

УДК 004.89:004.93

А. В. Ниценко

Государственное учреждение «Институт проблем искусственного интеллекта», г. Донецк
283048, г. Донецк, ул. Артема, дом 118 б

МЕТОД КОРРЕКЦИИ ФОНЕМНОЙ СЕГМЕНТАЦИИ РЕЧИ

A. V. Nitsenko

Public institution «Institute of Problems of Artificial intelligence», Donetsk
283048, Donetsk, Artema st., 118 b

SPEECH PHONEMIC SEGMENTATION CORRECTION METHOD

А. В. Ніценко

Державна установа «Інститут проблем штучного інтелекту», м. Донецьк
283048, м. Донецьк, вул. Артема, буд. 118 б

МЕТОД КОРЕКЦІЇ ФОНЕМНОЇ СЕГМЕНТАЦІЇ МОВИ

В статье описан способ коррекции автоматической фонемной сегментации, использующий фильтрацию сигнала цифровым полосовым фильтром и среднее значение вариации сигнала в качестве признака для определения вокализованности сегмента. Данный подход позволяет значительно уменьшить количество ошибок сегментации, связанных с особенностями произношения диктора.

Ключевые слова: распознавание речи, автоматическая сегментация речи, вариация.

This article describes a method of speech segmentation correction using digital band-pass filter and average variation value as a feature to determine segment vocalization. This method allows significantly reducing speaker-related segmentation errors.

Keywords: speech recognition, automatic speech segmentation, variation.

У статті описаний спосіб корекції автоматичної фонемної сегментації, який використовує фільтрацію сигналу цифровим смуговим фільтром і середнє значення варіації сигналу як ознака для визначення вокалізованості сегмента. Даний підхід дозволяє значно зменшити кількість помилок сегментації, пов'язаних з особливостями вимови диктора.

Ключові слова: розпізнавання мови, автоматична сегментація мови, варіація.

Во многих областях речевых технологий требуются устойчивые методы сегментации речи. В частности, сегментация является важным этапом при начальном обучении системы распознавания речи и при распознавании слитной речи. В настоящее время разработано и применяется большое количество различных методов сегментации речевых сигналов на слова, слоги и фонемы [1-8].

В работах [9-11] был предложен метод автоматической фонемной сегментации речевого сигнала. С помощью этого метода сигнал с заранее неизвестным фонемным составом сегментируется на участки гласных (тип сегмента «W»), голосовых согласных («C»), глухих фрикативных звуков («F») и аффрикат, глухих взрывных (паузообразных) звуков («P»). Эта процедура играет важную роль в процессе распознавания как отдельно произносимых слов, так и распознавания слитной речи. В частности, на нее опирается описанный в работе [12] метод распознавания слитно произносимых фраз с использованием глухих фрагментов. Для сегментации на вокализованные и невокализованные участки используется фильтрация полосовым фильтром и количество точек постоянства сигнала в качестве признака (точкой постоянства мы называем дискретный момент времени такой, что в следующий момент значение сигнала остается неизменным). Однако в ходе сегментации речи возникает определенная трудность при выделении участков звонких согласных звуков [б], [г], [д]. Для них количество точек постоянства у некоторых дикторов может быть относительно большим и при сегментации они могут быть ошибочно отнесены к невокализованным сегментам. После описанной сегментации сегменты этих звуков классифицируются как «P», хотя должны классифицироваться как «C». Это ведет к ошибкам при распознавании фраз. Поэтому **целью работы** является разработка эффективного способа коррекции подобных ошибок классификации путем дополнительной проверки сегментов «P» на наличие вокализованности. Для достижения данной цели в работе предлагаются признаки вокализованности сегмента и проводится исследование эффективности работы метода.

Пусть $x_i = x(t_i)$, $i=0, 1, \dots, N$ – дискретный звуковой сигнал. Его вариацией называется величина

$$V = \sum_{i=0}^{N-1} |x_{i+1} - x_i|.$$

Аналогично определяется вариация произвольного отрезка этого сигнала. Исследования показывают, что вокализованные сегменты «C» имеют более высокую среднюю вариацию (см. (1)), чем невокализованные «P». Поэтому можно предложить следующую проверку на вокализованность. Вначале весь речевой сигнал обрабатывается полосовым фильтром с полосой пропускания от 100 до 200 Гц. При этом используется цифровой нерекурсивный фильтр вида

$$y_n = \sum_{k=-N}^N a_k w_k x_{n-k},$$

где a_k – коэффициенты фильтра, $2N$ – порядок фильтра, w_k – коэффициенты сглаживающего окна, x_i – отсчеты входного сигнала. В качестве оконной функции

используется функция Ханна $w_k = \frac{1}{2} \left(1 + \frac{\cos \pi k}{N} \right)$. После применения фильтра производится нормирование полученного сигнала по амплитуде делением каждого отсчета на максимальную амплитуду всего сигнала и умножением на 255.

