Speech in Affective Computing

Chi-Chun Lee, Jangwon Kim, Angeliki Metallinou, Carlos Busso, Sungbok Lee, and Shrikanth S. Narayanan

Петрова Екатерина, Харитонова Алёна

О чем эта работа?

В этой статье речь рассматривается как основной способ общения людей для кодирования эмоций.

Основные аспекты речи:

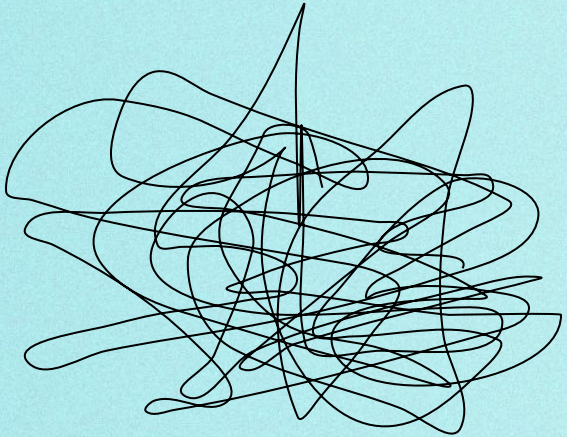
(1) Создание (воспроизводство) эмоциональной речи
(2) Извлечение акустических признаков для анализа эмоций
(3) Разработка распознавателя эмоций на основе речи

* Эта статья — глава из The Oxford Handbook of Affective Computing

Зачем эта работа

Звук голоса

Мимика и тело



Мы поняли, что нам хотели сказать

Немного о речи

Речь — это естественное и богатое средство общения, позволяющее людям взаимодействовать друг с другом

На речевой сигнал влияют многочисленные факторы:

- то, что говорится (лингвистическое содержание)
- кто это говорит (личность говорящего, возраст, пол)
- способ передачи сигнала (телефон, мобильный телефон, типы микрофонов)
- контекст, в котором генерируется речевой сигнал (акустика помещения и эффекты окружающей среды, включая фоновый шум)

Немного о речи

Человеческий речевой сигнал является результатом сложного и интегративного движения различных органов (например, голосовых связок)

→ мы попробуем измерить и количественно оценить внутреннее эмоциональное состояние человека, наблюдая за внешним аффективным и экспрессивным поведением

(то есть исследователи будут наблюдать за физиологическими проявлениями и стараться через это понять эмоциональное состояние человека)

Это что, биология?

Алгоритмы машинного обучения используются для обучения системы распознавания эмоций, чтобы находить связи между функциями речи, которые были определены, и основными метками эмоций

Часто исследования по производству речи проводятся в соответствии с теорией **«фильтра источника»** (Source–filter model, Fant, 1970)

Source–filter model представляет речь как комбинацию источника звука, такого как голосовые связки, и линейного акустического фильтра, голосового тракта. Она широко используется в ряде приложений, таких как синтез речи и анализ речи, благодаря своей относительной простоте.

Согласно этой модели производство речи состоит из двух компонентов:

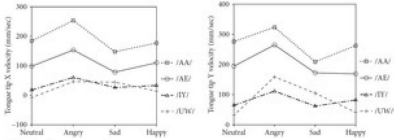
1. Исходные действия, которые генерируют поток воздуха
2. Фильтрация формирования голосового тракта, которая модулирует поток воздуха

Как это все выглядит

Модуляция голосовых связок приводит к изменению высоты тона (частоты вибрации голосовой складки), интенсивности (давления воздушного потока) и динамики качества голоса (степени апериодичности в результирующем голосовом цикле).

Взаимодействие между действиями источника голоса и элементами управления артикуляцией также способствуют модуляции звука речи.

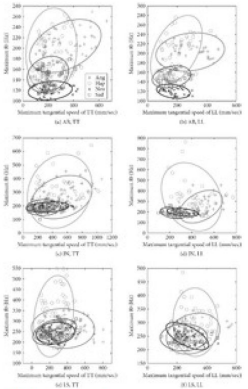
Большинство исследований эмоциональной речи были сосредоточены на акустических характеристиках результирующего уровня речевого сигнала, а не на непосредственном рассмотрении основных механизмов производства



Click to view larger

Fig. 12.1 Tongue tip horizontal (left) and vertical (right) movement velocity plots of four peripheral vowels as a function of emotion.

