



средства-data-preparation.html (дата обращения: 05.03.2021).

3 Осовский С. Нейронные сети для обработки информации [Текст]/ Осовский С. – М.: Научно-техническое издательство «Горячая линия–Телеком», 2017. 448 с.

4 Солдатова, О.П. Основы нейроинформатики. Самара, СГАУ, 2006. 131 с.

А.К. Алимуратов, П.П. Чураков

ПРИМЕНЕНИЕ МЕТОДА ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ ДЛЯ ИССЛЕДОВАНИЯ ВОКАЛИЗОВАННОЙ РЕЧИ В ЗАДАЧЕ ОБНАРУЖЕНИЯ СТРЕССОВЫХ ЭМОЦИЙ ЧЕЛОВЕКА

(Пензенский государственный университет)

Точность распознавания эмоций человека зависит от корректного выделения информативных параметров, отражающих эмоциональную составляющую речи. Анализ открытых источников отечественной и зарубежной литературы [1] выявил, что вопрос разработки высокоэффективного подходом выделения информативных параметров речи, релевантных нарушениям работы органов речевого аппарата остаются нерешенным и требует дальнейшей проработки. Данная статья посвящена исследованию вокализованной речи с применением уникальной адаптивной технологии анализа нестационарных сигналов - декомпозиции на эмпирические моды (ДЭМ) для обнаружения стрессовых эмоций человека [2].

Речь представляет собой нестационарный акустический сигнал сложной формы. Различные изменения в вегетативной нервной системе могут изменить речь человека. Речь человека состоит из вокализованных / невокализованных участков, участков пауз и дыхания. Информативные параметры вокализованной речи в полном объеме отражают нарушение работы органов речевого аппарата, вследствие эмоционального возбуждения человека. Вокализованная речь образуется в результате возбуждения голосового тракта, обусловленного вибрацией голосовых связок в области голосовой щели (глоттиса). Сила возбуждения во время глоттальной активности определяется в основном скоростью смыкания голосовых связок. Периодические колебания голосовых связок во время возбуждения голосового тракта называется основным тоном (ОТ). Величина обратная значению ОТ называется частотой основного тона (ЧОТ) и является важным информативным параметром речи.

ДЭМ представляет собой уникальную адаптивную технологию анализа нестационарных данных, не требующую никакой априорной информации для разложения сигнала на частотные составляющие [3]. Адаптивность ДЭМ позволяет эффективно применять ее для анализа естественных сигналов. Разложение с помощью ДЭМ обеспечивает извлечение из сигнала различных колебательных функций, называемых эмпирическими модами (ЭМ), каждая из кото-



рых имеет свой частотный диапазон. Среди всех модификаций (ДЭМ, МДЭМ, КМДЭМ и др.) наиболее перспективной в области речи является улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом (ПМДЭМАШ) [3]. Ниже представлено краткое математическое описание ДЭМ и ее модификаций с добавлением шума:

$$x(n) = \sum_{i=1}^I IMF_i(n) + r_I(n),$$

где $x(n)$ - анализируемый сигнал; n - дискретный отсчет времени в сигнале; $i = 1, 2, \dots, I$ - номер ЭМ; $IMF_i(n)$ - полученные в результате разложения ЭМ; $r_i(n)$ - конечный неделимый остаток (последняя мода).

$$x_j(n) = x(n) + w(n),$$

где $x_j(n)$ - зашумленные сигналы; $w_j(n)$ - белый шум;

$$x_j(n) = \sum_{i=1}^I IMF_{ji}(n) + r_{jI}(n),$$

$$IMF_i(n) = \sum_{j=1}^J \frac{IMF_{ji}(n)}{J},$$

$$r_I(n) = \sum_{j=1}^J \frac{r_{jI}(n)}{J},$$

где $j = 1, 2, \dots, J$ - количество реализаций белого шума.

Глоттальная активность относится к явлению, связанному с колебаниями голосовых связок во время образования вокализованной речи. Способ исследования вокализованной речи с использованием ДЭМ и ее модификаций сводится к выполнению следующих этапов:

- разложение сигнала на ЭМ (частотные составляющие сигнала);
- выделение ЭМ, содержащих периодическую информацию об источнике возбуждения голосового тракта;
- формирование комплексного сигнала, отражающего информацию о глоттальной активности;
- вычисление информативных параметров вокализованной речи.

Процесс выделения ЭМ, содержащих периодическую информацию об источнике возбуждения голосового тракта сводится к определению следующих параметров ЭМ: средней частоты; энергии и уровня смешивания.

