

Emotion Recognition From Expressions in Face, Voice, and Body: The Multimodal Emotion Recognition Test (MERT)

Tanja Banziger, Didier Grandjean, and Klaus R. Scherer
University of Geneva

Correspondence concerning this article should be addressed to Klaus R. Scherer, Swiss Centre for Affective Sciences, University of Geneva, rue des Battoirs 7, CH-1205 Geneva, Switzerland. E-mail: klaus.scherer@unige.ch

Emotion recognition ability has been identified as a central component of emotional competence. We describe the development of an instrument that objectively measures this ability on the basis of actor portrayals of dynamic expressions of 10 emotions (2 variants each for 5 emotion families), operation-alized as recognition accuracy in 4 presentation modes combining the visual and auditory sense modalities (audio/video, audio only, video only, still picture). Data from a large validation study, including construct validation using related tests (Profile of Nonverbal Sensitivity; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Japanese and Caucasian Facial Expressions of Emotion; Biehl et al., 1997; Diagnostic Analysis of Nonverbal Accuracy; Nowicki & Duke, 1994; Emotion Recognition Index: Scherer & Scherer, 2008), are reported. The results show the utility of a test designed to measure both coarse and fine-grained emotion differentiation and modality-specific skills. Factor analysis of the data suggests 2 separate abilities, visual and auditory recognition, which seem to be largely independent of personality dispositions.

Распознавание Эмоций по Выражениям Лица, Голоса и Тела: Мультимодальный Тест Распознавания Эмоций

Таня Банцигер, Дидье Гранжан, и Клаус Р. Шерер
Женевский университет

Способность распознавать эмоции считается центральным компонентом эмоциональной компетентности. Мы описываем разработку инструмента на основе изображения актерами динамического выражения 10 эмоций (2 способами для каждого из 5 семейств эмоций), объективно измеряющего эту способность, операционализированную как точность распознавания в 4 формах презентации (аудио/видео, только аудио, только видео, фотография). Представлены данные большого исследования валидности, включая конструктивную с использованием родственных тестов (Profile of Nonverbal Sensitivity; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979; Japanese and Caucasian Facial Expressions of Emotion; Biehl et al., 1997; Diagnostic Analysis of Nonverbal Accuracy; Nowicki & Duke, 1994; Emotion Recognition Index: Scherer & Scherer, 2008). Результаты показывают полезность теста, предназначенного одновременно для измерения грубого и тонкого дифференцирования эмоций и модально-специфичных навыков. Факторный анализ данных предлагает 2 автономные способности, визуальное и слуховое распознавание, которые, кажется, в значительной степени независимы от личностных диспозиций.

В виду малой разработанности диагностического инструментария в области тестирования эмоциональной компетентности (интеллекта), направленного на комплексное исследование данного феномена, авторы поставили перед собой **цель** сконструировать тест, который бы отражал способность людей распознавать эмоции с различными степенями точности (грубое распознавание - способность различить основные семьи эмоций, тонкое распознавание - способность распознавать эмоции в пределах каждой семьи) в различных коммуникативных каналах. Авторами делается особый акцент на требовании однородности тестового материала, предполагающей задействование одних и тех же людей для представления эмоций во всех модальностях, для сравнимости получаемых индексов, как показателей индивидуальных особенностей испытуемого, что позволило бы прояснить вопрос о соотношении и взаимовлиянии различных коммуникативных каналов на точность распознавания эмоций.

Актеры включенные в тест

Изображения эмоций в тесте были воспроизведены 10 профессиональными сценическими актерами (5 мужчин и 5 женщин). Все актеры были носителями немецкого языка. Они все закончили профессиональные действующие школы и регулярно работали на радио, телевидении, и сцене. Им заплатили за участие. Ни один актер не использовался дважды для одной и той же категории эмоций. Неравная представленность актеров, по мнению авторов, уменьшает возможность ассоциации определенного актера с данным набором категорий эмоции и использование правила исключения. Мужчины и женщины равно представлены в тестовых заданиях (по 15) MERT .

Эмоции включенные в тест

На основе классификации Экмана (см. Ekman, 1994) были выбраны четыре пары эмоций, которые состоят из двух вариантов для каждой из четырех "основных" семей, которые наиболее часто включаются в исследования эмоций: гнев, страх, печаль, и счастье (горячий гнев/холодный гнев, панический страх/беспокойство, отчаяние/печаль, восторг/счастье). Кроме того, были включены изображения отвращения (меньшая интенсивность) и презрения (большая).

Вербальное Содержание Заданий в Тесте

Использовались следующие два бессмысленных предложения с фонематически эквивалентной структурой более раннего исследования (Scherer, Banse, Wallbott, & Goldbeck, 1991): (a) "*Hut sandig pron you venzu,*" и (b) "*Fee gott laich jonkill gosterr.*" 16 изображений, включенных в MERT, предъявляют первое предложение, и 14 - второе, критерии выбора устраняют возникновение систематических ассоциаций между эмоциями и предложениями.

Выбор Тестовых Стимулов

Для создания MERT, было отобрано 30 изображений эмоций, снятых в контексте долгосрочной научно-исследовательской работы по распознаванию эмоций (Banse & Scherer, 1996; Scherer & Ellgring, 2007).

Первоначальный блок, из которого извлечены изображения MERT, включает 224 видеозаписи актеров(со звуком), которые были отобраны в 2 этапа. Экспертные оценки были даны сначала 12 студентами, отобравшими наиболее "аутентичные" и "правдоподобные" изображения тех категорий, которые предполагались актерами и исследователями. Окончательный выбор был сделан на основе оценок 12 наивными испытуемыми. Для MERT отобрали 30 записей из этого блока согласно следующим критериям: три изображения для каждой из 10 категорий эмоций; ни один актер не изображал одну и ту же эмоцию дважды; эквивалентная гендерная представленность и приблизительное равное устное содержание в во всем тесте.

Чтобы создать "статическое" подмножество стимулов для теста, были извлечены фотоснимки из видео. Отбор фотографий делался рекурсивным способом: один сотрудник, извлекал несколько выразительных фотографий из каждого видео, а остальные (все исследователи в психологии эмоций) отбирали самую выразительную из них. В случаях, когда на втором этапе ни одна фотография не была признана эмоционально выразительной - снова выбирались снимки из клипов. Фотоснимки, на которых были закрыты глаза или преобладала артикуляция – отсеивались, фотографии с открытым и закрытыми ртами - нет.

