УДК 004.934.2 DOI: 10.18698/2541-8009-2019-12-556

ОПРЕДЕЛЕНИЕ ЧАСТОТЫ ОСНОВНОГО ТОНА МЕТОДОМ ПОИСКА ПИКОВ В АМПЛИТУДНОМ СПЕКТРЕ С ДОБАВЛЕНИЕМ ВЕСОВЫХ ЧАСТОТНЫХ ФУНКЦИЙ

А.Б. Жукова zhukovaab@student.bmstu.ru
SPIN-код: 1923-4866

А.Л. Масленников amas@bmstu.ru
SPIN-код: 7087-6303

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

В распознавании речи выделяют задачи определения частоты основного тона и так называемых формантных частот. По значениям этих частот можно определять звуки языка — фонемы, произнесенные человеком. Существует несколько методов определения этих частот. В данной работе представлен метод определения частоты основного тона, основанный на известном механизме поиска пиков амплитудного спектра, но с добавлением сглаживающего фильтра и частотных весовых функций. Рассмотрены весовые функции двух видов: экспоненциальная и линейная. Предложенный метод применен для анализа точности определения частоты основного тона для шести испытуемых (трех мужчин и трех женщин). Результаты эксперимента показали, что существует интервал допустимых частот среза сглаживающего фильтра, а введение частотных весовых функций позволяет повысить надежность определения частоты основного тона.

Ключевые слова

Распознавание речи, частота основного тона, формантный анализ, спектр речевого сигнала, сглаживающий фильтр, весовые частотные функции, фильтр Савицкого – Голея, пики амплитудного спектра

Поступила в редакцию 18.10.2019

Введение. Определение частоты основного тона и соответствующих формант речевых сигналов является одной из задач в области распознавания речи. Вычисление данных частот используется для распознавания букв или слогов (строго говоря — фонем), произносимых человеком [1–3]. Частоты основного тона и формант определяют по спектру речевого сигнала, который ввиду физиологических особенностей голосового аппарата человека является гармоническим спектром [4–6].

Основным тоном речевого сигнала называют частоту, на которую приходится максимум мощности сигнала на частотах спектра, кратных частоте основного тона [3]. Другими словами, частота основного тона — это частота, через которую повторяются пики амплитудного спектра. Считают, что частота основного тона находится в интервале 65…350 Гц, помимо этого основной тон традици-

онно определяют по гласным звукам, а число формант, как правило, соответствует двум-трем [7, 8]. Частоты основного тона и формант связаны с пиками амплитудного спектра, поэтому для корректного определения формант необходимо сначала определить частоту основного тона и исключить ее из последующего формантного анализа.

Существуют несколько подходов к определению частоты основного тона, наиболее часто используют два из них: подход с использованием линейного предсказания (LPC—Linear Prediction Coefficients) и подход с непосредственным выделением пиков амплитудного спектра [9, 10]. В подходе с использованием линейного предсказания вычисляются передаточная функция и сглаженный спектр речевого сигнала. Он позволяет определить частоту основного тона с достаточно высокой точностью, однако его реализация в режиме реального времени существенно затруднена. Второй подход, основанный на непосредственном определении пиков амплитудного спектра, не является тривиальной задачей ввиду сложной формы спектра и наличия акустических и измерительных шумов при записи речевого сигнала. Стоит отметить, что под определением пиков понимается в первую очередь определение их частот, а не магнитуд.

В данной работе рассмотрен подход к непосредственному определению пиков амплитудного спектра с предварительным сглаживанием спектра и использованием весовых частотных функций.

Методы определения пиков амплитудного спектра. Определение частоты основного тона можно сформировать в виде оптимизационной задачи. В рамках этой задачи, например, методом полного перебора (с учетом разрешения по частоте) осуществляется поиск частоты, для которой произведение магнитуд пиков на кратных варьируемой частоте является максимальным:

$$f_0 = \arg\max_{f \in \Theta} \prod_k \left| S(j2\pi kf) \right|, \qquad (1)$$

где f_0 — частота основного тона; Θ — множество допустимых значений частот основного тона; $S(\cdot)$ — спектр сигнала.

