

Применение метода нелинейного отображения многомерных данных в задаче постановки правильного произношения звуков в составе отдельных слов

*Губочкин И.В.,
аспирант*



В статье предложен метод обучения произношению звуков в составе коротких слов и фраз с интерактивной визуализацией результатов. Его основу составляет совместное использование аппарата скрытых марковских моделей и нелинейного отображения многомерных данных. Приведён пример практического применения метода в задаче обучения произношению некоторых фонем в составе слов английского языка.

In the article the method of sounds pronunciation training in short words and phrases with interactive results visualization is suggested. Its foundation is the joint use of hidden Markov model approach and the multidimensional data nonlinear mapping. An example of method's practical application in the pronunciation training task for certain phonemes in English words is presented.

Введение

Системы компьютерного обучения языку (КОЯ) в настоящее время получают всё большее распространение. В этих системах компьютер обеспечивает немедленную реакцию на действия обучающегося и позволяет ему самостоятельно выбирать скорость изучения языка. Таким образом, реализация эффективной обратной связи между компьютером и человеком является важнейшей задачей при построении системы КОЯ. Для её решения разработано множество методов и подходов [1—4]. Среди них большой класс составляют методы, основанные на сравнении входного сигнала с некоторым эталоном в частотной или временной области. Основной недостаток систем, построенных по этому принципу, состоит в том, что даже при очень хорошем произношении входной сигнал и эталон могут иметь совершенно разные спектры или формы во времени [5].

Кроме того, получаемые результаты достаточно трудно интерпретировать, поскольку нет простого соответствия между артикуляционными движениями и отображаемыми результатами.

Описанные выше проблемы создают трудности при обучении языку с использованием средств автоматизированного контроля. Поэтому важно, чтобы ответ системы КОЯ был максимально понятным пользователю. Кроме того, обучающемуся необходимо знать не только тот факт, что он совершает ошибку, но также и её тип. Это позволит ему самостоятельно установить, что именно он делает неправильно. В [6] был предложен метод обучения произношению отдельных фонем с использованием нелинейного отображения данных Сэммона. Этот метод позволяет проводить сравнение произношения с множеством эталонных реализаций и диагностику наиболее распространённых ошибок. Представляемые результаты могут быть наглядно отображены на экране монитора компьютера. В настоящей работе даётся обобщение этого метода на случай тренировки произношения звуков внутри слов и коротких фраз.

Алгоритм отображения

Алгоритм отображения речевых сигналов состоит из двух этапов. На первом этапе происходит сегментирование входного сигнала на фонемы, соответствующие изучаемому слову, а на втором — отображение нужной фонемы.

