

# ГОЛОС И ПСИХОЛОГИЧЕСКИЕ СВОЙСТВА ЧЕЛОВЕКА: ОБЗОР СОВРЕМЕННЫХ ИССЛЕДОВАНИЙ

А.С. РАФИКОВА<sup>а</sup>, Е.А. ВАЛУЕВА<sup>б</sup>, А.С. ПАНФИЛОВА<sup>б</sup>

<sup>а</sup> Государственный академический университет гуманитарных наук, 119049, Москва, Мароно-  
вский пер., д. 26

<sup>б</sup> ФГБУН Институт психологии РАН, 129366, Москва, ул. Ярославская, д. 13, к. 1

## Voice and Psychological Characteristics: A Contemporary Review

A.S. Rafikova<sup>a</sup>, E.A. Valueva<sup>b</sup>, A.S. Panfilova<sup>b</sup>

<sup>a</sup> State Academic University for the Humanities, Maronovskiy Pereulok, 26, Moscow, 119049, Russian Federation

<sup>b</sup> Institute of Psychology, Russian Academy of Sciences, 13 build. 1, Yaroslavskaaya Str., Moscow, 129366, Russian Federation

### Резюме

В статье представлен обзор работ по изучению взаимосвязи психологических характеристик и акустических свойств голоса. Приводятся историческая справка развития данной области исследований, основные аудиохарактеристики речи, используемые для анализа голоса в современных исследованиях (частота основного тона, интенсивность и скорость речи), данные исследований об акустико-просодических характеристиках эмоций, намерений и личностных свойств (лидерские качества, харизматичность, диспозиции Большой пятерки). Показано, что негативные эмоции по голосу распознаются лучше, чем положительные, наиболее точно распознаются гнев и грусть. Рассматриваются аудиокорреляты таких измерений эмоций, как валентность и активация. Описаны акустико-просодические характеристики не только базовых эмоций (гнев, счастье, грусть, страх, отвращение, удивление),

### Abstract

This paper provides an overview of studies examining the relationship between psychological characteristics and properties of voice. The article presents a historical overview of studies in this field, the main audio characteristics of speech used for voice analysis in modern researches (pitch, intensity and speed), research data on the acoustic and prosodic characteristics of emotions, intentions and personality traits (leadership qualities, charisma, Big Five personality traits). It is shown that negative emotions are recognized by voice better than positive ones, anger and sadness are most accurately recognized. The audio correlates of such measurements of emotions as valence and activation are considered. Acoustic-prosodic characteristics are described not only for basic emotions

но и эмоций, имеющих менее интенсивный характер, – раздражение, уступчивость, безразличие. Так же как и эмоциональная речь, комбинацией акустических свойств обладают невербальные вокализации, которые тоже могут быть спрогнозированы на основе их физических свойств. Показано, что восприятие харизматичной речи обусловлено сочетанием просодических и лексических свойств речи и роль аудиальных характеристик в распознавании харизматичности ниже, чем при распознавании эмоций, где более заметную роль играют именно голосовые характеристики, нежели лексическое содержание. При оценивании черт Большой пятерки с помощью экспертных оценок у экстраверсии было обнаружено наибольшее число значимых корреляций с акустико-просодическими и аудиальными признаками. Рассматриваются исследования в области автоматического распознавания личности и эмоций по голосу с помощью машинного обучения и нейросетей, где в последнее время наблюдается интерес к данной теме. Современные технологии позволяют предсказывать личностные свойства с большой точностью, однако работы в этой области редко бывают основаны на той или иной теоретической модели. В заключение подчеркивается необходимость теоретического осмысления получаемых эмпирических результатов.

*Ключевые слова:* акустические свойства речи, Большая пятерка, личностные черты, эмоции, автоматическое распознавание личности, нейросети.

**Рафикова Антонина Семеновна** – научный сотрудник, Государственный академический университет гуманитарных наук, кандидат психологических наук.

Сфера научных интересов: психолингвистика, когнитивная психология.

Контакты: antoninaraf@yandex.ru

**Валуева Екатерина Александровна** – научный сотрудник, лаборатория психологии и психофизиологии творчества, ФГБУН «Институт психологии Российской академии наук», кандидат психологических наук.

Сфера научных интересов: когнитивная психология, интеллект, творчество.

Контакты: ekval@list.ru

(anger, happiness, sadness, fear, disgust, surprise), but also of emotions that are of a less intense nature - irritation, resignation, indifference. Just like emotional speech, nonverbal vocalizations have a combination of acoustic properties, which can also be predicted based on their physical properties. It is shown that the perception of charismatic speech is due to a combination of prosodic and lexical properties of speech, and the role of auditory characteristics in the recognition of charisma is lower than in the recognition of emotions, where voice characteristics play a more prominent role than lexical content. When evaluating the Big Five traits using expert assessments, extraversion was found to have the largest number of significant correlations with acoustic-prosodic and auditory traits. Automatic personality and emotion recognition (using machine learning and neural networks) is a research area, where we can see the burst of empirical studies. Computer technologies assure a high degree of accuracy in personality prediction, but such studies are rarely deeply theoretically grounded. In conclusion, the need for theoretical understanding of the empirical results obtained is emphasized.

*Keywords:* acoustic speech features, Big Five, personality traits, emotions, automatic personality recognition, neural networks.

**Antonina S. Rafikova** – Research Fellow, State Academic University for the Humanities, PhD in Psychology.

Research Area: psycholinguistics, cognitive psychology.

E-mail: antoninaraf@yandex.ru

**Ekaterina A. Valueva** – Research Fellow, Institute of Psychology, Russian Academy of Sciences, PhD in Psychology.

Research Area: cognitive psychology, intelligence, creativity.

E-mail: ekval@list.ru

**Панфилова Анастасия Сергеевна** — научный сотрудник, лаборатория психологии и психофизиологии творчества, ФГБУН «Институт психологии Российской академии наук», кандидат технических наук. Сфера научных интересов: когнитивная психология, методы машинного обучения. Контакты: panfilova87@gmail.com

**Anastasia S. Panfilova** — Research Fellow, Institute of Psychology, Russian Academy of Sciences, PhD in Engineering. Research Area: cognitive psychology, machine learning methods. E-mail: panfilova87@gmail.com

Исследования взаимосвязи психологических характеристик с особенностями голоса имеют почти столетнюю историю (Sapir, 1927). По замечанию Э. Сепира, «одно обстоятельство представляется интересным в отношении речи: с одной стороны, ее трудно проанализировать, с другой стороны — она во многом руководит нашими непосредственными впечатлениями... никто из нас не лишен полностью интуитивной способности исследовать личность, опираясь на восприятие речи» (Ibid., p. 892). В силу представления о бессмысленности атомистического анализа (Allport, Cantril, 1934), а также в силу отсутствия технических средств первые исследования взаимосвязи речи и психологических характеристик в начале XX в. велись с опорой на впечатление о голосе говорящего «в целом»: испытуемым предлагалось прослушивать аудиозаписи и оценивать разнообразные физические (возраст, рост, комплекция и др.), социальные (например, профессия, политические взгляды) и психологические (экстраверсия, доминантность и т.д.) характеристики говорящего. Впечатляющее по масштабу для того времени исследование (более шестисот испытуемых) было представлено в процитированной выше работе Г. Олпорта и Х. Кентрила. Авторы обнаружили, что люди дают согласованную оценку физических и в особенности психологических характеристик по голосу, однако часто это оценка не совпадает с объективными (полученными путем тестирования) данными о владельце голоса. Сходные результаты повторились еще в нескольких исследованиях (Eisenberg, Zalowitz, 1938; Stagner, 1936; Taylor, 1934). Общий вывод состоял в том, что уровень согласованности социальных суждений о человеке по голосу выше уровня их точности, а это свидетельствует о функционировании сильных стереотипов в межличностном восприятии.

