РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА «ВОКАЛЬНЫЙ ТРЕНЕР»

А.А. Алябушев, А.И. Куликов, М.А. Карпушин, С.Г. Левин

Алябушев А.А.
e-mail: al_der@ngs.ru
Куликов А.И.
e-mail: kulikov@nmsf.sscc.ru
Институт вычислительной математики и математической геофизики СО РАН
Карпушин М.А.
e-mail: homeplaner@yandex.ru
Новосибирский государственный университет
Левин С.Г.
ООО «Сигнатек»
e-mail: levin@signatec.ru

Аннотация
В работе рассматривается подход, основанный на выделении базовых инвариантов в структуре акустического сигнала, вводится понятие обобщённого акустического ядра. Этот подход является ключевым при обработке сигналов в рамках реализации программного комплекса «Вокальный тренер». Программный комплекс представляет собой обучающую систему, реализующую принцип обратной связи, для улучшения голосовых данных, в том числе с использованием гиперграватационных тренировок.

Введение

Была поставлена задача разработать программно-аппаратный комплекс для тренировки голоса. Постановка предполагает создание специализированного программного обеспечения, которое при условии использования определённой аппаратной реализации может решать задачу обработки звуковых данных в реальном времени. Выяснилось, что эффективность ряда известных подходы к обработке звука основывается на потере части информации сигнала. Современные же вычислительные машины позволяют получить более точную и полную информацию, и возникает задача её более эффективного использования. Для этого требуется изучить структуру звукового сигнала. Это даёт также возможность решить задачу выделения и удаления составляющих композитного сигнала, аналогичную задаче создания минус-треков.

Задача интерактивного обучения пению требует создания в составе программного комплекса модулей мелодического, фонетического и ритмического анализа. Сама работа с вокальной речью имеет некоторые особенности по сравнению с анализом звука и голоса в общем случае. Кроме того, одной из целей создания комплекса является изучение влияния гиперграватационного воздействия на голос человека [1]. Все эти особенности были учтены при разработке системы.

Система сигнал - источник - приёмник.

Первым шагом стало исследование структуры звукового сигнала. Для этого следует зафиксировать несколько базовых понятий.

Сигнал служит для связи источника сигнала с приёмником сигнала. Вне этой связи сигнал не существует.

Под *источником* сигнала будем понимать объект, состояние которого отображается сигналом. Таким источником может быть голос человека или музыкальный инструмент, пейзаж или деталь, обрабатываемая на станке.

Физическая форма сигнала принципиального значения не имеет. Это может быть электрический сигнал в проводе, электромагнитный сигнал в радио эфире или файл на носителе.

Приёмником сигнала – может быть как орган человеческого восприятия, так и компонент технологического оборудования.

Сигнал может исходить от объекта непосредственно или же посредством некоторого датчика. Приёмник некоторым образом интерпретирует сигнал. Таким образом, в явном или неявном виде мы имеем кодирование сигнала. Для хранения сигнала или для его передачи по линиям связи кодирование обязательно. В рассматриваемых задачах цифровой обработки кодирование будет двоичным.

Если мы имеем кодированный сигнал, тогда перед использованием он должен быть декодирован.

Кодирование и сжатие.

Двоичный сигнал на входе канала может представлять собой поток данных, передаваемых с весьма большой скоростью. Хранение такого сигнала требует значительного объёма носителя. Это оправдывает применение алгоритмов кодирования и декодирования, уменьшающих объём передаваемых и хранимых данных.

В общем случае сигнал не может быть сжат. К примеру, невозможно сжать без потерь случайный сигнал. Для того, чтобы сигнал можно было сжать, необходимо использовать гипотезы, накладывающие ограничения на источник и (или) приёмник сигнала. Примеры:

- Сжатие изображений с использованием предположения, что достаточно малые отличия между соседними точками изображения глазу незаметны - *гипотеза относительно приёмника* сигнала, сжатие с потерями.
- Сжатие изображений с использованием предположения, что на однородных или градиентных участках числа, описывающие состояния соседних точек отличаются мало - *гипотеза относительно внешних свойств источника*, сжатие без потерь.
- Распознавание чёрного текста на белом фоне – позволяет игнорировать цветовой шум и использовать шаблоны (шрифт). Это *гипотеза относительно внутренних свойств источника*. Аналогично предположение, что звук издан голосовым аппаратом человека, позволяет использовать вокодер [2].

