155,23007	164,45048	156,55759	135,98154	115,71090	55,386924	-11,924595
2	0	101,067295	-82,644637	-27,227068	56,907504	117,77445	128,625	155,23007	164,45048	156,55759	135,98154	115,71090	55,386924	-11,924595	
3	0	83,1809379	165,892986	56,907504	117,77445	128,625	155,23007	164,45048	156,55759	135,98154	115,71090	55,386924	-11,924595		
4	0	64,8256914	222,733079	310,894514	128,625	155,23007	164,45048	156,55759	135,98154	115,71090	55,386924	-11,924595			
5	0	139,552141	209,827219	294,450575	348,350133	164,45048	156,55759	135,98154	115,71090	55,386924	-11,924595				
6	0	145,001527	211,269637	247,282837	330,276059	349,677647	135,98154	115,71090	55,386924	-11,924595					
7	0	71,7174953	182,457146	247,095121	248,610352	301,807119	308,830955	55,386924	-11,924595						

РИС. 4. Моделювання процесу проходження звукової вибірки динамічно обмеженої довжини в слайді

Висновки. Таким чином у статті представлено адаптивний метод сегментації мовних сигналів, що побудований з використанням принципів перетворення Гільберта – Хуанга і ефекту стоячої хвилі. Середовище моделювання було створене на базі вільнопоширюваного програмного продукту Delphi7 фірми Borland.

СПИСОК ЛІТЕРАТУРИ

1. 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.
2. 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.
3. Georg von Bekesy. Concerning the Pleasures of Observing, and the Mechanics of the Inner Ear; Nobel Lecture, December 11, 1961. http://www.neurosci.info/courses/systems/Nobels/1961 von Bekesy/bekesy-lecture.pdf.
4. Семотюк М.В., Безвербний І.А. Адаптивний алгоритм виділення фонем у мовному сигналі. *Комп'ютерні засоби, мережі та системи.* 2017. № 16. С. 14–19.

REFERENCES

1. 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.
2. Davydov A.G., Lobanov B.M. Ispoljzovanie periodi4nosti re4evogo signala pri fonemnoj segmentacii re4i. Doklady BGUIR. 2006 APRELJ–IUNJ № 2 (14), http://ssrlab.by/wp-content/uploads/2006/12_100229_1_ 57873.pdf.
3. Georg von Bekesy. Concerning the Pleasures of Observing, and the Mechanics of the Inner Ear; Nobel Lecture, December 11, 1961. http://www.neurosci.info/courses/systems/Nobels/1961 von Bekesy/bekesy-lecture.pdf.
4. Semotjuk M.V., Bezverbnyi I.A. Adaptyvnyi alhorytm vydilenya fonem u movnomu sygnali. *Kompjuterni zasoby, merezi ta systemy.* 2017. № 16. С. 14–19.

Одержано 25.10.2019