Вычисление средней частоты ЭМ осуществляется по формуле;

$$F_{IMF} = \frac{\sum_0^{F_s/2} f \times S_{IMF}(f)}{\sum_0^{F_s/2} S_{IMF}(f)},$$

где F_{IMF} - средняя частота ЭМ, со спектром мощности $S_{IMF}(f)$; F_s - частота дискретизации речевого сигнала. Вычисление спектра мощности осуществляется с использованием быстрого преобразования Фурье с размерностью N (N - длина сигнала).

В процессе вокализованной речи ЧОТ взрослого человека (мужчин и женщин) находится в диапазоне частот от 80 до 400 Гц. Используя эти данные, можно выделить ЭМ, которые содержат периодическую информацию об источнике возбуждения голосового тракта. То есть моды, средние частоты которых находятся вне диапазона 80 - 400 Гц, в дальнейшей обработке не используются.



Приближая работу способа к функционалу слухового аппарата, для сжатия амплитуды сигнала в большом динамическом диапазоне применяют логарифмирование энергии:

$$LE_{s,i} = \log_2 \left(\sum_{n=1}^N (IM_{s,i}(n))^2 \right),$$

где $LE_{s,i}$ - логарифм энергии ЭМ фрагмента речевого сигнала; s - номер фрагмента.

Известно, что вокализованная речь имеет большую энергию, чем невокализованная речь. Поэтому моды, содержащие энергию, меньшую на 90 % от общей энергии сигнала, в дальнейшей обработке не используются.

Уровень смешивания для пары последовательных ЭМ (Degree of Mode Mixing, DMM), обозначает сходство частот i -ой и $(i + 1)$ -ой моды и определяется как:

$$DMM_i = \left[1 - \frac{F_i + F_{i+1}}{F_{i/2}} \right] \times 100\%,$$

где DMM_i - уровень смешивания мод; F_i и F_{i+1} - средние частоты i -ой и $(i + 1)$ -ой моды.

Если улучшенная ПМДЭМАШ обеспечивает почти идеальный диадический набор фильтров, то значение уровня смешивания для пары последовательных ЭМ будет близко к 0 %. Если моды будут отличаться по частоте, то значение уровня смешивания будет отрицательным, что указывает на отсутствие эффекта смешивания мод. Если моды имеют одинаковую среднюю частоту, то значение уровня смешивания будет 100% и это означает, что информация о возбуждении голосового тракта, вследствие смыкания голосовых связок, распределена между ними.

В соответствии с полученными результатами осуществляется формирование комплексного сигнала, отражающего информацию о глоттальной активности. Формирование представляет собой суммирование мод, содержащих периодическую информацию об источнике возбуждения голосового тракта.

Для проведения исследования сформирована тестовая выборка из 100 мультигармонических сигналов, представляющих сумму нескольких гармонических составляющих ОТ с заранее известной частотой. В рамках исследования каждый тестовый сигнал подвергался частотному модулированию в диапазоне 0 - 2,5 Гц/мс с шагом 0,5 Гц/мс. В качестве критериев оценки эффективности исследования использовались коэффициент грубых ошибок (Gross Pitch Error, *GPE*) и средний коэффициент мелких ошибок (Mean Fine Pitch Error, *MFPE*). В таблице 1 представлены усредненные результаты вычисления ЧОТ 100 комплексных сигналов, сформированных суммированием вокализованных ЭМ, полученных с помощью различных модификаций декомпозиции.

В соответствии с полученными данными можно сделать вывод, что наилучшие результаты вычисления ЧОТ достигаются при исследовании вокализованной речи с помощью улучшенной ПМДЭМАШ. В том числе и при больших значениях модуляции ЧОТ. Это объясняется тем, что при использо-



вании улучшенной ПМДЭМАШ, сформированный комплексный сигнал отражает максимальное количество периодической информации об источнике возбуждения голосового тракта. Справедливо отметить, что остальные методы декомпозиции с добавлением шума, также обеспечивают приемлемые результаты вычисления ЧОТ. Однако значения коэффициентов грубых (*GPE*) и мелких ошибок и (*MFPE*) больше по причине известных недостатков. Наихудшие результаты достигаются при использовании ДЭМ.