При решении данной оптимизационной задачи возникает сложность, связанная с тем, что спектр аудиосигнала, как правило, определен до частоты 22,4 кГц, однако в речевых сигналах на частотах выше 2 кГц присутствуют в основном только акустические и измерительные шумы. В результате этого магнитуды спектров, вычисленные с учетом составляющих на высоких частотах, могут быть очень малыми величинами, как следствие, минимизируемый функционал может принять весьма малое значение, что приведет к ошибке определения частоты основного тона. Для решения этой проблемы можно принять следующие меры. Во-первых, ограничить диапазон частот спектра, которые используются для расчета указанного в уравнении (1) произведения. В данной работе в качестве подобного ограничения выбрано значение 800 Гц, превышающее макси-

мальную границу интервала частот основного тона чуть более чем в 2 раза. Во-вторых, сама частота основного тона, т. е. f_0, ищется в некотором диапазоне, например в интервале 65…350 Гц, соответствующем диапазону определения частоты основного тона.

Однако, несмотря на указанные выше меры, вероятность «обнуления» целевой функции остается достаточно высокой. Для решения этой проблемы можно использовать сглаживание спектра, например, фильтр Савицкого – Голея [11, 12]. Сглаживание спектра позволяет исключить «обнуление» целевой функции в случае случайного нулевого значения в спектре сигнала. Пример исходного и сглаженного спектра с частотой среза сглаживающего фильтра $f_{\text{ср}} = 200$ Гц представлен на рис. 1.

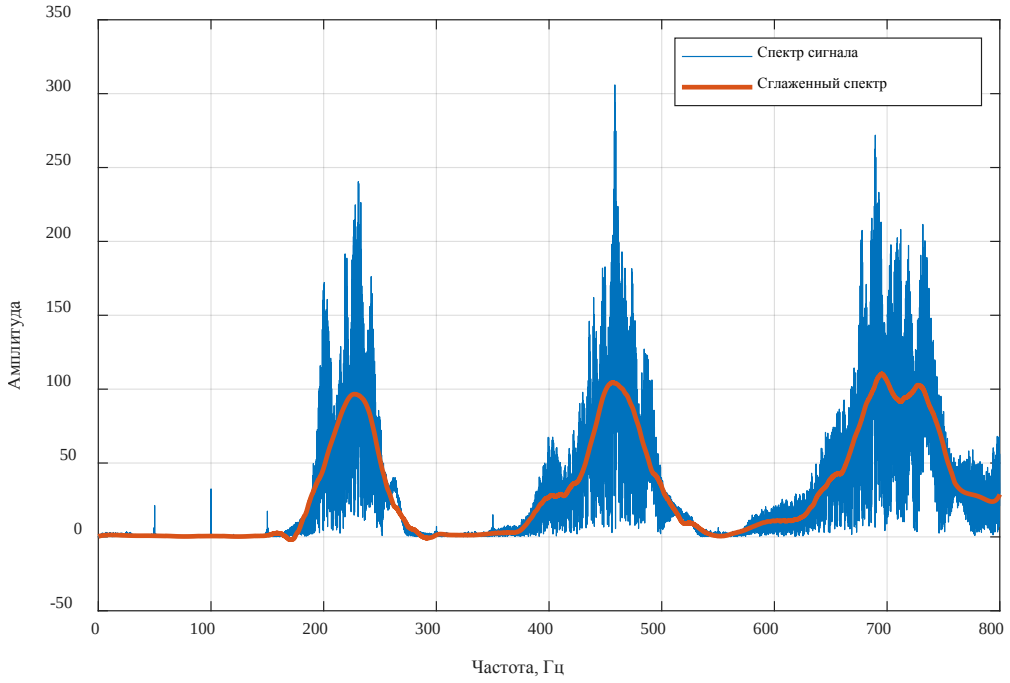

Рис. 1. Пример исходного и сглаженного спектра

Помимо этого можно заменить в постановке исходной оптимизационной задачи, сформулированной в уравнении (1) произведение на сумму, а как следствие исключить возможность «обнуления» целевой функции:

$$f_0 = \arg\max_{f \in \Theta} \sum_k \left| S(j2\pi kf) \right|. \qquad (2)$$

Стоит отметить, что при наличии сильно выраженных акустических и измерительных шумов пик с максимальной магнитудой может ошибочно оказаться на другой частоте, как проиллюстрировано на рис. 2.

