УДК 004.934

Г. Лань, А.С. Фадеев, А.Н. Моргунов

Синтез фрагментов голоса человека на основе модели реконструкции частотных спектров

Разрабатывается методика параметрического синтеза фонем голоса человека на основе аналитического описания отдельных гармоник. Представлена методика анализа спектра и спектрограмм оригинальных фонем для получения основных амплитудно-частотных характеристик составляющих сигнала, предложен алгоритм реконструкции речевого сигнала на основе полученных наборов параметров. Предложена методика оценки качества синтезированных элементов речи.

Ключевые слова: синтез речи, спектр, спектрограмма, фонема, гармоника, амплитудно-частотная характеристика.

doi: 10.21293/1818-0442-2021-24-2-14-20

Большинство исследований, связанных с синтезом речи человека, в основном используют один из двух подходов: синтез на основе готовых фрагментов и параметрический синтез [1]. Первая группа методов включает методы синтеза на основе кодирования и редактирования речевого сигнала. Недостатками данных методов являются значительная зависимость от размера и качества звуковой библиотеки, высокие требования к вычислительным ресурсам, возможность синтезировать только сегменты речи с ограниченным словарным запасом и возникающие сплайсинг-разрывы в результате синтеза.

Вторая стратегия – система гармонического синтеза – основывается на синтезе акустических параметров звукового сигнала, например кодирование с линейным прогнозированием (LPC). Требования к размерам звуковой библиотеки для параметрического синтеза существенно ниже, чем для методов первой группы, система способна адаптироваться к широкому диапазону ритмических характеристик, чем достигается хорошая естественность синтезированной речи, и значительно менее требовательна к вычислительным ресурсам. При формантном синтезе обычно используется несколько формантных фильтров, а синтез отдельных речевых сигналов осуществляется путем установки значений параметров каждого фильтра. Основными недостатками этого метода являются множество параметров и высокая сложность процесса их настройки [1, 2].

В связи с использованием гармонических характеристик предлагаемый в настоящей работе метод не требует настройки фильтров, все значения параметров можно получить на основе анализа спектра отдельных фонем и использовать полученные гармонические параметры для предложенной модели синтеза речи.

Одним из основных аспектов синтеза человеческой речи является качество выходного сигнала, от чего зависит пригодность использования конкретной технологии синтеза в коммерческих целях. Чаще всего в задачах синтеза речи человека применяются TTS-алгоритмы (Text to Speech) [3, 4].

В отрасли информационно-коммуникационных технологий существует целый ряд методов, позволяющих произвести оценку качества сформированного сигнала. Основными метриками, по которым определяется качество сигнала, являются оценки разборчивости, естественности и мультимодальности речи. Сами методы оценки качества можно разделить на объективные и субъективные. В проведенных исследованиях использовалась методика средней экспертной оценки (MOS, или mean opinion score). MOS-метод является субъективным, рекомендован стандартом Р.800 МСЭ-Т для оценки качества передачи звука в телефонных сетях [5, 6].

В рамках задачи синтеза речи MOS-метод применяется для выяснения надежности используемого решения и качества синтезируемого звукового сигнала. Основная задача в разработке систем синтеза речи – разработать систему и методику генерации человеческой речи, результаты которой будут таковы, что человек на слух не смог бы отличить сгенерированный сигнал от сигнала, что был записан реальным индивидом.

Опираясь на опыт предыдущих исследований, в данной работе рассматривается методика параметрического синтеза отдельных фонем русского языка на основе гармонических параметров, рассматривается возможность создания универсальных аналитических моделей, позволяющих описать и в дальнейшем синтезировать фонемы голоса человека.

Синтез гласных звуков

Согласно [7] гласный звук состоит из одного слога, слог состоит из одной или нескольких фонем. Фонема – минимальная смысловозначительная единица языка, не имеющая самостоятельного лексического или грамматического значения, служащая для различения и отождествления значимых единиц языка: морфем и слов [8].