Таблица 1 - Результаты вычисления ЧОТ комплексного сигнала, отражающего информацию о глоттальной активности

| Модуляция частоты, Гц/мс | Модификации декомпозиции | | | | | Критерии оценки |
|--------------------------|--------------------------|------|-------|---------|--------------------|-----------------|
| | ДЭМ | МДЭМ | КМДЭМ | ПМДЭМАШ | Улучшенная ПМДЭМАШ | |
| 0 | 0 | 0 | 0 | 0 | 0 | GPE |
| | 6,10 | 1,56 | 1,35 | 1,10 | 0,70 | MFPE |
| 0,5 | 0 | 0 | 0 | 0 | 0 | GPE |
| | 7,56 | 2,65 | 2,34 | 1,21 | 0,93 | MFPE |
| 1,0 | 0 | 0 | 0 | 0 | 0 | GPE |
| | 9,30 | 3,25 | 2,81 | 2,54 | 1,75 | MFPE |
| 1,5 | 0 | 0 | 0 | 0 | 0 | GPE |
| | 10,21 | 4,50 | 3,90 | 3,00 | 2,38 | MFPE |
| 2,0 | 2,05 | 0 | 0 | 0 | 0 | GPE |
| | 12,20 | 5,10 | 4,67 | 4,54 | 3,88 | MFPE |
| 2,5 | 7,60 | 5,30 | 4,20 | 3,50 | 2,70 | GPE |
| | 17,20 | 7,10 | 6,50 | 6,20 | 5,32 | MFPE |

При эмоциональном возбуждении вибрации голосовых связок характеризуются нерегулярностью, возникающей вследствие неполного смыкания при вокализованной речи. При крайне высоком и низком возбуждении изменение ЧОТ может достигать 30 - 40 % от номинального значения, соответствующего нейтральному эмоциональному состоянию. Процесс модулирования ЧОТ в диапазоне 0 - 2,5 Гц/мс с шагом 0,5 Гц/мс в рамках исследования имитировал нерегулярность вибрации голосовых связок. Полученные результаты вычисления ЧОТ позволяют сделать вывод, что представленный способ исследования вокализованной речи, с помощью улучшенной ПМДЭМАШ, может успешно тестироваться для обнаружения стрессовых эмоций человека.

Литература

1. Schuller B.W. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing / B.W. Schuller, A.M. Batliner // New York: Wiley. - 2013. - P. 344.
2. Алимуратов А.К. Помехоустойчивый алгоритм определения просодических характеристик речевых сигналов для систем оценки психоэмоциональ-



ного состояния человека / А.К. Алимуратов, А.Ю. Тычков, П.П. Чураков, Д.В. Артамонов // Известия высших учебных заведений. Поволжский регион. Технические науки. - 2019. - № 3(51). - С. 3 - 16.

3. Colominasa, M. A. Improved complete ensemble EMD: a suitable tool for biomedical signal processing / M. A. Colominasa, G. Schlotthauera, M. E. Torres // Biomed. Signal Proces. - 2014. - Vol. 14. - P. 19 - 29.

М.Г. Алкеев, А.А. Лобанков

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ АВТОМАТИЧЕСКОГО АНАЛИЗА И ПРИНЯТИЯ РЕШЕНИЙ НА ФОНДОВОЙ БИРЖЕ

(Самарский университет)

Индустрия вокруг цифровой валюты за последние пару лет сильно трансформировала привычные для нас рынки. Ввиду высокой скорости развития информационных, а также телекоммуникационных технологий, появился новый подход к торговле цифровыми валютами и ценными бумагами, который заключается в использовании различных автоматизированных программных систем [1]. Использование такого рода программных продуктов наиболее часто встречается на рынках цифровых валют, это связано с большей его волатильностью относительно других рынков, например, рынком ценных бумаг [2].

Использование автоматизированных торговых систем обеспечивает следующие преимущества:

- Многозадачность, позволяет одновременно работать с множеством валютных пар и ставить в короткий промежуток времени большое количество ордеров;
- Точность работы, это исключает появление различных ошибок, связанных с вычислениями;
- Высокая скорость работы, позволяет моментально реагировать на движение рынка;
- Возможность круглосуточной работы;
- Гибкость, система может быть отмасштабирована или расширена под определенные нужды;
- Строгое соблюдение используемой стратегии и как следствие отсутствие эмоций, что зачастую может исключить нежелательные действия.

В то же время использование автоматизированных торговых систем сопряжено с рядом недостатков:

- Неспособность реагировать на большой спектр внешних факторов, резкие изменения в движениях рынка могут отрицательно сказаться на работе торгового алгоритма;
- Сложность разработки и уязвимость к логике функционирования.