Компьютерный Интерфейс

Для представления оцифрованных стимулов испытуемым был создан компьютерный интерфейс с 10 альтернативными категориями эмоций в форме кнопок, нажимаемых на экране. Инструкции испытуемым также предъявлялись на экране. Они были коротки и главным образом требовали, чтобы участники сосредоточились на стимулах и выбрали лучшее определение эмоции после каждого предъявления. Участникам сообщали, что MERT - тест невербальной чувствительности.

30 тестовых стимулов предъявлялись 4 различными способами: аудио/видео, только видео, только аудио и только фотография. Получившиеся 120 стимулов давались в одной из двух неизменных случайных последовательностей (для уменьшения потенциальных эффектов порядка на уровне анализа группы и оценки надежности ретеста). Последовательность никогда не включала тот же самый стимул (в других модальностях). Случайная последовательность исключала более двух

последовательных изображений одного и того же актера, одной и той же эмоции и способа ее презентации (www.affectivesciences.org/MERT).

Тесты, Используемые для проверки конструктивной валидности

DANVA

DANVA (DANVA2-AF and DANVA2-AP; Nowicki & Duke, 1994) включает 24 аудиозаписи и 24 фотографии выражений лиц (предъявляемых на 2с). Стимулы изображают одну из 4 категорий эмоций (гнев, страх, радость/счастье, печаль) с двумя уровнями интенсивности (слабый или сильный). Тест включает три изображения для каждой комбинации эмоции и интенсивности. В роликах участвуют молодые американцы (не профессиональные актеры); каждый из них появляется максимум два раза в отобранных изображениях. Аудио-стимулы сделаны двумя профессиональными актерами (один мужчина, одна женщина). Эмоции изображаются в следующем предложении: "я сейчас выхожу из комнаты, и я вернусь позже."

PONS

Тест *PONS* (Rosenthal et al., 1979) составлен из 20 аудио/видеозаписей. Один актер (молодая женщина, сотрудник исследовательского коллектива) изображает 20 отношений (например, "пытается обольстить кого-то", "произносит молитву", "восхищается природой", "выражает ревнивый гнев"). Отношения классифицируются или как доминирующие, или как покорные, или положительные или отрицательные. Каждую запись показывают 11 различными способами: 1 только лицо; 2 только тело (от шеи до коленей); 3 лицо и туловище (голова/лицо и тело вниз до талии, включая жесты руками); 4 только речь, пропущенная через фильтр нижних частот (без картинки); 5 случайно соединенная речь (Scherer, 1971) (без картинки); 6 - 11 комбинаций трех визуальных записей с двумя записями голоса/речи. Эти 220 изображений представлены в неизменном порядке, и испытуемые обязаны выбрать одно из двух потенциальных отношений для каждого изображения. Одна альтернатива соответствует исследовательской интенции, а другая выбрана случайным образом из оставшихся 19.

JACFEE

Данный тест использовался с инструкциями Мацумото (Matsumoto, LeRoux, Wilson-Cohn, and Rarogue et al. 2000). Тест включает 56 фотографий выражений лиц. 14 Кавказских мужчин, 14 Кавказских женщин, 14 азиатских мужчин, и 14 азиатских женщин, изображают одну из семи основных эмоций с конфигурациями лицевых особенностей (Ekman and Friesen, 1975). Эти семь эмоций маркировали удивление, печаль, гнев, счастье, страх, отвращение, и презрение. В тесте каждая картинка показывалась на 200 миллисекунд между двумя показами на 500 миллисекунд того же самого ничего не выражающего лица. Авторы данного исследования выбрали экспозицию в 0.2 с, которая достаточна, чтобы осознанно распознать выражение лица. Эта экспозиция - самая длинная, используемая авторами этого теста. Эти 56 последовательностей, "нейтральный-выразительный-нейтральный", представлены в неизменном порядке. Участник обязан выбрать один из семи альтернативных ответов.

ERI

Тест *ERI* (Scherer & Scherer, 2008) использует изображения теста Экмана и Фризена Лицевого Аффекта (PFA; Ekman and Friesen, 1976), серия из 30-пунктов/5-категорий (гнев, страх, радость, печаль, отвращение) включает субтест распознавания выражений лица (FACIAL-INDEX) и субтест аудио изображений эмоций, отобранных из большого числа записей, созданных четырьмя немецкими профессиональными радио-актерами (VOCAL-INDEX) (см. Scherer et al., 1991). Оба теста не ограничены по времени, представляются автоматически на экране компьютера.

Личностные опросники

Чтобы исследовать влияние личностных черт, связанных с эмоциональной чувствительностью, участники заполнили следующие личностные опросники: (а) Большая пятерка НЕО 5-факторный опросник (NEO-FFI; Нейротизм [N], Экстраверсия [E], Открытость [O]; RoUand, Parker, & Stumpf, 1998), (b) шкала, разработанная Carver and White (1994), оценивающая межличностную чувствительность Поведенческой Активации/Угнетения систем (BIS-BAS) основанных на модели Грея; (c) Опросник Беспокойства (State-Trait Anxiety Inventory -STAI), измеряющий беспокойство как черту (Spielberger & Sydeman, 1994); и (d) Потенциал Управления (Management Potential -Mp) специальная шкала Калифорнийского Опросника Индивидуальности (CPI; Gough, 1984), состоящая - из эмпирически отобранных пунктов, которые успешно делят группы людей с различными степенями руководящей ответственности.

Участники

Студенты-психологи, 63 женщины и 9 мужчин, со средним возрастом 22 года ($SD=4$ года), участвовавшие в обмен на зачет. Им также обещали и дали персонифицированную обратную связь по их результатам для всех невербальных тестов чувствительности (MERT, DANVA, ERI, IACFEE, and PONS).

Процедура

Все тесты и опросники в этом исследовании были выполнены на индивидуальных компьютерах в компьютерном классе. Компьютеры были оборудованы стандартными наушниками. Аудио параметры были одинаковым на всех компьютерах. Участники проходили тестирование в группах из 10 (минимум) - 20 (максимум) человек. Каждый участник прошел 3 сессии:

В первой сессии MERT и опросники.

Во второй сессии (на следующей неделе), тех же самых участников просили закончить другие тесты невербальной чувствительности, включая PONS ($N = 68$ участников из первоначального объединения), ERI Vocal ($N = 68$) и and Facial ($N = 72$), DANVA ($N = 70$), и IACFEE ($N = 67$). Наконец, ретестовая сессия MERT была организована спустя приблизительно 6 недель после первой сессии (с 67 участниками от первоначального числа). Последовательности представления стимулов MERT в тестовой и ретестовой серии для каждого из участников не совпадали.

