



Озвучивание SMS-сообщений, отправляемых на стационарные телефоны

*Людовик Т.В.,
кандидат технических наук
Робейко В.В.*

Тексты SMS-сообщений отличаются спонтанностью, экспрессивностью, открытым характером лексикона, а также зачастую многоязычностью и неправильной орфографией. В данной статье описывается опыт использования системы синтеза украинской речи для озвучивания SMS-сообщений, посылаемых на стационарные телефоны. Описывается подход, позволяющий синтезировать как нейтральную, так и выразительную речь различной коммуникативной ориентации. Результаты проведённого тестирования свидетельствуют о том, что потенциальные пользователи предпочитают, чтобы SMS озвучивались выразительным женским голосом.

SMS texts are characterized by spontaneity, expressivity, open lexicons, multilinguality, and incorrect spelling. In this article we describe the experience of applying a TTS system for Ukrainian language to reading SMS texts sent to fixed phones. We describe the approach allowing to synthesize both neutral and expressive speech with distinct communicative orientation. Test results show that people prefer SMS texts synthesized with an expressive female voice.

Введение

В данной статье излагается опыт, приобретённый в рамках внедрения системы синтеза украинской речи [1] в службу, обеспечивающую автоматическое озвучивание SMS-сообщений.

Сервис SMS2Voice (SMS2Fix), разработанный немецкой компанией, которая занимается информационными коммуникационными технологиями [2],

позволяет пользователям сетей различных мобильных операторов посылать SMS обычного формата на стационарные телефоны. Данный сервис внедрён в Украине украинской компанией Global Message Services Ukraine (GMSU). На данный момент обслуживаются три языка: украинский, русский и английский. Соответственно, задействованы три различных синтезатора речи. Для русского и английского языков используются разработки известных иностранных компаний. Для украинского языка используется синтезатор речи, созданный украинской компанией CyberMova с учётом разработок Международного научно-учебного центра информационных технологий и систем (г. Киев).

Статья посвящена приложению системы синтеза украинской речи к озвучиванию специфических текстов (SMS) в специфических условиях (однократное прослушивание абонентом сообщения, озвученного незнакомым голосом). Мы не будем рассматривать такие важные и сложные проблемы, как автоматическое распознавание языков и транслитерация текстов SMS латиницей, поскольку первая из этих задач решается немецкими коллегами, а вторая – совместно с ними. Также мы не будем подробно останавливаться на решениях, связанных с синтезом речи на русском языке. Однако будут затронуты проблемы, общие для украинского и русского языков в связи с их тесным переплетением в мобильном пространстве Украины.

Типичный пример озвученного SMS-сообщения выглядит следующим образом:

Добрий вечір. (*Добрый вечер.*)
Ви отримали SMS-повідомлення. (*Вы получили SMS-сообщение.*)
Я прийду до дому з нивестой.
Відправник (*Отправитель*): +38 097 9897931.
Дата відправлення (*Дата отправки*): 12.10.2009. 11:00.
На все добре. (*Всего доброго.*)

По приведённому примеру можно судить о том, насколько проблема озвучивания текстов усложняется наличием орфографических и иных ошибок: правильный текст SMS должен выглядеть так: «Я прийду додому з нареченою» («Я приду домой с невестой»).

Исследованный материал

Материал в виде реальных SMS, отправленных в период с 17 мая 2008 года по 6 июня 2008 года, был анонимизирован путём исключения данных об отправителях и получателях. Таким образом, анализировались лишь тексты сообщений. Экспертом было выделено пять категорий SMS-текстов в соответствии с языком, на котором они написаны. В таблице 1 представлено распределение проанализированных SMS-сообщений.

В дальнейшем проводился анализ первой и второй категорий. Следует отметить, что так называемый украинско-русский язык не является ни украинским языком с «русскими» включениями, ни наоборот. Анализ реальных SMS свидетельствует о том, что в Украине распространён «суржик» – гибрид украинского и русского языков, усложняющий синтез речи на обоих языках и требующий изучения специалистами.