Последний подход даёт высокую степень сжатия, однако накладывает жёсткие ограничения на источник.

В целом, в практике кодирования сигналов различного рода, наблюдаются следующие закономерности.

Во-первых, более сильные гипотезы относительно приёмника дают большую степень сжатия за счёт удаления части сигнала, которая считается несущественной для его интерпретации приёмником. Здесь наилучшие результаты дают более точные модели восприятия – ярким примером этому является технология сжатия Mpeg Layer [3].

Во-вторых, более сильные гипотезы относительно внешних свойств источника позволяют получить большую степень сжатия сигнала за счёт ограничения пространства изменения сигнала и, таким образом, области применения. Это приводит к специализации алгоритмов сжатия.

Наибольший эффект дают гипотезы относительно внутренних свойств источника, поскольку многие параметры реальных источников звука меняются относительно медленно. К примеру – тембр гитары или рояля. Сжатие звука рояля можно свести к описанию его тембра и партитуры. Однако для сжатия звука симфонического оркестра придётся иметь модели всех инструментов оркестра. И, главное, *выделить* в звуке оркестра *звучание каждого инструмента*.

Для решения этой задачи можно предложить следующие гипотезы:

1. *Источник звука представляет собой совокупность конечного количества независимых объектов*. С учетом того, что звуки смешиваются аддитивно, вклад отдельных объектов можно рассматривать как независимый.

2. *Каждый источник имеет инвариантный на некотором отрезке времени набор свойств*. Это позволяет человеческому слуху выделять и удерживать в фокусе внимания такие объекты.

3. Существует *универсальный набор свойств* источников звука, частные реализации которого позволяют описать любой конкретный источник звука на некотором отрезке времени. Этот набор свойств следует формально описать.

4. Существует *процедура выделения этого набора* из акустического сигнала.

5. Существует процедура разделения акустического сигнала на конечное количество таких наборов.

Использование такой структуры, *обобщённого акустического ядра*, позволит сочетать эффективность сжатия и универсальность применения.

Применяя данный подход в рамках поставленной задачи, ограничим класс исследуемых сигналов. Сигнал вокальной речи характеризуется *высотой* – неким атрибутом слухового восприятия, находящимся в монотонной зависимости от его частотных свойств. Сигналы, которые можно так охарактеризовать, назовём *мелодическими*. Формальное определение мелодического сигнала будет следующим:

- сигнал является периодическим на рассматриваемом промежутке;
- величина, обратная периоду сигнала, лежит в диапазоне звуковых частот.

Обобщённое акустическое ядро.

Основная идея предлагаемого подхода состоит в математическом описании некоторого инвариантного набора свойств, удовлетворяющего указанным выше гипотезам.

Приведём конкретный пример реализации такого подхода.

Пусть F_s – частота дискретизации сигнала, и исследуемый отрезок сигнала содержит $2N$ отсчётов. Исследуется мгновенный дискретный спектр сигнала, поэтому N выбирается в соответствии с частотой дискретизации исходя из двух соображений:

- отрезок времени должен быть коротким, чтобы внутренние свойства источника сигнала (главным образом те, которые определяют частоту) изменились слабо за этот промежуток времени,
- для содержательности спектральной картины отрезок сигнала должен содержать достаточное количество отсчётов.

Обозначим через $X = \{X_n\}_{n=0}^{2N-1}$ результат дискретного преобразования Фурье выбранного отрезка исходного сигнала. В дальнейшем будут рассматриваться только первые $N+1$ коэффициентов спектра, поскольку ДПФ обладает известным свойством: коэффициент X_k для $k = 1,\ldots,N-1$ комплексно сопряжён с коэффициентом X_{2N-k} [2].

Пусть $\delta = \dfrac{F_s}{2N}$. Далее через *trunc(a)* будет обозначаться целая часть числа *a*, через *frac(a) – дробная часть*.