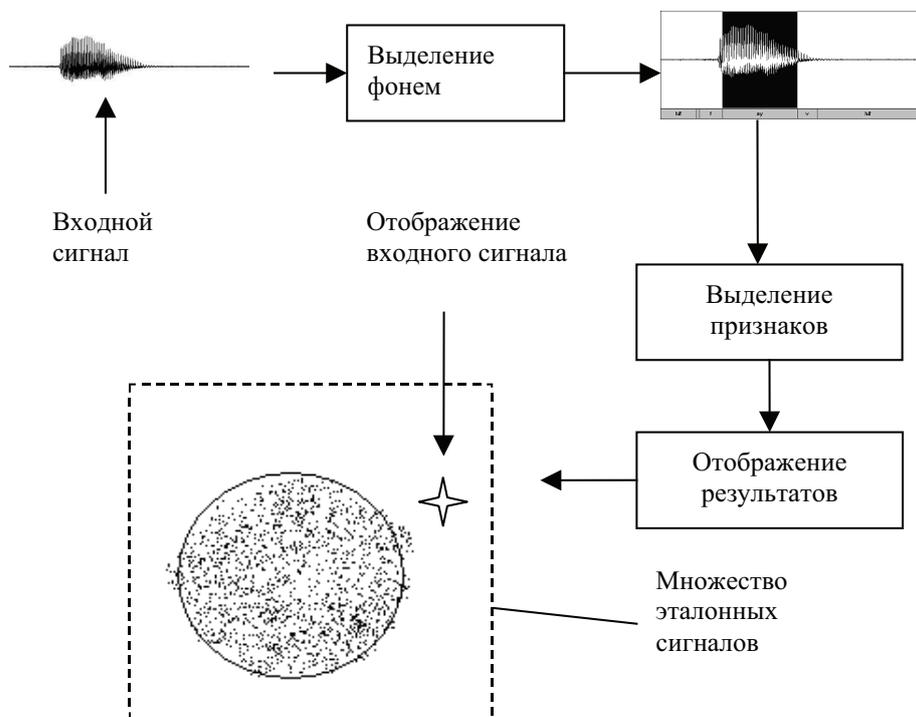


Рис. 1. Схема работы алгоритма

Для осуществления первого этапа наиболее перспективным представляется использование аппарата скрытых марковских моделей (СММ) [7]. После предварительной тренировки моделей отдельным фонем на обучающей выборке они составляются в слова или короткие фразы. Полученные составные модели в дальнейшем используются для определения, при помощи алгоритма Витерби, местоположения границ отдельных фонем. Таким образом, в результате выполнения первого этапа алгоритма мы получаем фонетическую разметку входного сигнала.

Рассмотрим теперь второй этап. Выделив во входном сигнале интервал, соответствующий интересующей нас фонеме, мы можем вычислить её вектор признаков и произвести его отображение на плоскости относительно множества эталонных сигналов. На рисунке 1 представлена схема работы алгоритма.

Если задача разметки речевого сигнала на фонемы при известной транскрипции и вычисления векторов признаков в настоящее время в основном решена, то в части представления результатов до сих пор ведутся интенсивные исследования. Одним из перспективных подходов в этом направлении следует признать метод нелинейного отображения многомерных данных, предложенный Сэммоном [8], который заключается в следующем.

Обозначим через δ_{ij} расстояние между векторами x_i и x_j исходного пространства размерности n , а через d_{ij} — расстояние между векторами y_i и y_j в пространстве отображения размерности q , $q < n$. Тогда суммарная ошибка с учётом нормирующего множителя будет равна:

$$E = \frac{1}{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \delta_{ij}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{(\delta_{ij} - d_{ij})^2}{\delta_{ij}}, \quad (1)$$

где N — число векторов множества входных данных X .

Для того чтобы получить оптимальное в смысле сохранения расстояния отображение, найдём минимум E (1) по y_j для случая использования в качестве δ_{ij} евклидова расстояния. Метод наискорейшего спуска приводит к следующему рекуррентному уравнению:

$$y_i(l+1) = y_i(l) - \alpha \frac{\partial E}{\partial y_j} = y_i(l) + [2\alpha / (\sum_{i=1}^{N-1} \sum_{j=i+1}^N \delta_{ij})] \sum_{\substack{j=1 \\ j \neq i}}^N [(\delta_{ij} - d_{ij}) / (\delta_{ij} d_{ij})] \cdot [y_i(l) - y_j(l)]. \quad (2)$$

Здесь α — настраиваемый параметр, l — номер итерации. Вычисление соотношения (2) прекращается в момент выполнения условия

$$E(l+1) - E(l) < \varepsilon, \quad (3)$$

где ε — некоторая константа, $\varepsilon > 0$. На практике значения ε обычно устанавливаются в интервале $10^{-4} \dots 10^{-9}$.

Для случая добавления в исходное множество X ещё одного вектора данных воспользуемся следующей процедурой [6]. Обозначим через Δ матрицу расстояний между векторами множества X :

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1N} \\ \delta_{21} & \delta_{22} & & \delta_{2N} \\ \vdots & & \ddots & \vdots \\ \delta_{N1} & \delta_{N2} & \dots & \delta_{NN} \end{bmatrix}. \quad (4)$$

После выполнения отображения в пространство размерности q по алгоритму (1)—(3) мы получим множество векторов Y , каждый элемент которого содержит в себе координаты некоторой точки в этом пространстве. Расстояния между точками в пространстве отображения при этом будут определяться как

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1N} \\ d_{21} & d_{22} & & d_{2N} \\ \vdots & & \ddots & \vdots \\ d_{N1} & d_{N2} & \dots & d_{NN} \end{bmatrix}. \quad (5)$$

Тогда при добавлении ещё одного вектора x_{N+1} матрица расстояний Δ будет выглядеть следующим образом:

$$\Delta^* = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1N} & \delta_{1N+1} \\ \delta_{21} & \delta_{22} & & \delta_{2N} & \delta_{2N+1} \\ \vdots & & \ddots & \vdots & \vdots \\ \delta_{N1} & \delta_{N2} & \dots & \delta_{NN} & \delta_{NN+1} \\ \delta_{N+11} & \delta_{N+12} & \dots & \delta_{N+1N} & \delta_{N+1N+1} \end{bmatrix}. \quad (6)$$

