

ИСТОРИЧЕСКИЕ ПРЕДПОСЫЛКИ СОЗДАНИЯ СИСТЕМ СИНТЕЗА БЕЛОРУССКОЙ РЕЧИ

Дан ретроспективный обзор исследований и разработок в области синтеза речи по тексту, проведенных в Беларуси за период до 2000-го года и составивших основу работ лаборатории распознавания и синтеза речи ОИПИ НАН Беларуси по созданию в последующие годы белорусскоязычного синтезатора речи.

Введение

Специальное подразделение ЮНЕСКО, занимающееся сохранением языков мира как живых представителей культурного наследия планеты, отнесло белорусский язык к «потенциально угрожаемым языкам» (http://www.helsinki.fi/~tasalmin/europe_index.html). Если современная тенденция кардинально не изменится, то в ближайшие десятилетия мы можем потерять еще один живой язык Земли – славянский белорусский язык, являющийся, к тому же, государственным, как никакой другой в этой категории европейских языков*.

В свете этого факта значимость создания белорусскоязычного синтезатора речи трудно переоценить. В нём программным образом аккумулированы и сохранены для будущих поколений: лексика белорусского языка – орфография, акцентуация, словоизменение – фонетика и правила чтения, интонация и звуки современной белорусской речи. В то же время современным пользователям компьютеров и Интернета этот программный продукт сможет помочь в овладении и совершенствовании знания и навыков белорусской речи, а также использовать его в различных прикладных системах.

В статье дан ретроспективный обзор исследований и разработок в области синтеза речи по тексту, проведенных в Беларуси за период до 2000-го года и составивших основу работ лаборатории распознавания и синтеза речи ОИПИ НАН Беларуси по созданию в последующие годы белорусскоязычного синтезатора речи. В статье даются ссылки в основном на оригинальные работы автора, непрерывно руководившего лабораторией на протяжении многих лет её существования. Из работ других сотрудников лаборатории приведены в основном лишь те, которые либо написаны в соавторстве, либо выполнены под руководством автора. Таким образом, в приведенном списке литературы упомянуты практически все научные сотрудники, внесшие определённый вклад в осуществление генерального направления научных исследований лаборатории на протяжении последних 25 лет прошлого века.

Под синтезом речи в данной работе понимается процедура превращения входного орфографического текста в звучащую речь. При этом необходимо решить три основные задачи:

- синтез речевого сигнала,
- синтез фонетических и
- синтез просодических характеристик речи.

Ниже рассмотрено в исторической последовательности решение перечисленных задач, а также описаны разработанные на этой основе различные модели синтеза речи по тексту, экспериментальные и промышленные системы синтеза речи, их внедрение и применение в различных отраслях.

*) **Potentially endangered languages Europe:**

- | | |
|----------------------------------|----------------|
| 1. Belorussian | 6. Piedmontese |
| 2. (Lowland) Scots | 7. Ligurian |
| 3. Low Saxon (Low German proper) | 8. Lombard |
| 4. Galician | 9. Emilian |
| 5. Francoprovençal | 10. Corsican |

1. Синтез речевого сигнала.

Первая, пока ещё примитивная модель синтезатора русской речи ФОНЕМАФОН-1 “заговорила” в начале 70-х гг., и успех в её создании был связан прежде всего с разработкой принципов формантного синтеза речевых сигналов. Разработанные принципы легли в основу авторских свидетельств “Формантный синтезатор речи” [1] и “Формирователь импульсов тонального возбуждения” [2]. В дальнейшем появилась усовершенствованная модель формантного синтеза речевых сигналов [3], а затем были оптимизированы характеристики формантных фильтров синтезатора речи последовательного типа [4]. Важную роль в создании промышленных синтезаторов речи сыграла разработка цифрового формантного синтезатора [5]. Ещё долгое время формантный синтезатор играл ключевую роль в системах синтеза речи по тексту, пока в конце 80-х - начале 90-х гг. не был предложен новый микроволновой метод синтеза речевых сигналов [6].

