Требования к алгоритмам скрытого встраивания информации в просодические параметры речи

Пономарь М.О.

Московский государственный университет им. М.В. Ломоносова, филологический ф-т.
Россия, 119899 Москва, Воробьевы горы, I гум. корпус.
Тел. (495) 939-26-01. E-mail: okri@philol.msu.ru

Использование языковой просодии позволяет скрытно внедрять информацию в речь путём модификации речевого сигнала, которая не обнаруживается в канале связи и на выходе системы ни на слух, ни с помощью инструментальных средств без сравнения с эталоном речи, которым владеет только передающая сторона. В представленных материалах рассматривается практический аспект этого подхода, заключающийся в оценке возможности технической реализации просодических методов сокрытия информации в речи. Это даёт возможность оценить подходящие для сокрытия информации параметры просодии и сформулировать требования к практическим алгоритмам для специализированных цифровых интегральных схем и вычислителей общего назначения. Обоснование этих требований позволяет произвести отбор существующих и разработку новых алгоритмов как для целей скрытой связи по открытым каналам, так и для целей создания аутентификационных меток для фонограмм.

Введение

Развитие речевой стеганографии на основе вариативности просодии со временем может приблизиться к тому рубежу, который отделит область научных исследований от разработки опытных и промышленных образцов. Уже в настоящее время в интересах скрытой связи разработан метод и опробован в лабораторных условиях алгоритм встраивания информации в некоторые параметры просодии [1]. Однако имеется потребность в увеличении количества используемых параметров и использования их комбинаторных возможностей в интересах увеличения пропускной способности скрытого канала связи. Многообразие параметров просодии и сложность их формализации являются препятствием для дальнейшего развития методов и практической реализации алгоритмов этого направления защиты информации. В связи с этим необходимо обосновать требования к просодическим стегоалгоритмам и их составным частям, провести отбор и оценить возможности их технической реализации.

Состав параметров просодической стегосистемы

Основой вариативности просодии является инвариант, позволяющий вносить в него такие изменения, которые не выходят за пределы допустимого отклонения от психоакустической нормы, а потому не заметны постороннему наблюдателю. Это — вариант ресинтеза речи. Для того, чтобы использовать просодию в качестве стегоконтейнера, потребуется решить три основных задачи [2].

Во-первых, установить, какие из наблюдаемых параметров просодии являются управляемыми. То есть для каждого параметра оценить, существуют ли в настоящее время технические и/или программные средства, позволяющие его инструментально измерять и модифицировать. Во-вторых, разработать метод внедрения скрываемых данных в управляемые просодические параметры. И, наконец, установить допустимые пределы модификации просодических параметров, при которых соблюдается требование скрытности. Принципиальные решения этих задач пока найдены только применительно к двум параметрам просодии — частоте основного тона и длительности сегментов речи [1,3,4].

В соответствии с основами традиционной фонетики [5,6] материальными акустическими средствами языковой просодии являются параметры, приведенные в табл. 1 (в таблицу не включены некоторые супрасегментные речевые средства, такие как артикуляция, эмоциональная окраска речи и другие, не имеющие пока средств для инструментальной оценки). На примере образца речи (рис. 1), обработанного в системе PRAAT [7] (в верхней части — волновая форма, в нижней — кривые частоты основного тона и интенсивности), показан подход к акустическому измерению параметров просодии. Инструментальная доступность для измерения является необходимым условием для использования параметра просодии в качестве стегоконтейнера. Таким образом, параметрами-контейнерами могут стать: уровень, контурность тона; интенсивность основного, второстепенного ударения; длительности и расположение пауз; уровень, длительность и интенсивность фразового акцента (P1-P11).

Таблица 1

Акустические средства языковой просодии

№	Группа	Параметр	Изменчивость	Размерность
P1	слоговые тоны	уровень тона	низкий-высокий	Гц
P2		контурность тона	восходяще-нисходящий	дб/с
P3	ритмические схемы слова и фразы	интенсивность основного ударения	меньше-больше	дб
P4		интенсивность второстепенного ударения	меньше-больше	дб
P5		длительность ударных гласных	меньше-больше	мс
P6		длительность безударных гласных	меньше-больше	мс
P7		длительность пауз	меньше-больше	мс
P8		расположение пауз	меньше-больше	мс
P9	фразовые тоновые интонации	уровень акцента	низкий-высокий	Гц
P10		длительность акцента	меньше-больше	мс
P11		интенсивность акцента	меньше-больше	дб