Размером набора частоты f называется величина

$$K_f = \max\left\{ i : trunc\left(\frac{if}{\delta}\right) \le N \right\}$$

Набором частоты f называется величина

$$M_f = \left\{ trunc\left(\frac{if}{\delta}\right) \right\}_{i=1}^{K_f}$$

Акустическим ядром частоты f *в мгновенном спектре* X назовём множество

$$A_f(X) = \left\{ X_{trunc\left(\frac{if}{\delta}\right)} \right\}_{i=1}^{K_f}$$

Построенную структуру можно наглядно проиллюстрировать следующим образом. Рассмотрим 46-миллисекундный отрезок записи исполнения *Арии из оркестровой сюиты №3 И.С. Баха* симфоническим оркестром. Частота дискретизации 44100 Гц. Вычисляется амплитудный спектр.

При рассмотрении распределения максимальных значений амплитуд гармоник на графике обнаруживается следующее: номера соответствующих спектральных коэффициентов разделяются на три группы так, что в каждой группе они оказываются равноудалёнными друг от друга. Это изображено на рис. 1, максимальные значения из разных групп обозначены маркерами различной формы в верхней части рисунка. Одновременное попадание одного максимума в несколько групп не исключается.

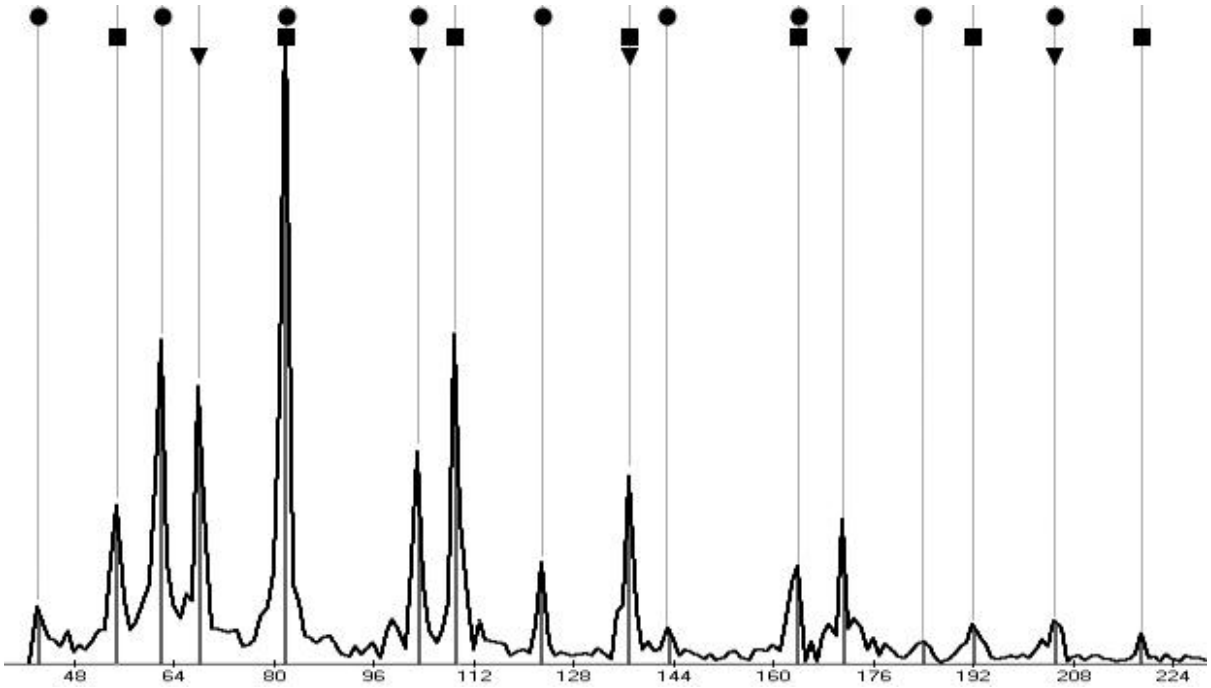

Рис. 1. Амплитудный спектр фрагмента музыкального произведения и группировка максимальных значений амплитуды.

Эти группы и будут представлять собой акустические ядра. Однако арифметически частоты указанных пиков не являются строго равноудалёнными. Это связанно с тем, что дискретный мгновенный спектр состоит из относительно небольшого числа отсчётов – 2048 в данном случае, и реальные значения частот не совпадают с узлами сетки, порождаемой дискретным преобразованием Фурье. Поэтому для анализа реальных сигналов полезно использовать обобщённые акустические ядра, в которые будут попадать не только гармоники с частотами, кратными базовой, но и несколько смежных, среди которых, с учётом дискретности, будут необходимые пики.

