АКУСТИЧЕСКИЙ ЖУРНАЛ, 2003, том 49, № 4, с. 567–569

КРАТКИЕ СООБЩЕНИЯ

УДК 534.78

МОДЕЛИРОВАНИЕ ГЛАСНЫХ ЗВУКОВ

© 2003 г. Д. И. Якушев, О. П. Скляров

Санкт-Петербургский научно-исследовательский институт уха, горла носа и речи
198013 Санкт-Петербург, ул. Бронницая, 9
E-mail: Scljarov@admiral.ru

Поступила в редакцию 12.04.2001 г.

Ранее предпринимались попытки математического моделирования гласных звуков не в частотной, а во временной области [6, 7]. Результаты этих исследований были подтверждены в работах [1, 5]. В настоящее время также проводятся подобные исследования [3, 4]. В предложенных моделях, характерных для линейного приближения, периоды основного тона (ОТ) аппроксимируются зависимостями в виде суммы затухающих синусоид, которая не содержит параметров нарастания звуковой волны. Предпринимались также попытки построения нелинейной модели спектра голосового источника [8].

В настоящей статье предлагается модель, включающая в себя, кроме формантной структуры гласных, характеристику как затухания, так и нарастания звуковой волны на каждом периоде основного тона путем введения квадратичной временной зависимости.

При моделировании была принята следующая аксиоматика:

1. Гласные звуки имеют периодическую структуру нарастания–затухания звука. Период нарастания–затухания звука называется периодом ОТ.

Заметим, что период ОТ не является периодом в строгом математическом понимании, т.к. не удовлетворяет условию: $f(t + T) = f(t)$.

2. Гласные звуки имеют формантную структуру.

При создании модели были приняты следующие допущения:

1. В результате многочисленных экспериментов с различными видами математических зависимостей в качестве модели нарастания–затухания звука на одном периоде ОТ была принята кривая: $t^2\exp(-bt)$, которая позволяет обеспечить максимальную степень сходства между моделью и оригиналом как в объективном, так и в субъективном смысле. Попытки использования иных зависимостей для процессов нарастания приводили к ухудшению оценок качества моделей.

2. Форманта – это отдельно взятая частота, т.е. синусоида, характеризующаяся частотой, амплитудой и фазой: $a\sin(2\pi ft + \varphi)$. При наложении на эту частоту характеристики нарастания–затухания форманта проявляется как усиление в области спектра сигнала, соответствующей этой частоте.

При этих допущениях была построена эмпирическая модель гласных звуков по периодам ОТ:

$$p(t) = t^2\exp(-bt)\sum_{i=1}^{n} a_i\sin(2\pi f_i t + \varphi_i), \qquad (1)$$

где t – время в пределах периода T основного тона;

$t = 0$ – соответствует времени начала возбуждения сигнала;

b – коэффициент, характеризующий ударение (сила голоса), чем больше b, тем сильнее выражено ударение;

n – количество формантных частот (определяется для каждого гласного);

a_i, f_i, φ_i – амплитуда, частота и фаза i-й форманты.

Коэффициенты a_i, f_i, φ_i, где i изменяется от 1 до n, рассчитываются для каждого периода ОТ отдельно. Параметры n и b – общие для всех периодов ОТ.

Метод позволяет производить как анализ сигнала (нахождение параметров модели), так и его синтез, т.е. воспроизведение модели как звука.

За меру близости между моделью и оригиналом принята величина $Q = (1 - D_{ост}/D)$, где $D_{ост}$ – остаточная, а D – исходная дисперсия сигнала [2]. Величина Q может принимать значения от 0 до 1 (полное сходство). Проведенные исследования показали соответствие принятой меры близости результатам слухового сравнения модели и оригинала. На некоторых периодах ОТ получено сходство $Q > 0.995$.

При отработке алгоритма расчетов было проведено моделирование нескольких десятков гласных фонем у разных дикторов. У различных дикторов на различных фонемах были получены меры сходства модели и оригинала, превышающие 0.9, что указывает на пригодность предложенной модели к исследованию гласных звуков.

При получении достаточно высокой меры сходства между моделью и оригиналом разработанный метод позволяет:

Частота F_0 ОТ и частоты f_i и амплитуды a_i формант в модели (1) для гласного [U] (фазы $\varphi_i = 0$)