Матрица расстояний \mathbf{D}^* в пространстве отображения определяется аналогично. В этом случае выражение (2) может быть переписано для расчёта только $N+1$ элемента отображения:

$$\mathbf{y}_{N+1}(l+1) = \mathbf{y}_{N+1}(l) + \left[2\alpha / \left(\sum_{i=1}^N \sum_{j=i+1}^{N+1} \delta_{ij} \right) \right] \sum_{\substack{j=1 \\ j \neq N+1}}^{N+1} [(\delta_{N+1j} - d_{N+1j}) / (\delta_{N+1j} d_{N+1j})] \cdot [\mathbf{y}_{N+1}(l) - \mathbf{y}_j(l)]. \quad (7)$$

Останов работы алгоритма (7) происходит при выполнении условия (3).

Программа экспериментальных исследований

Для экспериментальных исследований было сформировано две группы дикторов по четыре человека каждая. Первую группу составляли дикторы, свободно владеющие английским языком. Во вторую группу были включены студенты 1—5 курсов языкового вуза. В качестве тестового был выбран следующий набор слов английского языка: bit [blt], boot [bu:t], bottom ['bɒtəm], she

[i:], way [weɪ], winner ['wɪnə]. В этих словах встречаются фонемы, относящиеся к различным группам: фрикативы (/SH iy/), гласные (/b IH tcl t/, /b UW tcl t/), носовые (/b aa tcl t EM/, /w ih NX axr/), полугласные (/W eu/). Транскрипции слов в данном случае приведены в соответствии с обозначениями фонем в широко известной речевой базе ТИМІТ [9]. Каждое слово проговаривалось четыре раза дикторами из обеих групп. Для ввода речевого сигнала в ПК применялись специальные программные и аппаратные средства: динамический микрофон AKG D77 S и ламповый микрофонный предусилитель ART TUBE MP Project Series USB. Частота дискретизации встроенного АЦП была установлена на уровне 16 кГц — это общепринятая частота при обработке устной речи.

Реализация предложенного метода делится на три независимых во времени этапа. На первом этапе формируется база априорных данных по каждой реализации всех фонем. Для этого вычисляется вектор признаков, описывающий ту или иную реализацию. В [6] было показано, что наиболее предпочтительными для использования являются кепстральные коэффициенты линейного предсказания с неравномерным частотным разрешением, которые, в свою очередь, рассчитываются на основе коэффициентов линейного предсказания с неравномерным частотным разрешением [10]. Несмотря на то, что по своим свойствам они близки к широко известным мел-кепстральным коэффициентам (MFCC), их использование в задаче описания характеристик фонемы обусловлено прежде всего удобством практического применения.

Для получения коэффициентов линейного предсказания с неравномерным частотным разрешением сначала найдём коэффициенты автокорреляции r входного сигнала x .

$$\begin{aligned} W^{(0)} &= x \\ r(0) &= x^T x \\ W^{(m)}(n) &= -\lambda W^{(m-1)}(n) + W^{(m-1)}(n-1) + \lambda W^{(m)}(n-1), \quad 1 \leq n \leq N \\ r(m) &= x^T W \\ m &= \overline{1, p} \end{aligned} \quad (8)$$

Здесь N — объём выборки, p — порядок модели, λ — коэффициент деформации, определяемый как

$$\lambda \approx 1,0674 \left(\frac{2}{\pi} \tan^{-1}(0,06583 f_s / 1000) \right)^{1/2} - 0,1916, \quad (9)$$

где f_s — частота дискретизации в Гц.

