

ЭМОЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ РЕЧИ И ИХ СВЯЗЬ С АКУСТИЧЕСКИМИ ПАРАМЕТРАМИ

Проблемы невербальной коммуникации и, в частности, передача эмоционального содержания речи с помощью кодирования ее акустических параметров, чрезвычайно актуальны. Результаты по установлению этих взаимосвязей лежат в основе программ автоматического распознавания, кодирования и синтеза эмоций, моделирования эмоционально-окрашенной речи и др. Представлен обзор исследований этой проблемы, проводимых в течение более полутора веков. Описаны взаимосвязи эмоций с изменениями таких параметров, как частота основного тона, долговременный статистический спектр, темпо-ритмические характеристики и др.

Ключевые слова:

акустические параметры, невербальная коммуникация, речевой сигнал, эмоциональные характеристики.

Речевой сигнал предназначен прежде всего для воспроизведения смысловой (вербальной) информации. Однако он передает слушателю также и значительную невербальную (неречевую) информацию. Под термином невербальная коммуникация понимается «система неязыковых (несловесных) форм и средств передачи информации» [2]. Такая информация не всегда соответствует вербальной. В исследованиях, проведенных учеными А. Mehrabian и S. Ferris в 1967 году [13], в которых противопоставлялись три вида информации: вербальная, невербальная и мимическая, показали, что в случае противоречия вербальной и невербальной информации, восприятие основывается преимущественно на мимической информации (55%) и невербальной информации (38%). И по этим результатам, и по последующим исследованиям можно сказать, что невербальная информация оказывает большое влияние на восприятие.

Изучение различных видов невербальной информации имеет огромную научную ценность для автоматического распознавания речи, для разработки учебных методик изучения языка, для распознавания личности по голосу, для изучения коммуникативных процессов и т.д. К проблемам невербальной коммуникации обращаются многие современные ученые как в России, так и за рубежом. В.П. Морозов дает классификацию различных видов невербальной информации (эмоциональная, эстетическая, индивидуально-личностная, биофизическая, социально-групповая, психологическая, пространственная, медицинская и др.) [2]. Эмоциональная информация, характеризующая внутреннее состояние

человека в процессе общения (гнев, раздражение, счастье, печаль, страх, удивление, и мн. др.), является одной из важнейших. Она не только помогает пониманию вербальной информации, но и дополняет смысл, который приобретает эта информация в речи конкретного индивида.

Один из главных вопросов в изучении эмоционального содержания речи – восприятие одних и тех же эмоций людьми разных культур и национальностей. Так, E.G. Veier и A.J. Zautra [5, p. 166] проводили записи речевых фрагментов, исполненных американскими актерами с различным эмоциональным содержанием (гнев, грусть, кокетство, страх и безразличие). Потом некоторые из этих записей предъявлялись для прослушивания экспертам – людям разных национальностей (американцам, полякам и японцам). Результаты показали, что эмоции, с которыми произносились речевые фрагменты, распознаются с высокой долей правильных оценок экспертами разных национальностей. Похожие эксперименты проводились и другими учеными (см., напр. [4]), которые пришли к аналогичным результатам. Эксперименты с другими видами невербальной коммуникации, например, визуальной (по фотографиям людей, выражающих разные эмоции: радость, удивление, отвращение, печаль, страх и гнев) показали, что представители разных языковых культур, также имеют высокий процент совпадающих оценок [8]. Таким образом, можно считать, что эмоции универсальны, а способы их выражения можно считать общечеловеческими. Исследования, проведенные J. Pittam и K.R. Scherer показали, что вероятность распознавания эмоций по голосу составляет около 60%,

хотя она варьирует в зависимости от типа эмоций и уровня развития эмоционального слуха у человека (способность к определению эмоционального состояния говорящего по звуку его голоса [3]) [14, р. 190].

Выражение эмоций необходимо и в искусстве (актеры, певцы и др.), и в межличностном общении (напр., лекторам, педагогам, политикам). С помощью невербальной информации певцы выражают содержание произведения, актеры показывают свое отношение к представляемому ими герою, дикторы на радио передают свою оценку читаемого материала и т.д.

