



Об автоматическом определении эмоций по речи

Киселёв В.В.,

директор ООО «Речевые технологии», г. Минск

Возможность автоматически определять эмоции по голосу и речи человека необходима для развития успешных диалоговых систем. Идентификация эмоционального состояния человека востребована в телекоммуникационной сфере, в индустрии развлечений, обучении, медицине и других сферах. В статье представлен краткий обзор подходов к исследованию эмоционального состояния человека по его речи, а также приведены примеры реализованных программ для определения эмоций.

Abstract

Possibility of automatic detection emotions a voice and speech of speakers is necessary for the development of dialog systems. Identification of emotion of person is claimed in telecommunications, business solutions, training, medicine and other spheres. In this paper the review of some approaches to research of emotional condition of speaker speech and as examples of realized programs for detection emotions are resulted is presented.

Эмоции и речь тесно взаимосвязаны и играют огромную роль в общении, поэтому автоматическая и объективная диагностика эмоционального состояния человека по его речи представляет большой практический интерес. Возможность распознавать эмоции в речи важна как для исследования самой речи и эмоций, так и для улучшения качества обслуживания клиентов, например в контакт и call-центрах. Различные научные и коммерческие организации занимаются исследованием этого феномена.

Что же такое эмоции? Существует множество определений. Вот некоторые из них. Эмоции — сильные психические состояния, связанные обычно с возбуждением или высоким уровнем энергии и дающие начало чувствам и страстям. Также чувствами или эмоциями называют переживание человеком своего отношения к тому, что он познаёт и делает, к другим людям и к самому себе.

Эмоции бывают положительными или отрицательными. Удивление, эйфория, гнев, страх различаются по степени положительности либо отрицательнос-

ти. Эмоции дают нам информацию о том, как говорящий оценивает ситуацию и какие ответные реакции можно от него ожидать. Известно, что мысли и эмоции влияют на дыхание, выражение лица, положение тела, тон и темп голоса. Голос выражает любые сильные эмоции, он принимает музыкальный мелодичный характер, меняясь по громкости, тембру и высоте звука. Темп речи и её ритмическое членение с помощью пауз и логического ударения также имеют выразительное значение, помогающие уловить эмоции человека.

Важный канал для опознания эмоционального состояния человека — его речь. Она передаёт самые тонкие, деликатные эмоции. Скорость речи зависит от индивидуальных качеств и намерений говорящего. Тембр зависит от того, что говорит человек, какие чувства испытывает. Так, при раздражении тембр прерывисто-царапающий, при апатии — лениво-глухой, при радости — звонко-здоровый, при недоумении — оловянно-нерешительный, при гневе — прерывисто-разрывающий [1].

Т.В. Корнева и Е.Ф. Бажин ещё в 1977 г. установили, что различия в точности распознавания эмоций по голосу связаны в основном с модальностью эмоций [2]. Наименьшее количество ошибок при такой оценке испытуемые получили при идентификации гнева и ровного настроения. Средний балл их опознания в процентном соотношении составил соответственно 99,3 и 97,0. Другие эмоции оценивались хуже. Так, средний балл опознания сниженного настроения равнялся 75,8; тревоги — 81,4; апатии — 80,7; повышенного настроения — 79,5.

В.П. Морозов в 1991 г. ввёл термин «эмоциональный слух» — способность опознавать эмоции по речи и пению человека [3]. Между эмоциональным слухом и речевым слухом отсутствует корреляция. «Эмоциональная глухота» может встречаться и у людей с хорошо развитым восприятием речи. Любопытные данные были получены в отношении точности распознавания эмоций людьми разного возраста, пола и профессий. Испытуемые показали существенные различия в правильности понимания эмоций — от 10 до 95%. Выявлено, что музыканты и вокалисты обладают более развитым эмоциональным слухом. В связи с этим эмоциональный слух стал рассматриваться как один из критериев художественной одарённости, который стал использоваться на приёмных экзаменах в консерваторию.

Исследование В. Х. Манерова (1993) идентификации эмоций по речи показало, что основным признаком, используемым человеком при слуховом восприятии эмоционально обусловленных изменений речи, является степень речедвигательного возбуждения [4]. Определение вида эмоции, переживаемой говорящим, осуществляется слушающим менее успешно, чем определение степени эмоционального возбуждения. Наиболее точно опознаются базовые эмоции, затем удивление и неуверенность и хуже всего — презрение и отвращение. На точность опознания влияет способность диктора передавать в речи эмоциональные состояния. Существует тенденция лучшего распознавания положительных эмоций по сравнению с индифферентным и отрицательным эмоциональными состояниями.