Таким образом, фонемы являются наименьшими элементарными составляющими речи человека, синтез которых на основе аналитических моделей и наборов связанных параметров позволяет генерировать более крупные элементы речи: звуки, буквы и слова.

На рис. 1 и 2 изображены спектрограммы фонем двух гласных и двух согласных букв: «О», «Э» и «Ш», «М» соответственно. Очевидно, что согласные звуки имеют более широкий спектр, чем фонемы гласных букв. Особенно здесь стоит выделить шипящие, например, звук «Ш» – спектрограмма данной фонемы по своей форме больше похожа на спектрограмму шумового сигнала. Отсюда следует вывод, что аналитическая модель описания фонем согласных букв гораздо сложнее аналогичной модели для гласных звуков.

Рис. 1. Спектрограммы фонем гласных букв «О» и «Э»

Рис. 2. Спектрограммы фонем согласных букв «Ш» и «М»

Для дальнейшего синтеза частей фонем необходимо декомпозировать звук гласной буквы на простые компоненты частотного спектра – гармоники, получить аналитическое описание и исчерпывающий набор параметров, которые позволят создать математическую модель для синтеза каждой отдельной гармоники и фонемы сигнала. Гармоника – дополнительный тон, который по частоте всегда выше основного тона, причём строго кратно числам натурального ряда (т.е. выше по частоте в 2–5 и более раз) [9].

Спектральный анализ гласных звуков показывает, что на некоторых участках частотный спектр гармоник остаётся неизменным (см. рис. 1, 2). Для упрощения моделирования и синтеза отдельных гармоник можно применить следующую аналитическую модель на квазистационарных участках:

$$\varphi_i(t) = A_i \sin(2\pi \nu_i t), \quad (1)$$

где $\varphi_i(t)$ – временная функция гармоники i , A_i и ν_i – амплитуда и частота гармоники i соответственно. Аналитическая модель квазистационарного участка фонемы будет иметь вид

$$\Phi(t) = \sum_i \varphi_i(t), \quad (2)$$

Для определения параметров A_i и ν_i , а также числа отдельных гармоник используются оконное преобразование Фурье (ОПФ) и полученная на его основе спектрограмма записанного сигнала.

Спектрограмма звукового сигнала

Основным этапом предлагаемой методики является получение спектрограммы анализируемого сигнала, которая является результатом применения ОПФ:

$$STFT(\tau, \omega) = \int_{-\infty}^{+\infty} f(t)W(\tau-t)e^{-j2\pi\nu\tau}d\tau. \quad (3)$$

Здесь $W(\tau-t)$ – оконная функция; $f(t)$ – значение амплитуды исходного сигнала в момент времени t ; τ – положение окна; ν – частота компоненты. В виду малой продолжительности по времени записанного сигнала окно ОПФ было подобрано таким образом, чтобы в результате получить высокое частотное разрешение.

Квазистационарное состояние

Для чистоты эксперимента были использованы записи звуков отдельных фонем, сделанные одним человеком. Записанные звуки были произнесены на относительно разных частотах: низкие, высокие и «естественные» для человека и с различной эмоциональной окраской: подъём интонации, спад, без эмоционального произношения и пр.

Известно, что процесс воспроизведения звучания человеческого голоса нестабильный и непериодический [8], в данной работе анализируется только квазистационарный временной отрезок звучания фонемы (период выдержки). Значения параметров фонемы и их точность зависят от свойств выбранного временного отрезка. На рис. 3 продемонстрирована спектрограмма гласного звука «А», произнесенного с эмоциональным подъемом. Анализ рисунка показывает, что несмотря на нестационарность всего сигнала и каждой гармоники по отдельности, отдельные короткие временные отрезки продолжительностью 10^{-1} – 10^{-2} с являются квазистационарными и для них возможно применение описанной модели.