Source: Lee, Yildirim, Kazemzadeh, and Narayanan (2005).



Click to view larger

Fig. 12. 2 Example plots of the maximum tangential speed of critical articulators and the maximum pitch. A circle indicates that Gaussian contour with 2 sigma standard deviation for each emotion (red-Ang, green-Hap, black-Neu, blue-Sad). Different emotions show distinctive variation patterns in the articulatory speed dimension and the pitch dimension.

Source: Kim, Lee, and Narayanan (2010).

Современные инструментальные методы сбора артикуляционных данных включают:

- Ультразвук (Stone, 2005)
- Рентгеновский микроскоп (Fujimura, Kiritani,& Ishida, 1973)
- Электромагнитную артикулографию (EMA) (Perkell et al., 1992)
- Магнитно-резонансную томографию (MPT) (Narayanan, Alwan, & Haker, 1995; Narayanan,Nayak, Lee, Sethy, & Byrd, 2004)

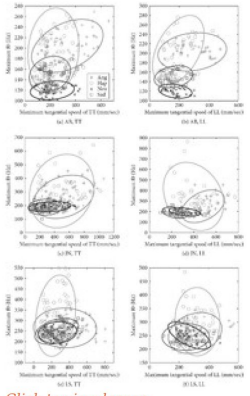
Плохие новости: испытуемым часто бывает сложно естественным образом выражать эмоции в этих средах сбора данных

Хорошие новости: систематические исследования показывают, что артикуляционные паттерны эмоциональной речи отличаются от нейтральной (неэмоциональной) речи

Lee, Yildirim,Kazemzadeh, & Narayanan, 2005

Исследование показало, что речевое производство эмоциональной речи больше связано с периферическими артикуляционными движениями, чем с нейтральной речью. Например, положение кончика языка, челюсти и губ более развито в эмоциональной речи, чем в нейтральной речи.

Грубо говоря: активная мимика связана с эмоциональной речью больше, чем с нейтральными интонациями



Click to view larger

Fig. 12. 2 Example plots of the maximum tangential speed of critical articulators and the maximum pitch. A circle indicates that Gaussian contour with 2 sigma standard deviation for each emotion (red-Ang, green-Hap, black-Neu, blue-Sad). Different emotions show distinctive variation patterns in the articulatory speed dimension and the pitch dimension.

Source: Kim, Lee, and Narayanan (2010).

Lee, Yildirim,Kazemzadeh, & Narayanan, 2005

Еще авторы этой работы провели исследование, где рассматривали четыре категории эмоций в качестве зависимой переменной: они предполагали, что по артикуляции можно определить саму эмоцию

Область глотки — для выражения радости

Счастье связано с большим подъемом гортани, чем гнев, нейтралитет и печаль

Сердитая речь вводит наибольшие модуляции скорости артикуляции, в то время как модуляции высоты тона были наиболее заметны для счастливой речи (Kim, Lee & Narayanan, 2010)

- Изменение артикуляционных положений и скорости, а также высоты тона и энергии в значительной степени связаны с силой восприятия эмоций в целом (Kim, Lee, & Narayanan, 2011)
- Голосовые качества — такие как резкий, напряженный, модальный, хриплый, шепчущий, скрипучий и слабый-скрипучий, а также их комбинации — связаны с аффективными состояниями с использованием синтезированной речи (Gobl & Chasaide, 2003).

Короче, интонации тоже важны

- Межъязыковая изменчивость включает неоднородное проявление эмоций и различия в структурах голосового тракта индивида (Lammert, Proctor, & Narayanan, 2013). Внутриговорящая вариативность обусловлена тем фактом, что говорящий может выразить эмоцию несколькими способами и зависит от контекста
- Инвариантный характер управления компонентами производства речи все еще остается неуловимым, что делает комплексное моделирование эмоциональной речи сложным и в значительной степени открытым.