Информация, используемая человеком при определении эмоций других людей, связана с так называемыми «когнитивными схемами эмоций», т.е. с установлением того набора признаков, с помощью которого можно судить о наличии той или иной эмоции. Сопоставление совокупности наблюдаемых признаков со схемой позволяет идентифицировать эмоцию.

При этом предполагается, что ни один из признаков жёстко не привязан к определённой эмоции, а её идентификация осуществляется на вероятной основе. Эмоции других людей



распознаются по внешним проявлениям эмоций: изменению речи и голоса, поведения, ответные реакции. Учитываются также antecedentes, т.е. то, что предшествует и является причиной эмоций: ситуация в её взаимодействии с имеющейся у человека целью [5].

Учёные университета Эль-Пасо (США) выбрали для изучения такое понятие, как **уровень уверенности высказывания**, т.е. насколько уверенно говорящий произносит то или иное высказывание. Идея исследования заключалась в том, чтобы создать модель прогнозирования уровня уверенности. Высказывания, на которых практиковалась модель, — это высказывания различных уровней уверенности, они взяты из речи носителей английского языка [6].

Уровень уверенности говорящего определяется **тоном и высотой голоса**:

- **явно высокий** — энтузиазм, радость, заинтересован и проявляет интерес;
- **высокий**, в широком диапазоне силы, тональности и высоты — гнев и страх, неуверенность;
- **чрезмерно высокий**, пронзительный — беспокойство;
- **мягкий и приглушённый**, с понижением интонации к концу каждой фразы — печаль, усталость;
- **форсирование звука** — напряжение, обман;
- в состоянии эмоционального возбуждения обычно *возрастает сила голоса*, изменяются его высота и тембр, но иногда сильное возбуждение может, наоборот, проявляться в *уменьшении силы голоса* (человек «шипит от ярости»).

Учёные Саутгемптонского университета (Великобритания) разработали компьютерные методы, позволяющие прогнозировать **ответную эмоциональную реакцию** говорящего. В ходе эксперимента было выявлено, что просодическая информация помогает в автоматическом определении степени раздражённости человека. Наиболее полную информацию о внутреннем психоэмоциональном состоянии человека может дать анализ его связной речи: расстановка логических ударений, скорость произнесения слов, конструкция фразы, наличие таких отклонений от нормы, как неуверенный или неверный подбор слов, обрывание фраз на полуслове, изменение слов, появление слов-паразитов, исчезновение пауз и т.д.

В результате было выделено:

- **быстрая речь** — очевидная взволнованность, страстное желание убедить или уговорить кого-то;
- **медленная речь** — высокомерие, усталость, угнетённое состояние;
- **прерывистая речь** — неуверенность;
- **лаконичность и решительность речи** — явная уверенность;
- **заикание** — напряжённость или обман;
- **нерешительность в подборе слов** — неуверенность в себе или намерение внезапно удивить чем-то;
- **появление речевых недостатков** (повторение или искажение слов, обрывание фраз на полуслове) — несомненное волнение, но иной раз и желание обмануть;
- **опускание речевых пауз** — напряжение;

- *слишком удлинённые паузы* — незаинтересованность или несогласие;
- *появление в речи пауз, заполняемых словами-паразитами* — нерешительность и затруднение в выражении мысли, поиск выхода из положения;
- *возрастание числа тривиальных наборов слов*, проговариваемых быстрее, чем обычно, — эмоциональное возбуждение, напряжение;
- *умолкание или скудость в словах* — обида.

Распознавание эмоционального состояния человека представляет огромный интерес. Проблема автоматического распознавания эмоционального состояния говорящего по голосу на данный момент не решена. Существующие системы различаются списками распознаваемых эмоций, типами используемых баз данных, акустическими параметрами и их производными, а также алгоритмами классификаторов; эти различия делают результаты распознавания напрямую несопоставимыми.

Так как эмоции и мысли влияют на дыхание, выражение лица, положение тела, тон и темп голоса, то определять эмоции можно по выражению лица, по речи и голосу. Уже созданы некоторые программы для определения эмоций по выражению лица. Так, например, учёные из Университета короля Хуана Карлоса (Испания) разработали систему, способную различать выражения лиц в режиме реального времени. На скорости 30 кадров в секунду программа анализирует выражение лица человека и классифицирует его в соответствии с шестью заложенными в неё шаблонами: гнев, отвращение, страх, счастье, печаль и удивление. Анализу может подвергаться как лицо целиком, так и его часть. Для идентификации выражения лица система использует базу данных Sohn-Kanade, содержащую 333 варианта выражения лиц различных людей. Вероятность совпадения с базой — 89%. Система может работать и в неблагоприятных условиях, на неё не влияет ни изменение освещённости, ни движение пользователя [7].