После изобретения хронографа взаимодействий (The Interaction Chronograph) (Chapple, 1949) стало возможно изучать некоторые характеристики речи объективными методами — фиксировались длина активной речи и пауз, а также их соотношение. Подробный обзор исследований по этой теме приведен в работе Р. Рамзая (Ramsay, 1968). В своих собственных исследованиях Рамзай установил связь интроверсии с длительностью пауз, интеллекта — с длительностью речи и отсутствие связей между особенностями речи и нейротизмом. Интересно также, что в его исследовании сложность задания не влияла на длительность речевых блоков (речь + пауза), но влияла на соотношение речи и пауз: в более сложных заданиях длительность речи сокращалась, а пауз — увеличивалась. Таким образом, автор делает вывод о существовании стабильных ритмических паттернов речи.

Параллельно с этим развивались субъективные методы «объективной» оценки отдельных характеристик голоса. Н. Маркел предложил протокол оценки голоса (Профиль качества голоса, Voice Quality Profile) по трем измерениям: высота (pitch), громкость (loudness) и темп (tempo) (Markel, 1965). В силу достаточно низкой согласованности экспертов по шкале высоты голоса, а также вследствие больших различий по этому параметру между мужчинами, женщинами и детьми он был исключен из усовершенствованной версии профиля (Markel et al., 1972). В этом исследовании авторы выделили четыре профиля голоса (громкий-быстрый, громкий-медленный, тихий-быстрый и тихий-медленный) и связали каждый из этих профилей с определенными психологическими качествами, полученными в результате тестирования по опросникам ММРІ и 16 РР.

На определенном этапе ранние исследования взаимосвязи голоса и психологических характеристик зашли в тупик, так как, опираясь на субъективные суждения, неизбежно сталкивались с обозначенной выше проблемой — восприятие голоса порождает стереотипы межличностного восприятия (Addington, 1968). До некоторых пор объективная оценка акустических параметров речи была доступна в крайне ограниченном варианте (Scherer, 1974), поэтому всплеск эмпирических исследований в этой области произошел с развитием современных компьютеров и программного обеспечения. В следующих разделах статьи мы постараемся кратко очертить основные подходы и достижения исследований на современном, «компьютерном» этапе.

### Акустические характеристики речи

Развитие специализированных программ для анализа аудиозаписей (например, PRAAT или платформа openSMILE) позволило исследователям перейти на другой уровень анализа речи. Основными параметрами анализа стали не субъективно воспринимаемые темп, тембр, высота и т.д., а характеристики самого аудиосигнала — так называемые низкоуровневые дескрипторы (Low Level Descriptors, LLD) и их функционал (Schuller et al., 2009). Наиболее часто в работах можно встретить следующие LLD.

*Форманты* — показатели, связанные с частотой голосового тона (резонанс речевого тракта, образующийся при произнесении различных гласных звуков), определяют восприятие тембра голоса. Форманту F0 называют частотой основного тона. Она связана с частотой колебания связок, характеризует воспринимаемую высоту голоса (pitch) и зависит от характеристик голосовых связок (размер и степень натяжения). Так как частота основного тона существенно различается у мужчин и женщин, в исследованиях, как правило, применяется стандартизация внутри групп.

*Число нулевых переходов* (zero crossing rate, ZCR) отражает количество изменений знака функции аудиосигнала, является показателем зашумленности и часто используется для различения речевых и неречевых сигналов.

*Среднеквадратическое значение громкостей* (root mean square energy, rms) характеризует распределение громких и тихих участков в записи, иногда интерпретируется как интенсивность (intensity) речи.

*Дрожание* (jitter) и *мерцание* (shimmer) — параметры голосового сигнала, отражающие вариабельность частоты основного тона и амплитуды голоса. Они передают субъективно воспринимаемое качество голоса и часто исследуются в контексте речевых патологий. Соотношение сигнала и шума (harmonics-to-noise ratio) также определяет качество голоса и субъективно характеризуется как степень хрипоты.

*Мел-частотные кепстральные коэффициенты* (MFCC) образуются через вычисление спектра всего сигнала и дальнейшее применение мел-фильтров с различным основанием для вычисления энергии каждого фильтра. Последующее логарифмирование и дискретное косинусное преобразование позволяют получить набор коэффициентов для каждого из фильтров соответствующего порядка. Для анализа в основном используются первые 13 мел-частотных кепстральных коэффициентов.

По каждой характеристике (F0, ZCR, RMS, MFCC и т.д.) можно дополнительно вычислить набор оценок распределения полученных характеристик: среднее (mean), стандартное отклонение (standart deviation), эксцесс (kurtosis), ассиметрия (skewness), максимальное значение (max), минимальное значение (min), размах (range), наклон линейной интерполяции, константа линейной интерполяции, среднеквадратическая ошибка линейной интерполяции.

Помимо LLD, в исследованиях часто используются темпоральные характеристики речи: скорость речи (отношение числа произнесенных слов к общему времени записи), средняя длина пауз (отношение длительности пауз к их количеству в записи), общая длительность речи и т.д.

В следующих разделах мы представим исследования, посвященные аудио-характеристикам эмоций, намерений, личностных качеств.

## Голос и эмоции

Гнев, счастье, грусть, страх, отвращение, приятное удивление — шесть базовых человеческих эмоций, каждая из которых имеет свою биологическую основу и экспрессивные качества, являющиеся универсальными для разных культур и языков (Ekman, 1992). Частота основного тона (высота голоса), интенсивность (сила звука) и скорость речи являются основными акустико-просодическими свойствами для распознавания эмоций (Juslin, Laukka, 2003; Preti et al., 2015).

Исследование акустической дифференциации эмоций показало, что для радости характерны высокие параметры темпа речи, силы звука и вариации высоты (Sobin, Alpert, 1999). Были выявлены дополнительные переменные, связанные с определенными эмоциями и повышающие акустическую дифференциацию между эмоциями: гнев и радость различались по параметрам продолжительности речи и паузы и вариации силы звука. Следует отметить, что гнев может быть коннотирован как «холодный» и «горячий» (Banse, Scherer, 1996), и высокий показатель вариабельности частоты основного тона может рассматриваться как специфическое свойство «горячего», но не «холодного»

гнева (Preti et al., 2015). Обнаружено, что страх оказался наименее точно распознаваемой по акустическим параметрам эмоцией (Sobin, Alpert, 1999). В целом негативные эмоции по голосу распознаются лучше, чем положительные. Согласно исследованиям (Öhman et al., 2001), эволюционные факторы объясняют повышенную чувствительность к одним эмоциям по сравнению с другими. Наиболее точно распознаются гнев и грусть (Pell et al., 2009; Liu, Pell, 2012). Гнев представляет собой потенциальную угрозу, поэтому хорошо распознаваем; выражение грусти, в свою очередь, является сигналом для членов группы о необходимости помощи, поддержки и защиты (Thompson, Balkwill, 2006). Отвращение и счастье продемонстрировали относительно низкий процент распознавания (Banse, Scherer, 1996; Castro, Lima, 2010; Pell, 2002). В нескольких исследованиях приятное удивление стало наиболее сложной для распознавания эмоцией (Pell et al., 2009; Liu, Pell, 2012), поскольку часто удивление интерпретируется как интенсивное выражение радости (Navas et al., 2004; Pell et al., 2009; Castro, Lima, 2010).