Затем, если в сегментации некоторый сегмент «P» находится между голосовыми сегментами (рис. 1), то на соответствующем участке сигнала, обработанного вышеописанным фильтром, вычисляется средняя вариация как среднее значение вариации на окнах сигнала длиной 256 отсчетов:

$$V_{cp.} = \frac{\sum_{j=0}^k \sum_{i=256j}^{256(j+1)-2} |x_{i+1} - x_i|}{k}, \quad (1)$$

где k – количество окон на проверяемом отрезке. Если вычисленное значение средней вариации на проверяемом отрезке превышает некоторое пороговое значение T , то данный участок сигнала считается вокализованным и относится к классу «С» (рис. 2).

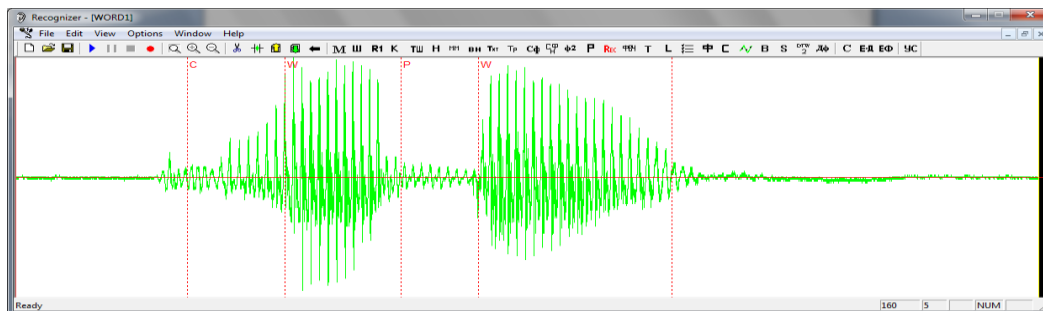


Рисунок 1 – Некорректная сегментация речевого сигнала для слова «вода»

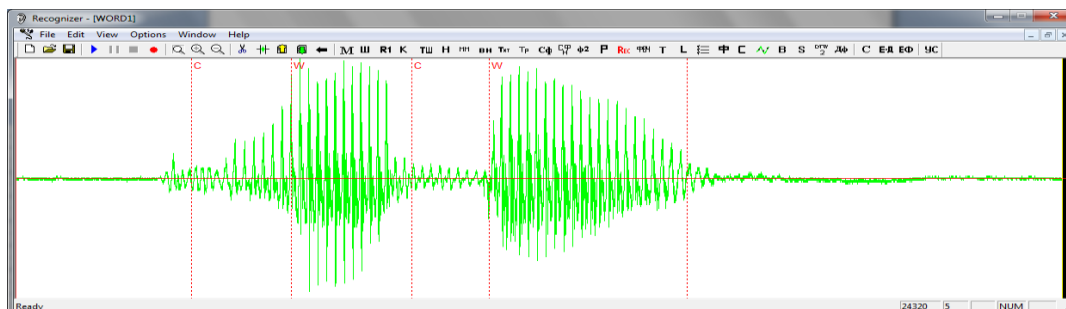


Рисунок 2 – Сегментация речевого сигнала для слова «вода» после коррекции

Пороговое значение вариации T должно подбираться с учетом индивидуальных особенностей диктора и микрофона.

Был проведён эксперимент, в котором участвовали 2 диктора. Эксперимент проводился со словарём, состоящим из 100 слов, содержащих согласные звуки [б], [г], [д] и их мягкие варианты. Для каждого диктора был создан банк речевых сигналов – результатов произнесения слов словаря. Запись производилась в условиях низкого уровня фонового шума. Параметры записи наборов слов: частота дискретизации – 22050 Гц; разрядность квантования – 8 бит; средняя длительность записанного слова – 2 с (включая окружающие слово паузы, длительностью около 0,3 с каждая). После записи ко всем сигналам применялся алгоритм сегментации с последующей коррекцией. Пороговое значение вариации в данном эксперименте составляло 60. В табл. 1 приведены некоторые примеры работы алгоритма коррекции над сегментацией, содержащей ошибочное определение вокализованных участков как невокализованных.