В виду технических причин и пропусков некоторыми испытуемыми одной или нескольких сессий часть данных была потеряна. Например, второе выполнение MERT было закончено 67 из 72 участников, которые прошли тест 6 неделями ранее.

Обработка Результатов

Вычисление Баллов Точности в Четырех Тестах

Точность определялась как пропорция правильных ответов, данных участником, к общему числу заданий в тесте. В случаях, когда это было возможно, также вычислялись субиндексы для отдельных каналов экспрессии.

Для PONS были сгруппированы два варианта аудио стимулов, чтобы вычислить единый индекс точности. Так же были вычислены индексы точности PONS для изображений лица и индекс, основанный на изображениях, комбинирующих лицо и голос.

Для DANVA, вычислены показатели точности для изображений лица и голоса.

Для MERT были вычислены четыре специфических коэффициента точности для следующих способов представления стимулов: (а) только видео; (б) только фотографии; (с) аудио только; и (д) комбинация видео и аудио.

В дополнение к общему показателю для MERT - индекс индивидуального распознавания эмоций (IR), был вычислен индекс семейного распознавания (FR), для которого ответ рассматривается как правильный, если сообщается о целевой эмоции или другом члене соответствующей семьи эмоций. Этот коэффициент представляет собой грубое распознавание, т.е., способность идентифицировать главные категории эмоции в отличительных выражениях.

Одним из источников несовместимости тестов является различное число вариантов ответов в тесте. Для решения этой проблемы был взят индекс пропорции, или pi (Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008; Rosenthal & Rubin, 1989). Pi дает среднюю точность, которая вычисляется как пропорция, независимо от того сколько вариантов ответа имел каждый пункт.

Согласованность

Для оценки согласованности была вычислена средняя корреляция для всех участников. Полученные значения по каждому способу были: аудио/видео $=.35$, аудио $=.38$, видео $=.40$, и фотоснимок $=.39$, приводя к среднему числу $.38$ для всех способов. Согласованность участников является большой, показывая, что паттерн ответов согласован для участников несмотря на индивидуальные различия (см. Zentner, Grandjean, & Scherer, 2008).

Надежность теста

В обеих сессиях MERT приняло участие 67 человек. Корреляции индексов между первой и второй сессией были следующими: аудио/видео $r = .55$, аудио $r = .56$, видео $r = .46$, и фотоснимок $r = .56$. Корреляция общих баллов достигла $r = .78$.

Точность Распознавания в Различных Модальностях

Таблица 1 показывает средние баллы точности по способам представления и эмоциям. Диапазон коэффициента FR (минимум =.63, максимум =.85) является приблизительно на треть меньше чем IR (минимум =.43, максимум =.76), предполагая, что люди расходятся меньше в грубом чем тонком распознавании, что подтвердили матрицы несоответствий, построенные отдельно для каждого из различных способов: большинство несоответствий встречается между членами семьи.

Дисперсионный анализ ANOVA точности по способу (аудио/видео, аудио, видео, и фотоснимок), уровню интенсивности/побуждения (высокий, низкий), и эмоциям (пять эмоций), как факторы показали следующие эффекты:

1. способ, $F(3, 213) = 119.16, p < .001$, динамические изображения лица (видео с или без голоса) распознавались лучше, чем фотоснимки или только аудио;
2. уровень интенсивности/побуждения, $F(1, 71) = 30.88, p < .001$, более интенсивные эмоции распознавались лучше, чем менее интенсивные;
3. и эмоция, $F(4, 284) = 78.68, p < .001$, как и в более раннем исследовании в этой области, есть значимые различия в точности распознавания для различных эмоций.

Все взаимодействия были значимыми Способ X Эмоций, $F(12, 852) = 41.48, p < .001$, Способ X Интенсивность, $F(3, 213) = 10.61, p < .001$, Эмоция X Интенсивность, $F(4, 284) = 65.75, p < .001$, и Способ X Эмоций X Интенсивность, $F(12, 852) = 25.67, p < .001$. Рисунок 1 показывает средние проценты правильных ответов для эмоций во всех четырех способах. Данные показывают четкие взаимодействия между способом представления и эмоциями.

Индивидуальные Различия в Точности

Кластерный анализ правильного распознавания различных эмоций для каждого участника для различных способов выявил три группы участников, выделенных на основе расстояний между кластерами (linkage distance > 2). Дисперсионный анализ ANOVA по способу (аудио / видео, аудио, видео, и фотоснимок) и группе (Группа 1, $N = 30$; Группа 2, $N = 23$; и Группа 3, $N = 19$), показал основное влияние способа, $F(3, 207) = 157.19, p < .001$, и группы, $F(2, 69) = 66.55, p < .001$, и эффект взаимодействия Способ X Группа, $F(6, 207) = 6.09, p < .001$. Средний баллы для этих трех групп приведены на Рисунке 2. Этот анализ показал, что Группа 2 показала наилучшие результаты, хотя точность распознавания аудио и фотоснимков была сравнительно низкой. Группа 1 сравнительно хорошо справилась с аудио/видео и видео заданиями, но значимо отличается от Группы 2 в распознавании звуковой информации, $F(1, 69) = 7.99, p < .01$, и фотографий, $F(1, 69) = 45.51, p < .001$. Результаты 3 Группы значительно ниже двух других, но при этом они повторяют профиль Группы 2. Что позволяет сделать предположение о том, что члены Группы 1 полагаются в большей степени на динамические сигналы в визуальной области, и оказываются в недоумении, когда доступны только статические выражения лица, как на фотоснимках.

Корреляции Индексов и Факторная Структура

В Таблице 2 представлены корреляции между четырьмя индексами MERT. Все они являются значимыми, указывая на значительное перекрытие в выполнении теста всеми четырьмя способами. В то же самое время, количество разделенной дисперсии не превышает 25 %, предполагая, что есть вполне достаточное пространство для индивидуальных различий.

Факторный анализ методом главных компонент суммарных оценок для четырех отдельных способов с 72 участниками дает один фактор с собственным значением > 1, который объясняет 55.5% дисперсии. Со 2-м фактором решение объясняет 74.6 % дисперсии. Нагрузки (после варимак вращения) приведены в Таблице 3. Эти результаты предполагают, что может быть два отдельных фактора компетентности распознавания эмоций, один визуальный и аудиальный. Интересно отметить, что аудио/видео индекс более высоко нагружен по аудиальному фактору, чем визуальному.