В исследовании Дж. Лискомба с соавт. (Liscombe et al., 2003) рассматривались два измерения эмоций: валентность и активация (Cowie, Cornelius, 2003). «Счастье» и «воодушевление» имеют положительную валентность, тогда как «гнев» и «грусть» — отрицательную. Под активацией имеется в виду степень возбуждения: «счастье» и «гнев» имеют позитивную активацию, «грусть» и «скука» — негативную. Исследования показывают, что для эмоций с позитивной активацией характерны высокие средняя частота основного тона, средняя энергия и более высокая скорость речи по сравнению с эмоциями с негативной активацией (Schroder et al., 2001; Tato et al., 2002; Liscombe et al., 2003). Спектральный срез, тип фразового акцента и пограничный тон могут использоваться для дифференциации валентности эмоций (Liscombe et al., 2003). Спектральный срез позволяет различать эмоции с позитивной активацией по валентности: с положительной валентностью — дружелюбность, счастье, воодушевление, с негативной — гнев, расстройство. Свойства частоты основного тона являются наилучшими предикторами гнева, уверенности, счастья, заинтересованности, грусти и расстройства, а спектральный срез служит наилучшим предиктором гнева, тревожности, скуки и дружелюбности (Ibid., 2003).

В исследовании П. Лаукки с соавт. (Laukka et al., 2011) были изучены акустические корреляты и автоматическое определение эмоциональной речи на примере аутентичной аффективной речи, записанной в ходе человеко-компьютерного взаимодействия в колл-центрах. Это отличает данное исследование от предыдущих, в которых подобные эксперименты проводились на примере постановочной речи. Объектом исследования стало человеко-компьютерное взаимодействие, те его эпизоды, в которых говорящие либо были раздражены, либо проявляли терпение/смирение.

Если гнев и грусть представляют собой очень интенсивные реакции на определенные прототипичные ситуации, часто называемые базовыми или модальными эмоциями, то в данном исследовании изучались эмоции, носящие гораздо менее интенсивный характер, — раздражение и уступчивость.

Раздражение и гнев — и уступчивость и грусть — в основе своей имеют сходные когнитивные механизмы (Scherer, Tran, 2001). Предыдущие исследования показали, что слушатели способны точно распознавать не только ярко выраженные эмоциональные категории, но и различные нюансы, степень интенсивности голосовых сигналов (Banse, Scherer, 1996; Juslin, Laukka, 2001). Но не было исследований, посвященных изучению восприятия и акустических коррелятов эмоциональной интенсивности в голосовых сигналах. Были обнаружены существенные акустические различия между речью, воспринимаемой как нейтральная и раздраженной/смирненной. Оценки, данные слушателями при оценивании раздражения, смирения, нейтральности и интенсивных эмоциях, коррелировали со следующими акустическими характеристиками: высота, интенсивность, форманты, источник голоса и временные характеристики речи. Автоматическое определение с использованием линейного дискриминантного анализа продемонстрировало результат, сопоставимый с показанным человеком при классификации высказываний, отражающих раздражение, смирение и нейтральность.

В исследовании М. Свертса и Дж. Хиршберг (Swerts, Hirschberg, 2010) были изучены просодические предикторы хороших или плохих новостей. Участников просили представить, что они информируют соискателя вакансии о том, принят ли он на работу. Исходные сообщения записанных фрагментов речи, не содержащие никаких лексических подсказок эмоционального содержания, были оценены слушателями исходя из предположения, хорошие или плохие новости будут следовать за этим фрагментом речи. Участников попросили оценить каждый фрагмент речи по десяти 5-балльным шкалам соответственно каждой характеристике речи: дружелюбие, сомнение, озабоченность, приветливость, убежденность, нервозность, грусть, приятность, заинтересованность или монотонность. Измерения показали, что исходные сообщения, полученные в эмоциональном состоянии «хороших новостей», можно отличить от полученных в состоянии «плохих новостей», исходя из субъективно воспринимаемых признаков голоса (звучит ли голос так, как будто говорящий улыбается), которые, в свою очередь, коррелируют с автоматически полученными акустическими/просодическими характеристиками.

Исследования Н. Мэдзлен с соавт. (Madzlan et al., 2014a, 2014b) были посвящены изучению просодических характеристик выражения различных видов отношения в видеоблогах — социально-аффективных состояний, которые видеоблогеры пытаются передать. Были изучены такие состояния, как радость, нетерпеливость, непринужденность, энтузиазм и разочарование. Результаты продемонстрировали возможность интерпретировать и прогнозировать сигналы социально-аффективных состояний, исходя из физических характеристик голоса.

В результате исследования (Mauchand, Pell, 2021), посвященного голосовым свойствам жалобы в речи, было показано, что жалующийся тон голоса имел большее влияние на восприятие участников, чем лингвистическое содержание высказываний. Жалоба выражает поиск поддержки и ощущения принадлежности к социальной группе, соответственно, с метапрагматической

точки зрения жалоба может быть квалифицирована как акт манипуляции без введения в заблуждение, наподобие других эмотивных актов, как убеждение, мотивирование или харизматичная речь: выражение интенции эмоции, которая регулирует диспозиционный аффект на слушающего и способствует социальной приобщенности (Pell, Kotz, 2021). Средняя частота основного тона стала главным голосовым свойством для определения жалобы в голосе.

В исследовании Д. Трусдейл и М. Пелла (Truesdale, Pell, 2018) изучались страсть и безразличие — крайние уровни эмоционального возбуждения — и нейтральность. Страсть — это гипертоническое состояние, которое объединяет интенсивно позитивные и негативные эмоциональные состояния (Vallerand, 2010; Truesdale, Pell, 2018). По сравнению с нейтральностью для страсти были характерны самые высокие уровни максимальной частоты основного тона, диапазона частоты основного тона, максимальной и средней интенсивности, тогда как безразличие передавалось более низкими измерениями по этим голосовым свойствам в сравнении с нейтральностью. Несмотря на то что было обнаружено просодическое совпадение между нейтральностью и безразличием, были выявлены и различия, например, показатели максимальной высоты тона и диапазона высоты тона были значительно ниже при выражении безразличия, чем нейтральности.

### *Невербальные вокализации эмоций*

Исследование невербальных вокализаций эмоций (Sauter et al., 2010) показало, что они так же, как и эмоциональная речь, обладают комбинацией акустических свойств и могут быть спрогнозированы на основе их физических свойств. Было исследовано восприятие 10 эмоций: ликование/триумф, веселье, удовлетворение, удовольствие, удивление, облегчение, гнев, отвращение, страх и грусть. Спектральные свойства, особенно спектральная переменчивость, играют такую же, если не более важную роль в восприятии эмоций из невербальных голосовых сигналов, как свойства частоты основного тона. Стоит отметить, что в исследованиях об эмоциональной речи не были представлены данные о спектральных изменениях (Banse, Scherer, 1996; Juslin, Laukka, 2001), возможно, потому что спектральная переменчивость речи скрывает определенные специфические изменения, связанные с эмоциями, или может препятствовать проявлению таких изменений (Sauter et al., 2010). Были выявлены определенные сходства между акустическими характеристиками, использованными в определении и распознавании эмоций, выраженных в речи и в невербальных вокализациях: оценки для гнева, грусти и отвращения имели общие акустические характеристики с данными, полученными в предыдущих работах об эмоциональной речи (Banse, Scherer, 1996; Juslin, Laukka, 2001). Но были обнаружены и различия, например, в оценках выражения страха. Это свидетельствует о том, что акустические сигналы, используемые в оценивании эмоций в речи и невербальных звуков, не идентичны.

