

Перцептивная роль амплитудной модуляции для идентификации дрожащего в русской речи

Орлова А.А.

Кузнецов В.Б.

ГОУ ВПО «Московский государственный лингвистический университет».
E-mail: asunehka@yandex.ru, kuvlad2007@yandex.ru

В работе исследуется перцептивная значимость формы амплитудной огибающей для распознавания дрожащего в русской речи. Экспериментально проверяется гипотеза о том, что кратковременное уменьшение амплитуды речевого сигнала на закрытой фазе вибранта (так называемый удар) не является основным слуховым признаком этого звука. Было показано, что идентификация дрожащего в последовательности VCV (ударение на втором гласном), вырезанной из различных слов четырёх дикторов, не ухудшается при сглаживании амплитудной огибающей. Оценки качества произнесения дрожащего в исходных и модифицированных стимулах на семибалльной шкале статистически достоверно не различались.

Как правило, при описании канонических артикуляторных и акустических характеристик вибранта в качестве основного признака указывается вибрантная смычка (закрытая фаза, удар), которая похожа на краткую (20–30 мс) звонкую смычку взрывного согласного, уверенно идентифицируемую как на осциллограмме, так и на спектрограмме. Однако наш опыт сегментации речевых баз данных [1, 2] показывает, что при артикуляции вибрантной смычки глубина амплитудной модуляции сигнала может значительно варьировать, что тем не менее не сказывается на оценке качества звучания согласного. Рис. 1 иллюстрирует это наблюдение.

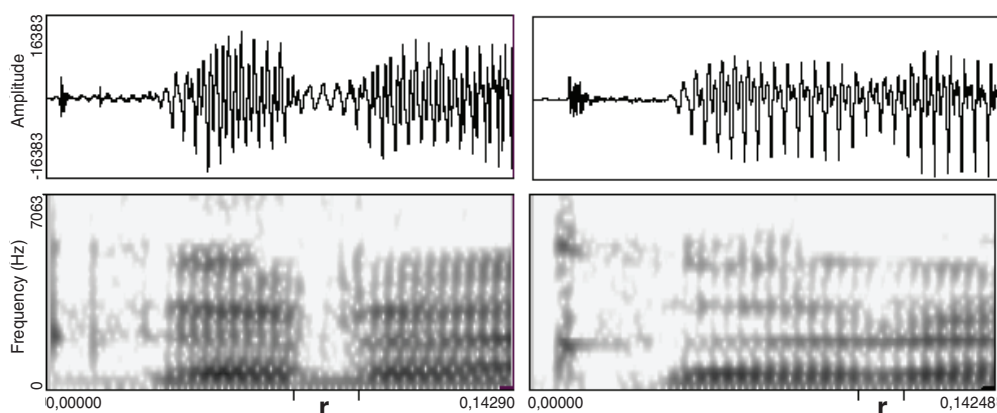


Рис. 1. Осциллограмма и спектрограмма фрагмента слова "карантин", демонстрирующие вариативность амплитудной модуляции дрожащего в произнесении двух дикторов женщин (r — обозначает удар)

Чтобы исследовать перцептивную значимость формы амплитудной огибающей речевого сигнала для восприятия дрожащего, был проведён эксперимент, в котором в качестве стимулов испытуемым предъявлялись естественные реализации твердого вибранта и те же самые реализации со сглаженной огибающей амплитуды на интервале вибрантной смычки. Испытуемые должны были идентифицировать дрожащий, а затем оценить качество его звучания.

МЕТОД

Речевой материал. Из записей каждого из четырёх дикторов (две женщины и двое мужчин, чтение списка предложений, частота дискретизации 22025 Гц, 16-разрядное квантование по уровню) было выбрано 4 слова, содержащие дрожащий в интервокальном положении перед разными ударными гласными. Из этих слов были вырезаны сегменты VCV с интересующим нас согласным. Таким образом, исключалась возможность идентификации вибранта путём распознавания слова. Из отрезков гласных звуков были удалены формантные переходы, несущие информацию о соседних согласных.

Наряду с естественными VCV стимулами применялись и их модификации, которая заключалась в сглаживании амплитудной огибающей на интервале удара. Пример исходного и модифицированного стимула приведён на рис. 2. Сглаживание, как правило, проводилось поперiodно и не затрагивало взрывной или фрикативный компонент закрытой фазы дрожащего.

В массив стимулов были включены также структуры VCV, в которых в качестве согласного выступали твердые [D] и [L], являющиеся наиболее фонетически близкими к дрожащему, о чем, в частности, свидетельствует их использование в качестве замены [R] при невозможности его произнесения. Слова, из которых вырезались стимулы с [D] и [L], были идентичными для всех дикторов. Каждый диктор был представлен двумя стимулами на каждый согласный.

Уровень интенсивности всех стимулов был пронормирован.