Проблема взаимосвязи акустических характеристик с эмоциональным содержанием речи представляет большой научный интерес, как в теоретическом плане, так и для решения различных прикладных задач, в том числе определения объективного состояния человека по звучанию его голоса в различных сферах деятельности, в частности, в психологии, в маркетинговом бизнесе, в криминалистике, в медицине. Особое значение решение этой проблемы имеет в современном искусстве звукорежиссуры. Звукорежиссер может влиять на эмоциональное наполнение речи, делая ее более насыщенной, или наоборот, менее выразительной. Необходимые знания в области психологии, музыкальной акустики и техники позволяют ему управлять акустическими характеристиками сигнала, с помощью которых он способен совершать качественные преобразования эмоционального содержания. Звукорежиссер способен и переосмысливать вербальную информацию, привнося в нее собственный эмоциональный контекст.

Распознаванием эмоций в речи и пении ученые начали заниматься в XIX в. Основы данной проблемы были заложены Ч. Дарвином. В своих наблюдениях он изучал сходства между выражением эмоций у человека и животных, а также – непосредственно вопросы передачи эмоций с помощью голоса как основного канала эмоциональных сигналов у животных и у человека. Дарвин отмечает, что «высота голоса имеет отношение к определенным состояниям чувства» [1, с. 83]. Современник Дарвина Herbert Spencer также обращается к проблеме голосового выражения эмоций. Он показывает, как человеческий голос может меняться в зависимости от ситуации – по громкости, по тембру, высоте звука и по длительности пауз [17].

Первые научные исследования взаимосвязи акустических параметров с эмоци-

ональным содержанием речи были приняты в первой половине XX в. Речевой сигнал кодируется и передается в виде акустического сигнала (звуковой волны). Основные акустические параметры речевого сигнала следующие: основная частота колебаний голосовых связок (частота фонации), формантная структура (расположение, амплитуда и ширина формантных областей), уровень звукового давления, динамика изменения частоты основного тона (т.е. мелодика речи), динамический спектр звука (оказывает существенное влияние на тембр голоса), темпо-ритмические особенности речи, энергетические характеристики звукового сигнала и их изменение во времени и др.

Эксперименты проводились прежде всего с «базовыми» эмоциями («гнев», «радость», «печаль», «страх», «раздражение»). К числу таких исследований принадлежат работы E.W. Scripture [15], M.R. Bonner [6], M. Isserlin [10], E.R. Skinner [16] и др. В 1939 г. G. Fairbanks и W. Pronovost [9] предложили наиболее полное на то время описание связи пяти базовых эмоций с акустическими характеристиками:

– «гнев» – частота фонации, а также первая форманта увеличиваются в отличие от нейтральной (безэмоциональной) речи, при этом артикуляция становится очень отчетливой;

– «радость» – частота фонации становится выше, чем в нейтральной речи, артикуляция остается обычной;

– «страх» – частота фонации снижается по сравнению с «гневом», контур ее изменения содержит резкие пики;

– «печаль» – имеют место малые вариации в частоте фонации, она монотонно спадает в конце фразы. Артикуляция становится медленной;

– «раздражение» – частота фонации снижается, по отношению к нейтральной речи, артикуляция сохраняется обычной.

Представленные результаты были первым этапом в исследованиях и отражали в основном связь эмоций с частотой фонации.

Во второй половине XX в. в ходе экспериментов удалось установить количественные связи эмоций в речи с ее акустическими характеристиками. С. Williams. и К. Stevens осуществили измерения основных акустических параметров речевого сигнала, записанных при чтении текстов актерами с различным эмоциональным содержанием («гнев», «печаль», «страх» и «нейтраль») [19]. Анализ полученных результатов позволил выявить следующие

основные закономерности, связанные: с изменением частоты основного тона, с изменением долговременного статистического спектра и с темпо-ритмическими характеристиками речи.

Изменения частоты основного тона могут быть определены с помощью следующих величин:

а) **Контур частоты основного тона – Fo cont.** (динамика изменения положения частоты основного тона за определенный отрезок времени, в данном примере – 1,5 сек). Для «нейтраль» характерна ровная и постоянная форма контура, для «гнева» – значительно более высокая и выпуклая, с резкими пиками, для «печали» – относительно плоская и несколько неустойчивая, для «страха» – подобная контуру «нейтраль», но очень изрезана, иногда с перепадом частоты (рис. 1).

б) Среднее значение частоты основного тона – **Fo mean.** Для «нейтраль» среднее значение частоты основного тона – 120 Гц, для «печали» оно понижается до 100 Гц, при «страхе» оно соответствует 140 Гц, при «ярости» поднимается до 200 Гц.