Рис. 3. Спектрограмма гласной буквы «А»

Рис. 4. Спектрограмма квазистационарного участка «А»

Получение спектра

Для определения количества гармоник (n), присутствующих в отдельных фонемах, и определения значения амплитуд (A_i) основных частот каждой гармоники были использованы амплитудно-частотные характеристики фонем речи, представленные спектром сигнала. Для построения спектра сигнала применено быстрое преобразование Фурье (БПФ) в пакете прикладных программ MATLAB:

$$Y(k) = \sum_{j=1}^N X(j) W_N^{(j-1)(k-1)}, \quad (4)$$

где $Y(k)$ – дискретный спектр сигнала; $W_N = e^{(-2\pi i)/N}$ – комплексная экспонента; $X(j)$ – дискретный сигнал анализируемой фонемы; j, k – номер отсчета в последовательностях $X(j)$ и $Y(k)$ соответственно; N – количество отсчетов.

Рис. 5. Спектр гласной буквы «У»

На рис. 5 показано, что основная энергия сигнала сконцентрирована в диапазоне частот 100–800 Гц,

а максимальной амплитудой обладает первая гармоника, имеющая частоту 286,5 Гц. Гармоники с частотами выше 800 Гц имеют малую амплитуду и практически неразличимы на фоне шумов. В логарифмическом представлении (рис. 6) пять гармоник (F1–F5) выражены более явно в отличие от линейной шкалы (см. рис. 5).

Рис. 6. Спектр гласной буквы «У» на логарифмической шкале

Вычисление значений параметров для отдельных гармоник

Для возможности синтеза отдельной гармоники i необходимо получить набор векторов F_i , содержащих параметры P_{ij} . Такие векторы составляют матрицу параметров M_Φ гармоник записанного звукового сигнала:

$$M_\Phi = [F_1 \ F_2 \ F_3 \ F_i \ F_n].$$

Для построения математической модели и реконструкции отдельных гармоник был предложен способ параметрического описания гармоник речи человека на основе полученных частотно-временных характеристик [4]. Очевидно, что амплитуда A_i и частота v_i являются основными параметрами, характеризующими поведение каждой гармоники гласного звука человеческой речи во времени. Для вычисления значений параметров каждой гармоники P_{ij} используются численные значения A_i и v_i вершин каждого всплеска спектра звукового сигнала F_i (рис. 6).

В таблице приведены параметры значений частоты и амплитуды каждой гармоники, полученные для некоторых гласных букв. Очевидно, что максимальной амплитудой гласного звука «У» обладают гармоники с индексами $i = 1$ и $i = 2$, что соответствует диапазону 200–600 Гц; для звука «А» – с индексами $i = 1–6$, и соответствуют диапазону 200–1600 Гц; для звука «О» с индексами $i = 1, 2$, и диапазон составляет 300–700 Гц; для звука Э – с индексами $i = 1–3$, $i = 6–10$ и $i = 12$, им соответствуют

диапазоны частот 300–1000 и 1800–3100 Гц и около 4000 Гц; для звука «И» максимальной амплитудой обладают гармоники с индексами $i = 1$ и $i = 9-14$, их диапазоны частот составляют 150–250 и 2900–4100 Гц. В среднем расстояние между гармониками состав-

ляет значение частоты первой гармоники. Но не все гармоники присутствуют в фонемах, амплитуда некоторых гармоник настолько мала, что они не присутствуют в спектре. Записанные гласные звуки не имели эмоциональной окраски.