2. Синтез фонетических характеристик речи

За время существования лаборатории сменилось три поколения систем синтеза речи по тексту, в основу которых были положены три существенно различных подхода к синтезу фонетических характеристик речи: фонемно-артикуляторно-формантный (1), фонемно-формантный (2) и фонемно-микроволновой (3). Толчком к появлению первого подхода послужило исследование коартикуляции на акустическом уровне [7], которое позволило осуществить текущее определение формантных частот по функциям движения артикуляторов [8]. В результате была разработана модель артикуляторного синтеза речи по печатному тексту [9], которая стала основой авторского свидетельства на синтезатор речи [10]. Второй подход удалось реализовать благодаря развитию акустической теории коартикуляции и редукции [11,12], созданию методики построения формантных портретов фонем [13], а также созданию алгоритмов синтеза формантных параметров [14] и вычислению фонемных портретов для синтеза речи по тексту [15]. Третий подход сформировался в начале 90-х гг. и получил название микроволнового синтеза речи по тексту [16].

3. Синтез просодических характеристик речи

Несмотря на исключительную важность просодических характеристик для качественного синтеза речи, сведения о закономерностях их поведения в русской речи были крайне скудными. Поэтому в начале 70-х гг. были проведены эксперименты по восприятию русской интонации односложной синтетической фразы [17], проведен анализ и синтез просодических характеристик двухсложного слова [18], разработаны правила синтеза просодических характеристик однослогных фраз [19] и сформулированы принципы автоматического синтеза интонационных структур [20]. В дальнейшем алгоритмы автоматического синтеза интонации по печатному тексту были усовершенствованы [21]. Это касалось алгоритмов синтеза по тексту мелодического и ритмического контуров [22], а также моделей синтеза мелодического контура русских и английских фраз [23]. Были разработаны алгоритмы интонирования текста [24] и многофакторная модель ритмики [25].

4. Модели синтеза речи по тексту

Разработанные методы синтеза речевого сигнала, а также методы синтеза фонетических и просодических характеристик речи позволили приступить к созданию целостных моделей синтеза речи по тексту. Первой такой моделью стал формантный синтезатор речи по последовательности

аллофонов [26]. Был разработан преобразователь графема-фонема для синтеза речи по орфографическому тексту [27], и вместе с моделью артикуляторно-формантного синтеза речи [28,29] он стал основой авторского свидетельства на устройство для синтеза речи [30]. Были заложены также лингво-акустические основы двуязычного синтеза речи [31,32] и разработан алгоритм синтеза многоязычной речи по тексту [33]. Результаты работ этого периода обобщены в кандидатской диссертации Б.В. Панченко [34] и в докторской диссертации Б.М. Лобанова [35]. Разработанные модели многоязычного синтеза речи по тексту с успехом демонстрировались на Всемирной выставке "TELECOM" в Женеве и на Международных конгрессах фонетических наук [36,37]. Они были положены также в основу двух международных проектов: с Дрезденским университетом [38,39] и с Московским институтом проблем передачи информации [40] в рамках Европейского фонда INTAS.

5. Промышленные системы синтеза речи

На базе разработанных моделей синтеза речи по тексту рядом организаций была предпринята попытка создания промышленных синтезаторов речи. В Московском НИИ ЭВТ разработан формантный синтезатор речи с фонемным управлением "Фонемафон-2" [41], а в Московском НИИ РП - его усовершенствованная версия для решения задачи вывода информации из ЭВМ в речевом виде [42]. В основе следующих поколений синтезаторов речи для вывода информации из ЭВМ "Фонемафон-3" и "Фонемафон-4" положены авторские свидетельства [43-45] и принципы построения синтезатора речи на базе программируемой ЭВМ [46]. Устройство речевого вывода информации "Фонемафон-3" описано в [47]. Синтезаторы речи серии "Фонемафон-4", разработанные в Калининградском ПО "Кварц", и их применение описаны в [48]. Следующая модель - "Фонемафон-5", выполненная в виде одноплатного модуля синтеза речи по тексту, разработана в Горьковском КБ "Квант" и описана в [49]. В [50] проанализированы состояние и перспективы разработки речевых устройств для интеллектуальных роботов связи и других приложений. Последняя, наиболее совершенная коммерческая программная модель микроволнового синтеза речи по тексту, нашедшая в последнее время самое широкое применение, описана в [51-53].