Таблица 1

Распределение текстов SMS-сообщений

Язык	Количество SMS (%)	Примеры
Украинский	1523 (14.1%)	Як справи? (Как дела?)
Украинско-русский («суржик»)	1262 (11.7%)	Умееш находить похидну с корня? (производную) Мама, готов шось поесть. Напиши SMS.
Русский	7640 (70.6%)	Я на работе.
Другие языки	274 (2.5%)	Tutto ok. E mille grazie
Нетекстовые SMS	126 (1.2%)	@1@2@3@

Проблемы синтеза речи, связанные со спецификой SMS-текстов

Наряду с проблемами повышения разборчивости (актуальной в условиях использования телефонных каналов связи) и естественности синтезированной речи, адаптация системы синтеза украинской речи к озвучиванию SMS потребовала решения дополнительных задач из области социо- и психолингвистики.

В таблице 2 перечислены основные проблемы, связанные с озвучиванием SMS-сообщений.

Как видно из примеров, многоязычность SMS-текстов может проявляться не только на уровне слов, но и на уровне букв, что существенно затрудняет автоматическое распознавание языков и расстановку словесного ударения, из-за того что неправильно написанные слова отсутствуют как в украинском, так и в русском словарях.

Как правило, SMS-тексты имеют выраженную коммуникативную направленность, которую желательно отразить при их озвучивании. Основным средством для этого является интонация. В настоящее время ставится задача моделирования декларативного, восклицательного и вопросительного коммуникативных типов в синтезированной речи. В частности, общий и специальный вопросы должны отличаться интонационно.

В отличие от проблемы выражения в озвучиваемых SMS эмоциональности и спонтанности, решение которой представляется делом будущего, проблемы, связанные с омографией, расшифровкой аббревиатур и сокращений, а также с отсутствием в текстах SMS разделительных знаков, должны решаться уже сегодня.

Таблица 2

Основные проблемы, связанные с озвучиванием SMS-сообщений

Специфические проблемы озвучивания SMS-сообщений	Примеры
Многоязычность	Вы хто? Сонечко, мы с тобой встречаемся? Прівэт бабуля Проверка св'язі ! Шо Робыш?
Неправильная орфография: а) намеренная; б) описки; в) неграмотность	Яаа люблю тильки тебее! Дякою. Візьми тіліфон. Яанемагу.позванити.поытому.номиру.
Коммуникативная направленность	Бабуся я тебе люблю! Вітаю зі святом! Батьки збираються? До тебе можна? Доброго дня!Як настроїй?А можна Інну до телефону? Нехай росте вам на радість! Хай!Як справи?Це Таня з Деревич це мій ном.пиши що там Руся,Ярик? Галя, нельзя так себя вести. Давай вставай. Давай когда?
Эмоциональность	это я!!! Ааааа!!!! Я заклеил гитару!! Поставил струны и она просто ВЕЛИКОЛЕПНО звучит! Алла!!! Шабаш Давай!!! Дима ты не послушный!!! Ж Д У !!! Хи-хи хи-хи.как делифки??? Игорь вставай!!! Лиля! Делай уроки! Немедленно!ясно???! Я їду з Дімою!!! де мое бабло вчорашне??? ЗРОЗУМІТИ ТЕБЕ! А ТИ ЗРОЗУМІЛА МЕНЕ??? Ты ведь сам хотел розстатса??? Я права??? АЛЕ ЦЕ ВЖЕ ЗАНАДТО! НЕ БРАТИ ТРУБКУ КОЛИ Я ДЗВОНЮ! ТИ ПЕРЕГИНАЄШ ПАЛКУ!
Отсутствие знаков пунктуации	Як дела що в тебе нового у мене всьо постарому привет вид дивчат. Лераятебялюблюіочуштобтиприехалазамной АтакВсеДобре. ГОВОРИ ГОВОРИ А ТЕБЕ КАК ОБ СТЕНКУ ГОЛОВОЙ ДУМАЙ И ВСПОМНИ
Омография и омофония	Бажаю добра і тепла. АНЮТКА,МЫ НЕКУДА НЕ ЕДИМ.
Аббревиатуры и сокращения	візьми 2 грн на завтра Вол. Ів. Дякую. Виба4. г.Одессы Что сег.собир.делать?