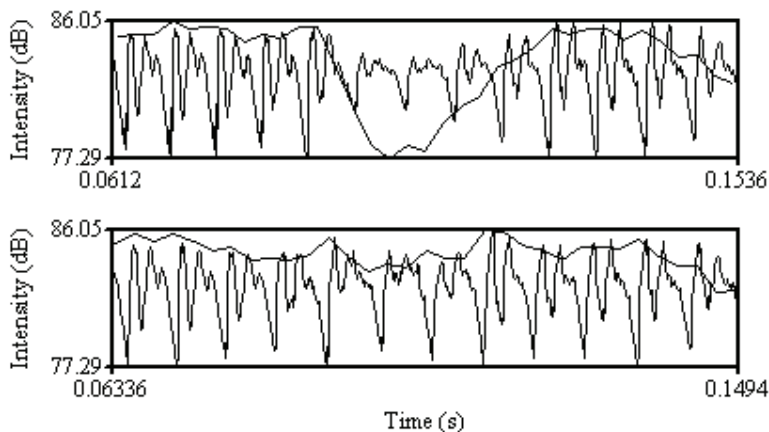


Рис. 2. Пример сглаживания амплитудной огибающей на интервале вибрантной смычки. Верхняя осциллограмма — исходный сигнал, нижняя — модифицированный. На осциллограмму наложена амплитудная огибающая

ПРОЦЕДУРА

Формирование тестовой последовательности стимулов, проведение эксперимента и предварительная обработка результатов были выполнены с помощью программного комплекса ASPECT [6].

Эксперимент состоял из двух тестов. В первом тесте испытуемые должны были идентифицировать предъявленный стимул как R, D или L. Исходные и модифицированные [R]-стимулы, стимулы с [D] и [L] повторялись 2 раза. Таким образом, случайная последовательность стимулов состояла из 96 элементов (4x12x2). Стимул воспроизводился дважды с интервалом 0,7 сек.

Во втором тесте испытуемые должны были оценить качество произнесения тех же самых стимулов, что и в первом тесте, используя семибалльную шкалу. Причем им сообщалось (текст на экране монитора) какой звук предъявляется. Исходные и модифицированные [R]-стимулы повторялись 3 раза. [D] и [L] стимулы повторялись 2 раза. Длина тестовой последовательности — 128 стимулов. Предъявляемый стимул проигрывался 2 раза (межстимульный интервал 0,7 сек) и испытуемый имел возможность повторно прослушать стимул.

Эксперимент проводился с помощью персонального компьютера. Для регистрации ответов испытуемых на экране монитора для первого теста формировались соответствующие три кнопки с символами звуков, для второго теста создавалась шкала со скользящим ползунком. Эксперимент проходил в тихом помещении, прослушивание стимулов проводилось в головных телефонах марки Sennheiser HD. 205.

В эксперименте принимали участие 11 испытуемых в возрасте 21-55 лет (из них трое мужчин). 6 испытуемых имели лингвистическое образование.

РЕЗУЛЬТАТЫ

Анализ данных первого теста показал, что испытуемые фактически безошибочно идентифицировали как исходные, так и модифицированные вибранты. Четыре из 11 испытуемых допустили 13 ошибок на разных стимулах. Причем модифицированные [R]-стимулы три раза идентифицировались как [D] и один раз как [L]. Исходные [R]-стимулы 4 раза идентифицировались как [D] и один раз как [L].

Данные второго теста обрабатывались отдельно для двух групп испытуемых. Это объясняется тем, что для оценки качества звучания дрожащего испытуемые использовали разные участки семибалльной шкалы: одна группа (5 человек) использовала всю шкалу, а другая (6 человек) — ограничилась интервалом от 3 до 6 баллов. На рис. 3 приведены гистограммы оценок исходного и модифицированного вибранта для двух групп. Использование критерия хи-квадрат для определения статистически достоверного различия этих распределений показало, что на уровне значимости $\alpha = 0,01$ распределения для исходных и модифицированных [R]-стимулов идентичны в обеих группах. Как можно видеть, испытуемые второй группы оценивали [R]-стимулы выше, чем испытуемые первой группы. Средняя оценка исходных и модифицированных [R]-стимулов в первой группе равна соответственно 4,92 и 4,55 баллов, во второй группе — 6,16 и 6,10 баллов.

При сравнении качества произнесения стимулов [D, L] с модифицированными [R]-стимулами статистически значимых различий не установлено. Средняя оценка для стимулов [D, L] равна 4,81 и 4,82 баллов соответственно, для модифицированного [R] — 4,49 баллов.