6. Системы с синтезированным речевым ответом

Вопросы разработки автоматизированных информационно-справочных телефонных систем с синтезированным речевым ответом впервые были рассмотрены в [54]. Разработаны проекты телефонной диалоговой системы "Абонент-АСУ МТС" [55], телефонного автоматического устройства для ИСС с речевым вводом-выводом [56], телефонного автосекретаря и речевой почты [57]. В монографии [58] обобщены многочисленные аспекты применения синтезированной речи в системах массового обслуживания. В середине 80-х гг. разработана и внедрена во многих городах СССР (от Бреста до Петропавловска-Камчатского) система автоматического информирования абонентов телефонной сети о задолженности за международные переговоры [59]. Опыт внедрения речевых процессоров в отрасли "Связь" обобщён в [60]. Для другой отрасли - Минжилкомхоза БССР разработана и внедрена интерактивная речевая система контроля исполнения и диспетчеризации производства на базе ЭВМ "Электроника-100-25" и синтезатора речи "Фонемафон-4Т" [61]. За последнее время, благодаря работам группы под руководством Г.В. Лосика., модель синтезатора речи [52] использована также в многочисленных информационных системах для слепых.

Заключение

В заключение хочется отметить, что автор обзора хорошо осознаёт не только значительность проделанной работы, но и грандиозность ещё нерешённых вопросов по проблеме синтеза речи. Для решения проблемы высококачественного синтеза речи в начале 2000-х годов положен лингво-акустический подход синтеза речи по тексту на основе компиляции речи из большого количества (свыше 1000) аллофонов и мультифонов, представленных в виде отрезков естественной речевой волны, а также на активном использовании морфологического и синтаксического анализа «читаемого» текста. Такой подход позволяет потенциально обеспечить высокую разборчивость,

естественность и выразительность звучания синтетической речи. Основы теории и практики разработки систем синтеза речи на основе этого подхода описаны в монографии [62].