Обобщённым акустическим ядром степени $p \in N$ *частоты* f *в мгновенном спектре* X назовём множество

$$A_f^p(X) = \bigcup_{i=1}^{K_f} \left(\bigcup_{k=trunc\left(\frac{if}{\delta}\right)-trunc\left(\frac{p-1}{2}\right)}^{\min\left(trunc\left(\frac{if}{\delta}\right)+trunc\left(\frac{p}{2}\right),N\right)} \{X_k\} \right)$$

Множество M_f содержит те номера спектральных коэффициентов, расстояние между которыми в единицах частоты близко к f, K_f – это количество таких коэффициентов. Множество $A_f(X)$ содержит точки спектра с номерами в M_f, а $A_f^p(X)$ представляет собой окрестность таких точек из p спектральных коэффициентов. Нетрудно видеть, что $A_f^1(X) = A_f(X)$.

Алгоритм использования описанного аппарата в прикладных задачах состоит в переборе всех акустических ядер с базовой частотой из диапазона $[\varphi,\psi]$, специфицируемого для конкретной задачи, в спектре сигнала на коротком промежутке, и выборе тех из них, что удовлетворяют определённым условиям.

Для определения этих условий введём понятие *функции релевантности*. В качестве такой функции может выступать любая функция вида $r\colon [\phi,\ \psi] \times \mathbb{R}_+^{n+1} \to \mathbb{R}$, которая достигает своего локального максимума для акустических ядер, генерируемых источниками мелодического сигнала. Аргументами функции являются частота f акустического ядра и коэффициенты $A_0,\ldots,A_N$ мгновенного энергетического спектра. Конкретный вид функции реализует некоторые гипотезы относительно особенностей анализируемого источника (источников) сигнала и деталей постановки задачи и может быть уточнён экспериментально.

Далее приводятся основные функции релевантности, которые были опробованы авторами к моменту написания статьи.

Первый пример – функция релевантности, в дальнейшем называемая амплитудной.

$$r_A(f, A_0, \ldots, A_N) = \frac{1}{K_f}\sum_{i=1}^{K_f}\left(A_{trunc(\frac{if}{\delta})}\left(1 - frac(\frac{if}{\delta})\right) + A_{trunc(\frac{if}{\delta})+1}frac(\frac{if}{\delta})\right)$$

Эта функция представляет собой среднее арифметическое величин, равных взвешенной сумме смежных коэффициентов энергетического спектра. Вес выбирается из соотношения частотных расстояний от точки if до соответствующих точек δk и $\delta k+1$. Амплитудная функция релевантности даёт хороший результат при анализе мелодического сигнала от одного источника.

В случае, когда источников несколько, применяется другая функция релевантности, более адаптированная к особенностям структуры сигналов от музыкальных инструментов. В дальнейшем она называется линейно взвешенной функции релевантности.

$$r_l(f, A_0, \ldots, A_N) = \frac{1}{K_f}\sum_{i=1}^{K_f} if\left(A_{trunc(\frac{if}{\delta})}\left(1 - frac(\frac{if}{\delta})\right) + A_{trunc(\frac{if}{\delta})+1}frac(\frac{if}{\delta})\right)$$

Возникновение под знаком суммы весового множителя, равного частоте спектрального коэффициента, происходит из следующего наблюдения. Энергетический спектр реального звукового сигнала в области высоких частот в среднем принимает значительно меньшие значения, чем в других частотных областях. Однако при исследовании спектров звучания некоторых музыкальных инструментов можно обнаружить, что свойство экстремальности акустического ядра не нарушается и в этой области частот, т.е. равноудалённые максимальные значения амплитуд гармоник присутствуют и там, лишь с той разницей, что порядки этих значений много меньше, чем в области средних и низких частот. Поэтому, если учитывать «вклад» спектральных коэффициентов из высокочастотного диапазона с возрастающим весом, можно точнее обнаруживать источник и локализовать его высоту. Такое предположение полностью подтвердилось на практике.

