

РАСПОЗНАВАНИЕ ЭМОЦИЙ ДИКТОРА ПО ЕГО ГОЛОСУ

Самочернов Н.Б.

Научный руководитель – к.т.н., доц. Омельченко С.В.

Харьковский национальный университет радиоэлектроники
(61166, Харьков, пр. Науки, 14, каф. Информационно-сетевой инженерии)
e-mail: nick.samochnov@gmail.com, тел. (068) 450-95-91

An algorithm for automatic emotion recognition from the speaker's voice has been developed. The classification efficiency for different acoustic features was estimated and a very small set of the most reliable characteristics was extracted in order to obtain a robust and quick emotion state classification. Using the classifier with quadratic kernel and this feature set provides the recognition accuracy of approximately 96 % between “anger” and “neutral” emotional states.

Key words: voice, emotions, emotional state, emotional speech.

Автоматическое распознавание эмоционального состояния окажется полезным в любой сфере человеческой деятельности, где требуется его оперативная оценка — в маркетинге, медицине, психологии, обеспечении безопасности и т. п.

Однако, проблема взаимосвязи эмоциональных состояний диктора с параметрами его голоса до сих пор полностью не решена. Основой алгоритма голосового анализа является модуль выделения информативных признаков речевого сигнала и классификатор, относящий звуковой фрагмент, согласно этим признакам, к тому либо иному эмоциональному классу. Соответственно, выделение новых, по возможности родственных человеческому восприятию, информативных признаков, а также поиск новых высокоэффективных техник классификации на текущий момент времени являются важнейшими задачами голосового распознавания эмоционального состояния. Чаще всего в целях дальнейшего анализа из аудио сигнала выделяют: различные параметры частоты основного тона и формант; кратковременную оценку мощности; темп речи (количество слов произносимых в единицу времени); контур основного тона.

На основе выделяемого набора информативных признаков строится классификатор, который обучается на предварительно подготовленном наборе звуковых фрагментов. Наиболее популярными техниками классификации являются следующие [1]: поиск ближайших соседей, метод опорных векторов, скрытые марковские модели, модель смеси нормальных распределений, модели на основе нечеткой логики, байесовские классификаторы максимума вероятности.

База включает 500 записей речи 10 дикторов (5 мужчин, 5 женщин), воспроизводящих набор дискретных эмоциональных состояний, называемых базовыми (гнев, раздражение, страх, радость, печаль, удивление и нейтральное состояние).

В качестве возможных информативных признаков был выделен ряд признаков: оценка мощности, частота основного тона (ЧОТ), асимметрия от медианы ЧОТ, линейные спектральные частоты, кепстральные коэффициенты, вычисленные по коэффициентам предсказания, статистики высшего порядка, энергетический оператор Тигера [2].

Количество записей в обучающей выборке при этом составляло 80 штук — 40 записей для нейтрального состояния, и 40 для состояния гнева. Тестирование алгоритма позволило выявить ряд информативных признаков, эффективность классификации эмоциональной речи по которым оказалась максимальной. Наиболее значимыми для принятия решения о принадлежности записи к классу нейтрального состояния и состояния гнева информативными признаками оказались частота основного тона, 2ой коэффициент линейных спектральных частот, вторая производная оценки мощности и эксцесс ошибки линейного предсказания.

Для выбранного набора параметров, оценки точности работы алгоритма классификации оказались следующими. Ошибка классификации в случае использования метода опорных векторов составила порядка 4 %. При применении модели смеси нормальных распределений соответствующее значение оказалось равным примерно 6 %.

Таким образом, использование выделенных информативных признаков позволило распознавать с точностью порядка 94–96 % в зависимости от используемого алгоритма классификации.

В состоянии гнева диктор издает звуки с более открытым речевым трактом, что приводит к возрастанию средней частоты первой форманты. В состоянии гнева существует влияние на характеристики частоты основного тона. Возрастает его медианное значение, скорость изменений, расширяется его диапазон. Так же, по отношению к ней, возрастают амплитуды второй и третьей формант, повышается неоднородность формантных контуров. Кроме частотных параметров голоса важную роль играют характеристики огибающей его энергии.

Увеличение числа распознаваемых эмоций, равно как и переход от модельных эмоциональных баз данных к реальным, необходимо ведет к возрастанию ошибки классификации. Точность распознавания одного из семи эмоциональных состояний эмоциональной речи порядка 89 % .

Список источников

1. Cornelius R. R. 1996. The Science of Emotion: Research and Tradition in the Psychology of Emotions.
2. Morrison D., Wang R., De Silva L. C. 2007. Ensemble Methods for Spoken Emotion Recognition in Call-Centres. Speech Communication, 4 : 98–112. Voice emotion classification: problems and solutions 195