С. Лима с соавт. (Lima et al., 2013) провели аналогичное исследование невербальных эмоциональных вокализаций на португальском языке на примере



8 эмоций: ликование/триумф, веселье, удовольствие, облегчение, гнев, отвращение, страх и грусть. Были выявлены специфические свойства, определяющие специфические эмоции в невербальных вокализациях. Как и в исследовании Д. Саутер с соавт. (Sauter et al., 2010), высокая средняя частота основного тона в значительной степени прогнозирует ликование/триумф и облегчение. Но в отличие от исследования Д. Саутер с соавт. (Ibid.), в котором было обнаружено, что вариабельность в спектральном центре масс прогнозирует удовольствие и страх, в результате исследования С. Лимы с соавт. таких данных не было получено, что может быть объяснено наличием незначительных различий в том, как вокализации воспроизводятся и воспринимаются людьми, говорящими на разных языках: исследование Д. Саутер и др. (Ibid.) было проведено на английском языке, Лимы с соавт. — на португальском.

Также были исследованы изменения в показателях распознавания невербальных эмоциональных вокализаций в зависимости от возраста (Lima et al., 2014). Было показано, что особенности обработки невербальных эмоциональных вокализаций варьируются в зависимости от возраста воспринимающего. Изменения, связанные с возрастом, не зависят от валентности эмоций, не могут быть объяснены ни снижением общих когнитивных способностей, ни разницей других индивидуальных показателей. Кроме этого, выявленные снижения показателей были отражены в различиях акустических сигналов, определяющих реакцию, это указывает на то, что старение связано с изменениями в правилах умозаключений, используемых в восприятии вокализаций.

### **Голос и намерения: аудиохарактеристики обмана**

Обман и правда выражаются в речи специфическими наборами акустико-просодических свойств. В исследовании С. Левитэн с соавт. (Levitan et al., 2018) были определены голосовые свойства правды и обмана в речи и голосовые свойства речи, воспринимаемой как правдивая или ложная. Исследование проводилось на английском языке, для одной группы участников родным языком был английский, для другой — китайский. Результаты продемонстрировали, что для обмана в речи были характерны более высокие значения максимальной частоты основного тона и максимальной интенсивности. Скорость речи была высокой в правдивой речи в «китайской» группе. Для речи, оцененной как ложная, были характерны большие максимальная высота и максимальная интенсивность. Высказывания, воспринимаемые как правдивые, отличались более быстрым темпом речи.

### **Голос и личностные качества**

#### *Аудиопредикторы уверенности*

Акустический анализ, проведенный с целью выявления перцептивно-акустических характеристик выражения уверенности и сомнения в устной речи, выявил, что уверенность проявлялась в высоких показателях диапазона

частоты основного тона, средней амплитуды и диапазона амплитуд, тогда как неуверенная речь характеризовалась высокими показателями средней частоты основного тона, медленной скоростью речи и более частыми паузами (Jiang, Pell, 2017). В исследованиях Х. Пон-Барри с соавт. (Pon-Barry, 2008; Pon-Barry, Shieber, 2011) уровень уверенности измерялся на уровне контекста, ключевого слова и целого высказывания. Экспериментатор задавал вопрос: «Как я могу добраться от Гарварда до Серебряной линии?». Говорящие видят неизменную часть ответа «Садитесь на красную линию до» (это называется контекстом). После нажатия клавиши вслед за контекстом появляются названия «Южный вокзал» и «Даунтаун-Кроссинг» (это ключевые слова). Были измерены корреляции между воспринимаемым уровнем уверенности и 60 просодическими характеристиками. Корреляции между средней оценкой и просодическими свойствами, выделенными из целых высказываний, свидетельствуют о том, что временные характеристики (например, процент тишины, общая тишина и т.д.) имели самую сильную корреляционную связь с воспринимаемым уровнем уверенности (Pon-Barry, 2008). Было показано, что одни просодические характеристики имеют сильные корреляционные связи на уровне целого высказывания, другие — на уровне контекста или ключевого слова. Это позволяет говорить о том, что модели машинного обучения, натренированные на просодических свойствах контекста и ключевого слова, могут быть эффективнее натренированных исключительно на целых высказываниях (Pon-Barry, Shieber, 2009). Уровень уверенности оценивали и сами говорящие, и группа слушателей, при этом часто обнаруживалось несоответствие между внутренними состояниями говорящих и субъективные представления о них.

### *Аудиопредикторы лидерских качеств*

В характеристиках речи могут проявляться личностные свойства человека. Целями исследований Д. Санчес-Кортес с соавт. (Sanchez-Cortes et al., 2011, 2012) стали анализ того, как лидер воспринимается в только сформированных малых группах, и автоматическое определение новых лидеров с использованием различных коммуникативных невербальных сигналов, извлеченных из аудио и видео. Участникам надо было выполнить задание, в ходе которого они должны были составить список из 12 предметов, в порядке необходимости в условиях авиакрушения зимой. Задание сначала выполнялось индивидуально, потом в группе для налаживания общения между участниками и выявления лидера. Далее участники должны были ответить на 17 утверждений, отражающих то, как они воспринимали каждого участника, включая самих себя. 16 утверждений оценивались по 5-балльной шкале. В том числе измерялись воспринимаемое лидерство (руководит группой, настаивает на своем мнении), воспринимаемое доминирование (доминирует, отстаивает себя), воспринимаемая компетентность (компетентный, умный, имеет большой опыт) и воспринимаемая симпатия (добрый, дружелюбный, доброжелательный). Измерения воспринимаемой симпатии и воспринимаемой компетентности

были включены для отражения двух аспектов лидерства: социоэмоционального и проблемноориентированного. Социоэмоциональный лидер заинтересован, в первую очередь, в хороших отношениях внутри группы, проблемноориентированный фокусируется на решении проблемы.

Результаты продемонстрировали, что новый лидер воспринимался другими членами группы как активный и доминирующий. Для восприятия компетентности самым информативным голосовым предиктором стала частота основного тона; для воспринимаемой симпатии — количество перебиваний. Было показано, что восприятие лидерства ассоциируется с человеком, который больше всего говорит и перебивает других, имеет большую вариабельность в тоне и интенсивности голоса. Количество перебиваний продемонстрировало среднюю корреляционную связь с восприятием компетентности. Как было показано в другом исследовании (Wentworth, Anderson, 1984), новые лидеры не обязательно должны быть самыми активными участниками общения, если они воспринимаются как достаточно компетентные.

#### *Аудиопредикторы харизматичной речи*

Результаты исследований показывают, что понятие харизматичной речи является достаточно сложным относительно значимости голосовых предикторов. С одной стороны, был выявлен ряд просодических свойств харизматичной речи: высокая скорость речи, высокая интенсивность, высокие показатели свойств частоты основного тона (Rosenberg, Hirschberg, 2005). Но восприятие харизматичной речи было обусловлено сочетанием просодических и лексических свойств речи. Результаты исследований, сравнивающие восприятие харизматичной речи между разноязыковыми группами, показывают, что есть как общие для разных культур, так и специфические акустико-просодические корреляты харизматичной речи (Biadys et al., 2007; Biadys et al., 2008). Во всех разноязычных группах восприятие харизмы коррелировало с высокой частотой основного тона, большой вариабельностью интенсивности, падающим интонационным контуром (Biadys et al., 2008).