Сила голоса b					2.11					
№ периода ОТ	F_0, Гц	f_1, Гц	f_2, Гц	f_3, Гц	f_4, Гц	a_1	a_2	a_3	a_4	Q
1	114	232	457	864	3210	6.9	3.5	0.2	0.0	0.99
2	114	230	454	855	3287	7.0	3.3	0.2	0.0	0.98
3	114	229	453	858	3288	7.3	3.4	0.2	0.0	0.99
4	114	229	452	854	3309	7.7	3.7	0.3	0.0	0.99
5	112	228	452	846	3273	8.1	3.6	0.3	0.0	0.98
6	112	227	450	843	3269	8.2	3.6	0.3	0.0	0.98
7	114	227	450	847	3291	8.3	3.6	0.3	0.0	0.98
8	114	229	455	848	3302	8.3	3.5	0.3	0.0	0.98
9	114	229	454	852	3303	8.5	3.7	0.3	0.0	0.99
10	115	230	458	855	3288	8.4	3.5	0.3	0.0	0.98
11	115	232	462	863	3335	8.4	3.8	0.3	0.0	0.98
12	115	232	458	862	3332	8.2	4.0	0.3	0.0	0.98
13	115	235	463	871	3327	8.5	4.3	0.3	0.0	0.98
14	115	234	460	866	3343	9.0	4.5	0.3	0.0	0.98
15	115	232	461	869	3246	9.2	4.8	0.4	0.0	0.98
16	114	231	456	859	3325	9.4	4.9	0.4	0.0	0.98
17	115	230	455	861	3238	9.6	5.0	0.4	0.1	0.98
18	115	232	458	864	3239	9.8	5.0	0.4	0.1	0.98
19	115	232	458	862	3243	9.9	5.0	0.4	0.0	0.98
20	116	233	459	867	3258	9.8	4.8	0.4	0.1	0.98
21	116	235	464	876	3249	9.8	4.8	0.4	0.0	0.99
22	115	234	462	870	3236	9.3	4.8	0.4	0.1	0.98
23	115	233	459	868	3239	8.9	5.1	0.4	0.1	0.98
24	116	234	460	871	3263	8.7	5.0	0.4	0.1	0.98
25	116	234	462	873	3265	8.6	4.9	0.4	0.1	0.98
26	119	238	464	707	3291	8.7	4.9	1.2	0.1	1.00
27	119	240	468	711	3274	8.7	4.8	1.2	0.0	1.00
28	119	242	471	710	3398	8.8	4.8	1.1	0.1	1.00
29	120	243	475	712	3259	8.9	4.7	1.0	0.0	0.99
30	120	243	479	722	3260	9.6	4.7	0.9	0.0	0.99

1. Производить разбиение сигнала на периоды ОТ с построением модели для каждого периода ОТ.

2. Анализировать состав формантных частот исследуемого сигнала.

3. Сопоставлять амплитудные характеристики выделенных формантных частот.

4. Получать числовые характеристики ударения.

5. Соотносить изменения характеристик нарастания–затухания звуковой волны и присутствующих в ней формантных частот с изменениями аудиторной оценки синтезируемого гласного.

В частности, предложенный метод исследования был использован для анализа гласных в речи ринолалика.

Например, для гласного [U] (без окружения) была получена модель, представленная в виде табл. 1.

Анализируя результаты моделирования, можно отметить следующее:

1. Получена высокая степень сходства модели и эксперимента на всех периодах ОТ: $Q > 0.98$.

2. Реализация гласного характеризуется подъемом частоты ОТ (F_0).

3. Первые две форманты указывают на качество гласного [U].

4. Присутствие третьей форманты в диапазоне от 700 до 900 Гц указывает на назализацию гласного, что характерно для гласных, произнесенных ринолаликами.

5. Амплитуды первых двух формант, определяющих качество гласного, на порядок и более превосходят амплитуды двух других формант.

В результате экспериментальных исследований была подтверждена применимость предложенного метода моделирования звуков к анализу гласных звуков. Представляется перспективной возможность синтеза звуков на основе предложенной модели. Эксперименты с частотным составом формант, их амплитудами, а также характеристиками ударения могут дать результаты в количественном описании других аудиторных характеристик гласных звуков (звонкость, хриплость и т.д.).

Предложенная модель включает в себя как формантную структуру гласных, так и характеристику возбуждения и затухания звука для каждого периода основного тона. Сходство между моделью и оригиналом может быть оценено не только с помощью математических методов, но и посредством органов чувств (слуха). Получаемые объективные и субъективные оценки близости модели и оригинала практически совпадают и характеризуют адекватное отражение моделью процессов фонации гласных звуков. Кроме того, на основании предложенной модели может проводиться исследование влияние изменения параметров возбуждения–затухания, а также изменения формантных частот на аудиторные характеристики звука.

СПИСОК ЛИТЕРАТУРЫ

1. *Акбулатов А.Ш. и др.* Вокодерная телефония. Методы и проблемы. М.: Связь, 1974. С. 475–477, 480–481.
2. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. М.: Наука, 1984.
3. *Kouznetsev V.* Spectral dynamics and classification of Russian vowels. XI session of the Russion acoustical society. Moscow, Novembdr 19–23, 2001. P. 439–442.
4. *Vorobev V.I., Davydov A.G.* Syntesis of speech-like signals. XI session of the Russion acoustical society. Moscow, November 19–23, 2001. P. 457–460.
5. *Лейтес Р.Д., Соболев В.Н.* Цифровое моделирование систем синтетической телефонии. М.: Связь, 1969. С. 54–55, 58–59.
6. *Manley H.J.* Analysis-synthesis of connected speech in terms of ortohonalised exponentially damped sinusoids // JASA. 1963. V. 35. № 4. P. 464–474.
7. *Pinson E.N.* Pitch-synchronous Time Domain Estimation of Formant Frequencies Bandwidth. // JASA. 1963. V. 35. № 8. P. 1264–1273.
8. *Скляров О.П.* Нелинейная модель спектра голосового источника // Акуст. журн. 1997. Т. 43. № 5. С. 583–588.