Computation of Affective Speech Features

Анализ данных о производстве речи предполагает, что сложное взаимодействие между активностью источника голоса и модуляциями голосового тракта, вероятно, лежит в основе того, как эмоциональная информация кодируется в форме речевого сигнала

- Люди значительно точнее оценивают эмоциональное содержание, чем просто угадывают на уровне случайности, слушая записи речи (Bachorowski, 1999)
- Шерер описал всеобъемлющую теоретическую модель производства-восприятия вокальной передачи эмоций и предоставил подробный обзор того, как каждый акустический параметр (например, высота тона, интенсивность, скорость речи и т.д.) соотносится с различной интенсивностью восприятия эмоций (Scherer, 2003); это классическое исследование было дополнительно расширено в руководстве по исследованию невербального поведения, посвященном вокальному выражению аффекта (Juslin & Scherer, 2005)

Эти исследования обработки эмоциональной речи людьми заложили основы для аффективных вычислений с использованием речи благодаря ее обширному научному обоснованию

Меры сигнала, связанные с просодией

- Основная частота (f)
- Краткосрочная энергия
- Скорость речи: скорость слога / фонемы

Измерения спектральных характеристик

- Кепстральные коэффициенты малой частоты (MFCCC)
- Энергетические коэффициенты банка фильтров Mel (MFBs)

Меры, связанные с качеством передачи голоса

- Дрожание
- Мерцание
- Отношение гармоник к шуму

Мини-вывод

Вышеупомянутый подход к обработке данных эффективен в различных задачах прогнозирования эмоций

> Но остается неясным,какие аспекты механизмов эмоционального производства-восприятия улавливаются с помощью этой техники. С вычислительной точки зрения

> Эффективный и надежный распознаватель эмоций в реальной жизни, построенный на основе этого подхода, может оказаться непрактичным и слишком дорогим.

> Будущие работы заключаются в разработке более информированных функций, основанных на понимании механизмов восприятия эмоциональной речи при сохранении надежной точности прогнозирования по сравнению с текущим подходом.

Распознавание и моделирование аффектов на основании речи

Распознавание и отслеживание эмоциональных состояний в человеческих взаимодействиях на основе устных высказываний требует:

- определение схемы кодирования соответствующих эмоций и их меток
- нормализация признаков
- разработка фреймворков МО для моделирования изменений эмоций в диалогах

Разметка эмоций для вычислений

- самоотчеты (чаще в аффективной науке)
- внешняя разметка (чаще в моделировании)
- внешняя разметка как категориальные эмоциональные состояния, и как непрерывные шкалы
- выбор типа разметки зависит от задачи и модели (направлена ли она на восприятие эмоций или их воссоздание)

Последние достижения в области разметки эмоций

- включение описания поведения
- представление эмоций как профилей, т.е. смеси эмоций
- описание эмоций с помощью естественных выражений родного языка

Нормализация акустических характеристик

Крайне важно нормализовать речь от влияний среды и записи

Используется оценка глобальных акустических параметров

говорящих и высказываний

Итеративная нормализация признаков

Вычислительная структура для распознавания эмоций

- категориальная разметка: support vector machine, decision tree, naive Bayes, hidden Markov model
- непрерывные шкалы в разметке: методы регрессии

1. Распознавание эмоции по одному высказыванию

- hierarchical tree-based approach
- обработка четких перцептивных различий акустических характеристиках в начальных точках дерева, а весьма неоднозначные оценки эмоций на листьях дерева
- уровни дерева - решение простых задач классификации

2. Контекстно-зависимое распознавание эмоций в устных диалогах

- эмоции каждого участника взаимодействия сглажены во времени и обусловлены эмоциональным состоянием другого говорящего
- динамическая Байесовская сеть
- контекстно-зависимая структура распознавания
- эмоциональное содержание прошлых и будущих наблюдений может предоставить дополнительную контекстную информацию, способствующую точности классификации эмоций текущих высказываний

3. Отслеживание непрерывно оцениваемых атрибутов эмоций

- описание эмоций как на непрерывный поток, а не на последовательность дискретных состояний
- отслеживание непрерывных уровней активации, валентности и доминирования участников
- смешанная модель Гаусса
- сопоставление набора наблюдаемых аудиовизуальных сигналов с лежащим в их основе эмоциональным состоянием по трем осям

Подведение итогов

- Все еще активная область исследований
- Трудность обобщения и масштабирования подходов
- Неясно, как эмоциональная информация кодируется в акустических волнах
- Сложности со связью баз данных