Кроме того, было показано, что харизматичный стиль речи не обусловлен напрямую результатами опросника по пятифакторной модели харизмы (активность, напористость, общительность, тревожность и депрессия) (Michalsky et al., 2020). Личность может влиять на определенные акустические характеристики, которые косвенно воздействуют на воспринимаемую харизму, но наличие харизматичных черт личности не является обязательным условием для харизматичной речи. В другом исследовании (Berger et al., 2017) харизматичность речи оценивалась с помощью шкалы Лайкерта по степени харизматичности, привлекательности, убедительности и мотивации. Было показано, что увеличение диапазона частоты основного тона и более быстрая речь в определенной степени характерны для харизматичной речи (Niebuhr et al., 2016; Rosenberg, Hirschberg, 2005). В отличие от предыдущих исследований, где была установлена позитивная корреляция между частотой основного тона и уровнем харизмы, в данном исследовании единственным

обнаруженным значимым эффектом основного тона было то, что чем выше основной тон, тем менее приятным воспринимался голос, что согласуется с исследованиями, посвященными изучению привлекательности голоса (Quené et al., 2016), но не харизмы. Это может свидетельствовать о том, что основной тон не является характеристикой, которая обуславливает восприятие речи как харизматичной, а скорее является дополнительным параметром или не играет большой роли в принципе. Можно сделать вывод, что роль аудиальных характеристик в распознавании харизматичности ниже, чем при распознавании эмоций, где более заметную роль играют именно голосовые характеристики, нежели лексическое содержание.

### *Аудиопредикторы личностных черт*

Уже в ранних работах было показано, что некоторые характеристики голоса могут быть связаны с личностными чертами. В частности, Э. Мэллори и В. Миллер (Mallory, Miller, 1958) обнаружили слабые, но статистически значимые корреляции интроверсии с высотой голоса, отсутствием резонанса, неуверенной манерой говорить. В других исследованиях продемонстрировано, что экстраверсия и интроверсия связаны с темпом речи (Feldstein, Sloan, 1984; Ramsay, 1968). Современные данные подтверждают и расширяют эти результаты, в основном используя в качестве модели личности Большую пятерку. Так, исследование Дж. Парка с соавт. (Park et al., 2020) продемонстрировало, что у экстравертов ниже среднее время реакции и общее время реакции. В исследованиях Дж. Бьела (Biel et al., 2011; Biel, Gatica-Perez, 2013), где применялись экспертные оценки качеств Большой пятерки, наибольшее число значимых корреляций было обнаружено у экстраверсии. Предикторами экстраверсии стали время говорения, средняя и максимальная энергии, средняя высота, малое число пауз в речи. В отношении других черт Большой пятерки были получены следующие корреляции: время говорения положительно коррелировало с добросовестностью и открытостью опыту, аудиальные свойства получили значительные корреляции со всеми воспринимаемыми личностными чертами, за исключением доброжелательности.

В других исследованиях доброжелательность ассоциируется с высоким голосом (Borkenau, Liebler, 1992), о чем свидетельствует положительная корреляция между доброжелательностью и средней высотой. Корреляция между добросовестностью и стандартными отклонениями энергии указывает на более высокий голосовой контроль. Наблюдаемые добросовестность, доброжелательность и открытость новому опыту оказались более ассоциированы с визуальными свойствами (Kenny et al., 1992).

Важно отметить, что точность прогнозирования личностных черт, измеренных самоотчетным и экспертным методами, может различаться (Mairesse et al., 2007). Например, и для наблюдаемой, и для самоотчетной экстраверсии просодическими маркерами стали варибельность интенсивности и средняя интенсивность. С другой стороны, для эмоциональной стабильности, измеренной по самоотчетам, характерны низкие варибельность интенсивности и

средняя интенсивность, тогда как при оценке внешних наблюдателей эти голосовые свойства не играли роли. Авторами выдвинуто предположение, что модель для определения черт личности должна переключаться с модели, построенной в результате оценок внешних наблюдателей, на модель, построенную на оценках, данных при самоотчете, потому что черты с высокой очевидностью (экстраверсия) более точно оцениваются внешними наблюдателями, тогда как черты с низкой очевидностью (эмоциональная стабильность) точнее оцениваются в результате самоотчета.

### **Автоматическое распознавание личности**

Среди современных трендов диагностики личностных особенностей можно выделить использование сетей глубокого обучения и комбинации анализа вербальных и невербальных компонентов речи вместе с анализом видео в части оценки динамики эмоционального состояния (Mehta, 2019).

Целью работы А. Сенгупта (Sengupta, 2018) было определение коэффициента ошибок при автоматическом распознавании личности с использованием моделей классификации для групп испытуемых, разделенных по принципу пола и родного языка (английский/китайский). Результаты показали, что коэффициент ошибок варьируется в зависимости от пола и родного языка и различается для разных свойств личности. Свойства энергии голоса оказались наиболее значимыми для точного распознавания личности по голосу.

Исследование Г. Ана с соавт. (An et al., 2016) было посвящено автоматическому распознаванию личности по голосовым характеристикам. Черты личности классифицировались на основе самоотчетов испытуемых. В результате исследования с наибольшей точностью была спрогнозирована открытость опыту, с наименьшей — экстраверсия. Эти результаты отличаются от тех, которые были получены в исследованиях, где черты личности оценивались другими людьми: нейротизм и экстраверсия прогнозировались с наибольшей точностью (Mairesse et al., 2007), а открытость новому опыту — с наименьшей (Mohammadi et al., 2010). Это может быть связано с тем, что экстраверсия и нейротизм ассоциируются со стереотипным речевым поведением и потому легче определяются по речи, тогда как открытость новому опыту менее стереотипна с точки зрения речевого поведения. В связи с этим открытость более точно идентифицируется при прохождении тестирования, чем при оценивании другими.

Исследования Дж. Мохамеди и А. Винчарелли (Mohammadi, Vinciarelli, 2012) и Т. Полжела (Polzehl, 2015) также показывают, что диапазон высоты тона, скорость речи, интенсивность, громкость, форманты или спектры могут предсказать элементы Большой пятерки.

В качестве методов на раннем этапе развития подхода использовались преимущественно регрессионный анализ и метод опорных векторов (SVM). С развитием методов нейронных сетей появились работы (An, Levitan, 2018), использующие многослойный перцептрон (MLP), дополненный моделью для анализа вербальной стороны речи (LSTM). В данной работе максимальное качество классификации (доля верных классификаций), которого удалось

достичь, следующее: открытость опыту (77%), добросовестность (63%), экстраверсия (64%), доброжелательность (61%), нейротизм (68%). Предложенный М. Карбонно с соавт. метод (Carbonneau et al., 2020), опирающийся на использование спектрограмм и SVM, позволил повысить эффективность распознавания доброжелательности до 65%, а нейротизма до 70% при уменьшении качества предсказания остальных показателей. Дальнейшие работы посвящены подходам к отбору характеристик и сравнению моделей нейронных сетей. Так, М. Таярани с соавт. (Tayarani et al., 2019) предложили использовать анализ вставок «ehm», «uhm» в речи и показали, что при сравнении Cascade Forward Neural Network (CFNN), Feed Forward Neural Networks (FFNN), Fuzzy Neural Networks (FNN), Generalized Regression Neural Networks (GRNN), k Nearest Neighbors (kNN), Linear Discriminant Function (LDF), Naive Bayes Classifier (NB), Support Vector Machines (SMV) и использовании подхода PCA-QEA к отбору характеристик классификатор LDF продемонстрировал значимое увеличение качества классификации для открытости опыту и экстраверсии.

### **Заключение**

Частота основного тона, интенсивность и скорость речи являются основными голосовыми свойствами для распознавания эмоций и психологических характеристик. Негативные эмоции по голосу распознаются лучше, чем положительные, наиболее точно распознаются гнев и грусть. При оценивании черт Большой пятерки с помощью экспертных оценок наибольшее число значимых корреляций с акустико-просодическими и аудиальными признаками было обнаружено у экстраверсии. При этом результаты могут существенно различаться в зависимости от того, какой метод использовался для определения личностных черт: например, при использовании оценок, полученных по результатам самоотчета, экстраверсия оказалась наименее точно прогнозируемой по свойствам голоса чертой, а наиболее точно прогнозируемой была открытость новому опыту.

