Электронный журнал «Техническая акустика»
http://www.ejta.org

2008, 15

В. В. Митянок

Полесский государственный университет, г. Пинск, Беларусь.
225710, г. Пинск, ул. Днепровской флотилии, 23. e-mail: mitsianok@mail.ru

Метод аппроксимации для определения числовых характеристик некоторых низкочастотных звуков человеческой речи

Получена 15.05.2008, опубликована 04.08.2008

Предложен метод распознавания звуковых сигналов, в основе которого лежит идея аппроксимации (почти) периодической функции набором мод с медленно (по сравнению с несущими частотами) меняющимися амплитудами. Метод применен к анализу 8 низкочастотных звуков человеческой речи, полученных от 8 респондентов. Создана база данных по амплитудам отдельных мод и на ее основе разработана система идентификации произнесенных звуков. Тестовые испытания созданной системы показали уровень распознаваемости звуков в 85–95 процентов.

Ключевые слова: распознавание речи, аппроксимация

ВВЕДЕНИЕ

Несмотря на многочисленные усилия на протяжении почти 50 лет, задача уверенного распознавания компьютером человеческой речи до сих пор не решена. Не решена также задача распознавания компьютером человека по его голосу, хотя известно, что хорошо знакомые между собой люди легко узнают один другого при разговоре по телефону. Тот факт, что к настоящему времени уже разработаны и используются так называемые дикторозависимые программы, которые достаточно уверенно распознают голос знакомого диктора, а голос незнакомого - почти не распознают, вовсе не означает, что имеет место распознавание диктора. Хорошая программа распознавания должна распознавать речь *любого* человека *без* предварительной тренировки. А если имеются образцы голоса диктора, то и идентифицировать его.

Хотя в целом проблема еще не решена, имеется ряд несомненных достижений. Полученные к настоящему времени достижения и еще нерешенные проблемы весьма полно изложены в [1]. В [2] можно найти обширный перечень направлений, тем и сопутствующих вопросов, имеющих отношение к проблеме распознавания звуков человеческой речи.

В настоящее время для распознавания речи чаще всего изучаются наборы различных признаков, характеризующих звуки, с использованием статистического анализа [3]. Используются уровни в спектральных полосах, формантные признаки, коэффициенты

линейного прогноза [3–5]. В [4] выдвинуто предположение об определяющей роли соотношения уровней мощности в определенных спектральных полосах речевого сигнала. Но ни один из возможных наборов признаков не имеет явного преимущества по сравнению с другими [5]. Поэтому некоторые из наборов используются совместно. Так, в [6] для идентификации звуков используются параллельно формантные и полосные признаки. Заинтересованный читатель может найти подробное освещение вопроса в [7].

Во многих системах распознавания речи требуется раздельное произнесение слов, с паузой между ними не менее 150 мс [8]. Недавно появились системы NaturallySpeaking фирмы Dragon Systems и ViaVoice фирмы IBM, не требующие таких пауз, но этот успех достигнут не за счет улучшенной математической проработки задачи, а за счет совершенствования компьютеров. Эти программы на первом этапе анализируют звук, чтобы отличить низкочастотные гласные от высокочастотных согласных. После этого результаты сравнивают с фонемами, и по результатам сравнения и происходит распознавание звука. Эти программы являются дикторозависимыми, то есть перед первым использованием программа должна «привыкнуть» к голосу диктора [9].

Надо полагать, что медлительность продвижения в задаче распознавания звуков связана с недостаточной математической проработкой задачи, желанием добиться скорейшего финансового успеха, что приводит к тому, что многие, существенные для распознавания звуков математические особенности звуковых кривых, до сих пор остаются неизученными. В этой связи особенно интригующим выглядит отмеченное в [10] обстоятельство, что «речевой сигнал достаточно хорошо воспринимается человеком даже в очень узкой полосе частот, причем расположенной в любой части речевого диапазона. Это свойство совершенно не соответствует механизмам обработки речи, принятым в системах автоматического распознавания». Совершенно очевидно, что без ответа на вопрос, почему это имеет место (а также на ряд других вопросов), дальнейший прогресс в решении проблемы распознавания звуков невозможен.