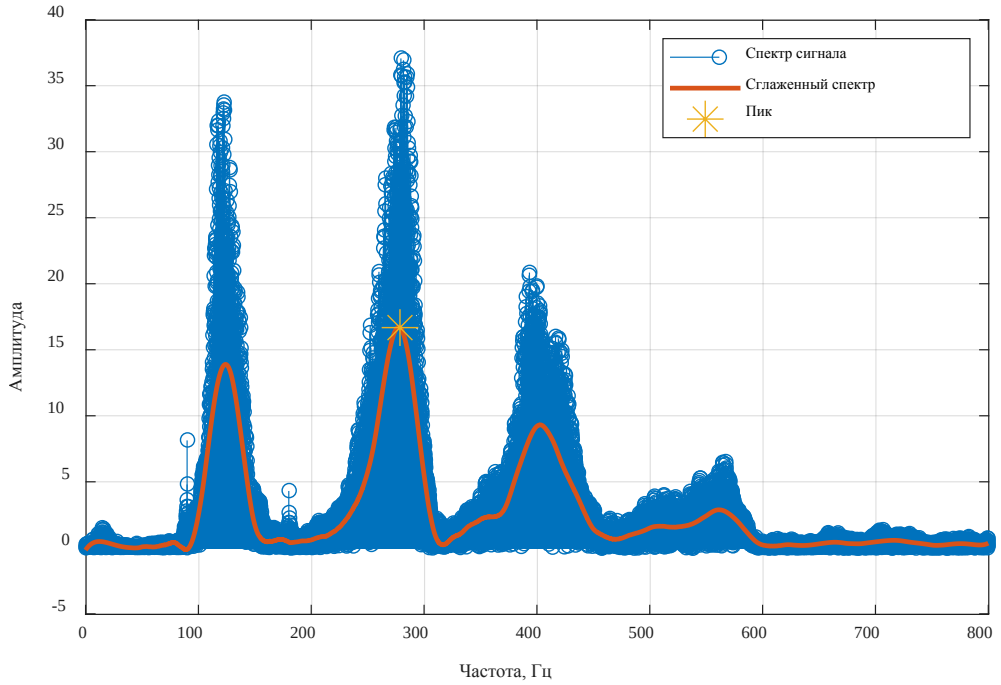

Рис. 2. Пример неправильного определения частоты основного тона

Другими словами, можно констатировать, что вклад магнитуд пиков с увеличением частоты следует уменьшить. Это можно осуществить путем добавления в оптимизационную задачу частотной весовой функции, а именно:

$$f_0 = \arg\max_{f \in \Theta} \prod_k \left| \omega(f)\ S(j2\pi kf) \right|;$$

$$f_0 = \arg\max_{f \in \Theta} \sum_k \left| \omega(f)\ S(j2\pi kf) \right|.$$

В данной статье рассматриваются два типа частотных весовых функций:1) экспоненциальная, которая определяется следующим образом:

$$\omega(f) = \exp(-\lambda f),$$

где $\lambda = 0{,}01$ — коэффициент затухания, а f — частота,

2) линейная, которая имеет следующий вид:

$$\omega(f) = af + b,$$

где $a = -b/800$, $b = 1$.

Параметры частотных весовых функций были подобраны эмпирически, а их выбор для обобщенного случая требует дополнительных исследований. Вид амплитудных спектров без использования частотных весовых функций и с их использованием представлен на рис. 3.