Использование нейросетей вернуло исследователей к старой идее оценки голоса «в целом» (Allport, Cantril, 1934), но на новом уровне. На вход нейросети подается несколько сотен характеристик голоса (аналогично тому, как слушатель воспринимает речь), а на выходе мы получаем классификацию психологических качеств (так же как слушатель дает оценки, иногда ошибочные, личности говорящего). Достоинством нейросети при этом является отсутствие у нее стереотипов межличностного восприятия, поэтому ее «решения» можно рассматривать как более надежные.

Представленный в настоящей статье обзор не является полным и исчерпывающим. За то время, пока наша статья, попав в редакцию, выйдет в печать, появится, вероятно, еще десяток статей по обсуждаемой теме. А к тому моменту, когда ее прочтет наш читатель, может быть, увидит свет еще сотня новых исследований. Область автоматического распознавания психологических характеристик по голосовым параметрам чрезвычайно востребована, поэтому мы видели нашу основную задачу в том, чтобы обозначить вектор ее

развития и поставить вопросы о необходимости теоретического осмысления получаемых результатов. Многие воодушевлены предсказательной силой, которой обладают машинное обучение и нейросетевые модели: точность их предсказаний иногда доходит до 100% (Jothilakshmi et al., 2017). Однако часто этими исследованиями занимаются специалисты-непсихологи, поэтому вопрос о смысле полученных данных остается за скобками.

Помимо вопроса о том, какие аудиохарактеристики речи связаны с различными психологическими характеристиками, важно задавать и вопрос о том, почему вообще какие-либо связи обнаруживаются. Первая приходящая в голову теоретическая модель заключается в том, что психологические качества определяют свойства голоса: например, быстрая речь экстравертов порождается их динамичным характером, а задумчивые интроверты вообще медлительны, поэтому и говорят медленно. Таким же образом могут формироваться и патологии голоса (Roy, Bless, 2000). Однако справедливы могут быть и другие гипотезы. Например, Э. Мэллори и В. Миллер (Mallory, Miller, 1958) предполагали, что особенности голоса (тесно связанные с мышечными реакциями) и соответствующие им личностные черты развиваются параллельно в результате комплекса реакций на определенные жизненные события. Так, например, ситуации подчинения, в результате попадания в которые развивается соответствующая черта характера, могут сопровождаться сжатием мышц, регулирующих голосовые связки, что приводит к формированию более высокого голоса, а сужение голосовых ходов при этом ведет к уменьшению резонансных свойств. Одна из самых ранних гипотез вообще удивительна: особенности голоса могут влиять на формирование характера. В книге «Основы речи» Х. Вулберт высказывает идею о том, что особенности (дефекты) голоса могут определять то, как общество воспринимает его носителя, и, как следствие, формирование его характера (Woolbert, 1920).

Помимо общих теоретических предположений о связи голоса и психологических свойств, разрабатываются и более частные гипотезы. Так, в работе А.С. Сильницкой (2016) обсуждается связь интонационных особенностей речи с особенностями характера и темперамента. Опираясь на теорию деятельности, автор разрабатывает идею о том, что интонация является средством для решения коммуникативных задач в деятельности общения. Темперамент и характер оказывают влияние на интонационные особенности речи, но это влияние опосредовано условиями, в которых протекает коммуникативная деятельность. Например, показано, что наличие или отсутствие собеседника, меняя смысловую регуляцию деятельности, изменяет и связи между психологическими характеристиками и особенностями голоса.

К сожалению, исследования, в которых обсуждаются теоретические гипотезы о связи аудиохарактеристик речи и психологических свойств человека, редки. Еще более редко эти гипотезы последовательно проверяются эмпирически. Работа в этом направлении представляется весьма перспективной и нужной, а пока нам все еще «благоразумнее было бы относиться к публикуемым результатам с английской сдержанностью, нежели с американской пылкостью» (McKelvey, 1953, p. 93).

## Литература

Сильницкая, А. С. (2016). *Отражение индивидуальных свойств личности в интонационных параметрах речи* [Кандидатская диссертация, Московский государственный университет имени М.В. Ломоносова]. <https://www.dissercat.com/content/otrazhenie-individualnykh-svoystv-lichnosti-nv-intonatsionnykh-parametrakh-rechi>

Ссылки на зарубежные источники см. в разделе *References*.

## References

- Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs*, 35(4), 492–503. <https://doi.org/10.1080/03637756809375599>
- Allport, G. W., & Cantril, H. (1934). Judging personality from voice. *The Journal of Social Psychology*, 5(1), 37–55. <https://doi.org/10.1080/00224545.1934.9921582>
- An, G., & Levitan, R. (2018). Lexical and acoustic deep learning model for personality recognition. *Proceedings Interspeech 2018*, 1761–1765. <https://doi.org/10.21437/Interspeech.2018-2263>
- An, G., Levitan, S. I., Levitan, R., Rosenberg, A., Levine, M., & Hirschberg, J. (2016). Automatically Classifying Self-Rated Personality Scores from Speech. *Proceedings Interspeech 2016*, 1412–1416. <https://doi.org/10.21437/Interspeech.2016-1328>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Berger, S., Niebuhr, O., & Peters, B. (2017). Winning over an audience – A perception-based analysis of prosodic features of charismatic speech. In *Proceedings of the 43rd Annual Conference of the German Acoustical Society* (pp. 1454–1457). SDU IRCA. <https://portal.findresearcher.sdu.dk/en/publications/winning-over-an-audience-a-perception-based-analysis-of-prosodic-features-of-charismatic-speech>
- Biadsy, F., Hirschberg, J., Rosenberg, A., & Dakka, W. (2007). Comparing american and palestinian perceptions of charisma using acoustic-prosodic and lexical analysis. In *INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27–31, 2007* (pp. 2221–2224). ISCA. [https://www.isca-speech.org/archive/interspeech\\_2007/i07\\_2221.html](https://www.isca-speech.org/archive/interspeech_2007/i07_2221.html)
- Biadsy, F., Rosenberg, A., Carlson, R., Hirschberg, J., & Strangert, E. (2008). A cross-cultural comparison of American, Palestinian, and Swedish. In *Speech Prosody 2008* (pp. 579–582). LBASS. <https://www.semanticscholar.org/paper/A-Cross-Cultural-Comparison-of-American-and-of-Biadsy-Rosenberg/f6195123089eddb2e3922374fcd87d3140123e3>
- Biel, J.-I., Aran, O., & Gatica-Perez, D. (2011). You are known by how you vlog: Personality impressions and nonverbal behavior in YouTube. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1). <https://ojs.aaai.org/index.php/ICWSM/article/view/14160>
- Biel, J.-I., & Gatica-Perez, D. (2013). The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Transactions on Multimedia*, 15(1), 41–55. <https://doi.org/10.1109/TMM.2012.2225032>
- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62(4), 645–657. <https://doi.org/10.1037/0022-3514.62.4.645>