В дальнейшем непосредственно для получения коэффициентов линейного предсказания с неравномерным частотным разрешением а воспользуемся автокорреляционным методом [11]:

$$\begin{aligned} E^{(0)} &= r(0) \\ k_i &= \left\{ r(i) - \sum_{j=1}^{i-1} \alpha_j^{(i-1)} r(i-j) \right\} / E^{(i-1)}, \quad 1 \leq i \leq p \\ \alpha_i^{(i)} &= k_i \\ \alpha_j^{(i)} &= \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \\ E^{(i)} &= (1 - k_i^2) E^{(i-1)} \\ \alpha_m &= \alpha_m^{(p)}, \quad 1 \leq m \leq p \end{aligned} \quad (10)$$

Кепстральные коэффициенты вычисляются по следующей формуле [7]:

$$\begin{aligned}c_m &= a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad 1 \leq m \leq p \\c_m &= \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad m > p\end{aligned}\tag{11}$$

В нашем эксперименте векторы признаков рассчитывались по реализациям фонем, взятым из речевой базы TIMIT, диалект DR1 (New England). Поскольку количество реализаций каждой фонемы составляло от сотни до нескольких тысяч, то векторы признаков предварительно подвергались векторному квантованию по методу k гармонических средних [12]. Выбор данного метода обусловлен тем, что он даёт меньшую ошибку квантования и быстрее сходится по сравнению с классическим методом k -средних. После векторного квантования производится отображение полученных результатов на двумерную плоскость по алгоритму (1)—(3), а в качестве меры расстояния δ_{ij} между векторами кодовой книги используется евклидова метрика.

На втором этапе производится оценка параметров монофонных СММ для реализаций фонем из обучающего подмножества TIMIT при помощи специализированного пакета НТК (Hidden Markov Model Toolkit) [13]. В качестве вектора признаков использовались 12 мел-кепстральных коэффициентов и логарифм энергии, а также их первая и вторая производные по времени (всего 39 элементов). Для всех фонем, кроме модели тишины, прототипы СММ имели 3 состояния, 39-элементный вектор признаков и 3 гауссовых смеси. Модель тишины имела одно состояние.

Начальное оценивание параметров моделей осуществлялось при помощи алгоритма Витерби. После этого была проведена их переоценка с использованием алгоритма Баума-Уэлча [7].

На третьем этапе обучающийся последовательно произносит реализации определённого слова. Каждая такая реализация сначала сегментируется на фонемы при помощи программы HVITE из пакета НТК. Затем по фрагменту речевого сигнала, который соответствует интересующей нас фонеме, вычисляется вектор признаков. Полученный вектор отображается на двумерную плоскость при помощи алгоритма (6)—(7). Таким образом, на плоскости кроме множества точек, соответствующих произношению фонемы, появляется ещё одна точка, которая характеризует расположение входного сигнала относительно данного множества. Учитывая это, диктор может корректировать своё собственное произношение, приближая его к эталонному.

Результаты эксперимента

В качестве параметров алгоритма визуализации были установлены следующие:

- порядок АР-модели $p = 8$;
- число кепстральных коэффициентов $m = 12$;

- параметр алгоритма отображения $\alpha = 0,4$;
- относительная ошибка отображения $\varepsilon = 10^{-5}$;
- максимальное количество векторов кодовой книги $k = 100$.

Начальные приближения y_0 задавались случайным образом в интервале от 0 до 1.

Для примера рассмотрим процесс работы алгоритма для слова «way», произнесённого одним из дикторов первой группы. На рисунке 2 показан результат определения границ фонем для одной из реализаций.

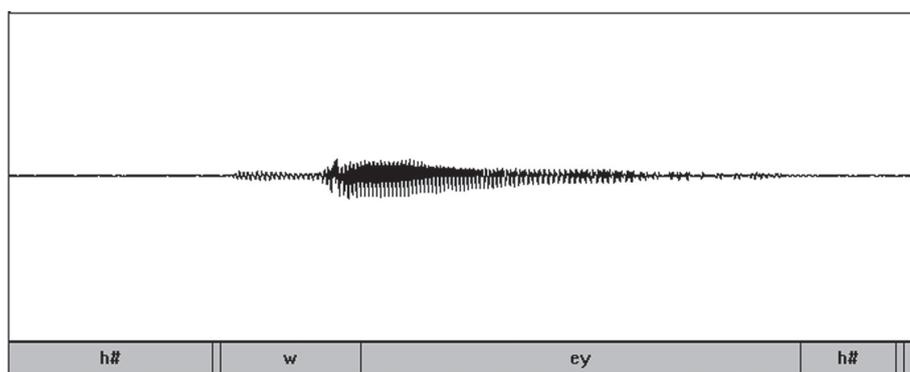


Рис. 2. Найденные границы фонем для слова «way»

После выделения границ фонем из входного сигнала выбирается фрагмент, соответствующий обучаемой фонеме. По нему рассчитывается вектор кепстральных коэффициентов с неравномерным частотным разрешением. На рисунке 3 на верхнем графике показан сигнал, соответствующий слову «way», на котором обозначены найденные границы фонемы /w/. На нижнем графике показаны значения коэффициентов вектора признаков (11), рассчитанных по выделенному участку сигнала.