Литература

- 1 А.С. СССР N 479107. Формантный синтезатор речи / Б.М. Лобанов. Заявл.19.02.73.
2. А.С. СССР N 492914. Формирователь импульсов тонального возбуждения / Б.М. Лобанов, Б.В. Панченко. Заявл. 7.06.74.
3. Гурьянов Н.И., Лобанов Б.М., Рыжиков В.В. Усовершенствованная модель формантного синтеза речевых сигналов // Автоматическое распознавание слуховых образов (АРСО-9). - Минск, 1976. – С. 18.
4. Бухтилов Л.Д., Гурьянов Н.И., Лобанов Б.М. Оптимизация характеристик формантных фильтров в последовательном синтезаторе речи // Автоматическое распознавание слуховых образов (АРСО-11). - Ереван, 1980. – С. 33.
5. Бухтилов Л.Д., Лобанов Б.М., Минкевич В.В., Первой Л.М. Цифровой формантный синтезатор // Автоматическое распознавание слуховых образов (АРСО-12). - Киев, 1982. – С. 445.
6. Лобанов Б.М. Микроволновой синтез речи // Автоматическое распознавание слуховых образов (АРСО - 16). – М., 1991. – С. 27-31.
7. Буданицкий Э.Г., Лобанов Б.М. Исследование коартикуляции на акустическом уровне // Автоматическое распознавание слуховых образов (АРСО-8), Ч. 2. - Львов, 1974. – С. 67.
8. Лобанов Б.М. Текущее определение формантных частот по функциям движения артикуляторов // Автоматическое распознавание слуховых образов (АРСО-8), Ч. 4. - Львов, 1974. – С. 91-93.
9. Лобанов Б.М., Панченко Б.В. Модель артикуляторного синтеза речи по печатному тексту // Автоматическое распознавание слуховых образов (АРСО-9). - Минск, 1976. – С. 27.
10. А.С. СССР N 533966. Синтезатор речи / Б.М. Лобанов. Заявл.15.1.75.
11. Лобанов Б.М. К акустической теории коартикуляции и редукции // Автоматическое распознавание слуховых образов (АРСО-11). - Ереван, 1980. – С. 25-29.
12. Lobanov B. On the Acoustic Theory of Coarticulation and Reduction. Proc. ICASP-82, Paris, 1982. , pp. 915-918.
13. Лобанов Б.М., Павлович Н.А. Методика построения формантных портретов фонем для синтеза речи по тексту // Автоматическое распознавание слуховых образов (АРСО-12). - Киев, 1982. – С. 417-421.
14. Лобанов Б.М., Марченков М.А. Алгоритмы синтеза формантных параметров по тексту // Автоматическое распознавание слуховых образов (АРСО-12). - Киев, 1982. – С. 414-416.
15. Аксютин И.В., Лобанов Б.М. Алгоритм вычисления фонемных портретов для синтеза речи // Автоматическое распознавание слуховых образов (АРСО-14). - Каунас, 1986. – С. 51-52.
16. Лобанов Б.М. Микроволновой синтез речи по тексту // Анализ и синтез речи. – Минск: Инст. техн. кибернетики АНБ, 1991. – С. 57-73.
17. Башкина Б.М., Лобанов Б.М. Восприятие русской интонации односложной синтетической фразы // Докл. Всесоюз. конф. "Анализ и синтез речи". – Минск: МГПИИЯ им. М. Тареза, 1972. – С. 16-20.
18. Башкина Б.М., Лобанов Б.М. Анализ и синтез просодических характеристик двухсложного слова // Докл. Всесоюз. конф. "Анализ и синтез речи". - Минск: МГПИИЯ им. М. Тареза, 1972. – С. 21-24.
19. Башкина Б.М., Лобанов Б.М. Синтез по правилам просодических характеристик односложных фраз // Автоматическое распознавание слуховых образов (АРСО-7). - Алма-Ата, 1973. – С. 123-126.
20. Лобанов Б.М. Принципы автоматического синтеза интонационных структур // Автоматическое распознавание слуховых образов (АРСО-10). - Тбилиси, 1978. – С. 158-160.
21. Бухтилов Л.Д., Лобанов Б.М., Минкевич В.В. Алгоритм автоматического синтеза интонации по печатному тексту // Автоматическое распознавание слуховых образов (АРСО-10). - Тбилиси, 1978. – С.132-133.
22. Лобанов Б.М., Марченков М.А. Алгоритмы синтеза по тексту мелодического и ритмиче-

ского контуров // Автоматическое распознавание слуховых образов (АРСО-12). - Киев, 1982. – С. 412-413.

23. Карневская Е.Б., Лобанов Б.М. Модели синтеза мелодического контура русских и английских фраз // Автоматическое распознавание слуховых образов (АРСО-12). - Киев, 1982. – С. 399-401.

24. Зимовина Г.В., Лобанов Б.М., Марченков М.А. Алгоритмы интонирования орфографического текста для синтезатора речи "Фонемофон-4" // Автоматическое распознавание слуховых образов (АРСО-13). - Новосибирск, 1984. – С. 139.

25. Карневская Е.Б., Лобанов Б.М. Многофакторная модель ритмики и ее реализация при синтезе речи по тексту // Автоматическое распознавание слуховых образов (АРСО-15). - Таллинн, 1989, - С.145-149.

26. Лобанов Б.М., Панченко Б.В. Формантный синтез речи по последовательности аллофонов // Тр. Всесоюзн. акустической конф. – М., 1973. – С. 27-28.