Наряду с линейно взвешенной функцией релевантности возникает ряд других функций, основанных на том же предположении. В дальнейшем такие функции называются взвешенными с весом $p : \mathbf{R} \to \mathbf{R}$ и имеют вид

$$r_p(f, A_0, \dots, A_N) = \frac{1}{K_f}\sum_{i=1}^{K_f} p(if)\left(A_{trunc(\frac{if}{\delta})}\left(1 - frac(\frac{if}{\delta})\right) + A_{trunc(\frac{if}{\delta})+1} frac(\frac{if}{\delta})\right).$$

Наиболее результативными вместе с линейно взвешенной функцией релевантности оказываются также пропорционально логарифмически взвешенная функция релевантности ($p(x) = x\ln(x)$) и линейно взвешенная функция релевантности со срезом частоты, для которой

$$p(x) = \begin{cases} x, x < f_{cp} \\ 0, x \geq f_{cp} \end{cases}$$

Здесь f_{cp} – значение частоты среза, т.е. такая частота сигнала, выше которой искомый источник никак себя не проявляет. Эта величина выбирается с ориентиром на искомый источник и позволяет с меньшим числом ошибок детектировать его высоту. Например, для некоторых традиционных вариантов звучания электрогитары f_{cp} принимает значение порядка 12 кГц.

Пример использования метода акустических ядер.

Одной из реализации предложенного метода является программное средство подбора партии солирующего инструмента.

Входными данными для программного средства выступают фрагмент сигнала в формате PCM WAV (частота дискретизации 44100 Гц, 16 бит на один отсчёт, один канал) и следующие настраиваемые параметры:

- границы частотного диапазона,
- размер фрагмента анализа,
- номер используемой функции релевантности,
- показатель разрежения.

Входной сигнал разбивается на фрагменты указанной длины, расположенные по порядку друг за другом. На основном этапе алгоритма производится последовательный анализ всех фрагментов

Автор или исполнитель	Композиция	Инструмент	Фрагмент, сек	Размер блока	Диапазон частот, Гц	Функция релевантности
B. Greb	Grebfruit	Голос	0:15-0:45	2048	90-300	Линейно взвешенная со срезом на 11 кГц
B. McFerrin	Don’t Worry Be Happy	Голос	0:00-0:28	2048	100-300	Линейно взвешенная со срезом на 10 кГц
J. Petrucci	Wishful Thinking	Электрогитара	0:02-0:27	2048	300-1000	Линейно взвешенная со срезом на 11 кГц
W. Hoffmann	Solveig’s Song	Электрогитара	0:54-1:10	2048	100-400	Линейно взвешенная
Л. В. Бетховен	Fur Elise	Фортепиано (партия правой руки)	0:00-0:20	4096	200-900	Пропорционально логарифмически взвешенная
И. С. Бах	Ария из оркестровой сюиты №3, Ре Мажор	Скрипка	0:00-0:36	4096	300-1000	Пропорционально логарифмически взвешенная
S. Vai	Building the Church	Электрогитара	0:00-0:10	256	250-600	Амплитудная
S. Vai	Building the Church	Электрогитара	0:19-0:40	2048	180-360	Линейно взвешенная со срезом на 12 кГц
T. Emmanuel	Footprints	Акустическая гитара	0:00-0:19	2048	220-700	Линейно взвешенная

Таблица 1. Таблица выбранных композиций и параметров тестирования программного средства подбора партий

Результатом работы является последовательность значений частоты для каждого фрагмента. Для удобства представления и редактирования последовательность представляется в виде множества пар частота-длительность, т.е. одинаковые подряд идущие значения частоты объединяются в один продолжительный блок.

Для тестирования был выбран ряд разнохарактерных музыкальных композиций. Основным критерием выбора композиций служило наличие в ней одноголосой мелодии с как можно меньшим числом пауз между нотами, исполняемой одним инструментом (голосом) либо группой инструментов и отчётливо различимой от остальных мелодий. В список были включены также и другие мелодии, в той или иной степени неудовлетворяющие такому критерию.

В большинстве композиций удалось подобрать партию солирующего или другого инструмента с незначительным числом ошибок, частично устраняемых автоматическими средствами. Эти примеры с указанием параметров и комментариями приведены в таблице 1.

Во всех примерах для осуществления подбора с наименьшим числом ошибок требуется как можно точнее определить все параметры процесса. Поиск нужных значений происходит опытным путём и занимает достаточно много времени.

Очевидно, задача получения минус-треков – звуковых сигналов, из которых удалено звучание некоторых источников (инструментов) – является в некотором смысле обратной к рассмотренной, и позволяет применять для её решения схожий процесс [4].