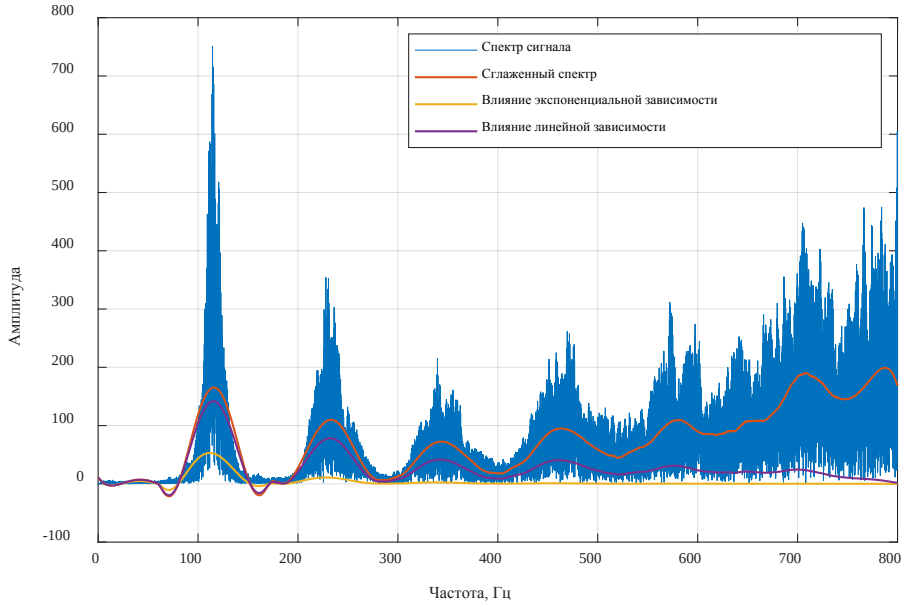

Рис. 3. Пример влияния весовых коэффициентов на спектр сигнала

Таким образом, алгоритм определения частоты основного тона описанным выше методом можно описать последовательностью действий:

1) вычисление амплитудного спектра сигнала;

2) исключение составляющих амплитудного спектра на частотах больше 2 кГц;

3) сглаживание полученного амплитудного спектра фильтром Савицкого – Голея;

4) применение частотной весовой функции к спектру;

5) решение оптимизационной задачи (вычисление суммы) для каждой частоты в интервале 50…400 Гц (с шагом, соответствующим величине разрешения по частоте) с исключением составляющих на частотах выше 800 Гц;

6) выделение частоты, для которой сумма получается максимальной.

Постановка эксперимента. Для апробации описанного выше метода определения пиков амплитудного спектра для определения частоты основного тона проведена серия практических экспериментов. В рамках одного эксперимента испытуемый произносит 10 раз подряд через продолжительные паузы повторяющуюся одну из фонем русского языка (запись сохраняется в один звуковой файл). Запись осуществлялась с использованием внешнего микрофона и программы, написанной в программной среде National Instruments LabVIEW, а обработка результатов осуществляется в MathWorks MATLAB.

Порядок обработки экспериментальных данных для каждого испытуемого следующий:

1) для каждого звукового файла определяется частота основного тона с различными частотами среза сглаживающего фильтра и при трех вариантах использования частотных весовых функций (без весовой функции, с экспоненциальной весовой функцией, с линейной весовой функцией);

2) высчитывается медиана и среднеквадратичное отклонение значения частоты основного тона, определенное по десяти записанным фрагментам;

3) выполняется сравнение полученных результатов.

В эксперименте приняло участие шесть испытуемых (трое женского пола и трое — мужского). Зависимость среднего значения (по десяти записанным фонемам) частоты основного тона от частоты среза сглаживающего фильтра Савицкого – Голея для всех шести испытуемых представлена на рис. 4. Можно заметить, что точность определения частоты основного тона (ее среднее значение по экспериментальной выборке) существенно уменьшается при значении частоты среза сглаживающего фильтра, превышающей 200…250 Гц.