Подход к созданию женского выразительного голоса

SMS на всех трёх языках озвучиваются методом Unit Selection [3], подразумевающим использование больших баз данных, содержащих отрезки реальной речи и отражающих как особенности голосов конкретных дикторов, так и стиль чтения, используемый диктором. В первом коммерческом варианте системы синтеза украинской речи использовался мужской голос ПАНАС с довольно монотонным стилем чтения художественной литературы. Считается, что для озвучивания SMS больше подходит женский голос, при этом синтезированная речь должна звучать в нейтральном стиле или в стиле «хорошие новости» [4–7].

Новый женский голос НАТАЛКА, используемый в настоящее время в рамках сервиса «SMS на стационарный телефон», отличается от мужского большей естественностью и выразительностью.

Как и ранее, для синтеза украинской речи применяется метод Unit Selection [3] в комбинации с индивидуализированными просодическими моделями.

Применяемый подход позволяет синтезировать естественно звучащую речь с индивидуальными интонацией и ритмикой.

Речевая база данных

Речевая база данных является одним из основных компонентов системы синтеза речи методом Unit Selection. Как правило, используется голос профессионального диктора, записанный в студийных условиях. В зависимости от сферы будущего применения синтезированной речи диктор читает различные подготовленные тексты (новости, прогнозы погоды и т.п.).

При создании голоса НАТАЛКА, предназначенного, в частности, для озвучивания SMS-сообщений, были отобраны и составлены тексты различных стилей. Чтение этих текстов профессиональным диктором-женщиной должно было отразить как различную коммуникативную направленность текстов (чтение художественной литературы, новостей, диалогов, SMS-сообщений), так и общепринятое произнесение телефонных номеров, дат, электронных адресов и т.п.

В таблице 3 описан акустический материал, послуживший основой речевой базы данных голоса НАТАЛКА.

Объединение в одной речевой базе данных акустических записей, неизбежно отличающихся степенью выразительности при использовании текстов разных типов, порождает дополнительную сложность при синтезе речи. С одной стороны, это позволяет синтезировать более выразительную речь; с другой стороны, затрудняется процесс поиска речевых элементов в базе данных.

Использованные тексты были затранскрибированы автоматически. Поскольку произношение диктора нормативное, результаты автоматического фонем-

Таблица 3

**Речевой материал, послуживший основой
для создания голоса НАТАЛКА**

Тип текста	Количество слов в тексте	Продолжительность записи для данного текста (мин.)	Продолжительность записи для данного текста (%)
Художественный текст «Алиса в Стране Чудес»	10602	130	65
Тексты SMS-сообщений	619	8	4
Диалоги	208	2	1
Новости	478	5	2
Украинский и английский алфавиты	61	2	1
Названия сайтов, даты и т.п.	280	5	2
Специально сконструированные фразы	3483	49	24
Всего	15731	201	100

ного транскрибирования не потребовали коррекции. Вручную в транскрипцию были вставлены паузы, добавленные диктором и не соответствующие знакам пунктуации. Полученная транскрипция была использована для автоматической фонемной сегментации речевого материала с помощью пакета программ НТК [8]. Границы аллофонов были откорректированы с помощью автоматизированных средств обработки речевых сигналов [9]. Дальнейшая сегментация выделенных аллофонов на периоды основного тона была произведена в автоматическом режиме с незначительной ручной коррекцией.

В речевой базе данных, насчитывающей около 80 000 аллофонов, каждый элемент аннотирован: идентификатором, именем, состоящим из трёх частей (имя предыдущей, текущей и последующей фонем), длительностью всего аллофона, уровнями интенсивности первой и второй половин, а также для гласных и звонких согласных – последовательностью длительностей периодов основного тона и количеством периодов. Таким образом, в аннотации речевой базы данных приведена исключительно объективная информация, т.е. отсутствуют символные просодические метки [10].

Предобработка текстов SMS-сообщений

Задача озвучивания SMS-сообщений потребовала существенного развития модуля предобработки текстов [2]. В настоящее время осуществляются такие преобразования:

- описание эмотиконов типа :) соответствующими словами («смайлик»);
- замена слов из списка «неприличная лексика» звуковым сигналом «би-и-п»;
- транслитерация SMS-сообщений, написанных латиницей;
- обработка обозначения дат, телефонных номеров, адресов web-страниц и т.п.;
- членение длинных предложений на синтагмы;
- расстановка словесных ударений;

- расшифровка аббревиатур и сокращений;
- преобразование чисел и символов в орфографический текст.