Таблица 1 – Примеры работы алгоритма коррекции сегментации

Слово	Транскрипция	Сегментация с ошибкой	Сегментация после коррекции
абака	[абака]	WPWPW	WCWPW
агама	[агама]	WPWCW	WCWCW
агаровый	[агаравыј]	WPWCWCWC	WCWCWCWC
агат	[агат]	WPWP	WCWP
адамов	[адамаф]	WPWCWP	WCWCWP
адат	[адат]	WPWP	WCWP
багажный	[багажныј]	CWPWCWC	CWCWCWC
белоголовый	[белагаловыј]	CWCWPWCWPWC	CWCWCWCWCWC
борода	[барада]	CWCWPW	CWCWCW
водолаз	[вадалас]	CWPWCWF	CWCWCWF
веди	[веди]	CWPW	CWCW
выбег	[выбек]	CWPWP	CWCWP
выбежать	[выбежат]	CWPWCWP	CWCWCWP
добыча	[дабычя]	CWPWFW	CWCWFW
ибо	[иба]	WPW	WCW
иго	[ига]	WPW	WCW

В результате эксперимента было установлено, что алгоритм коррекции смог исправить ошибочную классификацию вокализованных участков в 98% случаев.

Выводы

Предложен метод коррекции автоматической фонемной сегментации речевого сигнала, использующий фильтрацию сигнала цифровым полосовым фильтром и среднее значение вариации сигнала в качестве признака вокализованности сегмента. Метод коррекции позволяет значительно снизить количество ошибок при определении вокализованных/невокализованных сегментов, связанных с особенностями произношения диктора. Он был реализован в разрабатываемой с участием автора экспериментальной системе распознавания слитно произнесенных фраз, где продемонстрировал высокую эффективность.

Список литературы

1. Давыдов А. Г. Использование периодичности речевого сигнала при фонемной сегментации речи [Текст] / А. Г. Давыдов, Б. М. Лобанов // Доклады БГУИР. – 2006. – № 2 (14). – С. 69–74.
2. Жилияков Е. Г. Сегментация речевого сигнала на основе анализа особенностей распределения долей энергии по частотным интервалам [Текст] / Е. Г. Жилияков, Е. И. Прохоренко, А. В. Болдышев, А. А. Фирсова // Вестник Национального технического университета Харьковский политехнический институт. – 2011. – № 17. – С. 44–50.
3. Елистратов С. А. Сравнение параметров для выделения вокализованных сегментов и классификации гласных фонем [Текст] / С. А. Елистратов, М. А. Косенко, Е. Ю. Костюченко, А. А. Чичерин // Доклады ТУСУР. – 2012. – № 1. – С. 171–174.
4. Natarajan V. A. Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies [Текст] / V. A. Natarajan, S. Jothilakshmi // International Journal of Computer Applications. – 2012. – Vol. 56, № 15. – P. 24–27.
5. Beritelli F. Robust Voiced/unvoiced classification using fuzzy rules / F. Beritelli, S. Casale // 1997 IEEE workshop on speech coding for telecommunications proceeding – 1997. – P. 5–6.

6. Kim Y.-J. Automatic segmentation combining an HMM-based approach and spectral boundary correction [Text] / Y.-J. Kim, A. Conkie // Proceedings of ICSLP 2002. – 2002. – P. 145–148.
7. Toledano D. T. Automatic phoneme segmentation [Text] / D. T. Toledano, L. A. H. Gomez, L. V. Grande // IEEE Trans. Speech and Audio Proc. – 2003. – № 11. – P. 617–625.
8. Шелепов В. Ю. О распознавании первого звука в слитном речевом отрезке [Текст] / В. Ю. Шелепов, А. В. Ниценко // Проблемы искусственного интеллекта – 2015. – № 0(1). – С. 116–122.
9. Шелепов В. Ю. Сегментация речевого сигнала на основе предположения о его фонетическом составе [Текст] / В. Ю. Шелепов, А. В. Ниценко // Проблемы искусственного интеллекта. – 2016. – № 1. – С. 73–81.
10. Сегментация и диффонное распознавание речевых сигналов [Текст] / А. К. Бурибаева, Г. В. Дорохина, А. В. Ниценко, В. Ю. Шелепов // Труды СПИИРАН. – 2013. – № 31. – С. 20–42.
11. Шелепов В. Ю. Сегментация и диффонное распознавание речи [Текст] / В. Ю. Шелепов, А. В. Ниценко. – Донецк : ГУ ИПИИ, 2015. – 231 с.
12. Shelepov V. Ju. Recognition of the continuous-speech Russian phrases using their voiceless fragments / V. Ju. Shelepov, A. V. Nitsenko // Eurasian Journal of Mathematical and Computer Applications. – 2016. – Vol. 4, № 4. – P.19–24.