в) Диапазон изменения частоты основного тона – **Fo range.** Самый широкий диапазон изменения фундаментальной

частоты характерен для гнева – он составляет около 140 Гц, диапазоны изменения фундаментальной частоты при «нейтраль» и «печали» похожи – около 60 Гц, но среднее значение частоты основного тона для «нейтраль» выше, при «страхе» диапазон изменений фундаментальной частоты около 80 Гц.

Изменения долговременного статистического спектра. При «печали» в голосе наблюдается наименьшее количество обертонов, и резкое уменьшение высокочастотной энергии (до –10 дБ по сравнению с «нейтралью»), при «ярости» наибольшее количество обертонов и увеличение высокочастотной энергии (до 5 дБ по сравнению с «нейтралью»). При выражении «страха» долговременный статистический спектр похож на «нейтраль» (рис. 2).

Темпо-ритмические характеристики речи. Число звуков в секунду составляет 4,31 для нейтральной речи, 4,15 – для «ярости», 3,8 – для «страха» и 1,91 – для «печали». При «страхе» ударные гласные – короткие, при «ярости» увеличивается длительность согласных.

Таким образом, были получены следующие характеристики эмоциональных состояний:

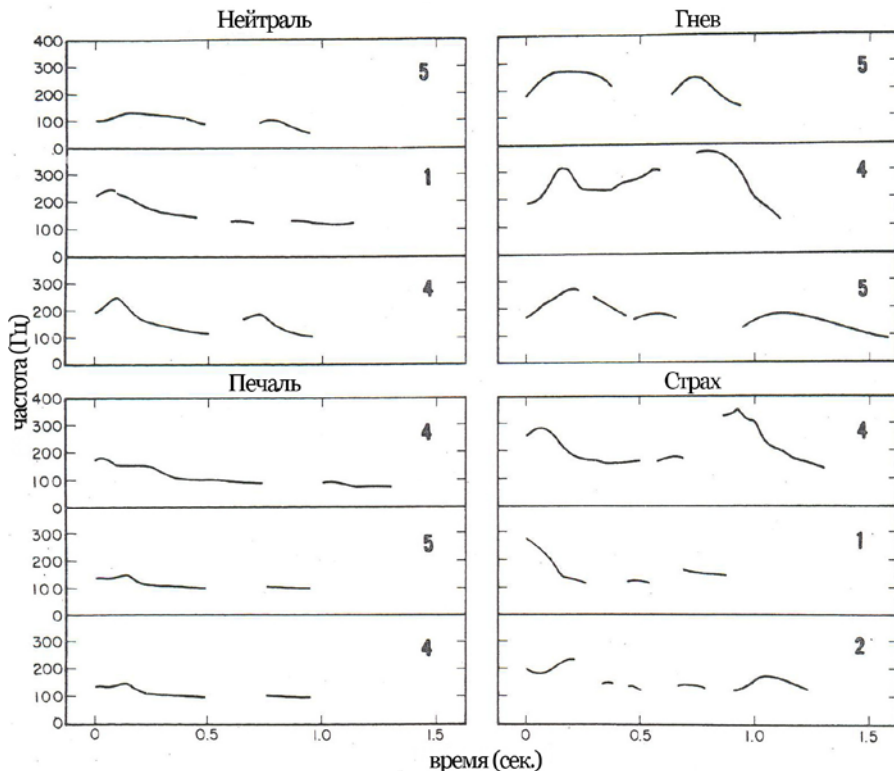


Рис. 1. Контур частоты основного тона для речи с разным эмоциональным содержанием.