Особое внимание было уделено расстановке ударений и расшифровке аббревиатур. Словари, используемые для этих операций, имеют открытый характер и динамически пополняются (обновляются). Тем не менее, вероятность ненахождения некоторых слов в словарях остаётся высокой из-за креативности или неграмотности отправителей SMS.

Расстановка ударений связана также со снятием омографии. Эта проблема продолжает оставаться сложной, поскольку система синтеза речи не производит семантический и прагматический анализ текстов.

Транскрибирование текстов SMS с учётом особенностей произношения диктора

Модуль транскрибирования был расширен введением правила вставки гортанной смычки перед фонетическими словами, начинающимися гласной фонемой. Гортанная смычка в этой позиции характерна для голоса НАТАЛКА. Кроме того, выполняя функцию граничного маркера [11], гортанная смычка способствует повышению качества синтезированной речи, улучшая её разборчивость.

Вычисление просодических характеристик SMS-сообщений

Вычисление длительности синтезируемых звуков осуществляется с помощью простой модели, параметрами которой являются средняя длительность фонемы, тип контекста, в котором она находится в синтезируемом высказывании, и набор коэффициентов длительности для данной фонемы. Модель длительности звуков индивидуализируется оффлайн на основе автоматического анализа аннотации речевой базы данных.

Разработанная модель интонирования близка к модели интонационных портретов акцентных единиц, предложенной Б.М. Лобановым [12].

Модель интонирования также была индивидуализирована с учётом особенностей голоса НАТАЛКА. На рисунке 1 приведены стилизованные интонационные контуры завершенности и незавершенности, используемые при синтезе речи голосами ПАНАС и НАТАЛКА. Интонационные фразы (синтагмы), представленные на рисунке, состоят из трёх акцентных групп (АГ). Представлен наиболее типичный случай, когда последняя акцентная группа синтагмы несёт синтагматическое ударение. Приведённые контуры нормированы. Ось абсцисс соответствует времени, а ось ординат – нормированным значениям частоты основного тона. Нормированное значение частоты основного тона «0» соответствует 150 Гц для голоса НАТАЛКА и 80 Гц для голоса ПАНАС. Максимальное значение «10» соответствует 375 Гц для голоса НАТАЛКА и 180 Гц для голоса ПАНАС. На рисунке видно, что изменения интонационных контуров голоса НАТАЛКА более резко выражены. Это свидетельствует о большей экспрессивности этого голоса.