Значения параметров гармоник гласных букв

Матрица параметров буквы «О»													
Гармоники	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	-
(v)	321	642	944	1265	1586	1889	2209	2548	2851	3172	3492	4116	-
(A)	0,62	0,41	0,50	1	0,65	0,42	0,35	0,06	0,03	0,02	0,03	0,005	-
Матрица параметров буквы «У»													
Гармоники	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	-
(v)	287	573	873	1146	1446	2320	2593	3179	3479	3776	6086	6372	-
(A)	1	0,12	0,02	0,003	0,003	0,001	0,001	0,001	0,001	0,002	0,001	0,004	-
Матрица параметров буквы «А»													
Гармоники	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
(v)	265	529	793	1058	1322	1587	1851	2116	2645	2915	3171	3435	3703
(A)	0,78	0,74	0,51	0,60	1	0,26	0,08	0,03	0,04	0,05	0,04	0,05	0,05
Матрица параметров буквы «Э»													
Гармоники	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
(v)	307	614	921	1244	1511	1858	2165	2472	2779	3087	3361	3717	4024
(A)	0,85	1	0,33	0,08	0,07	0,18	0,27	0,13	0,24	0,12	0,01	0,04	0,21
Гармоники	F_{14}	F_{15}	F_{16}	F_{17}	-	-	-	-	-	-	-	-	-
(v)	4331	4654	4954	5266	-	-	-	-	-	-	-	-	-
(A)	0,03	0,01	0,03	0,01	-	-	-	-	-	-	-	-	-
Матрица параметров буквы «И»													
Гармоники	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
(v)	303	598	901	1205	1499	1803	2106	2409	2713	3007	3311	3614	3917
(A)	1	0,03	0,03	0,01	0,01	0,01	0,01	0,02	0,05	0,1	0,07	0,07	0,03
Гармоники	F_{14}	F_{15}	F_{16}	F_{17}	-	-	-	-	-	-	-	-	-
(v)	4212	4524	9342	9924	-	-	-	-	-	-	-	-	-
(A)	0,04	0,01	0,01	0,01	-	-	-	-	-	-	-	-	-

Для удобства представления АЧХ в работе принято решение использовать нормализованные значения амплитуд всех гармоник отдельных фонем, при котором за единицу принимается значение максимальной амплитуды фонемы.

$$\varphi_i(t) = \frac{A_i}{\max(A_j)} \sin(2\pi\nu_i t), \quad (5)$$

Однако для возможности синтеза гармоник по формуле (2) необходимо выполнить нормализацию значений амплитуд фонем таким образом, чтобы выполнялось условие

$$\sum_{i=1}^n A_i = 1, \quad (6)$$

$$\varphi_i(t) = \frac{A_i}{\sum A_j} \sin(2\pi\nu_i t). \quad (7)$$

Генерация сигнала. Оценка качества синтезированного сигнала

В настоящей работе оценка качества предложенной методики для генерации звукового сигнала, имитирующего голос человека, производится на основе полученных параметров матриц.

$$\Phi(t) = \sum_{i=1}^n A_i \cdot \sin(2\pi\nu_i \cdot t), \quad (8)$$

где $\Phi(t)$ – временная функция синтезированного сигнала; $\sin(2\pi\nu_i t)$ – гармоника сигнала; ν_i – частота гармоники, A_i – значение амплитуды.

Качество синтетического сигнала оценивается путем сравнения АЧХ полученного сигнала и исходного. На рис. 7 представлены спектры оригинального и реконструированного сигнала гласных звуков; все звуковые записи сделаны одним и тем же лицом. Для большей наглядности приведенные спектры показаны в логарифмических осях на рис. 8.

Рис. 7. Спектры звукового сигнала «У»

Рис. 8. Спектры звукового сигнала «О» на логарифмической шкале: — оригинальный сигнал; ---- реконструированный сигнал

Особенности синтеза. Отличия гласных звуков

В настоящем исследовании были рассмотрены различные гласные звуки. На рис. 9 представлены спектры одного и того же гласного звука «А», синтезированного согласно описанной методике для разных звукозаписей диктора.

Рис. 9. Спектры звука «А», произнесенные на низкой (а) и высокой (б) частотах

Анализ спектров одного и того же звука, произнесенного человеком на разных частотах, равно как и спектров синтезированных по ним сигналов, показывает, что основными гармониками для записи на более низких частотах (рис. 10, а) являются гармоники с номерами 3 и 4, основная энергия звука сконцентрирована в диапазоне частот 350–3000 Гц.