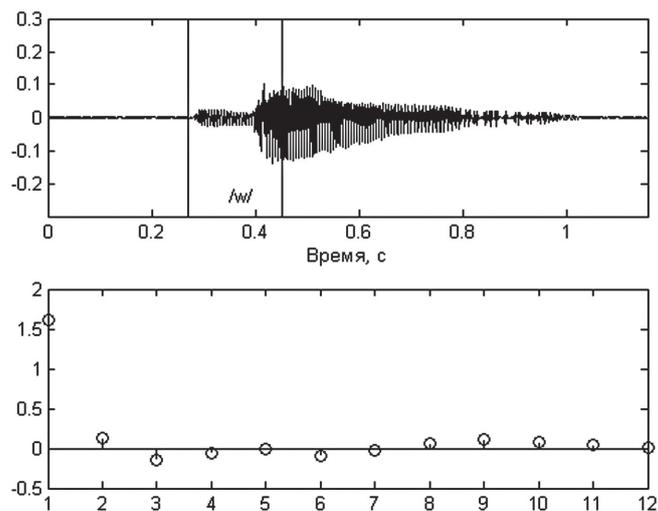


Рис. 3. Результаты выделения границ фонемы /w/

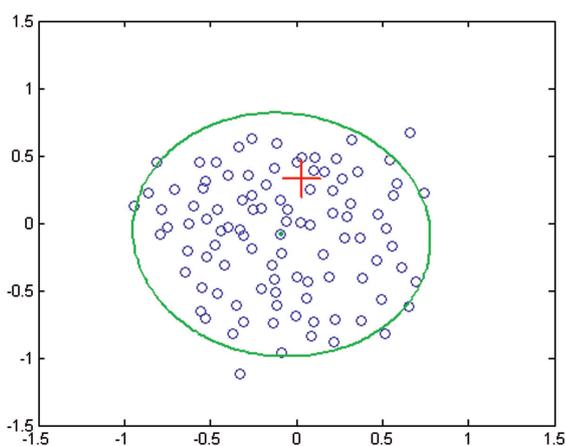


Рис. 4. Отображение реализации фонемы /w/ от диктора из первой группы

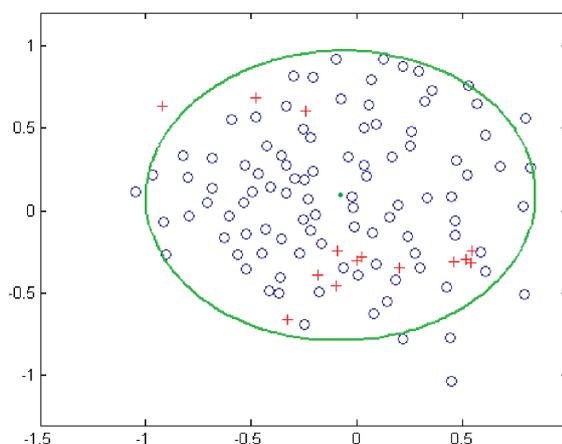


Рис. 5. Отображение реализаций эталонных фонем и фонем дикторов первой группы

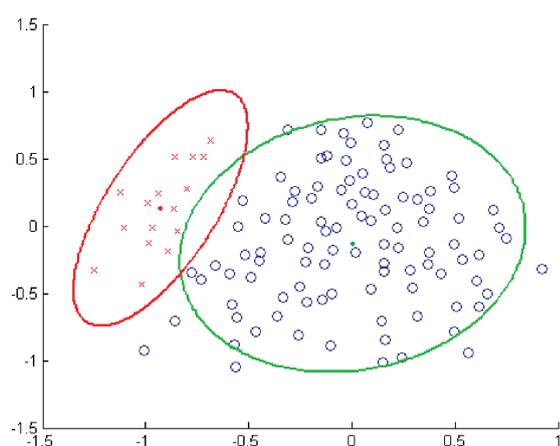


Рис. 6. Отображение реализаций эталонных фонем и фонем дикторов второй группы

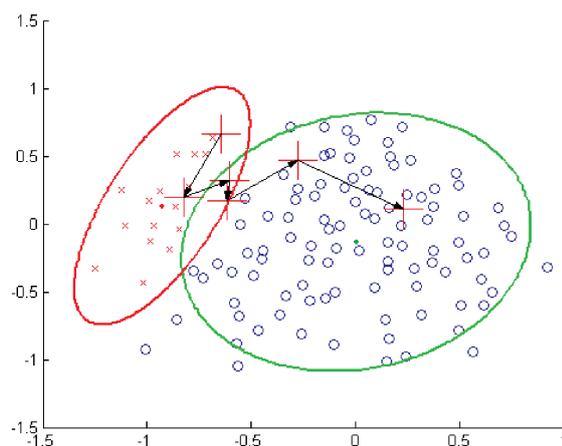


Рис. 7. Результаты обучения

На рисунке 4 показан результат отображения в двумерное пространство полученного вектора признаков при помощи алгоритма (6)—(7).

Здесь символами 'o' обозначены эталонные реализации фонем, а символом '+' обозначено положение вектора признаков входного сигнала. Эллипс ограничивает область эталонных сигналов. Поскольку диктор имеет хорошо поставленное английское произношение, произнесённая им фонема сразу же попала в область эталонных значений. Для подтверждения полученного результата на рисунке 5 показано отображение всех реализаций фонемы /w/ для дикторов первой группы.

Нетрудно видеть, что почти все из них попали в область эталонных значений. Аналогичные результаты получены и для остальных фонем из тестового

набора. Это позволяет говорить о том, что разработанный алгоритм является инвариантным к выбору эталонных сигналов.

Рассмотрим теперь результаты работу алгоритма для дикторов из второй группы, представленные на рисунке 6.

Здесь символами 'x' дополнительно показаны реализации фонемы /w/ для дикторов из второй группы. Видно, что они образуют компактный кластер (его границы показаны красной линией) вблизи границы области эталонов. Отсюда можно сделать вывод, что произношение дикторов, входящих во вторую группу, всё же существенно отличается от произношения носителей языка. Задачей обучаемого, таким образом, является изменить собственное произношение, чтобы переместить 'x' в область эталонных сигналов. На рисунке 7 показан пример такого обучения.