Анализ вокальной речи

В ходе экспериментов были сделаны записи продолжительных (5-7 секунд) произнесений гласных и сонорных звуков (А, О, У, Ы, И, Э, Л, М, Н) с различной высотой. В экспериментах участвовало 4 диктора: 3 девушки и один мужчина.

Записи делались с частотой дискредитации 44100 Гц, глубина буфера 16 бит. Далее строился дискретный спектр по 2048 отсчетам, в полученном спектре выделялись акустические ядра.

Наибольший интерес представляет акустическое ядро частоты, соответствующей высоте произносимого звука. На этапе экспериментов эта высота определялась на слух: дикторы произносили звуки с высотой, соответствующей какой-либо ноте. Частоты, соответствующие нотам, известны.

На рис. 2 приведён пример такого спектра. Здесь девушка исполняет ноту «До» перовой октавы (частота 261,6 Гц), пропевая при этом звук «а». На графике видны пики большой амплитуды на частотах, кратных указанной – они как раз и составляют обсуждаемое акустическое ядро.

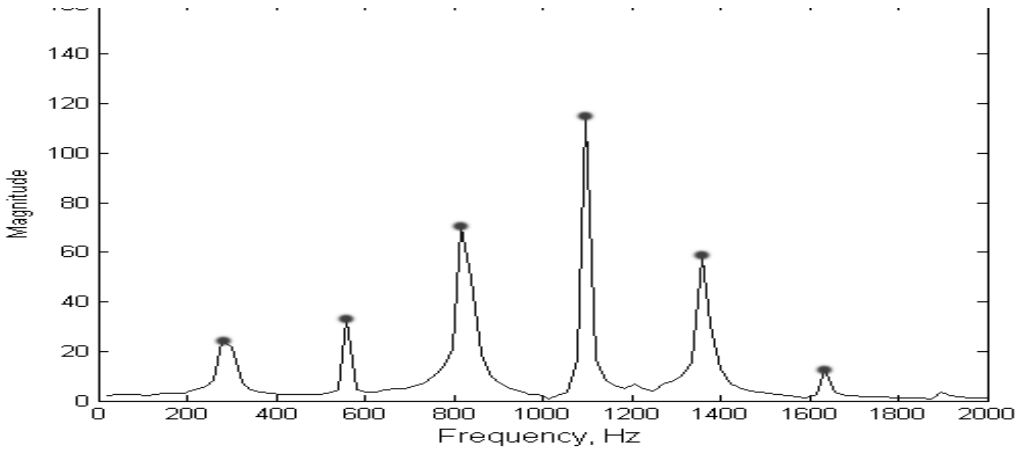

Рис.2. Пример спектра вокальной речи.

Посчитаны отношения между последовательными значениями энергий гармоник в полученных спектрограммах. Было замечено, что при произнесении одного звука (фонемы) одним диктором эти отношения сохраняются. По выявленным соотношениям можно определить, какой звук был произнесён. Полученные данные усреднялись, причем дисперсия в некоторых случаях достигала 30%.

Выявленное свойство оказалось полезным для анализа вокальной речи в сочетании с известными методами анализа речевых сигналов [4].

Программный комплекс «Вокальный тренер»

Разработанные методы используются в реализации проекта ”Вокальный тренер”. В качестве основной методики обучения вокалу, используемой в системе, выбран курс Сэта Риггза. Методика включает в себя комплекс упражнений, описанных в публикации её автора [5]. В оригинале курса упражнения представляют собой мелодии, исполненные на фортепиано, которые обучаемый должен повторить голосом. В реализации эти мелодии сохранены в виде набора параметров, получаемых методом акустических ядер. Порядок выполнения упражнений предлагается пользователю системой согласно методике, с учётом истории выполнений и прогресса обучаемого.