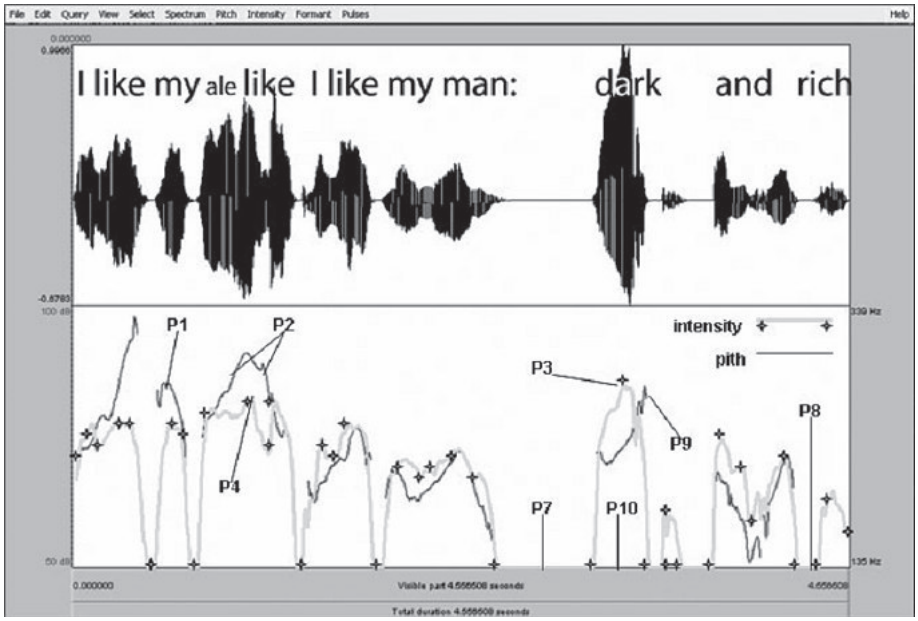

Рис. 1. Возможности измерения параметров просодии

Обоснование требований и оценка эффективности алгоритмов

Все перечисленные выше параметры являются физическим отражением просодических свойств речевого сигнала, не связанных непосредственно с его цифровый формой при передаче, обработке и хранении. Поэтому к ним применим общий подход к встраиванию информации и типовая последовательность алгоритмов: а) сегментации речевого потока; б) определения значения несущего параметра в каждом сегменте; в) вычисления нового его значения при помощи QIM-кодирования и шифрования; г) его модификации [1–4,8]. Однако сложность проектной разработки, технической реализации и область их применения (связь(С) и/или аутентификация(А)) существенно различаются. Например, для длительности паузы (Р7) все преобразования в сторону её увеличения достаточно просты и могут быть выполнены в реальном времени в сеансе связи. В то же время обработка фразовых интонаций (Р9-Р11) возможна только по завершении фразы. Возникающая при этом задержка неприемлема как с точки зрения скрытности, так и коммуникации. В разной стадии проработки находятся и методы сегментации, измерения, модификации параметров (таблица 2), различны их проектная и вычислительная сложность.

Таблица 2

Сегментация, измерения, кодирование, модификация, параметров просодии

парамеры/ алгоритмы	P1	P2	P3	P4	P7	P8	P9	P10	P11
сегментация	+	+	+	?	+	+	+	+	+
измерение	+	?	+	?	+	+	+	+	+
кодирование	+	+	+	?	+	+	+	+	+
модификация	+	?	+	+	+	?	+	+	+
область исп.	С,А	?	С,А	А	С,А	А	А	А	А

В приведённой таблице знаком «+» обозначено наличие, по крайней мере, формализованного подхода, модели, а иногда и программной реализации (часто без доступа к исходному коду) алгоритма. Знаком «?» — или полное отсутствие такого подхода (в фонетической

литературе обозначаемое как «очень сложно»), или возможность его реализации только в условиях полного отказа от управления другими параметрами.

Квантование речевого сигнала по времени и уровню является условием встраивания цифровых данных в просодические параметры. Квантование по времени — это сегментация речи на некоторые однородные временные участки. На них измеряемый параметр (тон, интенсивность, длительность, расположение) может быть описан простой функцией или правилом (звук-тишина, тональный-шумовой, подъем-спад, постоянство), а на границах этих участков происходят резкие изменения акустических характеристик. Очевидно, что для каждого параметра может быть разработано множество алгоритмов сегментации. Основное требование к ним — однозначность разбиения на сегменты до и после внедрения скрываемой информации и прохождения по каналу связи. Причём при одновременном, т.н. векторном внедрении данных в несколько параметров речевой поток или фонограмма может состоять одновременно из несколько наборов сегментов. Проектная и вычислительная сложность этих алгоритмов невелика. В настоящее время некоторые их них реализованы в средах C++, Matlab, аудиоредакторе Melodyne ©Celemony Software GmbH для разбиения речевого потока на синтагмы.