Пошаговый регрессионный анализ показателя аудио/видео представления показывает, что аудиальный показатель вводится на первом шаге ($r^2 = .196$), видео - на втором (увеличивающий r^2 до .302, аудио ($\beta = .348, p = .001$); видео ($\beta = .339, p = .01$); фотография не входит в уравнение. Таким образом, несмотря на в общем более низкую точность выполнения, слуховое впечатление, кажется, вносит больший вклад как компонент в комбинированном аудио-визуальном восприятии/умозаключении.

Личностные Корреляты Точности Распознавания

Значимые корреляции между индексами MERT и личностными чертами NEO-FFI или -Беспокойством (STAI) получены не были. Шкала BIS/BAS дала значимые положительные корреляции между субшкалой BAS Drive и аудио MERT (.263), видео (.314), и аудио/видео (.260), полученные показатели предполагают, что индексы MERT могут частично зависеть от "мотивации достижения" испытуемых, особенно учитывая высокий уровень трудности теста.

Дополнительно проведенный ANOVA дал следующие результаты. Главный эффект для Bas-D, $F(2, 69) = 9.00, p < .001, \eta^2 = .21$, подтверждает и уточняет значимые корреляции, показывая, что Группа 3 ($M = 2.9$) имеет значительно более низкие мотивационные показатели, чем Группы 1 (4.1) и 2 (3.9). Единственное различие между Группами 1 и 2, получено по шкале CPI Mr, $F(2, 69) = 3.26, p < .05, \eta^2 = .08$, показывая, что Группы 1 ($M = 49.1$) и 3 (47.3) значительно ниже чем Группа 2 (56.0) по шкале CPI.

Оценка Конвергентной Валидности Инструмента (Конструктивная Валидность)

В Таблице 4 представлена описательная статистика общих показателей точности для всех тестов. Средний сырой балл точности для MERT ниже чем среднее индексы точности в других тестах, что отражает более низкий уровень случайности и более высокую трудность заданий (см. Таблицу 1). Однако, если нивелировать влияние разного числа альтернатив в тестах посредством pi , MERT достигает тех же или даже лучших показателей чем другие тесты. Только у PONS, только с двумя альтернативами ответа, есть несколько более низкое pi по сравнению с другими тестами.

Таблица 5 представляет описательные статистики для показателей точности в различных модальностях, вычисленных для нашей выборки DANVA, PONS и ERI.

Таблицы 6, 7, и 8 демонстрируют конвергенцию индексов и модально специфичных субиндексов, полученных в других тестах. Для определения значимости использовался односторонний критерий, поскольку есть априорное предположение, что тесты способностей коррелируют друг с другом.

Таблица 6, представляя корреляции общих баллов по тестам, показывает, что общие показатели MERT, JACFEE и PONS значимо коррелируют ($r = .45$ к $.59$), указывая на то, что они вероятно все отражают основную компетентность невербального распознавания эмоций. Общий балл ERI коррелирует достаточно высоко с показателем MERT FR (вероятно из-за сопоставимого типа эмоций и альтернатив ответов) и с JACFEE (который также основан на фотоснимках). Баллы DANVA сходятся в меньшей степени с полученными в других тестах; только одна корреляция рядом с 10 % разделенное различие ($r = .30$ с JACFEE).

В таблице 7, представлены корреляции субшкал, основанных на динамических видео изображениях (выражения лица с/без голоса), которые показывают, что только баллы, полученные для динамических видео изображений в PONS и в MERT (выражение лица с голосом), значительно - коррелируют. Баллы, полученные в двух тестах на основе тех же самых динамических изображений лица, но без голосовых изображений дают значимых связей. Это наблюдение сначала кажется трудно объяснимым. Однако, это может быть связано с тем фактом, что в тесте PONS баллы, полученные для изображений лица со звуком, не коррелируют с баллами, полученными для изображений лица без звука (корреляция незначима), тогда как индексы для этих двух способов коррелируют в MERT (см. Таблицу 4).

Таблица 8 представляет корреляции индексов, основанных только на статических изображениях лица и на аудио. Как показано в Таблице 8, есть достаточно высокие значимые корреляции между баллами по фотоснимкам, полученных в MERT, ERI, и JACFEE. Аналогичные индексы DANVA снова не дали согласованных корреляций. Для коэффициентов, звукового распознавания (Таблица 8), MERT значимо коррелирует с DANVA и ERI. Соответствующие баллы DANVA, ERI, и PONS, не дают высоких корреляций друг с другом.

В целом, корреляции в Таблицах 7 и 8 слабее, чем корреляций в Таблице 6, несмотря на то, что модальности разделены. Это происходит вероятно из-за меньшего числа пунктов, включенных в вычисление баллов для различных способов представления внутри тестов, по сравнению с баллами, вычисленными на большом количестве пунктов для различных тестов, что соответственно снижает надежность.

В целом основываясь на данных можно признать конструктивную валидность MERT удовлетворительной.

Обсуждение

К недостаткам теста авторы относят неоднородное качество материала, который изначально был отскан на пленку и лишь спустя некоторое время оцифрован, которое, однако, по мнению исследователей, значимо не влияет на качество теста.

Продолжительность теста составляет 45 минут, что затрудняет его использование.

Гендерное соотношение не является удовлетворительным. Но, во-первых, данные по гендерным различиям в сфере распознавания эмоций утверждают, что эффект влияния этого аспекта очень мал, и для его выявления потребовалось бы масштабное исследование с привлечением большого числа участников, что является достаточно проблематичным; а, во-вторых, исследователями не ставилась задача разработки тестовых норм и исследовать систематические индивидуальные различия, включая гендерные и возрастные. Дополнительно проведенный анализ ANOVA не выявил влияния пола на тестовые показатели.

Критерий Точности

Можно было бы поднять проблему критерия правильности ответа. В этом тесте коммуникативные интенции актеров определяют выраженную эмоцию. Кроме того, используемые изображения были извлечены из ряда предварительно отобранных записей, чтобы гарантировать разумное распознавание аудио-визуальных стимулов (см. Banse & Scherer, 1996). Предварительный отбор пунктов является стандартной процедурой в исследовании при использовании изображений актеров, иногда критикуемый за смещение на результатов в определенном направлении). Однако, предварительный выбор заданий экспертами - стандартная процедура разработки большинства тестов, поскольку должна быть обоснована уместность пункта в контексте диагностируемого конструкта. Пока нормальные психометрические индикаторы свидетельствуют о приемлемой сложности заданий и их дискриминативности, фундаментальный критерий для диагностики индивидуальных различий достигнут. Кроме того, наши результаты показывают, что предварительный выбор на основе аудиовизуальных изображений не обязательно облегчает распознавание тех же самых изображений в других способах представления. Как обрисовано в общих чертах ранее, вполне может случиться так, что некоторые способы лучше передают определенные эмоции.