Определенные перспективы для распознавания речи имеет активно развиваемый в последнее время метод вейвлет-анализа (см., например, [11] и указанную там литературу), который позволяет находить тонкие особенности сигналов любого происхождения, очищать сигналы от шумов [12]. Однако этот метод имеет органический недостаток — соответствующие расчеты требуют непомерно больших затрат машинного времени. По этой причине представляется маловероятным, что реальное распознавание речи одного человека другим происходит с привлечением вейвлет-анализа.

Эти обстоятельства стимулируют проведение дальнейших исследований с целью выяснения тех *математических* особенностей звуковых кривых, которые существенны для распознавания как речи, так и говорящего. Задача настоящей статьи — изложение метода определения числовых характеристик звуков, основанного на решении задачи аппроксимации в противовес широко используемому преобразованию Фурье. Отметим, что во впечатляющем перечне, представленном в [2], возможности применения аппроксимации не упоминаются вообще.

1. ПРЕОБРАЗОВАНИЯ ФУРЬЕ

Большинство существующих систем распознавания человеческой речи основано на разложении звуков в спектр преобразованием Фурье [13]. Применительно к распознаванию реальных звуков здесь можно выделить две серьезные проблемы.

Первая — определение длительности отрезка звуковой кривой, отобранного для изучения. Вторая — нестрогая периодичность звуковой кривой, даже при произнесении долгих гласных звуков. Так, известно, что даже идеальная гармоническая кривая (сигнал)

$$y(t) = \sin(\omega_0 t), \tag{1}$$

будучи подвергнутой Фурье-преобразованию на ограниченном интервале изменения независимой переменной (времени) *t*, приводит к размазанному спектру, содержащему кроме главного, также и бесконечно большое количество побочных (фальшивых, то есть не соответствующих наличию излучения с соответствующей частотой) максимумов.

В самом деле, синус-преобразование функции (1) на симметричном интервале времени от $-t_0$ до $+t_0$ в соответствии с выражением

$$C_s(\omega) = \int_{-t_0}^{t_0} y(t)\sin(\omega t)dt \tag{2}$$

приводит к результату

$$C_s(\omega) = \frac{\sin((\omega_0 - \omega)t_0)}{(\omega_0 - \omega)} - \frac{\sin((\omega_0 + \omega)t_0)}{(\omega_0 + \omega)}. \tag{3}$$

(Косинус-преобразование здесь дает тождественный нуль). Анализируя выражение (3), нетрудно заметить, что зависимость $C_s(\omega)$ содержит бесконечное число локальных максимумов, главный максимум имеет место при $\omega = \omega_0$, остальные же являются побочными. По мере увеличения длительности сигнала (t_0) побочные максимумы сгущаются к главному, и спектр (3) становится все более четко выраженным.

Таким образом, с побочными максимумами можно бороться путем увеличения длительности сигнала, однако применительно к реальному звуку такое увеличение приводит к усугублению проблемы нестрогой периодичности. Дрейф частоты, нестабильность амплитуд практически не сказываются на спектре коротких сигналов, но, по мере роста их длительности, вносят все большие искажения в спектр.

Проиллюстрируем это обстоятельство графически на одном частном, но показательном примере. Пусть преобразованию Фурье подвергается сигнал с качающейся несущей частотой, описываемый выражением

$$y_i = \sin(i(0{,}1 + 0{,}01\sin(1 + 0{,}002i))). \tag{4}$$

Качания частоты сигнала (4) происходят вокруг ее среднего значения 0,1 с амплитудой 0,01 и частотой качаний, равной 0,002. Пусть длительность сигнала составляет 2000 точек ($i = 1…2000$). Преобразование Фурье сигнала (4) приводит к спектру, график которого показан ниже