- Carbonneau, M. A., Granger, E., Attabi, Y., & Gagnon, G. (2020). Feature learning from spectrograms for assessment of personality traits. *IEEE Transactions on Affective Computing*, *11*(1), 25–31. <https://doi.org/10.1109/TAFFC.2017.2763132>
- Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, *42*(1), 74–81. <https://doi.org/10.3758/BRM.42.1.74>
- Chapple, E. D. (1949). The Interaction Chronograph: its evolution and present application. *Personnel*, *25*, 295–307. <https://psycnet.apa.org/record/1950-00177-001>
- Cowie, R., & Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication*, *40*(1–2), 5–32. [https://doi.org/10.1016/S0167-6393\(02\)00071-7](https://doi.org/10.1016/S0167-6393(02)00071-7)
- Eisenberg, P., & Zalowitz, E. (1938). Judging expressive movement: III. Judgments of dominance-feeling from phonograph records of voice. *Journal of Applied Psychology*, *22*(6), 620–631. <https://doi.org/10.1037/h0059457>
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. <https://doi.org/10.1080/02699939208411068>
- Feldstein, S., & Sloan, B. (1984). Actual and stereotyped speech tempos of extraverts and introverts. *Journal of Personality*, *52*(2), 188–204. <https://doi.org/10.1111/J.1467-6494.1984.TB00352.X>
- Jiang, X., & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. <https://doi.org/10.1016/j.specom.2017.01.011>
- Jothilakshmi, S., Sangeetha, J., & Brindha, R. (2017). Speech based automatic personality perception using spectral features. *International Journal of Speech Technology*, *20*(1), 43–50. <https://doi.org/10.1007/s10772-016-9390-0>
- Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381–412. <https://doi.org/10.1037/1528-3542.1.4.381>
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, *62*(1), 88–97. <https://doi.org/10.1037//0022-3514.62.1.88>
- Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011). Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Computer Speech & Language*, *25*(1), 84–104. <https://doi.org/10.1016/j.csl.2010.03.004>
- Levitan, S. I., Maredia, A., & Hirschberg, J. (2018). Acoustic-prosodic indicators of deception and trust in interview dialogues. In *INTERSPEECH 2018* (pp. 416–420). ISCA. <https://doi.org/10.21437/Interspeech.2018-2443>
- Lima, C. F., Alves, T., Scott, S. K., & Castro, S. L. (2014). In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion*, *14*(1), 145–160. <https://doi.org/10.1037/a0034287>
- Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*(4), 1234–1245. <https://doi.org/10.3758/s13428-013-0324-3>
- Liscombe, J., Venditti, J., & Hirschberg, J. (2003). Classifying subject ratings of emotional speech using acoustic features. In *Proceedings of the 8th European Conference on Speech Communication and*

- Technology Interspeech 2003 – Eurospeech* (pp. 725–728). ISCA. <https://doi.org/10.7916/D8F76MWV>
- Liu, P., & Pell, M.D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, 44(4), 1042–1051. <https://doi.org/10.3758/s13428-012-0203-3>
- Madzlan, N. A., Han, J., Bonin, F., & Campbell, N. (2014a). Automatic recognition of attitudes in video blogs – prosodic and visual feature analysis. In H. Li & P. Ching (Eds.), *INTERSPEECH-2014* (pp. 1826–1830). ISCA. [https://www.isca-speech.org/archive/interspeech\\_2014/i14\\_1826.html](https://www.isca-speech.org/archive/interspeech_2014/i14_1826.html)
- Madzlan, N. A., Han, J., Bonin, F., & Campbell, N. (2014b). Towards automatic recognition of attitudes: Prosodic analysis of video blogs. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings 7th International Conference on Speech Prosody 2014* (pp. 91–94). ISCA. <https://doi.org/10.21437/SpeechProsody.2014-6.html>
- Mairesse, F., Walker, M., Mehl, M., & Moore, R. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 457–500. <https://doi.org/10.1613/jair.2349>
- Mallory, E. B., & Miller, V. R. (1958). A possible basis for the association of voice characteristics and personality traits. *Speech Monographs*, 25(4), 255–260. <https://doi.org/10.1080/03637755809375240>
- Markel, N. N. (1965). The reliability of coding paralanguage: Pitch, loudness, and tempo. *Journal of Verbal Learning and Verbal Behavior*, 4(4), 306–308. [https://doi.org/10.1016/S0022-5371\(65\)80035-4](https://doi.org/10.1016/S0022-5371(65)80035-4)
- Markel, N. N., Phillis, J. A., Vargas, R., & Howard, K. (1972). Personality traits associated with voice types. *Journal of Psycholinguistic Research*, 1(3), 249–255. <https://doi.org/10.1007/BF01074441>
- Mauchand, M., & Pell, M. D. (2021). Emotivity in the voice: Prosodic, lexical, and cultural appraisal of complaining speech. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.619222>
- McKelvey, D. P. (1953). Voice and personality. *Western Speech*, 17(2), 91–94. <https://doi.org/10.1080/10570315309373580>
- Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2019). Recent trends in deep learning based personality detection. *Artificial Intelligence Review*, 53(4), 2313–2339. <https://doi.org/10.1007/s10462-019-09770-z>
- Michalsky, J., Niebuhr, O., & Penke, L. (2020). Do charismatic people produce charismatic speech? On the relationship between the Big Five personality traits and prosodic features of speaker charisma in female speakers. In *Proceedings 10th International Conference on Speech Prosody 2020* (pp. 700–704). ISCA. <https://doi.org/10.21437/SpeechProsody.2020-143>
- Mohammadi, G., & Vinciarelli, A. (2012). Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing*, 3(3), 273–284. <https://doi.org/10.1109/T-AFFC.2012.5>
- Mohammadi, G., Vinciarelli, A., & Mortillaro, M. (2010). The voice of personality: mapping nonverbal vocal behavior into trait attributions. In A. Vinciarelli, M. Pantic, & A. Pentland (Eds.), *Proceedings of the 2nd International Workshop on Social Wignal Processing* (pp. 17–20). Association for Computing Machinery. <http://publications.idiap.ch/index.php/publications/show/1926>
- Navas, E., Hernáez, I., Castelruiz, A., & Luengo, I. (2004). Obtaining and evaluating an emotional database for prosody modelling in standard Basque. In P. Sojka, I. Kopeček, & K. Pala (Eds.), *Text, speech and dialogue* (Vol. 3206, pp. 393–400). Springer. [https://doi.org/10.1007/978-3-540-30120-2\\_50](https://doi.org/10.1007/978-3-540-30120-2_50)

