

ПЕРЦЕПТИВНО-АКУСТИЧЕСКИЙ АНАЛИЗ ЗВУЧАЩЕЙ РЕЧИ

УДК 81'342+004.934.2

П. А. Скрелин, А. О. Титюшина

г. Санкт-Петербург, Россия,

Санкт-Петербургский государственный университет

ПРОСОДИЧЕСКИЙ ТЕМБР: МЕТОДЫ АНАЛИЗА И МОДИФИКАЦИИ

В статье представлены разрабатываемая на кафедре фонетики методика анализа просодического тембра, способы, алгоритмы и инструменты определения его акустических характеристик, подходы к его моделированию, а также приведены первые результаты модификации тембра путем ресинтеза естественного сигнала и генерации искусственного сигнала с заданными свойствами.

Ключевые слова: просодический тембр; спектральный анализ; цифровая обработка сигнала; ресинтез.

P. A. Skrelin, A. O. Tityushina

St. Petersburg, Russia, St. Petersburg State University

PROSODIC TIMBRE: METHODS OF ANALYSIS AND MODIFICATION

The article presents the methodology of prosodic timbre analysis developed at the Department of Phonetics, algorithms and tools for determining its acoustic characteristics, approaches to its modeling, and also presents the first results of timbre modification obtained by resynthesis of a natural signal and generation of an artificial signal with specified features.

Keywords: prosodic timbre.

Тембр является одним из средств выражения эмоциональной окраски экспрессивной речи и одним из основных признаков звука [1, с. 53–56]. Задача уточнения акустических коррелятов просодического тембра, тем не менее, остается актуальной. В данном исследовании с помощью процедур модификации и ресинтеза речевого сигнала проводится исследование акустических свойств просодического тембра.

Процедуры модификации и ресинтеза заключаются в пересадке определенных просодических параметров с одного стимула на другой. Эта методика широко используется в решении целого ряда задач в исследованиях, проводимых на кафедре фонетики СПбГУ, в частности, для обнаружения перцептивной значимости различных акустических характеристик речевого сигнала и выявления параметров, отвечающих за эмоциональную окраску языковых единиц [2; 3; 4; 5]. Первые же эксперименты по модификации

тембра голоса, подтверждающие значимость этого параметра для определения социальных характеристик говорящего были проведены еще на аналоговом оборудовании в конце 80-х годов XX века [6, 7]

В настоящее время существуют модели клонирования речевого тембра, основанные на нейронных сетях [8]. У такого подхода есть ряд минусов – результат плохо контролируется, требует обширного материала и долгого времени обучения модели. С другой стороны, исследование физических параметров сигнала могло бы позволить найти признаки тембральных моделей, которые можно было бы изменять точными вычислениями без применения нейронных сетей.

На первом этапе настоящего исследования была проведена пересадка наиболее отличающихся спектральных составляющих от одного стимула на другой. Основой технологического решения послужили идеи, развивавшиеся несколько лет назад на кафедре фонетики СПбГУ [9, 10]. На Рис. 1 представлен результирующий спектр ударного гласного [а], получившийся при переносе наиболее характерных параметров от сигнала-донора на сигнал-акцептор: оба сигнала взяты из одинакового фонетического контекста и позиции во фразах, произнесенных с разными имитированными эмоциями, предварительно подтвержденными аудиторским экспериментом. Полученный результат позволяет сделать вывод, что изменение небольшого числа спектральных составляющих сигнала, приводит к изменению интонационной (тембральной) окраски речевого сигнала. Это приводит к предположению о возможности определения моделей тембральной окраски речевого сигнала, которые коррелировали бы с различными эмоциональными коннотациями у слушателей. Важным условием эффективности предложенной процедуры является точность определения спектральных характеристик при анализе и обработке сигналов не ниже нескольких герц.