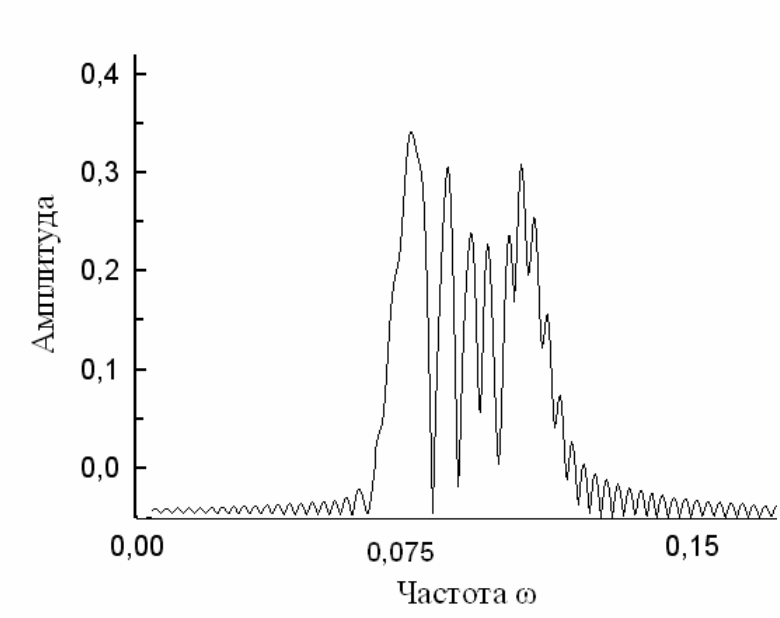

Рис. 1. Преобразование Фурье (спектр) сигнала с качающейся несущей частотой

Глядя на рис. 1, трудно распознать в сигнале наличие всего лишь одной, пусть даже и качающейся, частоты. Увеличение же длительности сигнала (4) приводит лишь к тому, что количество максимумов на графике спектральной кривой возрастает, они становятся все более резкими, и разобраться в этой картине становится все более затруднительным.

Все это, вместе взятое, означает, что при разложении в спектр преобразованием Фурье, отрезок реальной звуковой кривой, подлежащей анализу, должен быть с одной стороны, не слишком коротким, а с другой стороны — не слишком длинным. То есть, должен быть найден компромисс по длительности. Но, каким бы ни было компромиссное решение, распознаванию звуков будут мешать с одной стороны, как слишком малая, так, с другой стороны, слишком большая их длительность.

Отмеченные выше трудности являются принципиальными для применения преобразований Фурье. Но их можно избежать при ином подходе к проблеме. А именно, вместо разложения сигнала в спектр преобразованием Фурье следует решать задачу аппроксимации. При таком подходе длительность сигнала вообще не имеет никакого значения, а нестабильность параметров мод вносит лишь небольшие искажения в спектр. Рассмотрим этот подход подробнее.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть требуется аппроксимировать функцию $y_i = y(x_i)$, заданную своими значениями в избранных точках x_i, с помощью набора гармонических функций. Для упрощения положим $x_i = i$. С целью аппроксимации сначала рассмотрим невязку (функционал)

$$S = \sum_{i=1}^{n} (y_i - b_0 - \sum_{k=1}^{l_1} a_k \sin(\omega_k i) - \sum_{k=1}^{l_1} b_k \cos(\omega_k i))^2 , \quad (5)$$

где l_1 — количество аппроксимирующих гармоник (мод), ω_k — их частоты (считаются известными), n — количество используемых точек кривой звукового давления, a_k, b_k — соответственно амплитуды синус — и косинус мод, b_0 — нуль (среднее значение, начало отсчета) звуковой кривой. Индекс i нумерует те значения аргумента звуковой кривой, для которых известно звуковое давление.

Минимизация (5) по a_k, b_0, b_k, где $k = 1...l_1$, и последующее решение получившейся системы линейных алгебраических уравнений относительно a_k, b_0, b_k приводит к решению задачи аппроксимации. В частности, если аппроксимируемая функция y_i является суммой нескольких гармоник с постоянными амплитудами, частотами и фазами (причем высшие частоты необязательно кратны низшей), то, после решения задачи аппроксимации и нахождения амплитуд и начала отсчета и подстановки решения в (5) невязка достигает теоретического минимума, равного нулю, при совпадении наборов частот аппроксимируемой и аппроксимирующих функций. Причем *независимо от длительности сигнала*.

Таким образом, если известны частоты гармоник, то методом аппроксимации можно произвести разложение их суммы на исходные составляющие. Если же при такой же аппроксимируемой функции частоты аппроксимирующих функций известны с некоторой ошибкой, то, конечно, после решения задачи аппроксимации и подстановки решения в (5) остаточная невязка нулю не равна, но в этом случае существует следующая возможность. Для разложения сигнала можно выбрать несколько разных наборов частот, для каждого из них провести аппроксимацию, подсчитать остаточные невязки и в качестве «более правильного» выбрать тот набор частот, который обеспечивает наименьшую невязку. Осуществляя последовательный подбор (дрейф) частот, можно найти истинные частоты как частоты, обеспечивающие остаточную невязку, меньшую по сравнению с невязкой, соответствующей любому другому набору частот.