Действие

Изображения актеров, используемые в исследовании выражений эмоций, часто отмечаются как недостаточно реалистичные и правдоподобные. Можно оспорить это широко распространенное предубеждение. Например, тем фактом, что люди постоянно разыгрывают эмоции в действительности (применение правил показа или управление экспрессией в манипулятивных целях; Goffman, 1959; Griffiths, 2003), при этом достигая разумной степени подлинности и правдоподобности (см. также Banse & Scherer, 1996; Banziger & Scherer, 2007, 2008).

Личностные Корреляты Точности распознавания

Индекс BAS-D фактически предсказывает различия групп в точности распознавания. Предполагается, что выполнение тестов подвержено влиянию индивидуальной мотивации достижения. К сожалению, этот эффект почти количественно никогда не оценивается.

Различия групп, полученные по шкале CPI Mr также логично согласуются с успешностью выполнения MERT. Как было показано ранее (Scherer, 2007), эта субшкала CPI демонстрирует высокие значимые корреляции с эмоциональной стабильностью, регулированием эмоций, стрессоустойчивостью, функциональными копингами, и объективными показателями флюидного интеллекта, предполагая, что высокие оценки указывают на общее хорошее регулирование, эмоциональную компетентность, и "разумность".

Сложность Пунктов

В MERT 28 % пунктов с очень высокими коэффициентами распознавания (на них дается больше 80% правильных ответов) по сравнению с 52 - 62 % в четырех других тестах. Процент пунктов, распознаваемых ниже *случайного уровня*, определенного как функция числа альтернативных ответов в данном тесте, составлял 4 % (по сравнению с процентами в пределах от 0 % до 16 % для других тестов). Таким образом, в целом, уровень трудности пунктов MERT несколько

выше по сравнению с другими тестами, главным образом из-за меньшего числа "легких" заданий. Что хорошо, поскольку процент легких заданий превышающий 50 может ограничить способность теста различать широкий диапазон людей.

Учитывая цель построить тест, который в состоянии отличить целый диапазон основного континуума компетентности, со специальным акцентом на высоких уровнях навыка, необходимых, чтобы отличить тонкое различие между членами одной семьи эмоций, которые отличаются главным образом по интенсивности, наша оценка трудности пунктов и их дискриминативности дала удовлетворительные результаты, по сравнению с теми же самыми индексами для других тестов распознавания эмоций.

Надежность

Поскольку не возможно вычислить стандартную надежность пунктов психометрическом смысле, использовалась средняя корреляция согласованности оценок как индекс воспроизводимости. Результаты были удовлетворительными, показывая высокую согласованность по профилю различий точности для эмоции, несмотря на значительные индивидуальные различия в уровне точности.

Более высокая корреляция общих баллов вероятно отражает общее наблюдение, что значение, вычисленное на основании большего числа пунктов, более надежно в целом, чем для меньшего N . Средние показатели точности в ретесте были приблизительно на 3-5% выше чем для первой сессии. Учитывая, что участники улучшили свои баллы, что обычно наблюдается в ситуациях ретеста, вовлекающих компетентности научения. В данном случае, это - даже большая проблема, поскольку каждому изображению показывается четыре раза, хотя и в различных коммуникативных каналах. Поскольку три из этих четырех способов содержат сходные лицевые конфигурации, можно ожидать некоторые эффекты переноса и более высокий потенциал обучения.

Конвергентная и Конструктивная Валидность

Корреляции общих баллов и субиндексов MERT и других тестов, используемыми в этом исследовании, демонстрируют вполне достаточную конструктивную валидность для MERT.

Корреляции Субиндексов и Факторная Структура

Матрица корреляции и факторная структура MERT предлагают два отдельных фактора, один из которых отвечает за распознавание лица, а другой голоса. Этот концепт подтверждает более ранние данные, которые показывают сходную картину. Таким образом, тесты распознавания эмоции должны отдельно оценивать эти две компетентности, поскольку кажется маловероятным существование единственного мощного общего фактора.

Результаты MERT представленные в этой статье, полученные на общем материале с участием одних и тех же людей (что устраняет возможность низких корреляций в других тестах, таких как ERI и DANVA, из-за индивидуальных различий людей представленных в тестовом материале различных модальностей), предполагает, что способности правильно расшифровать и динамические сигналы и статические конфигурации в лица могут быть различными у разных людей. Так, в дополнение к группе испытуемых, которые были вообще менее квалифицированными чем лучшая группа, есть третья группа, которая, кажется, столкнулась только со сложностями в расшифровке статических визуальных стимулов реплики. Подход MERT позволит исследователям внести ясность в эти вопросы в будущем.

Тестовые баллы

MERT, как другие тесты, используемые в этом исследовании валидности, дает показатели точности, отражающие степень компетентности испытуемого в идентификации целевых эмоций, изображенных актерами и подтвержденных экспертным отбором заданий (см. Banse & Scherer, 1996). Однако, в случае MERT, множество IR показатель отражает реальную способность распознавания, а не только способность различить среди несколько категорий, как это часто бывает в случае использования нескольких основных эмоций. Последние приближается к показателю FR, который оценивает способность человека сделать грубую оценку качества выраженной эмоции. Суженный диапазон FR предлагает, что грубая способность дискриминации, которой вероятно способствуют случайные угадывания, является элементарной способностью, которая присутствует, по крайней мере до известной степени, у большинства людей. Это не удивительно, поскольку

грубая дискриминация - вероятно обязательный навык для любого вида социального взаимодействия. Однако, сомнительно, что это является хорошим основанием для диагностики - эмоционального интеллекта в смысле компетентности восприятия. Необходим более чувствительный инструмент, позволяющий измерить способность к более тонкому анализу. На основании на ранее представленных результатах, был получен индекс (IR/FR), который выражает потенциал для такого тонкого дифференцирования (тонкий индекс дифференцирования, SDI). Другими словами, чем выше этот индекс, тем выше способность человека пойти вне грубого дифференцирования в направлении тонкого распознавания. Например, для испытуемого с FR .85 и IR .79, SDI достигает.89 (предполагающий, что человек способен к прекрасной дифференциации в большинстве случаев), тогда как у испытуемого с FR .78 и IR .54 - SDI .69, указывает на больший разрыв между грубым и точным распознаванием. Эти индексы, как часть результатов MERT позволяют давать детализированную обратную связь о силе и слабостях каждого кандидата. Главное преимущество MERT состоит в том, что в дополнение к общим показателям (IR и FR), субиндексы позволяют дифференцировать способности распознавания эмоций по способу представления стимула. Рисунок 3 показывает субиндексы для всех участников исследования, отсортированных по аудио/видео показателю точности в порядке убывания. Хотя значения для звукового способа является вообще самым низким, есть люди с удивительно высокими показателями, приближающимися к аудио/видео показателю. Как было показано на группах, выделенных с помощью кластерного анализа, у некоторых людей, кажется, есть особые проблемы с правильным опознанием эмоций по фотоснимкам. Наконец, есть люди, видео показатели которых фактически превышают аудио/видео показатели, которые в общем случае дают самые высокие значения точности. Одним из возможных объяснений этого может быть то, что их интуитивное чувствование правильных визуальных сигналов (например, мимики) интерферируется ошибочными выводами слухового канала. Этот сильный эффект звуковых сигналов в комбинации с визуальными также подтверждается тем фактом, что аудио показатель предсказывает звуковое/видео более эффективно, чем делают визуальные. Поскольку это - первая попытка внесения ясности в данный вопрос, то она пока является только иллюстрацией, требуя большего всестороннего эмпирического исследования.