Создаются компьютерные программы, позволяющие определять эмоции по речи человека. Проводятся работы по компьютерному детектору эмоций по голосу (**Voice-Stress Analysis**) на основе анализа стресса. Такие современные системы находят применение в США в государственных и правоохранительных органах [8].

Создана ещё одна интересная компьютерная программа, позволяющая выявить и проанализировать в диалоге эмоциональное состояние собеседника по его речи — **детектор любви**. Научно доказано наличие глубинной связи между чувствами человека и особенностями его речи. Богатая палитра эмоций и оттенков настроения выражается в тончайших модуляциях нашего голоса. А эта компьютерная программа анализирует особенности голоса, исследует диапазон эмоций говорящего, определяет степень концентрации внимания, уровень смущения и волнения [9].

В 2006 году один из южнокорейских операторов запустил мобильный сервис анализа голоса, который основан на системе голосового анализа и действует как детектор эмоций, делая заключения **об уровне честности** участников разговора. В течение разговора анализируются различные звуки, которые попадают в микрофон абонента, и делается заключение об их эмоциональном статусе. В конце разговора абонент получает сообщение с графиком правдивости, где показан уровень стресса и число неточных ответов и попыток сменить тему. Происходит анализ, который учитывает, как определённая мозговая активность влияет на специфические особенности голоса. Это позволяет определить и измерить широкий спектр эмоций, используя различные оценки составляющих эмоций, строить оценку правдивости любого утверждения, сделанного участниками разговора [10].



Среди коммерческих организаций, активно использующих и разрабатывающих автоматические модули оценки эмоционального состояния, можно выделить такие компании, как Nemesysco Ltd. [11], Nice Systems Ltd [12], Центр Речевых технологий [13]. Модули нашли своё практическое применение в call-центрах при анализе разговоров как оператора, так и клиента. Как правило, компании не ограничиваются двумя эмоциональными состояниями. Например, компания Nemesysco Ltd. может опознавать до 16-ти эмоциональных состояний с различными числовыми значениями каждого состояния — от удовлетворённости, расстройства или злости до сомнения или неуверенности. Другие компании используют комплексный анализ по голосу и речи, применяя языко-зависимые технологии поиска ключевых слов.

Системы, распознающие эмоциональное состояние человека, могут быть применены в интерактивном телевидении, виртуальном обучении, при исследовании нарушений функций мозга, а также будут полезны людям, имеющим какие-либо речевые отклонения. Для развития успешных диалоговых систем необходимы исследования по выявлению эмоций человека по его речи. Понимание эмоций другого человека важно как для общения между людьми, так и при взаимодействии человека с системами искусственного интеллекта. Автоматическое распознавание речи и прогнозирование эмоций говорящего нашли бы активное применение, например, в телекоммуникационной сфере и индустрии развлечений, что помогло бы избегать конфликтных ситуаций и улучшить качество обслуживания клиентов.

Литература

1. Хаббард Л. Рон. Свободный человек // Способность. № 232.
2. <http://www.emotionlabs.ru/content/66/>.
3. <http://cons-help.com/63/>.
4. Джемс В. Психология. Часть II СПб: Изд-во К.Л. Риккера, 1911. С. 323–340.
5. Манёров В.Х., Шнейдер Е.М. Автоматическое распознавание эмоций по спектральным и интонационным признакам // Материалы доклада и сообщения 5-го Все-союзного совещания-симпозиума цикла «Акустика речи и слуха». Одесса, 1989.
6. Frijda N.H. (1986). The emotions. Cambridge: Cambridge University Press.
7. Jaime C. Acosta and Nigel G. Ward. Responding to User Emotional State by Adding Emotional Coloring to Utterances. In Twelfth International Conference on Spoken Language Processing. ISGA, 2009.
8. <http://www.voicestressanalysis.net/>
9. <http://www.membrana.ru/lenta/?26699>
10. <http://www.ukrpolygraph.org/2006/09/28/90>
11. <http://www.nemesysco.com>
12. www.nice.com
13. <http://www.speechpro.ru>