27. Лобанов Б.М., Панченко Б.В. Преобразователь графема-фонема для синтеза речи по орфографическому тексту // Автоматическое распознавание слуховых образов (АРСО-8), ч. 4. - Львов, 1974. – С. 15.

28. Budanitsky E., Lobanov B., Panchenko V. An Articulatory Model of Speech Synthesis. Proc. Of Open Seminar on Acoustic, Wroclaw, 1975. pp. 47-49.

29. Лобанов Б.М., Минкевич В.В. Модель артикуляторно-формантного синтеза речи по печатному тексту // Тез. докл. Республиканского симп. "Экспериментально-фонетические исследования речевого текста". - Минск: МГПИИЯ им. М. Тареза, 1977. – С. 28-31.

30. А.С. СССР N 459797. Устройство для синтеза речи / Б.М.Лобанов. Заявл.25.7.72.

31. Карневская Е.Б., Лобанов Б.М. Лингвоакустические основы двуязычного синтеза речи // Автоматическое распознавание слуховых образов (АРСО-11). - Ереван, 1980. – С. 119-122.

32. Lobanov B. Articulatory-Formant Speech Synthesis from Printed Text. Proc. of Franco-Sovietique Symposium on Speech. Paris, 1981. pp.73-75.

33. Лобанов Б.М., Марченков М.А. Алгоритм синтеза многоязычной речи по тексту // Автоматическое распознавание слуховых образов (АРСО-13). - Новосибирск, 1984. – С.140-141.

34. Панченко Б.В. Исследование формантных методов синтеза речи по орфографическому тексту: Дисс. канд. техн. наук. - Минск, 1983.

35. Лобанов Б.М. Исследование и разработка методов синтеза речи по тексту: Дисс. докт. техн. наук. - Рига, 1984.

36. Lobanov B. The Phonemophon Text-to-Speech System. Proc. of the XI International Congress of Phonetic Sciences, Tallin, 1987, pp. 61-64.

37. Lobanov B., Karnevskaya E. MW Speech Synthesis from Text. Proc. of the XII International Congress of Phonetic Sciences. Aix-en-Provence, France, 1991, pp. 406-409.

38. Lobanov B. Microwave Speech Synthesis from Text. Proc. of the 24 Fachkolloquim Informationstechnik, Dresden, 1991, pp. 118-120.

39. Lobanov B., Ivanov A., Kubashin A., Levkovskaya T. A Bilingual German / Russian Text-to-Speech System, Proceedings of the 3rd International Workshop "Speech and Computer" - SPECOM'98, St.-Petersburg, 1998, pp.327-330.

40. Boguslavsky I., Karnevskaya E., Lobanov B. Generation of Intonation and Accentuation of Synthetic Speech on the Basis of Morpho-Syntactic Knowledge. Proc.of International Workshop "Integration of Language and Speech", Moscow, 1995, pp. 11-28.

41. Лобанов Б.М., Панченко Б.В., Рождественская А. Формантный синтезатор речи с фонемным управлением // Автоматическое распознавание слуховых образов (АРСО-8), Ч. 4. - Львов, 1974. – С.15.

42. Бабин И.И., Лобанов Б.М., Панченко Б.В. Об одном решении задачи вывода информации из ЭВМ в речевом виде // Вопросы радиоэлектроники. Сер. ЭВТ. - N11. - М., 1974. – С. 110-124.

43. А.С. СССР N 485492.. Устройство для синтеза речи / Б.М. Лобанов, Б.В.Панченко, Г.С. Слуцкер. Заявл. 9.1.73.

44. А.С. СССР N 607211. Устройство для вывода речевой информации / Б.М. Лобанов, Б.В.Панченко. Заявл.11.7.75.