Из рисунка видно, что чем более чётко проговаривается фонема /w/ внутри слова, тем ближе к эталонному сигналу (и дальше от области неправильного произношения) располагаются её реализации.

В рассматриваемом эксперименте не проводилась какая-либо классификация ошибок произношения. Однако при наличии такой информации предложенный алгоритм легко обобщить на данный случай. Для этого на множестве точек двумерного отображения необходимо выделить области, соответствующие какой-либо типичной ошибке (аналогично рисунку 6). Рассматриваемый случай схематично показан на рисунке 8.

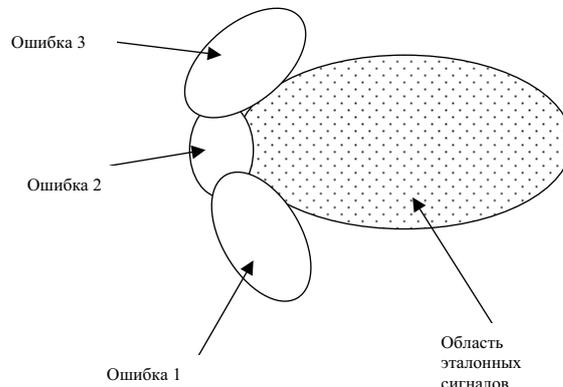


Рис. 8. Иллюстрация обобщения предложенного алгоритма на случай наличия нескольких типов ошибок произношения фонемы

Таким образом, попадая в ту или иную область, обучаемый может самостоятельно диагностировать проблемы в собственном произношении и при необходимости производить его коррекцию.

Численная оценка качества произношения

Рассмотренные выше результаты оценки произношения дают в первую очередь качественные результаты. Для более точного определения правильности произношения следует

задать некоторую его количественную оценку. Из практических соображений следует, что такая оценка должна обладать следующими двумя свойствами:

- все возможные значения оценки должны находиться в некоторой ограниченной области (например, на интервале $[0..1]$);
- принимать малые значения, в случае если получаемый результат выходит из области эталонных значений или же находится на её периферии;
- принимать значения, близкие или равные максимуму, в случае если получаемый результат находится в глубине области эталонных значений или близок к её центру.