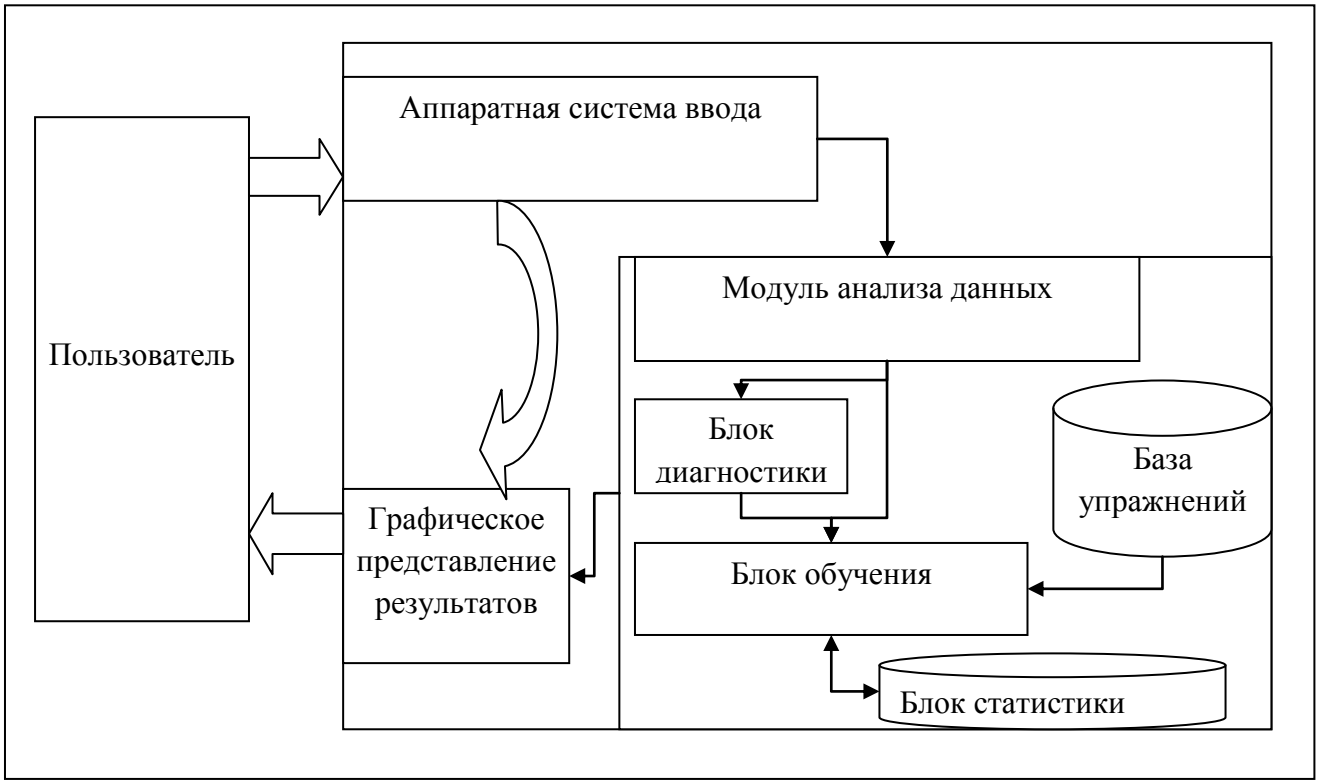

Рис.3. Схема работы комплекса

Функциональная схема комплекса представлена на рис. 3. Программный комплекс реализует принцип обратной связи. Эффективная обратная связь достигается за счёт визуализации результатов анализа сигнала, поступающего в систему через микрофонный вход звуковой карты.

Сам анализ распадается на контроль точности воспроизведения мелодической и речевой составляющей голосового сигнала, а так же его ритмических характеристик.

При мелодическом анализе сигнал понимается как мелодический, и его высота (т.е. частотная характеристика) сравнивается с требуемой в рамках упражнения. В реализации такой анализ выполняется методом акустических ядер.

При фонетическом анализе вокальной речи важным является определение участков произнесения гласных звуков, как несущих в себе основную информацию о высоте. Для распознавания гласных в речи используется формантный анализ [6, 7].

Ритмический анализ предполагает следующее. Во-первых, в каждом упражнении содержится информация о времени начала воспроизведения той или иной ноты. Система определяет момент фактического начала исполнения ноты, сравнивает эти значения и сохраняет результат для дальнейшего предоставления пользователю. Во-вторых, контролируется продолжительность исполнения ноты. Соответствующая информация так же сохраняется и предоставляется пользователю.

Упражнения, сохраняемые в базе, представляют собой набор параметров сигнала, который должен воспроизвести голосом обучаемый. База упражнений наполняется разработчиками. За сравнение текущих параметров с требуемыми отвечает блок обучения, в котором так же принимаются решения о дальнейших действиях системы в зависимости от результатов сравнения.

