

УДК 681.3

Б.Н. Епифанцев, М.С. Илющенко

Об информативности «модулирующих» функций речевых сообщений*

В целях упрощения алгоритмов идентификации речевых паролей и их дикторов в системах авторизированного доступа к информационным ресурсам выделены огибающие изменения амплитуды и частоты в речевых сообщениях и на их основе синтезированы новые сигналы с амплитудно-частотной модуляцией. Разборчивость синтезированных сигналов оказалась достаточно высокой, что дало основания считать целесообразным внести коррекции в существующие модели речеобразования.

Ключевые слова: информационная безопасность, идентификация речевых паролей, демодуляция сигналов, синтез речевых сообщений, разборчивость.

Задача по распознаванию речевых сообщений поставлена более 50 лет назад. Несмотря на огромные усилия, направленные на решение этой задачи, достигнутые успехи достаточно скромны. До сих пор нет адекватных моделей, описывающих структуру речи [1]. «Задача идентификации личности по речи в настоящее время существует только в постановочной части» [2], предлагаемые модели речевых сигналов не отражают базовых положений формирования и восприятия речи [3]. Целесообразно взглянуть на процесс речеобразования с «радиотехнической» точки зрения, представив речь в виде модулированного по частоте и амплитуде процесса. Основания для этого следующие.

Диапазон частот, в котором ведётся речевое общение, простирается от 80 Гц до 10 кГц. Формальная оценка плотности потока информации в этом диапазоне – 160 кбит/с. В литературе имеются данные, согласно которым в силу инерционности нервной системы максимальная скорость восприятия поступающей человеку информации не превышает 50 бит/с. Приведённые цифры говорят о большой избыточности речевых сигналов, обусловливающей их высокую помехоустойчивость до поступления на вход органов слуха. Частотная и фазовая модуляция – примеры представления сообщений для обеспечения помехоустойчивости.

В статье проанализированы изменения огибающих амплитуды и частоты речевых сообщений и даётся ответ об их информативности, а также о степени дублирования одной и той же информации фазовых и амплитудных характеристик этих сообщений. В экспериментах использовались микрофон Logitech Dialog 320 Black, звуковая карта SB Creative Live 5.1 формата HD Audio, процессор Pentium Dual-Core E2200 (2200 МГц), оперативная память Samsung DDR2-4GB.

Речевые сигналы подвергались преобразованиям:

$$A(t, \tau) = \int\limits_t^{t-\tau} f^2(t) dt, \quad N(t, \tau) = \int\limits_t^{t-\tau} n(t) dt, \quad (1)$$

где $f(t)$ – амплитуда речевого сигнала; $n(t)$ – число пересечений функции $f(t)$ с нулевой осью; τ – интервал интегрирования; t – время.

На рис. 1 приведён пример работы «демодулятора» (1) при воздействии на их вход речевого сигнала «оценить восстановленную речь» (осциллограмма в верхней части рисунка). Обращает на себя внимание непохожесть кривых $A(t, \tau)$ и $N(t, \tau)$. Коэффициент корреляции этих кривых во всех проведённых экспериментах изменялся в пределах от –0,2 до –0,5. Согласно существующим представлениям глухие звуки характеризуются высокой частотой и малой амплитудой, звонкие – наоборот. Поэтому следовало ожидать корреляцию функции (1) вблизи значения –1. Следовательно, $A(t)$ и $N(t)$ в информационном смысле не дублируют друг друга и для выделения информации в речевых сообщениях используются совместно.

Для подтверждения гипотезы об информативности функций $A(t, \tau)$, $N(t, \tau)$ они использовались в качестве модулирующих амплитуду и частоту синусоиды. Полученный

* Работа выполнена в рамках реализации программы «Научные и научно-педагогические кадры инновационной России на 2009–2013 годы», контракт № П215 от 22.07.09 г.

амплитудно-частотно-модулированный сигнал изображён на рис. 2 совместно с исходным речевым сигналом. Смоделированные таким способом звуки прослушивались и по известной методике оценивался процент правильно распознанных сигналов. Полученные результаты колебались от 60 до 72%.