Измерение в метрических единицах каждого параметра в сегменте является основой для его QIM-кодирования по уровню, являющегося основой внедрения данных [3] . Из всех параметров, перечисленных в табл. 1, только два имеют размерность Гц (P1,P9) — частоты основного тона (ЧОТ). Но именно определение ЧОТ является наиболее технически сложным в проектировании и реализации. Как указывается в [6], на то есть объективные причины — периодичность вокальных звуков со строгой физической точки зрения достаточно условна. Тем не менее, число разработок определителей ЧОТ постоянно растёт, что связано с потребностями вокодерных технологий, распознавания и синтеза речи. Поэтому для этого параметра есть возможность выбора алгоритма. Главное требование — минимальное время вычислений при приемлемой точности, так как вычислительные затраты изначально достаточно большие. Для остальных параметров проектная и вычислительная сложность алгоритмов измерений невелика.

Кодирование с квантованием просодических параметров (QIM) описывается достаточно простым в реализации алгоритмом, основанном на использовании заранее подготовленных кодовых таблиц [3]. Для обеспечения стеганографической стойкости алгоритм QIM комбинируется с поточным криптографическим алгоритмом [8]. Оба алгоритма могут быть эффективно реализованы в специализированных цифровых интегральных схемах.

Модификация просодических параметров тесно связана с их измерениями и имеет те же проблемы применительно к ЧОТ, а также дополнительные проблемы с модификацией длительности вокализованных сегментов. В настоящее время установлено, что с практической точки зрения наиболее целесообразно модифицировать такие просодические характеристики, как частота основного тона, интенсивность и длительность отрезков речевого сигнала, непосредственно модифицируя акустический сигнал как таковой, не используя параметрических моделей [9].

Заключение

Использование просодических параметров речи для скрытого встраивания информации является достаточно сложной задачей не только потому, что человеческий слух довольно точно определяет признаки искусственности

речи. Эта проблема решается определением пределов психоакустической нормы модификации параметров, обеспечивающей гарантированную скрытность. Необходимо оценить реализуемость алгоритмов и стегосистемы в целом. Приведённые данные показывают, что в настоящее время для встраивания информации в основном используются исследовательские алгоритмы, разработанные для других применений, что ограничивает их практические возможности. Недостаточно изучены вопросы искусственного воспроизведения контурных тонов, особенностей артикуляции, ритмических схем ударения, средств артикуляции, интонации и других характеристик просодии в интересах ресинтеза речи. Это объясняется, в том числе, и тем, что разработка математических моделей просодии, связывающих характеристики просодии в её классическом понимании с материальными, акустическими показателями звуков речи, до настоящего времени не имела такого важного практического применения, как задача защиты информации. И не только защиты информации, но и клонирования речи.

ЛИТЕРАТУРА

1. *Ponomar, Marina*. Data hiding in speech signals on the basis of the modification of segment pitch and duration // 19th International Congress on Acoustics ICA2007MADRID, 2–7 Sept. 2007, Madrid, Spain, 2007, CAS-03-023. P. 46-49.

2. *Пономарь М.О*. Использование вариативности речевой просодии при создании интеллектуальных систем защиты информации // Материалы III Всероссийской конференции студентов, аспирантов и молодых ученых «Искусственный интеллект: философия, методология, инновации», 11–13 ноября 2009 г. М.: МИРЭА. С. 374–377.

3. *Пономарь М.О*. Кодирование с квантованием несущих параметров речевых сигналов для скрытого встраивания данных // Сборник трудов XVII Международной конференции «Информатизация и информационная безопасность правоохранительных органов». М.: Академия управления МВД России, 2008, с. 394–396.

4. *Ponomar M.O*. On Acceptable Modification Limits of Electroacoustic Speech Signals for Data Hiding // Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP-2009, September 12-14, 2009, Kyoto, Japan), IEEE Computer Society, Los Alamitos, CA, USA, 2009. Pp. 551–554.

5. *Потапова Р.К*. Речь: коммуникация, информация, кибернетика: Учебное пособие. Изд. 2-е, доп. М.: Эдиториал УРСС, 2001. С. 276–285.

6. *Кодзасов С.В., Кривнова О.Ф*. Общая фонетика: Учебник. М.: Рос. гос. гуманит. ун-т. 2001. С. 183–194.

7. *Paul Boersma & David Weenink*. Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from http://www.praat.org/.

8. *Пономарь М.О*. Обеспечение стеганографической стойкости при встраивании данных в несущие параметры речи // Труды Всероссийской конференции «Проведение научных исследований в области обработки, хранения, передачи и защиты информации». Ульяновск, УлГТУ, 1–5 декабря 2009 г., секция 4, т. 4. С. 65–68.

9. *Бабкин А.В*. Особенности применения технологии TD-PSOLA для модификации характеристик вокальных аллофонов. // Труды международного семинара «Диалог'2000 по компьютерной лингвистике и ее приложениям». М., 2000.