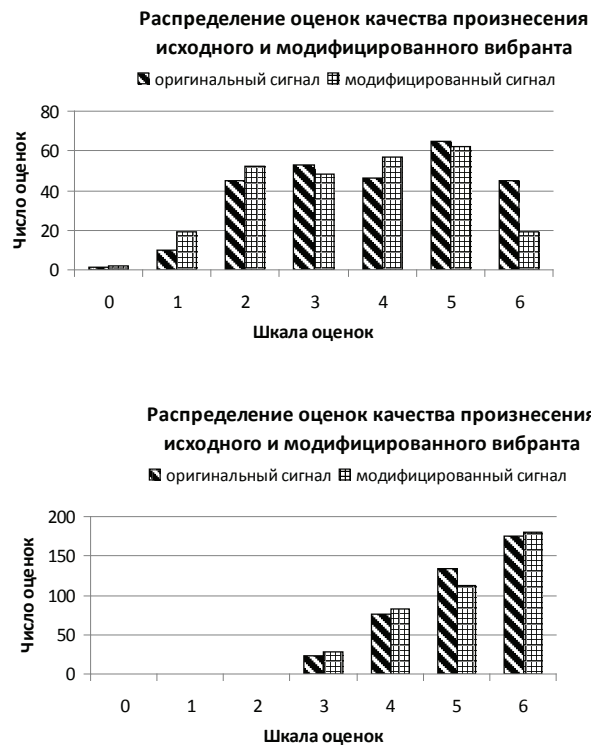


Рис. 3. Гистограммы распределения оценок качества произнесения исходных и модифицированных [R]-стимулов двумя группами испытуемых. Слева данные первой группы, справа — второй

ОБСУЖДЕНИЕ И ВЫВОДЫ

Данные эксперимента свидетельствуют о том, что кратковременное уменьшение амплитуды речевого сигнала на закрытой фазе вибранта не может выступать в качестве его основного признака. Сглаживание амплитудной огибающей не привело к статистически достоверным изменениям в точности идентификации и оценке качества звучания вибранта. Это, однако, не отрицает артикуляторной значимости вибрантной смычки. Речь идёт лишь об одном из её акустических проявлений.

При сглаживании амплитудной огибающей изменяется и соотношение спектральной энергии на вибрантной смычке и граничных участках гласного или вокалического компонента вибранта. Так, в одном случае в исходном сигнале амплитуда первой форманты на ударе была меньше, чем на предыдущем периоде гласного [a] на приблизительно 15 дБ, а в модифицированном сигнале эта разность составляла около 8 дБ¹. Однако это не отразилось на распознавании и оценке дрожащего.

Можно предположить, что существенные перцептивные признаки дрожащего связаны с более тонкими структурными изменениями частотно-амплитудных характеристик сигнала на переходных участках между гласным или открытой фазой вибранта и ударом [3, 4, 5, 8].

¹ Как известно, энергия первой форманты определяет в основном суммарную энергию спектра [7].

ЛИТЕРАТУРА

1. Кузнецов В.Б. Аллофонические реализации и акустические характеристики русского вibrанта // Акустика речи. Медицинская и биологическая акустика. Сборник трудов XIII сессии Российского акустического общества.. Т 3. М.: Геос, 2003. С. 30–33.
2. Кузнецов В.Б. О вокалическом компоненте вibrанта в русском языке // Тезисы IV Международной научной конференции «Фонетика сегодня: актуальные проблемы и университетское образование». М., 2003. С. 76–77.
3. Кузнецов В.Б. Об акустико-фонетических характеристиках твердого вibrанта в русском языке / В.Б. Кузнецов, Н.В. Бобров // Акустика речи. Медицинская и биологическая акустика. Сборник трудов XVI сессии Российского акустического общества. Т 3. М.: Геос, 2005. С. 57–61.
4. Кузнецов В.Б. Об акустико-фонетических характеристиках мягкого вibrанта в русском языке / В.Б. Кузнецов, Н.В. Бобров // Proc. XII-th Intern. Conf. "Speech and Computer", SPECOM 2007, Moscow, v. 2, 2007. P. 619–625.
5. Кузнецов В.Б. Односмычный вibrант — это вibrант или тэп. Данные анализа русской и испанской речи / В.Б. Кузнецов, А. Б. Памиес // Акустика речи. Медицинская и биологическая акустика. Сборник трудов XX сессии Российского акустического общества. Т 3. М.: Геос, 2008. С. 95–98.
6. Топровер Г.Л. Комплексная инструментальная система ASPECT для подготовки, проведения и обработки экспериментов с помощью ЭВМ в области восприятия речи / Г.Л. Топровер, В.Б. Кузнецов // Фонетика сегодня: актуальные проблемы и университетское преподавание. М., 1998. С. 70–72.
7. Фант Г. Акустическая теория речеобразования. М.: Наука, 1964. 284 с.
8. Kouznetsov V. Why is Russian trill so tricky to synthesize // Proc. of Intern. Workshop "Speech and Computer", SPECOM'2003, Moscow, 2003, P. 158–161.