Перевод: Новтоцкой-Власовой Е.
novototskaya@gmail.com

Table 1
Descriptive Summaries for MERT Accuracy Subscores

Variable	<i>M</i>	<i>SD</i>	<i>pi</i>
Total scores			
Individual recognition	.61	.07	.93
Family recognition	.75	.05	.96
Mode-specific scores (IR)			
Audio/video	.70	.09	.95
Audio	.49	.10	.90
Video	.69	.08	.95
Picture	.56	.10	.92
Emotion-specific scores (IR)			
Low-intensity arousal			
Anxiety	.42	.14	.87
Happiness	.73	.19	.96
Cold anger	.55	.27	.92
Sadness	.59	.11	.93
Disgust	.61	.19	.93
Total	.59	.18	.92
High-intensity arousal			
Panic fear	.56	.27	.92
Elation	.72	.18	.96
Hot anger	.87	.09	.98
Despair	.41	.22	.86
Contempt	.66	.11	.95
Total	.64	.17	.94

Note. *N* = 72. MERT = Multimodal Emotion Recognition Test; *pi* = proportion index; IR = individual emotion recognition.

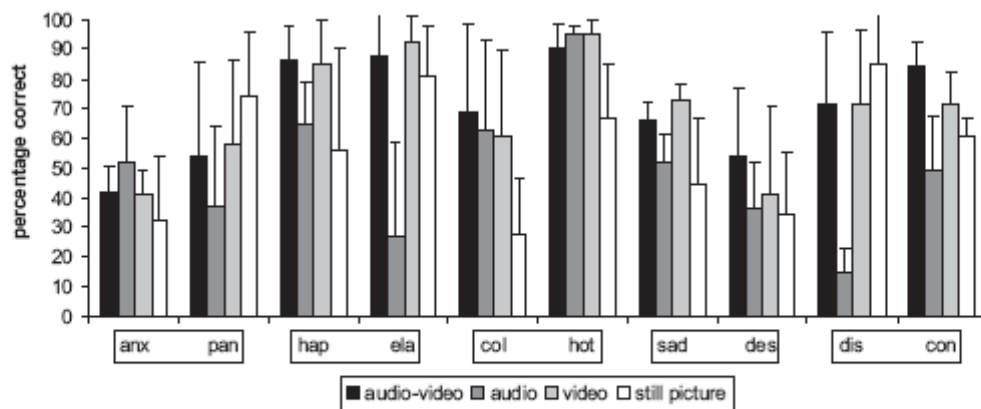


Figure 1. Mean accuracy scores for the basic emotion family pairs in four modes. anx = anxiety; pan = panic fear; hap = happiness; ela = elation; col = cold anger; hot = hot anger; sad = sadness; des = despair; dis = disgust; con = contempt.

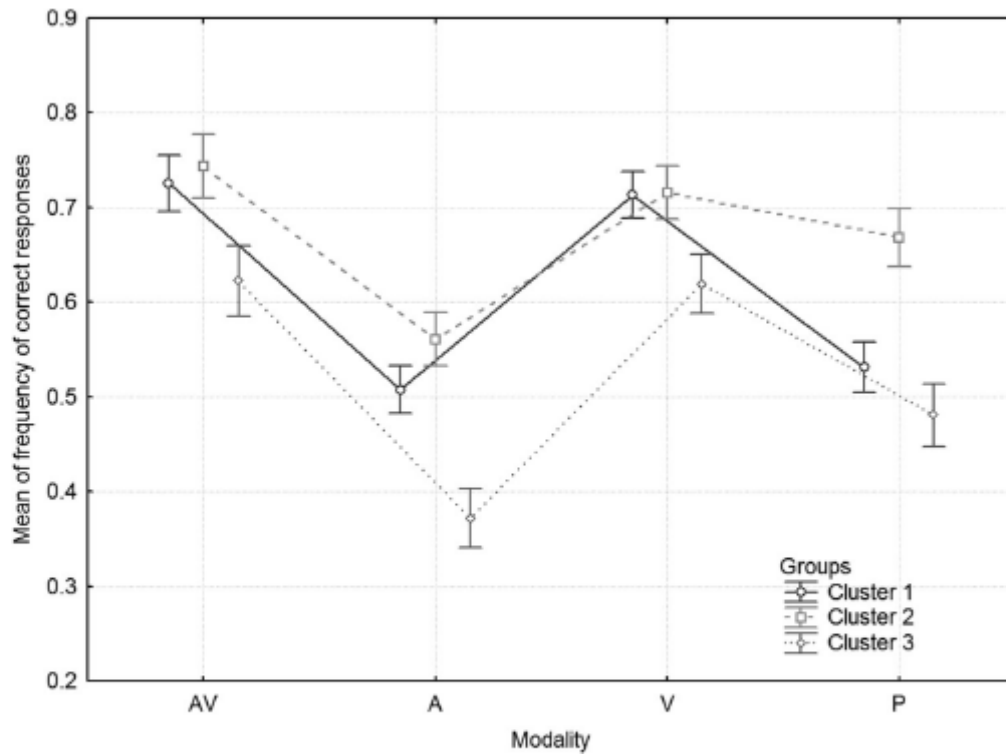


Figure 2. Mean frequency of correct responses by mode for the three clusters of participants. AV = audio/video; A = audio; V = video; P = still picture.

Table 2
Correlations Between Total Accuracy Subscores for Different Modalities in the Multimodal Emotion Recognition Test (MERT)

MERT	Audio-video	Video	Audio
Video	.436**		
Audio	.443**	.279*	
Still picture	.336**	.535**	.411**

Note. $N = 72$.