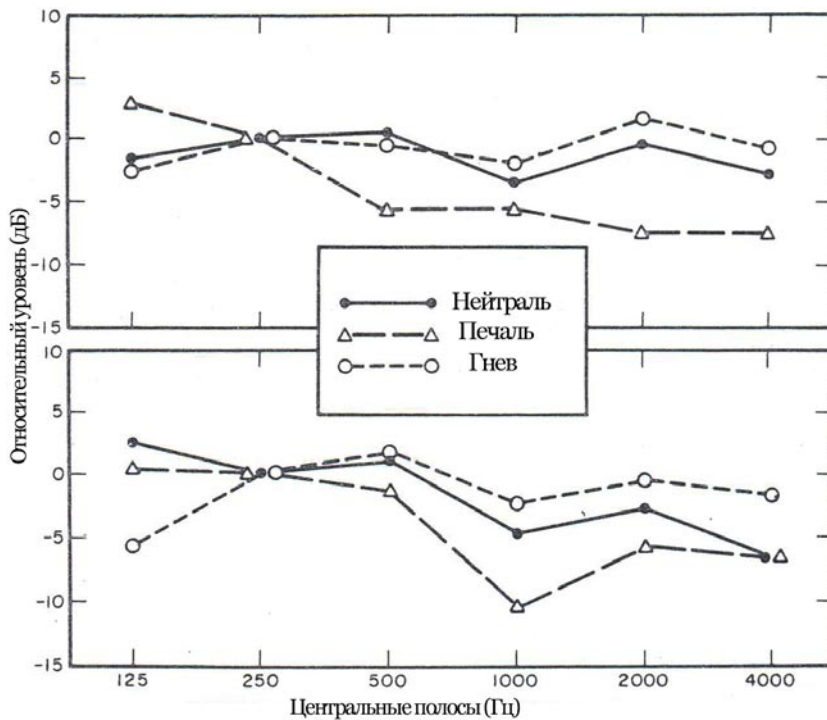


Рис. 2. Долговременный статистический спектр для речи с разным эмоциональным содержанием.

– *гнев* – высокое положение частоты фонации, на пол-октавы выше, чем при «нейтральности». Диапазон изменения фундаментальной частоты намного шире, чем при «нейтральности». В слогах с возрастающей интенсивностью – гласные имеют наивысшее среднее значение фундаментальной частоты (190 Гц), эти слоги имеют также слабо выраженную первую форманту и часто сформированы с небольшой голосовой иррегулярностью (нерегулярными колебаниями голосовых связок). Артикуляция отчетливая: гласные звуки произносятся открыто, поэтому имеют высокую первую форманту, согласные более закрытые. Длительность слогов в гневе всегда дольше.

– *страх* – среднее значение фундаментальной частоты ниже, чем для «гнева» и для некоторых голосов близко к «нейтральности», однако пики основной частоты значительно выше, чем при «нейтральности», и контур основной частоты около пиков иногда имеет необычную форму, возникает небольшая голосовая иррегулярность. Длительность слогов дольше, чем для «нейтральности» или «гнева».

– *печаль* – среднее значение фундаментальной частоты ниже, чем для «нейтральности» и диапазон изменений частоты основного тона достаточно узкий.

– *нейтраль* – спектрограммы голосов актеров показывают четкую, определенную структуру гласных, с небольшим шумом и иррегулярностью между формантами в высокочастотной области. Безударные согласные неакцентированы.

К началу XXI в. количество исследований в области эмоционального содержания речи и ее взаимосвязи с акустическими характеристиками значительно возросло. Ученые изучали естественные речевые фрагменты – образцы записи речи радиокomentаторов [18], отрывков телевизионных интервью [7], психотерапевтических сеансов. Исследования, проводимые на рубеже XX–XXI вв. значительно расширили набор акустических характеристик эмоциональной речи. В результатах исследований [12; 20] появляются такие параметры, как:

– *джиттер* – частотная модуляция основного тона. Измеряется на стабильном фрагменте продолжительного гласного звука. Глубина модуляции для нейтральной речи равна 1,9%, для гнева – от 2,8% до 9,5%, для счастья – 4,5%, для грусти – 1,8%.

– *шиммер* – амплитудная модуляция на основном тоне. Измеряется на стабильном фрагменте продолжительного гласного

звука. Глубина модуляции для нейтрала составляет 0,9 дБ, для гнева – от 1,2 дБ до 1,7 дБ, для счастья – 1,2 дБ, для грусти – 0,8 дБ.

Р. Juslin и Р. Laukka систематизировали работы по взаимосвязи эмоций с акустическими характеристиками речи – по каждой из них выделены эмоции и соответствующие акустические характеристики, рассматриваемые в исследованиях, уточнена методика, язык, на котором проводились исследования, соотношение участников экспериментов и слушателей/экспертов, характеристики исследуемого материала [11].

В настоящее время проводятся многочисленные эксперименты по описанию полного каталога эмоций. Данные основного набора акустических характеристик определенных эмоций постоянно пополняются, расширяется экспериментальная база, создаются новые программы для анализа акустических составляющих эмоционального содержания.