Описанный подход не вполне оправдан в тех случаях, когда по каким-то причинам параметры аппроксимируемой функции медленно (и, тем более, не медленно) зависят от времени. Это может случиться, например, в задачах распознавания реальных звуков, произнесенных человеком, хотя бы потому, что в силу физиологических причин голос человека слегка дрожит. Кроме того, при переходе от звука к звуку можно предвидеть

изменение как значений частот, так и их амплитуд, фаз. Если звуковую кривую обрабатывать по вышеописанной методике, то следует ограничиваться лишь фрагментами звуковых кривых небольшой длины, так как при большой длине фрагмента вызванные дрожаниями искажения накапливаются и в итоге приводят к значительным погрешностям в спектре звука.

Однако можно предвидеть, что при распознавании человеческой речи придется иметь дело и с достаточно длительными сигналами. Поэтому примем, что амплитуды мод могут медленно (по сравнению с частотами ω_k) меняться со временем. (Что касается медленного изменения частот и фаз, то они могут путем тривиальных тригонометрических преобразований трансформироваться в медленные изменения тех же амплитуд). Соответственно, вместо (5) рассмотрим невязку в виде

$$S = \sum_{i=1}^{n}(y_i - b_{0i} - \sum_{k=1}^{l_1} a_{ki} sin(\omega_k i) - \sum_{k=1}^{l_1} b_{ki} cos(\omega_k i))^2 + \alpha \sum_{i=1}^{n-1}(b_{0,i} - b_{0,i+1})^2 + \\ \alpha \sum_{k=1}^{l_1}\sum_{i=1}^{n-1}(b_{k,i} - b_{k,i+1})^2 + \alpha \sum_{k=1}^{l_1}\sum_{i=1}^{n-1}(a_{k,i} - a_{k,i+1})^2. \quad (6)$$

В отличие от (5), в (6) амплитудам и нулю звуковой кривой разрешено меняться от точки к точке, соответственно у них появился второй индекс (*i*), указывающий на номер точки на кривой звука. Однако чтобы эти изменения не были слишком резкими, в невязку вводится группа членов, учитывающая скачки амплитуд и дрейфующего нуля от точки к точке. Неотрицательное число α показывает относительную важность различных слагаемых. Если $\alpha = 0$, то можно добиться идеальной аппроксимации звуковой кривой, правда, ценой того, что амплитуды и нуль аппроксимирующей функции могут оказаться резко меняющимися от точки к точке. Если же $\alpha = \infty$, то амплитудам запрещено изменяться, что фактически эквивалентно (5) и что во многих случаях не соответствует физическому содержанию задачи. Истина, как правило, находится где-то посередине. Поэтому α будем выбирать так, чтобы вклад в остаточную невязку (6) от членов содержащих, и от членов не содержащих α, был примерно одинаков. Несколько забегая вперед, отметим, что для этого лучше всего выбрать $\alpha = 10$, но варьирование α в интервале (1…100) практически не сказывается на окончательных результатах исследований.

Предположим, что набор частот ω_k известен. Для нахождения минимума невязки (6) требуется приравнять нулю производную *S* по каждой из независимых переменных a_{ki}, b_{0i}, b_{ki} где $k = 1...l_1$, $i = 1...n$. Таким образом, общее число уравнений равно

$$n_1 = (2l_1 + 1)n. \quad (7)$$

Дифференцирование (6) по $b_{0,l}$ и приравнивание производной нулю приводит к уравнениям

$$b_{0l}(1+2\alpha)-\alpha b_{0,l-1}-\alpha b_{0,l+1}+\sum_{k=1}^{l_1}a_{kl}\sin(\omega_k l)+\sum_{k=1}^{l_1}b_{kl}\cos(\omega_k l)=y_i,\ l=1\ldots n. \quad (8)$$

Дифференцирование по $a_{m,l}$ — к уравнениям

$$\sum_{k=1}^{l_1}a_{k,l}\sin(\omega_k l)\sin(\omega_m l)+\sum_{k=1}^{l_1}b_{k,l}\cos(\omega_k l)\sin(\omega_m l)+2\alpha a_{m,l}-\alpha a_{m,l-1}-\alpha a_{m,l+1}=y_l\sin(\omega_m l), \quad (9)$$

где $m=1\ldots l_1, l=1\ldots n$.