* $p < .05$ (two-tailed). ** $p < .001$ (two-tailed).

Table 3
Factor Structure of the Multimodal Emotion Recognition Test, Principal Component Analysis of Total Accuracy Scores per Mode

Mode	Component 1	Component 2
Video	.887	
Still picture	.799	
Audio		.889
Audio/video	.326	.736

Note. Varimax rotation with Kaiser normalization, factor loadings of four presentation modes (loadings smaller than .3 are not displayed).

Table 4
Descriptive Summaries of Accuracy Scores for the Four Tests

Test	<i>M</i>	<i>SD</i>	Options	<i>pi</i>
DANVA	.77	.054	4	.91
ERI	.71	.059	5	.91
JACFEE	.76	.097	7	.95
MERT	.61	.071	10	.93
PONS	.77	.046	2	.77

Note. *N* = 72 (except JACFEE, *N* = 69). *pi* = proportion index; DANVA = Diagnostic Analysis of Nonverbal Accuracy; ERI = Emotion Recognition Index; JACFEE = Japanese and Caucasian Facial Expressions of Emotion; MERT = Multimodal Emotion Recognition Test; PONS = Profile of Nonverbal Sensitivity.

Table 5
Descriptive Summaries of Accuracy Subscores in the Other Tests

Test	Mean	<i>SD</i>	<i>pi</i>
DANVA ^a			
Face	.83	.071	.94
Voice	.71	.085	.88
PONS ^b			
Face	.81	.074	.81
Voice	.62	.084	.62
Face and voice	.84	.061	.84
ERI ^c			
Facial	.75	.062	.92
Vocal	.67	.087	.89

Note. For DANVA, the two subscores were obtained on independent portrayals (not produced by the same senders), whereas for the Multimodal Emotion Recognition Test and PONS the same portrayals were presented repeatedly in different modes. *pi* = proportion index; DANVA = Diagnostic Analysis of Nonverbal Accuracy; PONS = Profile of Nonverbal Sensitivity; ERI = Emotion Recognition Index.

^a *N* = 72. ^b *N* = 70. ^c Facial *n* = 72, vocal *n* = 68.

Table 6
Correlations Between Total Scores in Different Tests

Test	MERT IR	MERT FR	DANVA	ERI	JACFEE
MERT FR	.854*				
DANVA	.217*	.224*			
ERI	.277*	.383**	.159		
JACFEE	.455**	.433**	.302**	.303**	
PONS	.505**	.420**	.144	.087	.589**

Note. *Ns* vary between 65 and 72 (see the Method section). MERT = Multimodal Emotion Recognition Test; IR = individual recognition; FR = family recognition; DANVA = Diagnostic Analysis of Nonverbal Accuracy; ERI = Emotion Recognition Index; JACFEE = Japanese and Caucasian Facial Expressions of Emotion; PONS = Profile of Nonverbal Sensitivity.

* *p* < .05 (one-tailed). ** *p* < .001 (one-tailed).

Table 7
Correlations Between Scores Based on Facial Dynamic Portrayals (With and Without Voice)

Test	PONS dynamic facial	PONS dynamic face and voice
MERT dynamic video	.181	.087
MERT dynamic audio/video	.235*	.335**

Note. $N = 68$. PONS = Profile of Nonverbal Sensitivity; MERT = Multimodal Emotion Recognition Test.

* $p < .05$ (one-tailed). ** $p < .001$ (one-tailed).

Table 8
Correlations of Scores Based on Still Pictures of Facial Portrayals or Vocal Portrayals

Test	MERT		DANVA		ERI	
	Photo	Audio	Facial	Vocal	Facial	Vocal
DANVA facial ^a	.224*					
ERI facial	.296***		.146 ^b			
JACFEE	.325***		.278**		.300***	
DANVA vocal		.323**				
ERI vocal		.274* ^d		.013 ^e		
PONS vocal		.252* ^e		.235 ^a		.032 ^e

Note. MERT = Multimodal Emotion Recognition Test; DANVA = Diagnostic Analysis of Nonverbal Accuracy; ERI = Emotion Recognition Index; JACFEE = Japanese and Caucasian Facial Expressions of Emotion; PONS = Profile of Nonverbal Sensitivity.

^a $N = 70$. ^b $N = 72$. ^c $N = 67$. ^d $N = 66$. ^e $N = 68$.

* $p < .05$ (one-tailed). ** $p < .001$ (one-tailed).