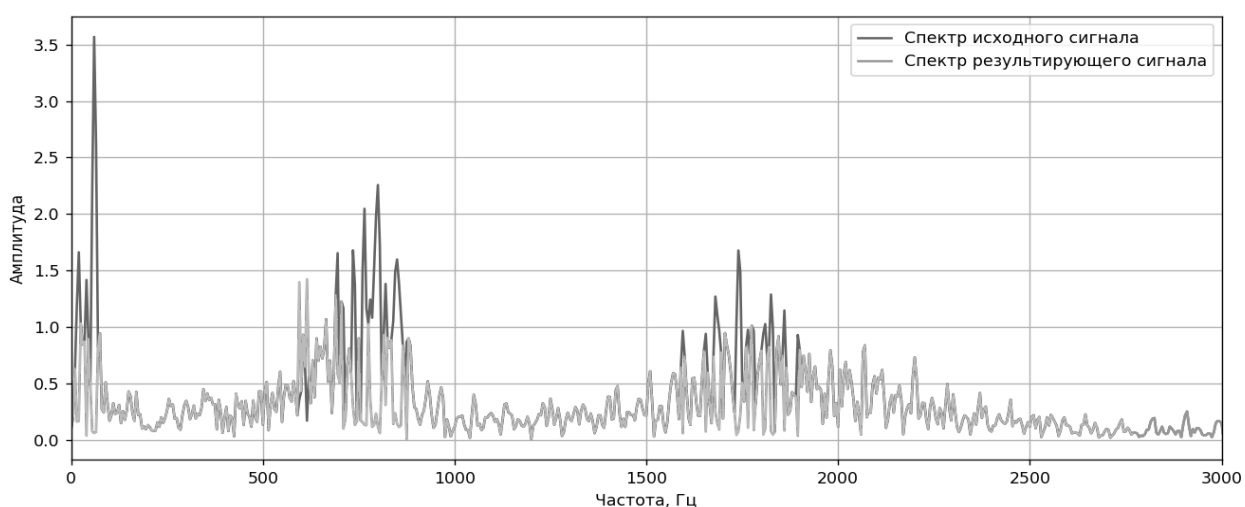


Рис. 1. Наложение исходного спектра ударного гласного [а] до его модификации на спектр результирующего сигнала, воспроизводящего параметры сигнала-донора.

Следующим шагом, основываясь на технологии ресинтеза, был создан искусственный сигнал с передачей тембральных характеристик сигнала-донора синтезированным гласным и с сохранением согласных аллофонов, параметров длительности и мелодики звуков сигнала-акцептора. Для эксперимента была использована фраза «*Посмотри на эту фотографию*», реализованная профессиональным актером, воспроизводящим эмоциональное состояние, заданное окружающим микротекстом или поставленной задачей: скучающее (скрипучий голос) или капризное (противный). Как и на предыдущем этапе, для синтеза гласных аллофонов использовались лишь наиболее яркие спектральные характеристики сигнала-донора (см. Рис. 2). Полученный сигнал хорошо отражал его тембральную окраску, но содержал некоторые артефакты, выдающие его искусственность, связанные с отсутствием координации и синхронизации тембра с другими просодическими и сегментными характеристиками, сохранившимися в исходном сигнале.