Дифференцирование по $b_{m,l}$ — к уравнениям

$$\sum_{k=1}^{l_1}a_{k,l}\sin(\omega_k l)\cos(\omega_m l)+\sum_{k=1}^{l_1}b_{k,l}\cos(\omega_k l)\cos(\omega_m l)+2\alpha b_{m,l}-\alpha b_{m,l-1}-\alpha b_{m,l+1}=y_l\cos(\omega_m l), \quad (10)$$

где $m=1\ldots l_1, l=1\ldots n$.

3. ПРИМЕНЕНИЕ К РАСПОЗНАВАНИЮ РЕАЛЬНЫХ ЗВУКОВ

Система уравнений (8)–(10) — суть система линейных алгебраических уравнений относительно амплитуд мод и дрейфующего нуля, и методы решений таких систем хорошо известны. При практическом применении вышеописанной методики изучались звуки Л, М, Н, А, О, Ы, У, Э, образцы которых были получены от 4-х мужчин и 4-х женщин. Особенность исследования состояла в том, что звуки произносились респондентами в составе слогов. При записи образцов респондент произносил слово, составленное из 8 одинаковых слогов. Каждый слог состоял из одного согласного звука и одного гласного — всего 15 вариантов слогов. Полученные звуковые кривые слов расщеплялись на отдельные моды в соответствии с вышеописанной методикой, затем отдельные моды просматривались и визуально определялись те участки, где их характер принципиально менялся соответственно смене произносимого звука. Таким способом было проведено выделение отдельных звуков из слова. Не каждая из мод была пригодна для этой цели, однако для расщепления слова на звуки достаточно было найти лишь одну, наиболее характерную, моду. Список таких мод, наиболее удачно подходящих для разделения слов на звуки приведен ниже в табл. 3. После разделения слова на звуки каждый из участков выделенного звука обрезался с каждой стороны на 20 процентов своей длины с целью нивелирования переходных явлений.

Цель этой части исследования состояла в нахождении интенсивностей мод для каждого из изучаемых звуков и нахождении их средних значений и средних квадратичных отклонений. Так как человек обладает способностью распознавать речь независимо от ее громкости в широком диапазоне интенсивностей, то усредненные по всем образцам интенсивности мод следует нормировать на громкость звука. Именно нормированные величины подвергались математической обработке.

Рассматривались разные варианты количества мод — от 6 до 13 и разные варианты наборов их частот. Во всех случаях частоты ω_k каждого из наборов выбирались

равноотстоящими, рассматривалось 11 вариантов выбора нижней (базовой) частоты, все остальные частоты были кратны базовой. Ниже все численные значения частот даются в долях от частоты $\omega_0 = \frac{192000}{2\pi} \approx 30558\,\text{Гц}$, условно принятой за единицу. Базовая частота табулировалась в интервале от 0,0058 до 0,0068 с шагом 0,0001. Оказалось, что при переборе базовых частот в указанном выше диапазоне результаты менялись незначительно, изменения результатов были меньшими, нежели их среднее квадратичное отклонение в пределах одного любого звука, полученного от любого из респондентов.

Далее, при смене респондента полученные данные по относительным интенсивностям мод, приведенных в таблице 3 (см. ниже), менялись на величину, сопоставимую с их естественным среднеквадратичным отклонением. Поэтому, если стоит задача распознавания речи независимо от диктора, то имеет смысл усреднить полученные результаты по всем респондентам. Напротив, интенсивности некоторых из мод, не входящих в таблицу 3, при смене респондента менялись весьма заметно. Вероятно, изучение именно таких мод позволит решить проблему идентификации человека по его голосу.

Приведем таблицу относительных интенсивностей мод, имеющих место при произнесении 8 звуков, усредненных по всем образцам и по всем респондентам при количестве мод, равном 13 и при базовой частоте, равной 0,0058.