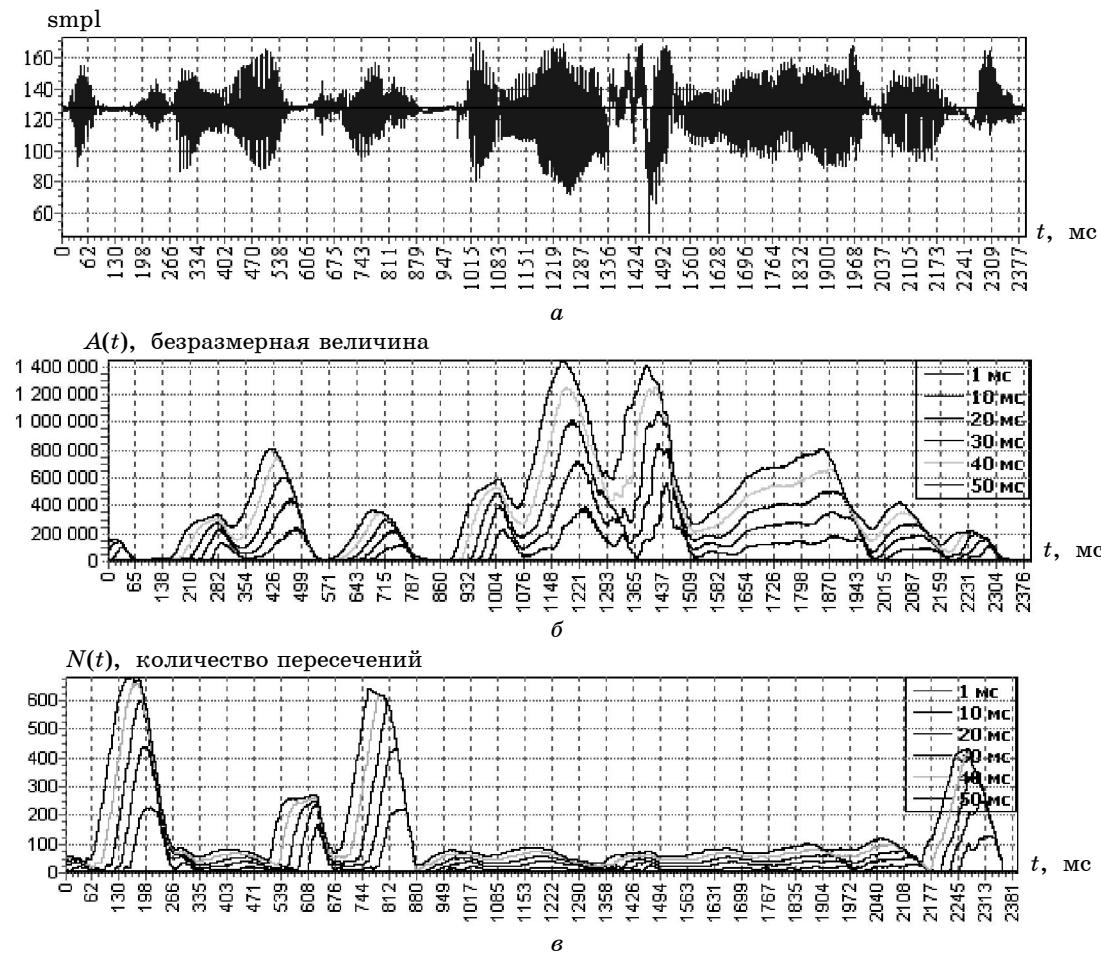


Рис. 1. Изображение речевого сигнала – *а*; его преобразования амплитудным «демодулятором» – *б*; изменения числа переходов через нулевую ось исследуемого сообщения – *в*