Для звукозаписей на более высоких частотах (см. рис. 10, б) основные гармоники — 2 и 3, основная энергия сконцентрирована в диапазоне 500–8000 Гц, значение базисной частоты (б) выше, чем (а).

Рис. 10. Спектрограммы звука «А», произнесенные на низкой (а) и высокой (б) частотах

Для оценки качества полученных синтезированных звучаний было принято решение использовать корреляцию оригинального и синтезированного сигналов [17]. Были вычислены коэффициенты корреляции Пирсона для спектра и спектрограммы каждого оригинального и синтезированного звучания каждой фонемы. Коэффициент корреляции Пирсона для оригинального и синтезированного звучания звука А составил: 0,7139 (спектр) и 0,9338 (спектрограмма); для звука «И»: 0,7349 (спектр) и 0,9858 (спектрограмма); для звука «О»: 0,8453 (спектр) и 0,9691 (спектрограмма); для звука «У»: 0,9343 (спектр) и 0,9914 (спектрограмма); для звука «Э»: 0,8649 (спектр) и 0,9694 (спектрограмма). Численные значения показывают, что корреляции спектров оригинальных и синтезированных звучаний являются существенными (0,5–0,8), а спектрограммы оригинальных и синтезированных звучаний являются сильно коррелированными (0,8–1).

Заключение

В данной работе предложена аналитическая модель описания и синтеза квазистационарных фрагментов отдельных фонем некоторых букв. При помощи спектра и спектрограммы оригинальных фонем получены основные амплитудно-частотные характеристики речи человека и численные значения параметров для модели. Предложена методика, позволяющая синтезировать сигнал фонем человека на основе полученных параметров, а также методика оценки качества синтезированных элементов речи. В работе также приведены некоторые выявленные особенности частотных характеристик наборов гармоник для звуков, произнесенных на разных относительных частотах.

В данной работе также приведена оценка качества полученных синтезированных звучаний с использованием коэффициента корреляции Персона. В результате вычисления корреляции спектра и спектрограмм синтезированных и оригинальных звучаний для разных фонем показано, что корреляция между оригинальными и синтезированными сигналами является высокой.

Методика синтеза речи на основе гармонических характеристик отличается от популярных классических методов гармонического синтеза. Кроме того, предложенная методика может быть применена для исследования и моделирования нестационарных фрагментов фонем голоса человека с изменяющимися частотами тонов, например, при вокальном исполнении или при наличии эмоциональной окраски голоса.