Исследования, проводимые более полтора веков и посвященные взаимосвязи акустических характеристик с эмоциональным содержанием речи и пения, привели к появлению множества программ автоматического распознавания, кодирования и синтеза эмоций, моделирования эмоционально-окрашенной речи. Поиск количественной связи эмоциональных характеристик голоса с акустическими параметрами, которыми они кодируются, является важным аспектом дальнейшего развития электронных искусств и, в частности, главным инструментом, позволяющим звукорежиссеру управлять эмоциональностью речи и пения. Основные задачи дальнейших исследований в этой области – получение точных и непротиворечивых результатов экспериментов, расширение круга исследуемых эмоций, создание новых программ позволяющих детально анализировать все акустические составляющие речевого сигнала и разработка полных русских эмоциональных баз.

Список литературы:

1. Дарвин Ч. О выражении эмоций у человека и животных. – СПб.: Питер, 2001.
2. Морозов В.П. Искусство и наука общения: невербальная коммуникация. – М.: ИП РАН, Центр «Искусство и наука», 1998.
3. Морозов В.П. Эмоциональный слух человека // Журнал эволюционной биохимии и физиологии. – 1985, № 6.
4. Экман П. Психология эмоций. Я знаю, что ты чувствуешь / 2-е изд. / Пер. с англ. – СПб.: Питер, 2010.
5. Beier E.G., Zautra A.J. Identification of vocal communication of emotions across cultures // Journal of Consulting and Clinical Psychology. Vol. 39, Issue 1. – 1972, August.
6. Bonner M.R. Changes in the speech pattern under emotional tension // Amer. J. Psychol. – 1943, N 56. – 262–273.
7. Douglas-Cowie E., Cowie R., Schroder M. A new emotion database: considerations, sources and scope / ISCA Workshop on Speech & Emotion. – Northern Ireland, 2000. – P. 39–44.
8. Elfenbein H.A., Ambady N. On the universality and cultural specificity of emotion recognition: A meta-analysis // Psychological Bulletin. – 128, (2002b). – P. 205–235.
9. Fairbanks G., Pronovost W. An experimental study of the pitch characteristics of the voice during the expression of emotion // Communication Monographs. Vol. 6. – 1939, Issue 1. – P. 87–104.
10. Isserlin M. Psychologisch-phonetische Untersuchungen // Zeitschrift für die gesamte Neurologie und Psychiatrie. Vol. 94. – 1925, N 1. – P. 437–448.
11. Juslin. P.N., Laukka P. Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? // Psychological Bulletin. Vol. 129. – 2003, N 5. – P. 770–814.
12. Leitman D.I., Wolf D.H., Ragland J.D., Laukka P., Loughhead J., Valdez J.N., Javitt D.C., Turetsky B.I., Gur R.C. «It's not what you say, but how you say it»: a reciprocal temporo-frontal network for affective prosody // Frontiers in human neuroscience. – 2010. – Интернет-ресурс. Режим доступа: http://www.frontiersin.org/Journal/FullText.aspx?ART_DOI=10.3389/fnhum.2010.00019&name=human_neuroscience&x=y
13. Mehrabian A., Ferris S. Inference of attitudes from nonverbal communication in two channels // Journal of Consulting Psychology. – 1967, N 31. – P. 248–252.
14. Pittam J., Scherer K.R. Vocal expression and communication of emotion / Handbook of emotions. – 1993.
15. Scripture E.W. A study of emotion by speech transcription // Vox. – 1921, N 31.
16. Skinner E.R. A calibrated recording and analysis of the pitch, force, and quality of vocal tones expressing happiness and sadness // Communication Monographs. Vol. 2. – 1935, Issue 1. – P. 81–137.
17. Spencer H. The Origin and Function of Music // Fraser's Magazine. – 1857, N 56. – P. 396–408.
18. Trouvain J., Barry W.J. The prosody of excitement in horse race commentaries // Proc. ISCA-Workshop on «Speech and Emotion» Newcastle. – Northern Ireland, UK, 2000. – P. 86–91.
19. Williams C.E., Stevens N.K. Emotions and Speech: Some Acoustical Correlates // J. Acoust. Soc. Am. Vol. 52. – 1972, Issue 4B. – P. 1238–1250.
20. Whiteside. S.P. Simulated emotions: an acoustic study of voice and perturbation measures / 5th International Conference on Spoken Language Processing, Sydney, Australia. – 1998. – Интернет-ресурс. Режим доступа: <http://www.shlrc.mq.edu.au/proceedings/icslp98/PDF/AUTHOR/SL980153.PDF>