Можно предложить множество способов определения оценки, обладающей указанными выше свойствами. Один из наиболее простых способов сделать это — использовать функцию распределения вероятностей расположения координат точек, полученных в результате отображения эталонных сигналов на двумерную плоскость. Таким образом, оценка качества произношения определяется следующим образом:

$$\rho(Y) = 1 - [P(M+|M-Y) - P(M-|M-Y)]. \quad (12)$$

Здесь P — функция распределения вероятностей расположения координат точек отображения, M — математическое ожидание, Y — координаты отображения входного сигнала.

Для возможности практического применения приведённой выше оценки (12) требуется задать функцию плотности вероятности $\rho(Y)$. Такая функция должна быть определена на всём множестве значений Y и иметь симметричную форму относительно значения математического ожидания M . Подобным требованиям удовлетворяет широко известная функция плотности вероятности нормального распределения в её двумерном варианте. Для подтверждения корректности применения нормального закона при описании распределения значений координат точек отображения был проведён следующий эксперимент.

Для каждой фонемы из тестового набора было произведено отображение её реализаций на двумерную плоскость с помощью алгоритма (1)—(5). Затем каждое полученное множество точек было проверено на соответствие нормальному закону с помощью универсального критерия Дурника-Хансена (Doornik-Hansen) [14] на уровне значимости 0,05. Полученные результаты приведены в таблице 1.

Таблица 1

Значения решающей статистики Дурника-Хансена

Фонема	/SH/	/IH/	/UW/	/EM/	/NX/	/W/
E_p	0,3029	0,7013	0,8174	0,9678	0,3726	0,1843

Здесь E_p — значения решающей статистики, используемой в критерии.

Из таблицы видно, что, по крайней мере, на выбранном уровне значимости тест был пройден для всех фонем (значения решающей статистики E_p превы-

шают уровень значимости). Данный результат позволяет сделать вывод о корректности использования нормального закона для описания распределения координат точек отображения.

Воспользуемся теперь оценкой (12) для определения качества произношения дикторов из первой и второй тестовых групп. Для этого по множеству координат точек отображения эталонных реализаций выбранной фонемы r производилось оценивание математического ожидания $\bar{\mathbf{M}}_r$ и ковариационной матрицы $\bar{\Sigma}_r$:

$$\begin{aligned}\bar{\mathbf{M}}_r &= \frac{1}{N_r} \sum_{i=1}^{N_r} \mathbf{y}_{i,r}, \\ \bar{\Sigma}_r &= \frac{1}{N_r - 1} \sum_{i=1}^{N_r} (\mathbf{y}_{i,r} - \bar{\mathbf{M}}_r)^T (\mathbf{y}_{i,r} - \bar{\mathbf{M}}_r),\end{aligned}\tag{13}$$

где N_r — число реализаций r -й фонемы, $\mathbf{y}_{i,r}$ — вектор координат i -й точки отображения. Затем по реализациям фонем, полученных от дикторов из тестовых групп, рассчитывалась оценка $\rho(Y)$. Полученные значения, усреднённые по всем дикторам первой и второй групп, показаны в таблицах 2 и 3 соответственно.

Таблица 2

Оценка качества произношения дикторов первой группы

Фонема	/SH/	/H/	/UW/	/EM/	/NX/	/W/
$\rho(Y)$	0,2736	0,2707	0,2775	0,4194	0,3010	0,3168

Таблица 3

Оценка качества произношения дикторов второй группы

Фонема	/SH/	/H/	/UW/	/EM/	/NX/	/W/
$\rho(Y)$	0,3662	0,1776	0,2968	0,3258	0,2976	0,2902

Представленные в таблицах результаты наглядно свидетельствуют о том, что дикторы с хорошо поставленным английским произношением получают более высокие оценки. Недостаточно сильная разница в значениях $\rho(Y)$ (и даже превышение усреднённого качества произношения дикторов второй группы над результатами дикторов первой группы для фонем /SH/ и /UW/), по-видимому, обусловлена недостаточными размерами выборок и большой вариативностью результатов внутри каждой группы.