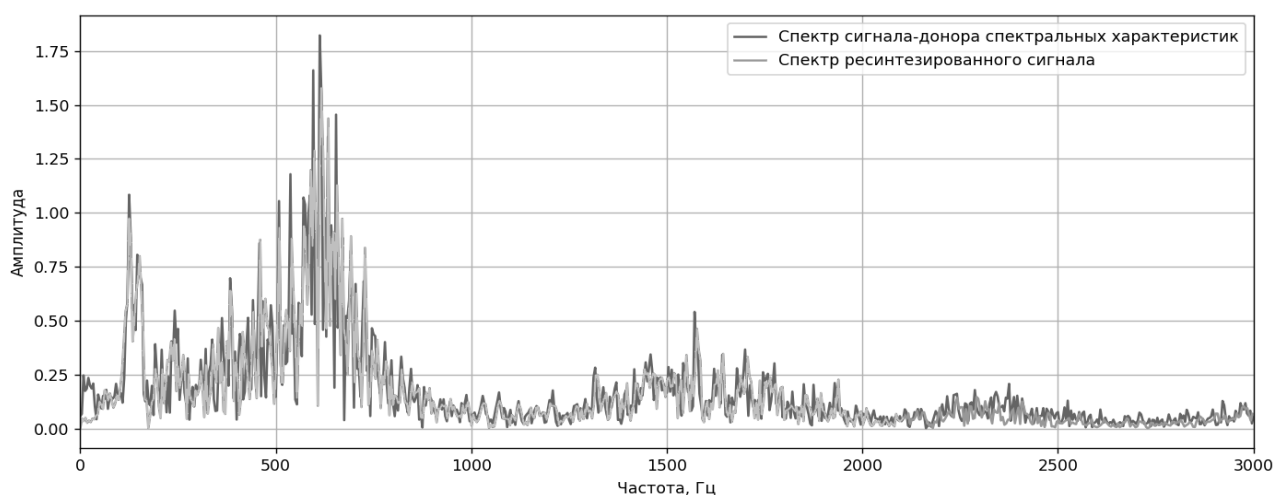


Рис. 2. Наложение спектра ресинтезированного ударного гласного [а] на спектр сигнала-донора спектральных характеристик.

При построении ресинтезированного сигнала аналогичным методом, но с одновременной модификацией параметров частоты основного тона и длительностей гласных звуков сигнала-акцептора по модели целевого сигнала, получился результат хорошего качества с тембральной окраской, соответствующей характеристикам сигнала-донора (см. Рис. 3).

Таким образом, в результате анализа ресинтезированных модифицированных сигналов и их сравнения с исходными стимулами, был сформирован алгоритм переноса тембральных характеристик гласных с одного стимула на другой. По всей видимости, с помощью подобной методики можно создать инструмент для построения моделей тембральных контуров, связанных с передачей различных эмоций. Для проверки эффективности предложенного алгоритма предполагается серия перцептивных тестов. В последние несколько лет на кафедре фонетики СПбГУ были созданы несколько речевых корпусов, содержащих высказывания, эмоциональная окраска которых была

подтверждена в ходе аудиторских экспериментов. Этот материал планируется использовать для переноса тембральных характеристик уже атрибутированных ранее стимулов друг на друга, чтобы в ходе аудиторских экспериментов оценить изменение восприятия их эмоциональной окраски.