References

1. Davydov A. G., Lobanov B. M. Ispol'zovanie periodichnosti rechevogo signala pri fonemnoj segmentacii rechi [The speech signal periodicity utilization in phoneme speech segmentation]. *Doklady BGUIR*, 2006, no. 2 (14), pp.69-74.
2. Zhiljakov E. G., Prohorenko E. I., Boldyshev A. V., Firsova A. A. Segmentacija rechevogo signala na osnove analiza osobennostej raspredelenija dolej energii po chastotnym intervalam [Segmentation of speech signals based on analysis of distribution of shares on energy frequency band]. *Vestn. Khar'k. politekhn. in-ta*. [Bulletin of the Kharkov Polytechnic Institute], 2011, no.17, P. 44–50.
3. Elistratov S. A., Kosenko M. A., Kostjuchenko E. Y., Chicherin A. A. Sravnenie parametrov dlja vydelenija vokalizovannyh segmentov i klassifikacii glasnih fonem [Comparison of parameters for the selection and classification of voiced phonemes segments]. *Doklady TUSUR* [Proceedings of TUSUR University], 2012, no. 1, pp. 171 - 174.
4. Natarajan V. A., Jothilakshmi S. Segmentation of Continuous Speech into Consonant and Vowel Units using Formant Frequencies. *International Journal of Computer Applications*, 2012, vol. 56, no. 15, pp. 24 – 27.
5. Beritelli F., Casale S. Robust Voiced/unvoiced classification using fuzzy rules. *1997 IEEE workshop on speech coding for telecommunications proceeding*, 1997, pp. 5-6.
6. Kim Y.-J., Conkie A. Automatic segmentation combining an HMM-based approach and spectral boundary correction. *Proceedings of ICSLP 2002*, 2002, pp.145-148.
7. Toledano D.T., Gomez L.A.H., Grande L.V. Automatic phoneme segmentation. *IEEE Trans. Speech and Audio Proc*, 2003, No. 11, pp. 617-625.
8. Shelepov V. Ju., Nitsenko A. V. O raspoznavanii pervogo zvuka v slitnom rechevom otrezke [On recognition of the first sound in continuous speech fragment]. *Problemy iskusstvennogo intellekta* [Problems of Artificial Intelligence], 2015, no. 0(1), pp. 116-122.
9. Shelepov V. Ju., Nitsenko A. V. Segmentacija rechevogo signala na osnove predpolozhenija o ego foneticheskom sostave [Segmentation of speech signal on the assumption about its phonetic structure]. *Problemy iskusstvennogo intellekta* [Problems of artificial intelligence], 2016, no.1, pp. 73-81.
10. Buribaeva A.K., Dorohina G.V., Nitsenko A.V., Shelepov V.Ju. Segmentacija i difonnoe raspoznavanie rechevyh signalov [Segmentation and diphone recognition of speech signals]. *Trudy SPIIRAN* [SPIIRAS Proceedings], 2013, no. 31, pp. 20–42.
11. Shelepov V. Ju., Nitsenko A.V. *Segmentacija i difonnoe raspoznavanie rechi* [Segmentation and diphone recognition of speech], Donetsk, GUIPII, 2015. 231 p.
12. Shelepov V. Ju., Nitsenko A.V. Recognition of the continuous-speech Russian phrases using their voiceless fragments. *Eurasian Journal of Mathematical and Computer Applications*, 2016, Vol. 4, No.4. pp.19–24.

RESUME*A. V. Nitsenko**Speech Phonemic Segmentation Correction Method*

Background: Automatic speech segmentation is required in many fields of speech technologies. It has an important role in speech recognition training and continuous speech recognition. Segmentation is also used to detect the proper start and end points of speech events. Today there are many segmentation methods, which can segment speech into words, subwords, syllables and phonemes.

Materials and methods: This article describes a method of speech segmentation correction using digital band-pass filter and average variation value as a feature to determine segment vocalization.

Results: The method of speech phonemic segmentation correction is suggested. The accuracy and stability of the correction algorithm are evaluated. Experiments show that segmentation correction algorithm can correct misclassification of voiced consonant regions as unvoiced ones in 98% cases.

Conclusion: This method allows significantly reducing speaker's specific pronunciation-related segmentation errors due to misclassification of voiced regions as unvoiced ones. The proposed approach is implemented in experimental spoken phrases recognition software demonstrating high effectiveness.

Статья поступила в редакцию 25.01.2017.