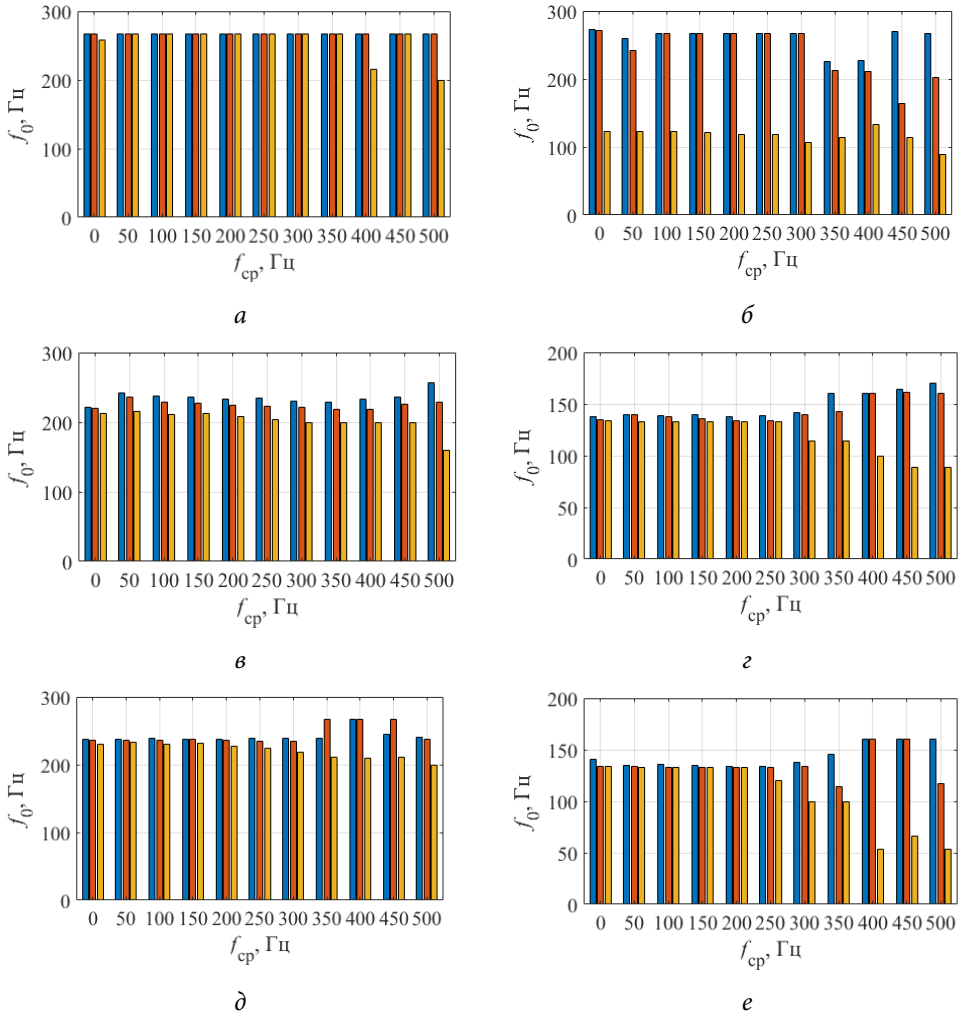

Рис. 4. Зависимость среднего значения частоты основного тона от частоты среза сглаживающего фильтра:

а, *в*, *д* — женские голоса; *б*, *г*, *е* — мужские голоса;

Графические зависимости, иллюстрирующие среднеквадратичное отклонения частоты основного тона в зависимости от частоты среза сглаживающего фильтра, представлены на рис.5. Можно констатировать, что применение линейной частотной функции для мужских голосов может привести к существенному улучшению точности определения част основного тона (по сравнению со случаем без