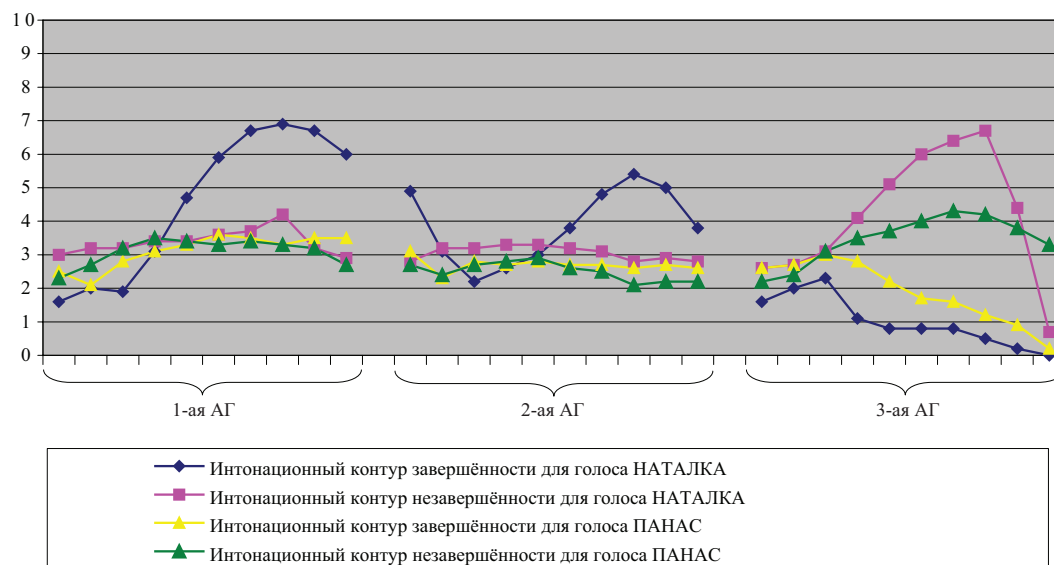


Рис. 1. Стилизованные интонационные контуры завершенности и незавершенности для женского голоса НАТАЛКА и для мужского голоса ПАНАС

Поиск в базе данных

Описанные выше модули вырабатывают целевое описание озвучиваемого текста – его фонемно-просодическую транскрипцию, в которой указана последовательность фонем, их длительности и их интонационные контуры в виде значений длительности периодов основного тона.

Алгоритм выбора аллофонов (Unit Selection) из речевой базы данных НАТАЛКА отличается от предыдущего варианта, предназначенного для голоса ПАНАС. В частности, используются дополнительные критерии выбора, связанные с интенсивностью (громкостью) аллофонов. Первоначальный алгоритм выбора работал локально, т.е. учитывалась не вся цепочка аллофонов синтагмы, а только текущий выбираемый аллофон и аллофон, выбранный на предыдущем шаге. Для голоса НАТАЛКА был разработан новый вариант алгоритма Unit Selection, приближенный к классической схеме [3].

Генерация акустического сигнала

Генерация акустического сигнала представляет собой конкатенацию речевых отрезков, соответствующих аллофонам, выбранным из речевой базы данных. При синтезе речи голосом ПАНАС конкатенация осуществляется с просодической модификацией ударных гласных (изменением их длительности и контура основного тона). Аллофоны голоса НАТАЛКА не модифицируются.

Результаты

Было проведено формальное тестирование с целью определить, какой из голосов (ПАНАС или НАТАЛКА) звучит качественнее (естественнее). В частности, обоими голоса-

ми было озвучено 33 SMS-текста, т.е. было синтезировано 66 стимулов. Эксперимент проходил в три этапа. На первом этапе аудиторами были 5 человек – сотрудников МНУЦИТиС, на втором – 33 преподавателя и студента-лингвиста (специальность «украинский язык и литература»), на третьем – 4 сотрудника компании Global Message Services Ukraine. Всем аудиторам предлагалось прослушать синтезированные стимулы, предъявляемые в случайном порядке, и выставить каждому стимулу оценку от 0 (очень плохо) до 5 (очень хорошо).

Результаты тестирования приведены на диаграмме (рис. 2). Все три группы аудиторов отдали предпочтение женскому голосу НАТАЛКА.

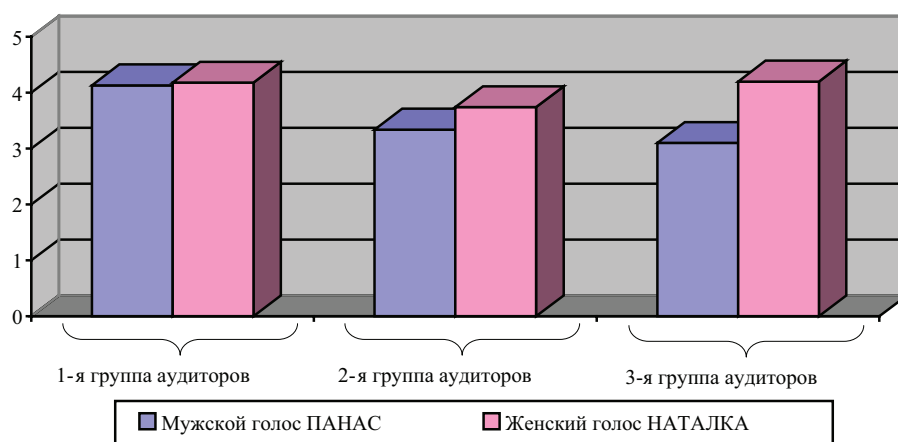


Рис. 2. Результаты тестирования естественности озвученных SMS

Заключение

Опыт применения системы синтеза речи к озвучиванию SMS-сообщений свидетельствует о том, что традиционная для синтеза речи проблема повышения естественности звучания продолжает оставаться актуальной. При этом сложно найти компромисс между нейтральным и выразительным (эмоциональным) стилями озвучивания. Очевидны также социолингвистические аспекты проблемы озвучивания SMS-сообщений, связанные с многоязычностью и широким распространением «суржика». Отдельной проблемой является открытый характер словарей, требующий постоянного мониторинга.