Литература

1. Zhang X. Digital Speech Processing and matlab Simulation (Second Edition). – Beijing: Publishing House of Electronics Industry, 2010. – 311 p.
2. Song Z. Application of MATLAB in Speech Signal Analysis and Synthesis. – Beijing: Beihang University Press, 2013. – 308 p.
3. Лобанов Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск: Белорусская наука, 2008. – 316 с.
4. Рыбин С.В. Синтез речи: учеб. пособие. – СПб.: Ун-т ИТМО, 2014. – 92 с.
5. Оценка качества синтезированной речи: проблемы и решения / А.И. Соломенник, А.О. Таланов, М.В. Соломенник, О.Г. Хомицевич, П.Г. Чистиков // Известия вузов. Приборостроение. – 2013. – Т. 56, № 2. – С. 38–42.
6. Способы оценки субъективного качества речи, материал из Habr [Электронный ресурс]. – Режим доступа: <https://habr.com/en/post/177099>, свободный (дата обращения: 11.05.2021).
7. Phoneme (Lexicon of Linguistics). Universiteit Utrecht [Электронный ресурс]. – Режим доступа: <https://lexicon.hum.uu.nl>, свободный (дата обращения: 11.05.2021).
8. Кодзасов С.В. Общая фонетика / С.В. Кодзасов, О.Ф. Кривнова. – М.: РГГУ, 2001. – 592 с.
9. Гармоника (музыка): материал из википедии [Электронный ресурс]. – Режим доступа: [https://ru.wikipedia.org/wiki/Гармоника_\(музыка\)](https://ru.wikipedia.org/wiki/Гармоника_(музыка)) (дата обращения: 26.06.2021).
10. Лань Г. Методика реконструкции фонем голоса человека / Г. Лань, А.Н. Моргунов // Вестник современных исследований. – 2018. – Т. 25, № 10-3. – С. 130–135.
11. Rabiner L. Theory and applications of Digital Speech Processing / L. Rabiner, R. Schafer // Journal of applied Mechanics. – 2010. – Vol. 1, № 30. – P. 445–447.
12. Фролов А.В. Синтез и распознавание речи. Современные решения / А.В. Фролов, Г.В. Фролов [Электронный ресурс]. – Режим доступа: <https://frolov-lib.ru/books/hi/index.html> (дата обращения: 11.05.2021).
13. Немченко В.Н. Введение в языкознание: учебник для вузов. – М.: Дрофа, 2008. – 703 с.
14. Вендина Т.И. Введение в языкознание. – М.: Высшая школа, 2001. – 288 с.
15. Фирсанков А.А. Система идентификации и синтеза динамических частотных составляющих сложных сигналов // Молодёжь и современные информационные технологии // Сб. трудов XV Междунар. науч.-практ. конф. студентов, аспирантов и молодых учёных. – Томск: НИ ТПУ, 2018. – С. 65–66.
16. Айфичер Э. Цифровая обработка сигналов. Практический подход: пер. с англ. / Э. Айфичер, Б. Джервис. – 2-е изд. – М.: Вильямс, 2008. – 992 с.
17. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов. – 10-е изд., стереотип. – М.: Высшая школа, 2004. – 479 с.

Лань Гунбо

Аспирант отд. информационных технологий (ОИТ)
Национального исследовательского
Томского политехнического университета (НИ ТПУ)
Ленина пр-т, 30, г. Томск, Россия, 634050
Тел.: +7-923-433-76-97
Эл. почта: langongbo@qq.com

Фадеев Александр Сергеевич

Канд. техн. наук, доцент ОИТ НИ ТПУ
Ленина пр-т, 30, г. Томск, Россия, 634050
Тел.: +7-923-457-95-15
Эл. почта: fas@tpu.ru

Моргунов Александр Николаевич

Аспирант Института вычислительных технологий
Сибирского отделения Российской академии наук (СО РАН)
Академика Лаврентьева пр-т, 6,
г. Новосибирск, Россия, 630090
Тел.: +7-999-499-73-63
Эл. почта: toruk010994@gmail.com

Lan G., Fadeev A.S., Morgunov A.N.

Synthesis of Human Voice Fragments Based on Frequency Spectra Reconstruction

This article details the development of methods for the synthesis of phonemes of the human voice based on the analytical description of individual formants. A technique for analyzing the spectrum and spectrograms of original phonemes to obtain the main amplitude-frequency characteristics of the signal components is presented. An algorithm to reconstruct a speech signal based on the obtained sets of parameters is proposed. A technique to assess the quality of synthesized speech elements is described.

Keywords: speech synthesis, spectrum, spectrogram, phoneme, formant, amplitude-frequency characteristic.

doi: 10.21293/1818-0442-2021-24-2-14-20

References

1. Zhang X. [Digital Speech Processing and matlab Simulation (Second Edition)]. Beijing, Publishing House of Electronics Industry, 2010, 311 p.
2. Song Z. [Application of MATLAB in Speech Signal Analysis and Synthesis]. Beijing, Beihang University Press, 2013, 308 p.
3. Lobanov B.M., Tsurulnik L.I. *Kompyuternyy sintez i klonirovaniye rechi* [Computer synthesis and speech cloning]. Minsk, Belarusian Science, 2008, 316 p. (in Russ.).
4. Rybin S.V. *Sintez rechi. Uchebnoye posobiye* [Synthesis of speech. Textbook.]. St. Petersburg, ITMO University, 2014, pp.21–25 (in Russ.).
5. Solomennik A.I., Talanov A.O., Solomennik M.V., Khomitseich O.G., Chistikov P.G. *Otsenka kachestva sintezi-*

rovannoy rechi: problemy i resheniya [Assessing the quality of synthesized speech: problems and solutions]. *News of Higher Educational Institutions. Instrument Engineering*, 2013, vol. 56, no. 2, pp. 38–42 (in Russ.).