45. Лобанов Б.М., Панченко Б.В., Минкевич В.В. Фонемно-формантный синтез речи для вывода информации из ЭВМ // Тр. инст. техн. кибернетики АН БССР. - Минск, 1978. – С. 151-160.
46. Бойкевич А.М., Пивоваров В.М., Сережкина Т.Г., Лобанов Б.М., Минкевич В.В., Панченко Б.В. Принципы построения синтезатора речи на базе программируемой ЭВМ // Автоматическое распознавание слуховых образов (АРСО-11). - Ереван, 1980. – С.143-145.
47. Лобанов Б.М., Минкевич В.В., Панченко Б.В. Устройство речевого вывода информации "Фонемафон-3" для ЭВМ // Управляющие системы и машины. - N 2.– М., 1982. – С. 33-37.
48. Афанасьев В.П., Лобанов Б.М. Синтезаторы речи серии "Фонемафон" и их применение // Проблемы практического использования систем автоматического распознавания и синтеза речи. – Л., 1983. – С. 6-7.
49. Зимицкий Ю.С., Лобанов Б.М. Одноплатный модуль синтеза речи по тексту "Фонемафон-5" // Автоматическое распознавание слуховых образов (АРСО-14). - Каунас, Ч. 2. - 1986. – С. 54-58.
50. Лобанов Б.М. Состояние и перспективы разработки речевых устройств для интеллектуальных роботов связи // Электросвязь. - N 8. – М., 1988. – С. 43-45.
51. Лобанов Б.М. Программная модель микроволнового синтеза речи по тексту. // Автоматическое распознавание слуховых образов (АРСО-14). - Москва, 1991. – С. 82-84.
52. Иванов А.Н., Лобанов Б.М. Синтезатор речи ФОНЕМАФОН для САПР на базе IBM PC // Тез. докл. конф. "Теория и методы создания интеллектуальных САПР". - Минск, 1992. – С. 29-30.
53. А.С. СССР N 1683063. Способ компиляционного синтеза речи и устройство для его осуществления / Б.М Лобанов. Заявл. 08.06.1991.
54. Красносельский Н.И., Лобанов Б.М. Вопросы разработки автоматизированной информационно-справочной телефонной системы с синтезированным речевым ответом // Докл. Всесоюзн. конф. по проблемам совершенствования проектирования радиосистем и их элементов. - Минск, 1975. – С. 35-37.
55. Лобанов Б.М., Панченко Б.В., Усов Л.П. Проект телефонной диалоговой системы "Абонент-АСУ МТС" // Автоматическое распознавание слуховых образов (АРСО-9). - Минск, 1976. – С. 23.
56. Лобанов Б.М., Первой Л.М. Телефонное автоматическое устройство для ИСС с речевым вводом-выводом // Автоматическое распознавание слуховых образов (АРСО-12). - Киев, 1982. – С. 582-584.
57. Аксютин И.В., Левков Е.Я., Лобанов Б.М., Первой Л.М. Телефонный автосекретарь и речевая почта // Автоматическое распознавание слуховых образов (АРСО-13). - Новосибирск, 1984. – С. 165-166.
58. Кучеров В.Я., Лобанов Б.М. Синтезированная речь в системах массового обслуживания. – М.: Радио и связь, 1983. – 130 с.
59. Лобанов Б.М., Панченко Б.В. Система автоматического информирования абонентов телефонной сети о задолженности за международные переговоры / Экспресс информация. Серия: Эксплуатация средств связи. Вып. 6. - М., 1986. – С. 4-6.
60. Лобанов Б.М., Панченко Б.В., Первой Л.М., Усов Л.П. Опыт внедрения речевых процессоров в отрасли "Связь" // Автоматическое распознавание слуховых образов (АРСО-14), Ч. 2. - Каунас, 1986. – С. 56.
61. Войнило В.В., Безнис В.М., Афанасьев В.П., Лобанов Б.М. Интерактивная речевая система контроля исполнения и диспетчеризации производства на базе "Электроника-100-25" и "Фонемофон-4Т" // Автоматическое распознавание слуховых образов (АРСО-14), Ч. 2. - Каунас, 1986. – С. 59-60.
62. Лобанов, Б. М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник // Минск: Белорусская Наука, 2008. – 342 с