**О возможности повышения эффективности автоматического интонационного анализа речи**

Марьев А.А.

*Работа выполнена при поддержке РФФИ, проект № 10-06-00110а*

За прошедшие полтора-два десятилетия существенно возрос интерес к системам автоматического интонационного анализа речи (АИАР). Среди областей применения подобных система стоит отметить такие, как: распознавание психологического, физиологического или эмоционального состояния оператора по его речи; идентификация типа произношения (акцента); разработка речевых интерфейсов, допускающих эмоционально окрашенное взаимодействие с компьютером. До настоящего времени не предложено достаточно общего и успешного подхода к решению задачи АИАР, поэтому отыскание возможностей повышения эффективности АИАР является в настоящее время актуальной задачей.

Часто задача АИАР может быть представлена в классификационной формулировке: по речевому фрагменту необходимо определить один из априорно известных типов интонации. Количество априорной информации в задачах АИАР, как правило, недостаточно для прямого применения статистически оптимальных методов распознавания. В частности, до настоящего времени не вполне изучен характер связей объективных характеристик речи как акустического колебания с ее интонационными характеристиками. Важно также отметить, что восприятие интонации субъективно, и в этом смысле задача автоматического интонационного анализа является некорректной.

Вследствие перечисленных сложностей в системах АИАР используются классификаторы различных типов [1], обучаемые на некотором множестве заранее классифицированных по интонационному признаку речевых фрагментов (обучающей выборке).

Достаточно общим подходом к определению информативных признаков речевого сигнала является определение первоначального множества параметров (признаков) речевого сигнала, из которых затем отбираются наиболее информативные в некотором смысле. Первоначальное множество признаков формируется разработчиком на основе результатов известных исследований и из эвристических соображений. В это множество могут быть включены параметры различных моделей речевого сигнала, например частота основного тона (модель квазипериодического стационарного случайного процесса), коэффициенты отражения (модель линейного предсказания), параметры огибающей (модель случайного процесса, нестационарного по дисперсии) и др. Для одновременного измерения выбранных параметров может потребоваться организация измерений в нескольких масштабах времени. Обычно различают кратковременный (short-term, окно ~$10^{-2}$ с) и долговременный (long-term, окно длиной до единиц с) анализ речевого сигнала. Для приведения результатов кратковременного анализа к масштабу долговременного вычисляются статистические характеристики кратковременных оценок.

Для решения задачи автоматического распознавания эмоций по речи автором была разработана система, структурная схема которой изображена на рис. 1. Система была реализована в виде программы для ЭВМ, пригодной как для обработки речевого сигнала в масштабе реального времени, так и для обработки звукозаписей. Источником сигнала (ИС) в данном случае является поток цифровых аудиоданных.

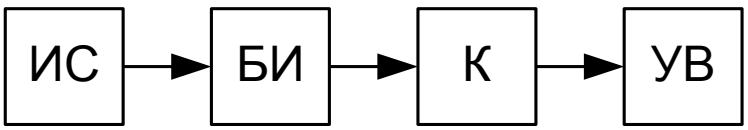


Рис.1 – Структурная схема системы автоматического интонационного анализа речи: ИС – источник сигнала, БИ – блок измерений, К – классификатор, УВ – устройство вывода.

Из первоначального множества в 878 признаков [3] были отобраны 16 наиболее информативных: среднее значение (СЗ) 5-го коэффициента линейного предсказания; СЗ 2-го, 11-го, 36-го и 39-го коэффициентов отражения в модели линейного предсказания; СЗ 4-го, 5-го, 9-го

и 25-го мел-частотных кепстральных коэффициентов (MFCC); СЗ и коэффициент вариации центроиды мгновенного амплитудного спектра (MAC); коэффициент вариации коэффициента асимметрии MAC; СЗ коэффициента вариации MAC; относительный размах вариации СЗ производной от огибающей сигнала в фазе атаки; относительная длительность вокализованных звуков и коэффициент пиковости (пик-фактор) речевого сигнала.

Для повышения надежности распознавания автором на основе информационного подхода был разработан классификатор, оптимальный в смысле принципа максимума информации (ПМИ) [4]. Данный принцип был предложен Г.А. Голицыным [5] и позволяет в рамках информационной модели биологического организма в явном виде определить целевую функцию адаптационного поведения.

Формирование множества записей для обучения системы АИАР само по себе представляет трудную задачу. Однако в настоящее время существует несколько крупных баз записей эмоциональной речи, среди которых одной из наиболее широко используемых исследователями является Берлинская база записей эмоциональной речи EmoDB [6]. Она содержит 495 речевых фрагментов, начитанных 10-ю дикторами обоего пола, демонстрирующих: злость (96 записей); страх (67); скуку (81); отвращение (46); радость (64); нейтральное состояние (79); огорчение (62).

Для оценки надежности распознавания семи перечисленных эмоциональных состояний был поставлен эксперимент, в ходе которого системе по очереди предъявлялись записи из обучающей выборки. Перед предъявлением очередной записи система обучалась на материале выборки, из которой исключалась распознаваемая запись. Таким образом, система на этапе обучения не получала информации о распознаваемой записи.

В ходе эксперимента была достигнута средняя вероятность верного распознавания семи эмоциональных состояний 0,71, что превосходит ряд известных разработок. Так в работе [7] сообщается о средней вероятности верного распознавания семи эмоций около 0,55, при этом также использовалась база EmoDB.

Достигнутые результаты подтверждают перспективность информационного подхода к решению задачи классификации интонаций речи с использованием ПМИ, в то же время в разработанной системе возможности данного подхода реализованы не в полной мере, что означает возможность дальнейшего повышения эффективности разработанной системы автоматического интонационного анализа.

Литература

1. К.А. Астапов, Применение вейвлет-преобразования для сокращения области значения искусственных нейронных сетей на примере задачи распознавания речи // Инженерный Вестник Дона, №1 2009 г. http://ivdon.ru/magazine/archive/n1y2009/105/

2. А.А. Марьев, В.П. Рыжов, Выбор признаков в задачах распознавания эмоциональных состояний оператора по речевым сигналам //Материалы Всероссийской научной конференции "Актуальные проблемы современности: человек, общество, техника" - часть 2 - Таганрог: Изд-во ТТИ ЮФУ, 2012 С. 31-36

3. Марьев А.А. Метод интерпретации результатов измерений параметров речевого сигнала в задачах диагностики психоэмоционального состояния человека по его речи // Инженерный Вестник Дона, №4 2011 г. http://ivdon.ru/magazine/archive/n4y2011/538/ 6с.

4. Голицын Г.А. Информация и творчество: на пути к интегральной культуре - М.: "Русский мир", 1997. - 304 с.

5. Berlin Database of Emotional Speech, http://pascal.kgw.tu-berlin.de/emodb/

6. Theodoros Iliou, Christos-Nikolaos Anagnostopoulos, Classification on Speech Emotion Recognition-A Comparative Study, International Journal on Advances in Life Sciences, vol. 2 no 1 & 2, 2010. pp. 18-28.