Таблица 1. Интенсивности 13 мод для 8 звуков, нормированные на сумму интенсивностей мод и усредненные по всем образцам и по всем респондентам

	А	О	Ы	У	Э	Л	М	Н
1	0,207	0,279	0,601	0,664	0,271	0,445	0,808	0,778
2	0,104	0,191	0,174	0,179	0,119	0,347	0,111	0,138
3	0,116	0,197	0,022	0,072	0,208	0,053	0,013	0,020
4	0,224	0,167	0,009	0,039	0,084	0,065	0,019	0,009
5	0,069	0,109	0,007	0,012	0,031	0,042	0,010	0,006
6	0,131	0,020	0,006	0,004	0,026	0,010	0,003	0,005
7	0,082	0,008	0,010	0,003	0,027	0,005	0,004	0,005
8	0,022	0,004	0,010	0,002	0,038	0,003	0,006	0,005
9	0,012	0,004	0,016	0,002	0,068	0,002	0,004	0,003
10	0,009	0,003	0,028	0,002	0,043	0,002	0,003	0,002
11	0,007	0,003	0,041	0,003	0,025	0,003	0,004	0,003
12	0,007	0,005	0,041	0,006	0,025	0,004	0,007	0,005
13	0,010	0,010	0,035	0,013	0,034	0,009	0,009	0,021

Приведем также средние квадратичные отклонения величин, представленных в таблице 1.

Таблица 2. Средние квадратичные отклонения относительных интенсивностей 13 мод для 8 звуков, усредненных по всем образцам и по всем респондентам

	А	О	Ы	У	Э	Л	М	Н
1	0,037	0,043	0,075	0,099	0,039	0,084	0,184	0,156
2	0,025	0,049	0,054	0,043	0,027	0,114	0,025	0,028
3	0,033	0,039	0,015	0,041	0,063	0,025	0,006	0,006
4	0,073	0,072	0,004	0,029	0,030	0,033	0,009	0,005
5	0,022	0,055	0,004	0,009	0,010	0,019	0,003	0,004
6	0,047	0,007	0,004	0,003	0,008	0,004	0,002	0,002
7	0,044	0,002	0,007	0,002	0,012	0,002	0,002	0,003
8	0,010	0,001	0,009	0,001	0,019	0,001	0,003	0,002
9	0,006	0,001	0,012	0,001	0,034	0,001	0,001	0,001
10	0,004	0,001	0,021	0,001	0,022	0,001	0,001	0,001
11	0,003	0,001	0,023	0,001	0,010	0,001	0,002	0,001
12	0,003	0,001	0,021	0,003	0,010	0,001	0,004	0,002
13	0,004	0,003	0,015	0,005	0,011	0,003	0,005	0,009

Список мод, наиболее удачно подходящих для разделения слогов на отдельные звуки, приведен в следующей таблице

Таблица 3. Список номеров мод, наиболее удачно подходящих для разделения слогов на отдельные звуки

	А	О	Ы	У	Э
Л	3,4,6	3,4	10	3	3,5
М	1,3,4	1,3,4,5	10	3,4,5	1,3,4
Н	1,3,4	1,3,4,5	10	3,4	3,4

Используя эти результаты можно создать систему, позволяющую идентифицировать вышеприведенные звуки. Была разработана компьютерная программа, в которой анализировались короткие, длительностью в 1–2 тысячных долей секунды, образцы звуков. Образцы в соответствии с вышеприведенной методикой расщеплялись на моды, относительные амплитуды которых сравнивались с приведенными в табл. 1. По результатам сравнения и проводилась идентификация. Обнаружилось следующее.

1. Идентификация звука является наиболее успешной при использовании мод, приведенных в табл. 3. Использование других мод — менее эффективно. При совместном же использовании мод из табл. 3. и иных наблюдается некоторое улучшение надежности распознавания, но оно незначительно. В отдельных же случаях при совместном использовании нескольких мод надежность распознавания может даже ухудшиться.

2. Надежность распознавания мало зависит от смены респондента, изменение надежности распознавания может объясняться статистическим разбросом данных.

3. Напротив, усреднение данных по нескольким образцам звучания того же самого звука приводит к существенному улучшению надежности распознавания.

Данные по надежности распознавания (в процентах) приведем в нижеследующей таблице, каждый столбец которой соответствует заданному числу образцов, отобранных для усреднения.