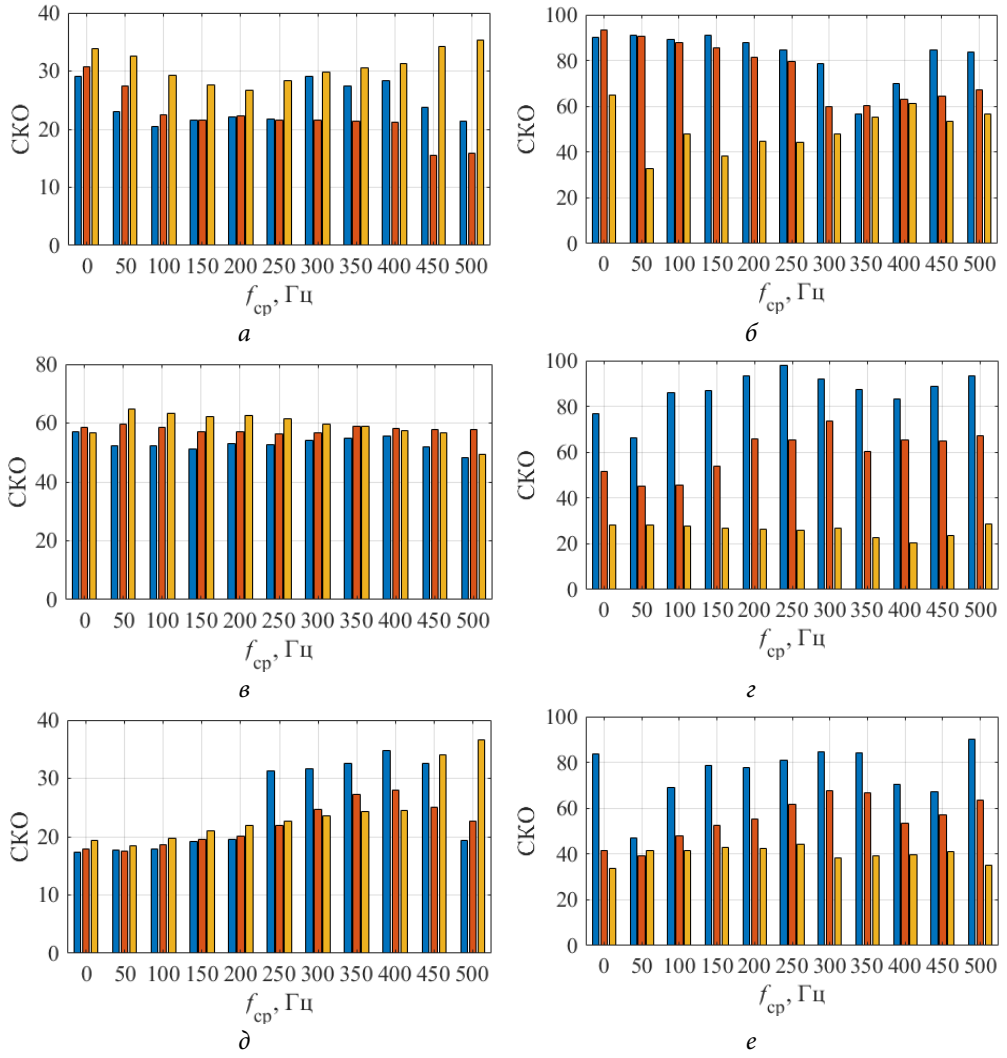

Рис. 5.Зависимость среднеквадратичного отклонения частоты основного тона от частоты среза сглаживающего фильтра:

а, *в*, *д* — женские голоса; *б*, *г*, *е* — мужские голоса;

использования или при использовании экспоненциальной частотной весовой функции), где под улучшением понимается минимальность дисперсии оценки частоты основного тона. Для женских голосов однозначный выбор весовой ча-

стотной функции сделать нельзя, однако более устойчивые результаты получаются при использовании экспоненциальной весовой функции.

Выводы. В данной статье представлен метод определения частоты основного тона в речевом сигнале через решение задачи оптимизации с добавлением процедуры сглаживания (с использованием фильтра Савицкого – Голея) амплитудного спектра, а также частотных весовых функций. Рассмотрены также ряд дополнительных ограничений, накладываемых на решение оптимизационной задачи.

Описанный метод был применен к обработке экспериментальных данных для шести испытуемых. Результаты показали, что для мужских голосов лучше применять линейную весовую функцию, а для женских голосов сделать однозначный выбор не представляется возможным. Однако применение экспоненциальной весовой функции для женских голосов в целом позволяет получить более робастный результат.

Литература

[1] Бондаренко М.Ф., Работягов А.В., Щепковский С.В. Распознавание речи: этапы развития, современные технологии и перспективы их применения. *Бионика интеллекта*, 2010, № 2(73), с. 164–168.

[2] Ortega-García J., González-Rodríguez J. Overview of speech enhancement techniques for automatic speaker recognition. *Proc. ICSLP'96*, 1996, vol. 2, pp. 929–932.

[3] Плотников В.Н., Суханов В.А., Жигулевцев Ю.Н. Речевой диалог в системах управления. М., Машиностроение, 1988.

[4] Деркач М.Ф., ред. Динамические спектры речевых сигналов. Львов, Вища школа, 1983.

[5] Сорокин В.Н. Теория речеобразования. М., Радио и связь, 1985.

[6] Чистович Л.А., Венцов А.В., Гранстрем М.П. и др. Физиология речи. Восприятие речи человеком. Л., Наука, 1976.

[7] Лабутин В.К., Молчанов А.П. Слух и анализ сигналов. М., Энергия, 1967.

[8] Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. Киев, Наукова думка, 1987.

[9] Злоказов В.Б. Метод для автоматического поиска пиков в гамма-спектрах. Дубна, ОИЯИ, 1981.

[10] Маркел Д.Д., Грэй А.Х. Линейное предсказание речи. М., Связь, 1980.

[11] Savitzky A.A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 1964, vol. 36, no. 8, pp. 1627–1639. DOI: 10.1021/ ac60214a047 URL: https://pubs.acs.org/doi/abs/10.1021/ac60214a047

[12] Savitzky A.A. A historic collaboration. *Anal. Chem.*, 1989, vol. 61, no. 15, pp. 921A–923A. DOI: 10.1021/ac00190a003 URL: https://pubs.acs.org/doi/10.1021/ac00190a003

Жукова Анастасия Борисовна — студентка кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Масленников Андрей Леонидович — ассистент кафедры «Системы автоматического управления», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:
Жукова А.Б., Масленников А.Л. Определение частоты основного тона методом поиска пиков в амплитудном спектре с добавлением весовых частотных функций. *Политехнический молодежный журнал*, 2019, № 12(41). http://dx.doi.org/10.18698/2541-8009-2019-12-556