- Niebuhr, O., Voße, J., & Brem, A. (2016). What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of Steve Jobs tone of voice. *Computers in Human Behavior, 64*, 366–382. <https://doi.org/10.1016/j.chb.2016.06.059>
- Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381–396. <https://doi.org/10.1037/0022-3514.80.3.381>
- Park, J., Lee, S., Brotherton, K., Um, D., & Park, J. (2020). Identification of speech characteristics to distinguish human personality of introversive and extroversive male groups. *International Journal of Environmental Research and Public Health, 17*(6). <https://doi.org/10.3390/ijerph17062125>
- Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: some preliminary findings on a new battery of tests. *Brain Cognition, 48*(2–3), 499–504. <https://doi.org/10.1006/brcg.2001.1406>
- Pell, M. D., & Kotz, S. A. (2021). Comment: The next frontier: Prosody research gets interpersonal. *Emotion Review, 13*(1), 51–56. <https://doi.org/10.1177/1754073920954288>
- Pell, M., Paulmann, S., Dara, C., Allasseri, A., & Kotz, S. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37*(4), 417–435. <https://doi.org/10.1016/j.wocn.2009.07.005>
- Polzehl, T. (2015). *Personality in speech assessment and automatic classification*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-09516-5>
- Pon-Barry, H. (2008). Prosodic manifestations of confidence and uncertainty in spoken language. In J. Fletcher, D. Loakes, R. Goecke, D. Burnham, & M. Wagner (Eds.), *INTERSPEECH-2008* (pp. 74–77). ISCA. [https://www.isca-speech.org/archive/interspeech\\_2008/i08\\_0074.html](https://www.isca-speech.org/archive/interspeech_2008/i08_0074.html)
- Pon-Barry, H., & Shieber, S. M. (2009). The importance of sub-utterance prosody in predicting level of certainty. In M. Ostendorf, M. Collins, S. Narayanan, D. W. Oard, & L. Vanderwende (Eds.), *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers* (pp. 105–108). Association for Computational Linguistics. <https://dash.harvard.edu/handle/1/4729250>
- Pon-Barry, H., & Shieber, S. M. (2011). Recognizing uncertainty in speech. *EURASIP Journal on Advances in Signal Processing, 2011*, Article 251753. <https://doi.org/10.1155/2011/251753>
- Preti, E., Suttora, C., & Richetin, J. (2015). Can you hear what I feel? A validated prosodic set of angry, happy, and neutral Italian pseudowords. *Behavior Research Methods, 48*(1), 259–271. <https://doi.org/10.3758/s13428-015-0570-7>
- Quené, H., Boomsma, G., & van Erning, R. (2016, May 31 – Jun 3). *Attractiveness of male speakers: Effects of voice pitch and of speech tempo* [Poster session]. Speech Prosody 2016, Boston, USA. <https://doi.org/10.21437/SpeechProsody.2016-223>
- Ramsay, R. W. (1968). Speech patterns and personality. *Language and Speech, 11*(1), 54–63. <https://doi.org/10.1177/002383096801100108>
- Rosenberg, A., & Hirschberg, J. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. In N. Mamdani (Ed.), *Proceedings of the 9th European Conference on Speech Communication and Technology, Interspeech 2005* (pp. 513–516). ISCA. [http://www.cs.columbia.edu/~julia/files/rosenberg\\_hirschberg\\_05.pdf](http://www.cs.columbia.edu/~julia/files/rosenberg_hirschberg_05.pdf)
- Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: A foundation for future research. *Journal of Speech, Language, and Hearing Research, 43*(3), 737–748. <https://doi.org/10.1044/jslhr.4303.737>

- Sanchez-Cortes, D., Aran, O., Mast, M., & Gatica-Perez, D. (2011). *Detecting emergent leaders in small groups using nonverbal behavior*. Idiap Research Institute. [http://publications.idiap.ch/downloads/papers/2011/Sanchez-Cortes\\_IIIEETMM\\_2011.pdf](http://publications.idiap.ch/downloads/papers/2011/Sanchez-Cortes_IIIEETMM_2011.pdf)
- Sanchez-Cortes, D., Aran, O., Mast, M., & Gatica-Perez, D. (2012). A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia*, 14, 816–832. <https://doi.org/10.1109/TMM.2011.2181941>
- Sapir, E. (1927). Speech as a personality trait. *American Journal of Sociology*, 32, 892–905. <https://doi.org/10.1086/214279>
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, 63(11), 2251–2272. <https://doi.org/10.1080/17470211003721642>
- Scherer, K. R. (1974). Voice quality analysis of American and German speakers. *Journal of Psycholinguistic Research*, 3(3), 281–298. <https://doi.org/10.1007/BF01069244>
- Scherer, K.R., & Tran, V. (2001). Effects of emotion on the process of organizational learning. In M. Dierkes, A. B. Antal, J. M. Child, & I. Nonaka (Eds.), *Handbook of organizational learning and knowledge* (pp. 369–392). Oxford University Press. [https://www.researchgate.net/publication/255862191\\_Effects\\_of\\_emotion\\_on\\_the\\_process\\_of\\_organizational\\_learning](https://www.researchgate.net/publication/255862191_Effects_of_emotion_on_the_process_of_organizational_learning)
- Schroder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In P. Dalsgaard (Ed.), *Proceedings Eurospeech 2001, Scandinavia, 7th European Conference on Speech Communication and Technology* (pp. 87–90). Center for Personkommunikation. [https://www.researchgate.net/publication/221481168\\_Acoustic\\_Correlates\\_of\\_Emotion\\_Dimensions\\_in\\_View\\_of\\_Speech\\_Synthesis](https://www.researchgate.net/publication/221481168_Acoustic_Correlates_of_Emotion_Dimensions_in_View_of_Speech_Synthesis)
- Schuller, B., Steidl, S., & Batliner, A. (2009). The INTERSPEECH 2009 Emotion Challenge. In M. Uther, R. Moore, & S. Cox (Eds.), *Proceedings INTERSPEECH 2009* (pp. 312–315). ISCA. [https://www.researchgate.net/publication/224929671\\_The\\_Interspeech\\_2009\\_Emotion\\_Challenge](https://www.researchgate.net/publication/224929671_The_Interspeech_2009_Emotion_Challenge)
- Sengupta, A. (2018). *Intergroup variability in personality recognition* [Master's thesis, The Graduate Center, The City University of New York]. CUNY Academic Works. [https://academicworks.cuny.edu/gc\\_etds/2733/](https://academicworks.cuny.edu/gc_etds/2733/)
- Sil'nickaya, A. S. (2016). *Otazhenie individual'nykh svoystv lichnosti v intonacionnykh parametroh rechi* [Reflection of individual personality traits in the intonational parameters of speech] [PhD dissertation, Lomonosov Moscow State University, Moscow]. <https://www.dissercat.com/content/otazhenie-individualnykh-svoistv-lichnosti-nv-intonatsionnykh-parametrakh-rechi>
- Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, 28(4), 347–365. <https://doi.org/10.1023/A:1023237014909>
- Stagner, R. (1936). Judgments of voice and personality. *Journal of Educational Psychology*, 27(4), 272–277. <https://doi.org/10.1037/h0057086>
- Swerts, M., & Hirschberg, J. (2010). Prosodic predictors of upcoming positive or negative content in spoken messages. *The Journal of the Acoustical Society of America*, 128(3), 1337–1345. <https://doi.org/10.1121/1.3466875>
- Tato, R., Santos, R., Kompe, R., & Pardo, J. M. (2002). Emotional space improves emotion recognition. In *Proceedings of International Conference on Spoken Language Processing* (Vol. 3, pp. 2029–2032). ICSLP. <https://www.semanticscholar.org/paper/Emotional-space-improves-emotion-recognition-Tato-Santos/e6e350f18e0af270bc515a05c8a0f6729cabac5c>

- Tayarani, M., Esposito, A., & Vinciarelli, A. (2019). What an “Ehm” leaks about you: Mapping fillers into personality traits with quantum evolutionary feature selection algorithms. *IEEE Transactions on Affective Computing*, 1–15. <https://doi.org/10.1109/TAFFC.2019.2930695>
- Taylor, H. C. (1934). Social agreement on personality traits as judged from speech. *Journal of Social Psychology*, 5(2), 244–248. <https://doi.org/10.1080/00224545.1934.9919452>
- Thompson, W., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica*, 2006(158), 407–424. <https://doi.org/10.1515/SEM.2006.017>
- Tuesdale, D. M., & Pell, M. D. (2018). The sound of passion and indifference. *Speech Communication*, 99, 124–134. <https://doi.org/10.1016/j.specom.2018.03.007>
- Vallerand, R. J. (2010). On passion for life activities: The dualistic model of passion. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 42, pp. 97–193). Academic Press. [https://doi.org/10.1016/S0065-2601\(10\)42003-1](https://doi.org/10.1016/S0065-2601(10)42003-1)
- Wentworth, D. K., & Anderson, L. R. (1984). Emergent leadership as a function of sex and task type. *Sex Roles: A Journal of Research*, 11(5–6), 513–524. <https://doi.org/10.1007/BF00287475>
- Woolbert, C. H. (1920). *The fundamentals of speech: A behavioristic study of the underlying principles of speaking and reading. A text book of delivery*. Harper & Brothers. <https://archive.org/details/cu31924027197163/page/n11/mode/2up>