Таблица 4. Надежность распознавания одного из 8 звуков в зависимости от количества образцов, отобранных для усреднения. Усреднения проводились по 1,10 и 100 образцам. Каждый вариант усреднения отражен в соответствующей колонке

Звук	1	10	100
А	62	81	95
О	62	81	91
Ы	55	70	83
У	48	69	86
Э	61	77	94
Л	51	66	84
М	55	64	89
Н	40	61	88

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Анализируя данные, приведенные в табл. 4, можно обратить внимание на следующее:

Надежность распознавания улучшается при увеличении числа образцов, взятых для усреднения, но никогда не достигает 100 процентов. Это наводит на мысль: *не существует критериев, позволяющих однозначно распознавать произнесенные звуки*. Человек распознает звуки *по вероятности*. Скорее всего, в мозгу человека производится усреднение нескольких коротких, длительностью не более нескольких миллисекунд, отрезков звучания. Известно, что речь, произнесенная скороговоркой, воспринимается хуже, чем произнесенная в нормальном темпе, а если темп скороговорки превышает 20 букв в секунду, то речь не распознается вообще. Вероятно, причиной этому служит не недостаточное быстродействие человеческого мозга, а недостаток статистического материала. Подтверждением этому служит то общеизвестное обстоятельство, что человек способен «дорисовывать» звуки, произнесенные в составе слов в условиях плохой слышимости. Но если человек способен «дорисовывать» звуки в условиях плохой слышимости то, надо полагать, он способен делать это и в условиях хорошей слышимости, поднимая приведенные в табл. 4. значения надежности от 83–95 процентов до 100 процентов. «Дорисовка», вероятно, происходит путем лингвистического контроля. Опять же, подтверждением этому служит то общеизвестное обстоятельство, что иногда, даже в условиях хорошей слышимости человек воспринимает речь другого человека с ошибками, но при этом он ***уверен*** в том, что все расслышал правильно.

ВЫВОДЫ

Таким образом, несмотря на то, что вероятность распознавания звука пока что не достигает 100%, предложенный метод можно считать перспективным, поскольку, во-первых, он достаточно прост в математическом плане, во-вторых, надежность распознавания звуков повышается по мере увеличения объема статистического материала. Имеются перспективы применения метода аппроксимации к решению задачи идентификации человека по голосу.

Благодарю Л. Денскевич и А. Власова за ряд ценных замечаний и советов.

ЛИТЕРАТУРА

1. Галунов В. И., Лобанов Б. М., Загоруйко Н. Г. Синтез и распознавание речи. Труды XIV сессии Российского акустического общества, 2004.
2. http:// intsys.msu.ru/invest/speech/research. Интеллектуальные системы. Официальный сайт кафедры МТИС и лаборатории проблем теоретической кибернетики механико-математического факультета МГУ.
3. Kuhl P. K., Iverson P. Linguistic experience and the «perceptual magnet effect». In Strange W. Speech perception and linguistic experiment, p. 121–154.
4. Галунов В. И., Гарбарук В. И. Акустическая теория речеобразования и системы фонетических признаков. Материалы международной конференции «100 лет экспериментальной фонетике в России», 1–4 февраля 2001 г.
5. Галунов В. И., Соловьев А. Н.. Современные проблемы в области распознавания речи. Информационные технологии и вычислительные системы, №2, 2004.
6. Kraft D. Speech perception. J. Phonetics, vol. **7**, p. 279–312, 1979.
7. http://www.auditech.ru; http://www.smartphone.ru; http://www.summatech.ru; http://www.sakrament.com; http://www.speechpro.ru; http://www.opencom.ru; http://www.istrasoft.ru/speech.html
8. Elinek F. Разработка экспериментального устройства, распознающего раздельно произнесенные слова. ТИИЭР, т. 73, №11, с. 91–99, 1985.
9. http://www.digest.univers.cv.ua/cnp_start.html
10. Галунов В. И. Современные речевые технологии. http://g-klishin.narod.ru/works.html
11. Дьяконов В. Абраменкова И. МАТЛАБ. Обработка сигналов и изображений. Специальный справочник. С.-П.: Питер, 2002.
12. http://www.prodav.narod.ru/wavelet
13. Зигмунд А. Тригонометрические ряды, т.1. Перев. с англ. М.: Мир. 1965.