Литература

1. Lyudovyyk T., Sazhok M. Unit Selection Speech Synthesis Using Phonetic-Prosodic Description of Speech Databases, Proc. International Conference «Speech and Computer» (SPECOM'2004), St.-Petersburg (Russia), 2004. P. 594–599.
2. Lyudovyyk T., Brozinski S., Noner M., Robeiko V., Sazhok M. Speech Synthesis Applied to SMS reading, Proc. of the 13th International Conference «Speech and Computer: SPECOM'2009». – St.-Petersburg, Russia, 2009. P. 300–305.

**Людювик Т.В., Робейко В.В. Озвучивание SMS-сообщений,
отправляемых на стационарные телефоны**

3. *Hunt A., Black A.* «Unit selection in a concatenative speech synthesis system using large database», Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 1996. P. 373–376.
4. *Shinsuke S., Jinfu N., Ranniery M., Keiichi T., Minoru T., Tomoki Toda, Hisashi Kawai and Satoshi Nakamura.* Communicative Speech Synthesis with XIMERA: a first step, Proc. of 6th ISCA Speech Synthesis Workshop, Bonn, Germany, August 22–24, 2007 (CD-ROM proceedings).
5. *Fernandez R., Ramabhadran B.* Automatic exploration of corpus-specific properties for expressive text-to-speech: A case study in emphasis. In Proceedings of the 6th ISCA Workshop on Speech Synthesis, pages 34–39, Bonn, Germany, 2007.
6. *Eide E., Aaron A., Bakis R., Hanza W., Picheny M., Pirelli J.* A corpus-based approach to <AHEM/> expressive speech synthesis, in Proceedings of 5th ISSW, June 2004. P. 79–84, Pittsburgh, USA.
7. *Theune M., Meijs K., Heylen D.K.J., Ordelman R.J.F.* Generating Expressive Speech for Storytelling Applications, IEEE transactions on audio, speech and language processing, 2006, 14 (4). P. 1137–1144.
8. *Young S., Evermann G., Kershaw D., Moore G., Odell J., Ollason D., Valtchev V., Woodland P.* The HTK Book, Cambridge University Engineering Department, 2002.
9. *Сажок М.М.* Автоматизовані засоби формування баз даних і знань для озвучення українських текстів: Автореф. дис. канд. техн. наук: 05.13.06. К.: МННЦІТІС, 2004. 20 с.
10. *Silverman K., Beckman M., Pitrelli J.* ToBI: a Standard for Labeling English Prosody, Proc. of the International Conf. on Spoken Language Processing. Banff, 1992. Vol.2. P. 867–870.
11. *Кривнова О.Ф.* Ларингализация как граничный маркер в устной речи, Труды XVI сессии Российского акустического общества РАО-16. М. 2005.
12. *Лобанов Б.М., Цирульник Л.И.* Компьютерный синтез и клонирование речи. Минск: Белорус. Наука, 2008. 337 с.

Людювик Татьяна Владленовна

Кандидат технических наук.

Работает старшим научным сотрудником Международного научно-учебного центра информационных технологий и систем (г. Киев, Украина). Закончила Киевский государственный университет по специальности «Прикладная лингвистика». Главные научные интересы лежат в области синтеза речи и экспериментальной фонетики.

Робейко Валентина Васильевна

Закончила Институт филологии Киевского национального университета им. Т. Шевченко. Специалист в области современного украинского языка. Работает младшим научным сотрудником в Международном научно-учебном центре информационных технологий и систем (г. Киев, Украина), аспирантка. Область интересов: синтез и распознавание речи, фонетика.