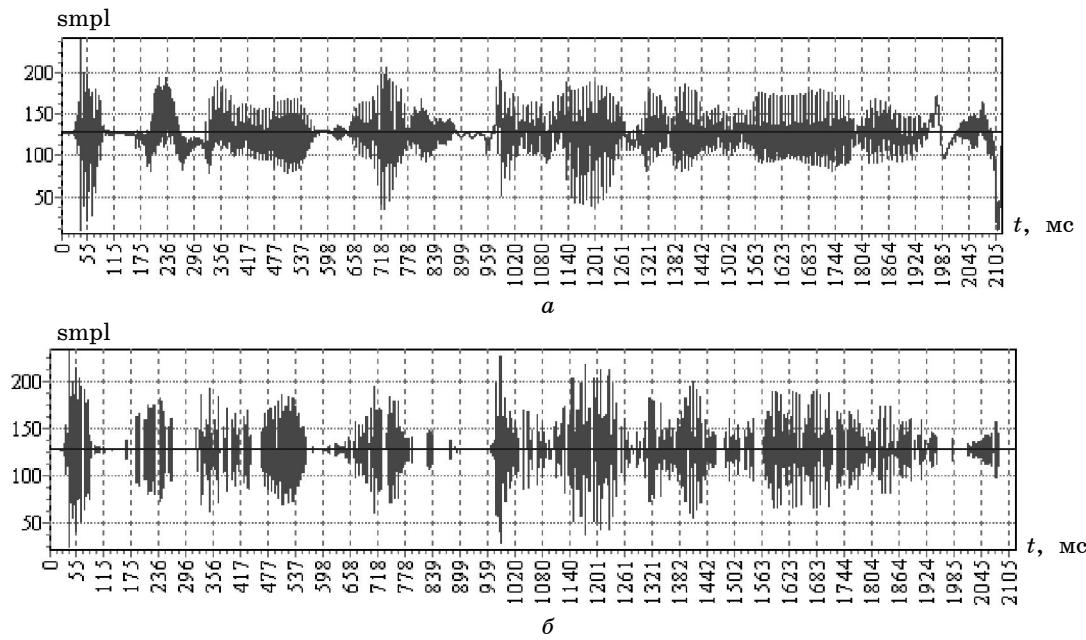


Рис. 2. Осциллограмма речевого сигнала – *а*; и восстановленного с использованием полученных из него модулирующих функций – *б*

Зарегистрированный процент распознавания восстановленных сигналов говорит скорее в пользу высказанной гипотезы, чем против неё. Есть отличия от реального механизма речеобразования в описанной процедуре. В голосовых звуках импульсы основного тона подвергаются амплитудной и времязимпульсной модуляции. Носителем для глухих звуков является случайный процесс, передаваемая информация закладывается через изменения спектральной плотности и её интенсивности. Использование реальных носителей вместо синусоиды позволит окончательно закрыть вопрос об информативности параметров используемых носителей и функциях, приводящих к изменению этих параметров в процессе речеобразования.

Литература

1. Бондаренко В.П. Математическое и программно-алгоритмическое обеспечение в задачах идентификации и распознавания речи / В.П. Бондаренко, А.А. Конев, Р.В. Мещеряков, А.А. Шелупанов // Вестник Сиб. гос. аэрокосмического ун-та. Спецвыпуск «Системная интеграция и безопасность». – Красноярск: СибГАУ, 2006. – С. 11–14.
2. Женило В.Р. Специальная техника и информационная безопасность: учеб. пособие / В.Р. Женило и др.; под ред. В.Н. Кирина. – Т. 1. – М.: Академия управления МВД России, 2000. – 780 с.
3. Ролдугин С.В. Модели речевых сигналов для идентификации личности по голосу / С.В. Ролдугин, А.Н. Голубинский, Т.А. Вольская // Радиотехника. – 2002. – № 11. – С. 79–81.

Епифанцев Борис Николаевич

Доктор техн. наук, профессор, зав. каф. информационной безопасности
Сибирской государственной автомобильно-дорожной академии (СибАДИ), г. Омск
Тел.: (+7-381-2) 25-04-80
Эл. адрес: epifancev_bn@sibadi.org

Илющенко Михаил Сергеевич

Студент 5-го курса каф. информационной безопасности
Сибирской государственной автомобильно-дорожной академии (СибАДИ), г. Омск
Тел.: +7-950-330-53-30
Эл. адрес: super_tm@mail.ru

B.N. Epifancev, M.S. Ilyushchenko

About informativeness of «modulating» functions of speech messages

With a view of simplification of algorithms of identification of speech passwords and their speakers in systems authorized access to informational resources bending around changes of amplitude and frequency in speech messages are selected and on their basis new signals with peak-frequency modulation are synthesized. Articulation of the synthesized signals has appeared enough high that has given the grounds to consider expedient to bring correction in existing models speech production.

Keywords: informational safety, identification of speech passwords, demodulation of signals, synthesis of speech messages, articulation.