Выводы

Полученные в ходе экспериментальных исследований результаты позволяют сделать вывод об эффективности разработанного алгоритма в решении задачи обучения произношению. Кроме того, в метод легко может быть добавлена возможность диагностики наиболее часто встречающихся ошибок путём добавления во множество априорных

данных реализаций фонем, соответствующих какой-либо типичной ошибке произношения. Зная порядковые номера таких реализаций, на двумерном отображении легко можно выделить области, соответствующие различным классам ошибок (см. рисунок 8).

Ещё одно преимущество представленного алгоритма состоит в том, что, в случае если предварительно определён язык (и/или диалект), произносительные транскрипции могут быть определены независимо от диктора. Эта возможность является особенно актуальной для практического применения в реальном учебном процессе.

Также предложена численная оценка качества произношения и рассмотрена методика её вычисления.

Дальнейшее развитие метода может быть направлено на повышение точности диагностики ошибок. Здесь может быть выделено два пути. Первый — учёт ошибок, связанных с неправильной длительностью произношения отдельных фонем и ударениями, поскольку предложенный алгоритм анализирует, в основном, различия в спектральной области. А второй путь — это совместное использование акустических признаков и информации, получаемой из видеоизображения обучаемого [15].

Литература

1. Akahane-Yamada R., McDermott E. Computer-based second language production learning by using spectrographic representation and HMM-based speech recognition scores. Proceedings of ICSLP, Sydney, Australia 1998.
2. Nouza J. Training speech through visual feedback patterns. Proceedings of ICSLP Sydney, Australia 1998.
3. Molholt G. Computer-assisted instruction in pronunciation for Chinese speakers of American English. TESOL Quarterly 22, 91—111, 1988.
4. Tsubota Y., Kawahara T., and Dantsuji M. CALL system for Japanese students of English using pronunciation error prediction and formant structure estimation. In InSTIL 2002 Advanced Workshop, 2002.
5. Neri A., Cucchiari C., Strik H. Feedback in Computer Assisted Pronunciation Training: technology push or demand pull? Proceedings of ICSLP 2002, Denver, USA, P. 1209—1212.
6. Губочкин И.В. Алгоритм визуализации речевых сигналов для интерактивного обучения правильному произношению // Речевые технологии/ 2008. №3, С. 72—80,
7. Rabiner L.R., Juang B.-H. Fundamentals of speech recognition. Prentice Hall, Englewood Cliffs, NJ, 1993.
8. Sammon J.W. A non-linear mapping algorithm for data structure analysis. *IEEE Trans. Computers*, CC-18(5):401—409. 1969.
9. William M.Fisher, George R.Doddington, and Kathleen M.Goudie-Marshall. «The DARPA Speech Recognition Research Database: Specifications and Status», Proceedings of DARPA Workshop on Speech Recognition. P. 93—99, Feb. 1986.
10. Härmä A. Frequency-warped autoregressive modeling and filtering. Dissertation for the degree of Doctor of Science in Technology. Department of Electrical and Communications Engineering, Helsinki University of Technology, Espoo, Finland, 2001.

Губочкин И.В. Применение метода нелинейного отображения многомерных данных в задаче постановки правильного произношения звуков в составе отдельных слов

11. Марпл С.Л.-мл. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 584 с.
12. Bin Zhang. «Generalized K-Harmonic Means — Boosting in Unsupervised Learning», Hewlett-Packard Labs, Technical Report HPL-2000-137, 2000.
13. Young *et al.*, Hidden Markov Model Toolkit v3.1 reference manual, Technical report, Speech group, Cambridge University Engineering Department, December. 2001.
14. Doornik J.A. and Hansen H. An omnibus test for univariate and multivariate normality. Discussion Paper W4&91. Nuffield College, Oxford, UK, 1994.
15. Vicsi K. An Overview of Speech Training Methods Based on Multi Modal Feedback. IOS Press, Amsterdam. P. 1—10, 2004.
16. Фукунага К. Введение в статистическую теорию распознавания образов: Пер. с англ. М.: Наука, 1979. 368 с.

Губочкин Иван Вадимович

Инженер-программист ООО «МФИ-Софт», аспирант кафедры математики и информатики Нижегородского государственного лингвистического университета им. Н.А. Добролюбова. Область научных интересов — автоматическая обработка речевых сигналов. Автор девяти научных работ.
E-mail: jhng@yandex.ru.