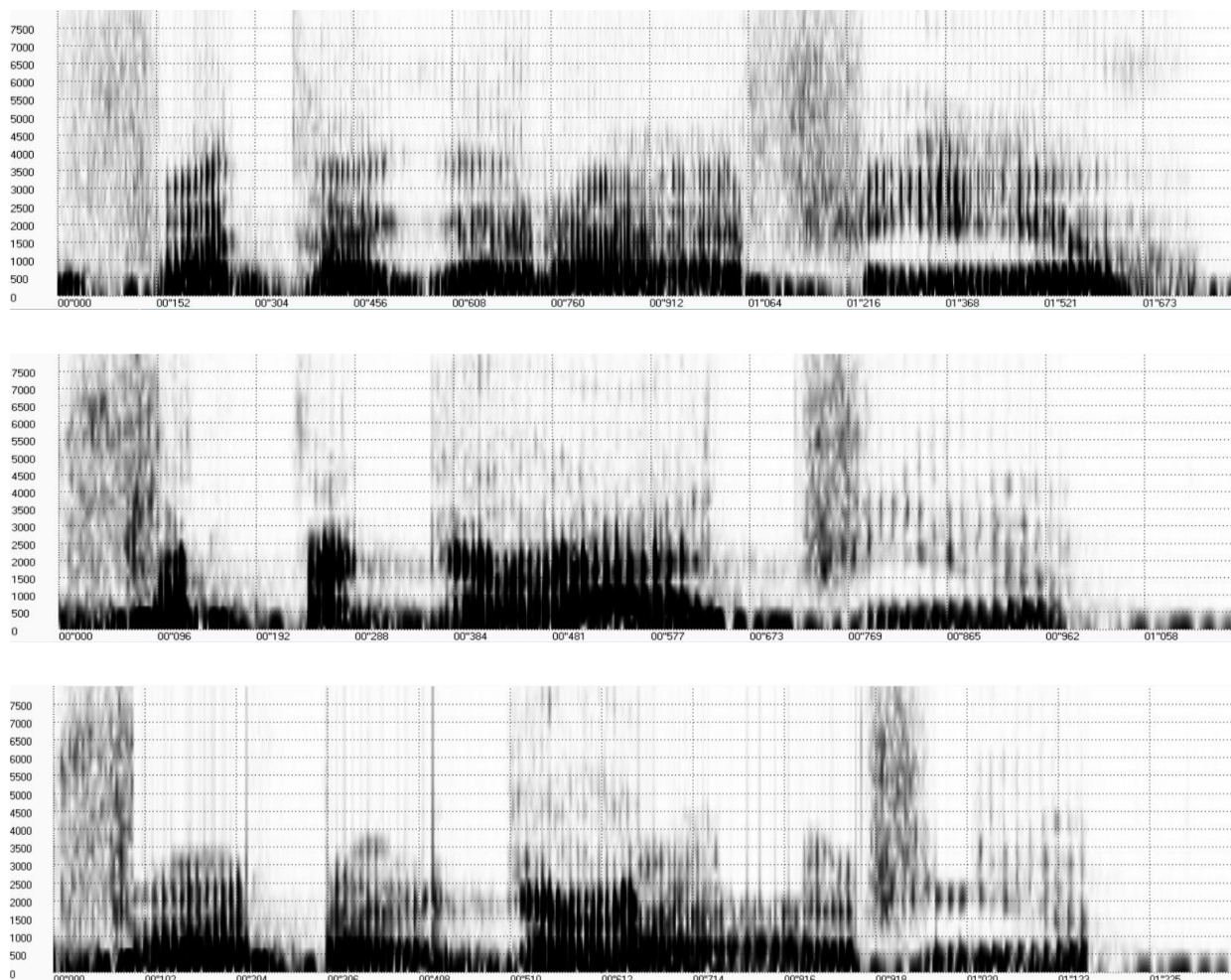


Рис. 3. Сонограммы сигнала-донора, исходного сигнала-акцептора и сигнала, синтезированного с учетом длительностей гласных аллофонов и изменений мелодического контура.

ЛИТЕРАТУРА

1. *Светозарова Н. Д.* Интонационная система русского языка. Л., 1982. 76 с.
2. Анализ через синтез: установление акустических ключей иронии с помощью модификаций звукового сигнала / П. А. Скредин, У. Е. Кочеткова, В. В. Евдокимова, Д. Д. Новоселова // Анализ разговорной русской речи (АРЗ-2023) : труды десятого междисциплинарного семинара / У. Е. Кочеткова, П. А. Скредин. Санкт-Петербург : Скифия-принт, 2023. С. 55–59.
3. Using audio signal modification to validate the acoustic features relevance / U. Kochetkova, P. Skrelin, V. Evdokimova, D. Novoselova // Записки научных семинаров Санкт-Петербургского Отделения Математического Института им. В. А. Стеклова РАН. 2023. – Р. 86–101.

4. Interplay of Visual and Acoustic Cues of Irony Perception: A Case Study of Actor's Speech / U. Kochetkova, V. Evdokimova, P. Skrelin, R. German, D. Novoselova // Artificial Intelligence and Natural Language / V. Malykh, A. Filchenkov (eds.). CCIS ed. Cham : Springer Nature, 2022. Vol. 1731. P. 82–94.
5. Prosodic Features of Verbal Irony in Russian and French: Universal vs. Language-Specific / U. Kochetkova, P. Skrelin, R. German, D. Novoselova // Speech and Computer : 24th International Conference, SPECOM 2022. Gurugram, India, November 14–16, 2022 Proceedings.