6. *Sposoby otsenki sub'yektivnogo kachestva rechi* [Methods for assessing the subjective quality of speech]. Available at <https://habr.com/en/post/177099>, free (Accessed: May 10, 2021) (in Russ.).

7. [Phoneme] (Lexicon of Linguistics) Available at <https://lexicon.hum.uu.nl>, free. (Accessed: May 12, 2021).

8. Kodzasov S.V., Krivnova O.F. *Obshchaya fonetika* [General Phonetics], M.: RGGU, 2001, 592 p. (in Russ.).

9. *Garmonika (muzika)* [Harmonics (music)] Available at: [https://ru.wikipedia.org/wiki/Гармоника_\(музыка\)](https://ru.wikipedia.org/wiki/Гармоника_(музыка)), free. (Accessed: June 26, 2021) (in Russ.).

10. Lan G., Morgunov A.N. *Modelirovaniye i sintez fonem glasnykh bukv* [Reconstruction Method of Human Voice Phones]. *Bulletin of Modern Studies* vol. 25, no. 10-3, pp. 130–135 (in Russ.).

11. Rabiner L.R. Schafer Theory and applications of Digital Speech Processing, *Journal of Applied Mechanics*, 2010, vol. 1, № 30, pp. 445–447.

12. Frolov A.V., Frolov G.V. *Sintez i raspoznavaniye rechi. Sovremennyye resheniya* [Synthesis and speech recognition. Modern solutions] (in Russ.). Available at <https://frolov-lib.ru/books/hi/index.html>, free (Accessed: May 11, 2021) (in Russ.).

13. Nemchenko V.N. *Vvedeniye v yazykoznaniye: uchebnyk dlya vuzov* [Introduction to Linguistics: A Textbook for High Schools]. Moscow, Bustard, 2008, 703 p. (in Russ.).

14. Vendina T.I. *Vvedeniye v yazykoznaniye* [Introduction to Linguistics]. Moscow, Higher School, 2001, 288 p. (in Russ.).

15. Firsankov A.A. *Sistema identifikatsii i sinteza dinamicheskikh chastotnykh sostavlyayushchikh slozhnykh signalov* [System for identification and synthesis of dynamic frequency components of complex signals]. Youth and modern

information technology. *Proceedings of the XV International Scientific and Practical conference for Students, Graduate Students and Young Scientists*. Tomsk: National Research Tomsk Polytechnic University, 2018, pp. 65–66 (in Russ.).

16. Ifeachor E.C., Jervis B.W. [Digital Signal Processing: A Practical Approach]. Upper Saddle River, Prentice Hall, 2008, 992 p. (in Russ.).

17. Gmurman V.E. *Teoriya veroyatnostey i matematicheskaya statistika: Uchebnoye posobiye dlya vuzov* [Probability theory and mathematical statistics: Textbook for universities], Moscow: Higher school, 2004, 479 p. (in Russ.).

Gongbo Lan

Postgraduate student, Division for Information Technology, Tomsk Polytechnic University (NS TPU)
15B, Usov st., Tomsk, Russia, 634034
Phone: +7-923-433-76-97
Email: langongbo@qq.com

Aleksandr S. Fadeev

Candidate of Science in Engineering,
Division for Information Technology, NS TPU
30, Lenin pr., Tomsk, Russia, 634050
Phone: +7-923-457-95-15
Email: fas@tpu.ru

Alexander N. Morgunov

Postgraduate student, Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences (ICT SB RAS)
6, Academician M.A. Lavrentiev pr., Novosibirsk, Russia, 630090
Phone: +7-999-499-73-63
Email: toruk010994@gmail.com