В блок диагностики заносятся данные о соответствии параметров сигнала состоянию голосового аппарата человека. Параметры текущего сигнала поступают так же сюда, и на их основе блок диагностики может управлять условиями выполнения упражнений в блоке обучения.

В блоке статистики сохраняется информация о выполнении упражнений и состоянии голоса каждого обучаемого. Это позволяет отслеживать изменение голоса со временем, а так же влияет на выбор текущих упражнений блоком обучения.

Использование гипергравитационных тренажёров при тренировке голоса

В рамках исследований, проводимых в Самарском Государственном Медицинском Университете (СамГМУ) было показано, что кратковременное импульсное гипергравитационное воздействие тренажёра Power Plate на тело человека вызывает динамическое изменение скорости воздушного потока в дыхательных путях с частотой 30-40 Гц. Были выявлены реакции релаксации гладких мышц мелких дыхательных путей в этих условиях. С участием профессиональных певцов было показано положительное влияние гипергравитационной физической нагрузки на голос [1].

Создание комплекса «Вокальный тренер» ведётся в сотрудничестве с кафедрой нормальной физиологии СамГМУ, со стороны которой предлагаются упражнения для развития голоса, выполняемые на тренажёре Power Plate.

Был поставлен вопрос об обработке голосовых данных, получаемых во время гипергравитационной тренировки. Для его решения проведены эксперименты по анализу голоса человека при такой нагрузке и в покое. Дикторы в обозначенных условиях произносили одни и те же звуки. Было показано, что в спектрах сигналов, записанных при использовании тренажёра, амплитуды гармоник больше, чем у соответствующих гармоник в спектрах сигналов, записанных в покое, а так же имеется значительный пик на частоте стимуляции. При этом общее соотношение амплитуд гармоник сохраняется, и результаты применяемых методов анализа (например, выделение акустического ядра в спектре голоса) остаются адекватными.

Заключение

Предложенная выше технология, связанная с разработкой программного комплекса «Виртуальный тренер», является достаточно универсальной и позволяет решать в рамках единого подхода задачи различной направленности.

Примерами таких задач являются: выделение голоса из шума, обнаружение “заданных” источников звука, непосредственно сжатие звуковой информации и т.д. Соответственно, конкретные методы использования структуры акустического ядра могут быть различны.

В рамках программного комплекса реализуется система обратной связи, позволяющая организовать эффективное обучение владению голосом. Этот же подход, путём создания специальных наборов упражнений, может быть применён как для обучения пению, так и в логопедии и обучении иностранным языкам.

Ещё одно направление разработок – изучение влияния гиперграватационных тренировок на голос человека. Разработанные подходы позволяют создать программный инструментарий для проведения исследований, связанных с физиологией голосового аппарата [4].

ЛИТЕРАТУРА

1. Пятин В.Ф., Широлапов И.В. Однократная вибрационная нагрузка значительно увеличивает скорость экспираторного воздушного потока у человека // Вестник ТГУ. Серия «биология и экология», 2009. №1, С. 38–42.

2. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ./ Под ред. М.В. Назарова и Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 496 с.

3. Brandenburg K., Popp H. An introduction to MPEG Layer-3 // EBU Technical Review. – 2000.

4. Алябушев А.А., Карпушин М.А., Кузьмин А.В., Куликов А.И., Левин С.Г. Сравнительный анализ голосовых данных человека при импульсном гиперграватационном воздействии и в покое // Ершовская конференция по информатике, Труды НПО 20011, Рабочий семинар «Наукоемкое программное обеспечение», Новосибирск, 27 июня – 1 июля, 2011, С. 19-24.

5. Риггз С. Пойте как звёзды // Сост. и ред. Дж.Д.Каррателло.— СПб.: Питер, 2007.— 120 с.

6. Аграновский А.В., Леднов Д.А., Репалов С.А. Метод текстонезависимой идентификации дикторов на основе индивидуальности произношения гласных звуков. — Акустика речи и прикладная лингвистика. Ежегодник Российского акустического общества. Выпуск 3. М., 2002, с. 103–115.

7. Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных. // Информационные процессы. – 2004. - №2, том 4 - с. 202-220.