VOICE PITCH FREQUENCY DETECTION VIA SPECTRUM PEAKS SEARCH WITH ADDITIONAL FREQUENCY WEIGHT FUNCTIONS

A.B. Zhukova zhukovaab@student.bmstu.ru
SPIN-code: 1923-4866

A.L. Maslennikov amas@bmstu.ru
SPIN-code: 7087-6303

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

In speech recognition problems two main tasks are related to determination of the voice pitch frequency and formants frequencies. Using those frequencies, the exact phoneme could be recognized with some probability. In this paper the method of determination voice pitch frequency is described. Method is based on known idea of finding spectrum peaks but with addition of spectrum smoothing and frequency weighted functions of two types. Those types are exponential and linear. The described method was applied to the set of experiments with three men and three women. The results shown that there is exists the critical cut-off frequency of the smoothing filter and that the incorporating frequency weighted function increase accuracy of the voice pitch frequency determination.

Keywords

Speech recognition, voice pitch frequency, formants analysis, voice spectrum, smoothing filter, frequency weighted functions, Savitzky-Golay filter, spectrum peaks

Received 18.10.2019

References

[1] Bondarenko M.F., Rabotyagov A.V., Shchepkovskiy S.V. Speech recognition: stages of development, modern tech-nologies and prospects of their application. *Bionika intellekta* [Bionics of Intelligence], 2010, no. 2(73), pp. 164–168 (in Russ.).

[2] Ortega-García J., González-Rodríguez J. Overview of speech enhancement techniques for automatic speaker recognition. *Proc. ICSLP'96*, 1996, vol. 2, pp. 929–932.

[3] Plotnikov V.N., Sukhanov V.A., Zhigulevtsev Yu.N. Rechevoy dialog v sistemakh upravleniya [Speech dialogue in control systems]. Moscow, Mashinostroenie Publ., 1988 (in Russ.).

[4] Derkach M.F., ed. Dinamicheskie spektry rechevykh signalov [Dynamical spectrum of speech signals]. L'vov, Vishcha shkola Publ., 1983 (in Russ.).

[5] Sorokin V.N. Teoriya recheobrazovaniya [Speech production theory]. Moscow, Radio i svyaz' Publ., 1985 (in Russ.).

[6] Chistovich L.A., Ventsov A.V., Granstrem M.P., et al. Fiziologiya rechi. Vospriyatie rechi chelovekom [Speech physiology. Speech perception by a human]. Leningrad, Nauka Publ., 1976 (in Russ.).

[7] Labutin V.K., Molchanov A.P. Slukh i analiz signalov [Hearing and signal analysis]. Moscow, Energiya Publ., 1967 (in Russ.).

[8] Vintsyuk T.K. Analiz, raspoznavanie i interpretatsiya rechevykh signalov [Analysis, recognition and interpretation of speech signals]. Kiev, Naukova dumka Publ., 1987 (in Russ.).

[9] Zlokazov V.B. Metod dlya avtomaticheskogo poiska pikov v gamma-spektrakh [Automated search method for gamma-spectral peaks]. Dubna, OIYaI Publ., 1981 (in Russ.).
[10] Markel J.D., Gray A.H. Jr. Linear prediction of speech. Springer, 1976. (Russ. ed.: Lineynoe predskazanie rechi. Moscow, Svyaz' Publ., 1980.)
[11] Savitzky A.A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem*., 1964, vol. 36, no. 8, pp. 1627–1639. DOI: 10.1021/ac60214a047 URL: https://pubs.acs.org/doi/abs/10.1021/ac60214a047
[12] Savitzky A.A. A historic collaboration. *Anal. Chem*., 1989, vol. 61, no. 15, pp. 921A–923A. DOI: 10.1021/ac00190a003 URL: https://pubs.acs.org/doi/10.1021/ac00190a003

Zhukova A.B. — Student, Department of Automatic Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Maslennikov A.L. — Teaching Assistant, Department of Automatic Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:
Zhukova A.B., Maslennikov A.L. Voice pitch frequency detection via spectrum peaks search with additional frequency weight functions. *Politekhnicheskiy molodezhnyy zhurnal* [Politechnical student journal], 2019, no. 12(41). http://dx.doi.org/10.18698/2541-8009-2019-12-556.html (in Russ.).