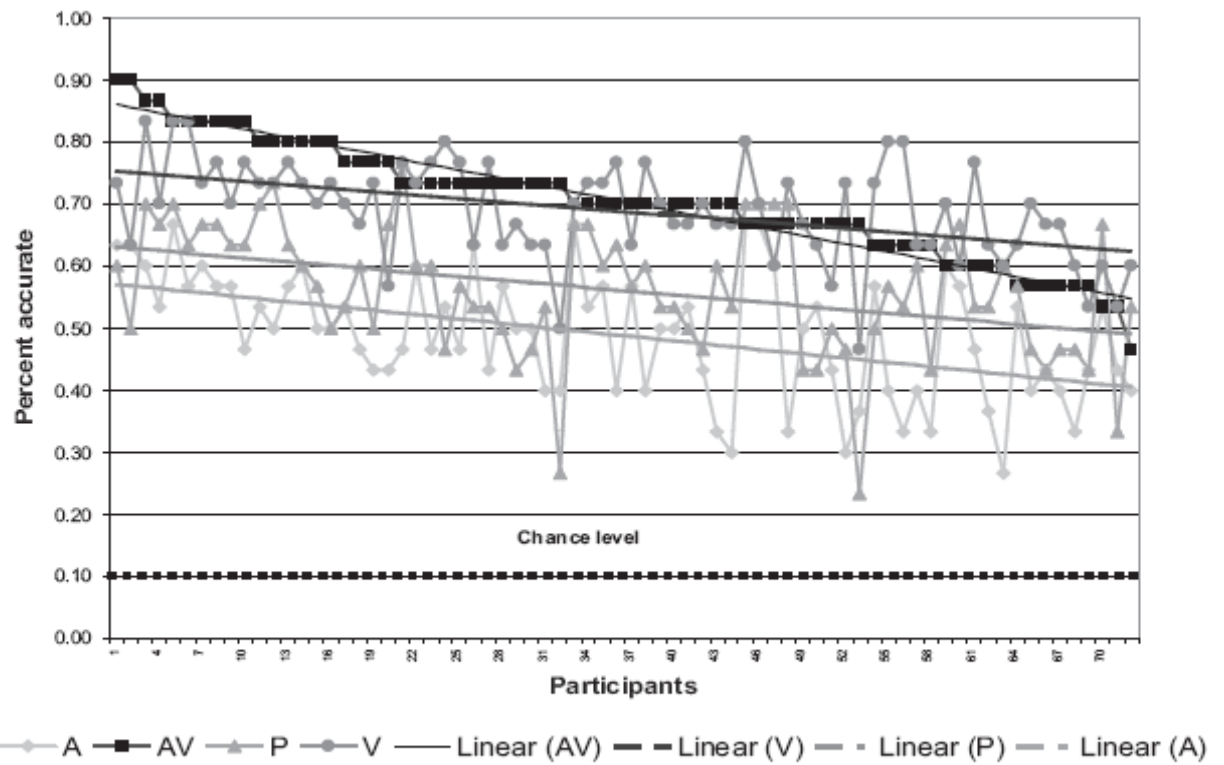


Figure 3. Plot of the mode-specific accuracy subscores for all participants. AV = audio/video; A = audio; V = video; P = still picture.

Supplemental material

Table 1

Proportion of Items in the Four Tests With Respectively Low and High Difficulty

	MERT	PONS	JACFEE	DANVA	ERI
Proportion of items recognized by more than 80% of the participants	28%	60%	52%	62%	53%
Proportion of items recognized below chance level	4%	16%	0%	4%	7%
Chance level	10%	50%	14%	25%	20%

Note. MERT = Multimodal Emotion Recognition Test; PONS = Profile of Nonverbal Sensitivity; JACFEE = Japanese and Caucasian Facial Expressions of Emotion; DANVA = Diagnostic Analysis of Nonverbal Accuracy; ERI = Emotion Recognition Index.

Table 2

MERT Average Scores and Standard Deviations for Test and Retest in All Modalities

	Audio/video		Audio		Video		Still picture	
	1st test	retest	1st test	retest	1st test	retest	1st test	retest
Average score	.70	.73	.49	.54	.69	.72	.56	.61
Standard deviation	.09	.09	.10	.10	.08	.09	.11	.11

Note. $N = 67$. MERT = Multimodal Emotion Recognition Test.

c) Video Mode

Selection	Target emotions (correct answers) – dynamic facial portrayals										Total
	anx	pan	hap	ela	col	hot	sad	des	dis	con	
Anxiety	41	22	3	0	0	0	3	5	2	0	8
Panic fear	15	58	0	1	0	0	0	41	0	0	12
Happiness	0	0	85	2	0	0	0	0	0	0	9
Elation	0	0	6	93	0	0	0	0	0	0	10
Cold anger	21	5	2	0	61	5	1	1	2	16	11
Hot anger	2	1	0	0	4	95	0	2	0	3	11
Sadness	7	3	0	1	3	0	73	9	7	3	11
Despair	9	11	0	1	2	0	21	41	1	1	9
Disgust	0	0	0	1	2	0	1	0	72	6	8
Contempt	4	1	3	0	27	0	0	0	15	71	12
Total	100	100	100	100	100	100	100	100	100	100	100

d) Still Picture Mode

Selection	Target emotions (correct answers) – still pictures										Total
	anx	pan	hap	ela	col	hot	sad	des	dis	con	
Anxiety	32	9	0	0	5	1	10	5	0	1	6
Panic fear	44	75	0	4	0	16	7	38	0	0	18
Happiness	0	0	56	4	3	0	0	0	0	0	7
Elation	0	0	41	81	0	0	0	0	0	0	12
Cold anger	6	1	0	2	27	7	6	2	3	6	6
Hot anger	1	0	0	1	4	66	1	4	5	2	8
Sadness	5	5	0	0	27	0	44	14	0	7	10
Despair	5	9	0	1	5	0	22	34	0	2	8
Disgust	1	1	1	6	10	5	3	1	85	21	14
Contempt	4	0	1	0	18	4	6	1	6	61	10
Total	100	100	100	100	100	100	100	100	100	100	100

Note. anx = anxiety, pan = panic fear, hap = happiness, ela = elation, col = cold anger, hot = hot anger, sad = sadness, des = despair, dis = disgust, con = contempt.

Figure 1

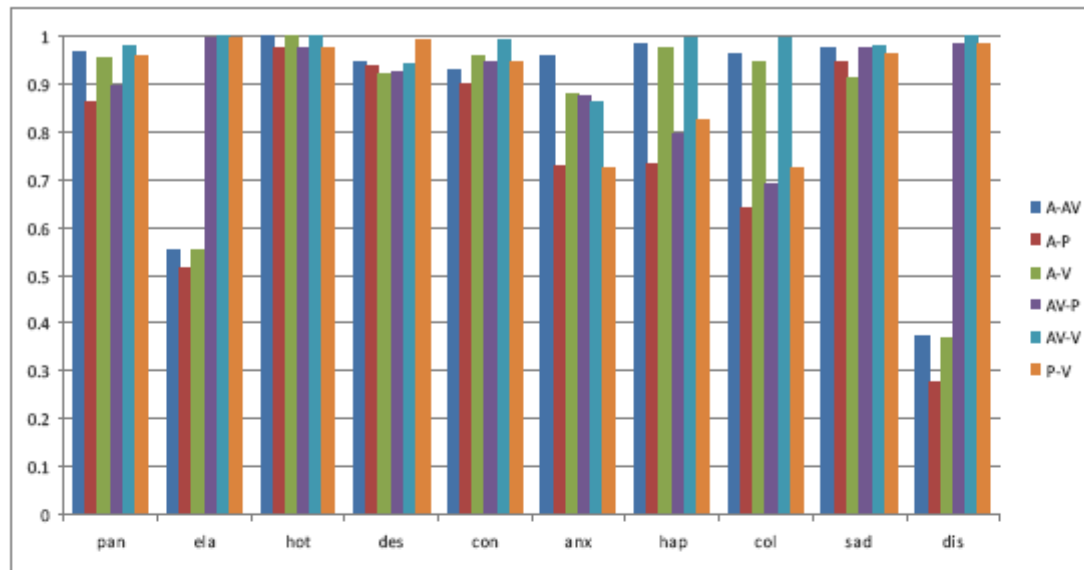


Figure 1. Profile correlations between confusion patterns for the different modes and for the emotional categories (AV = audio/video, A = audio, V = video, P = still picture, anx = anxiety, pan = panic fear, hap = happiness, ela = elation, col = cold anger, hot = hot anger, sad = sadness, des = despair, dis = disgust, con = contempt).

Figure 2

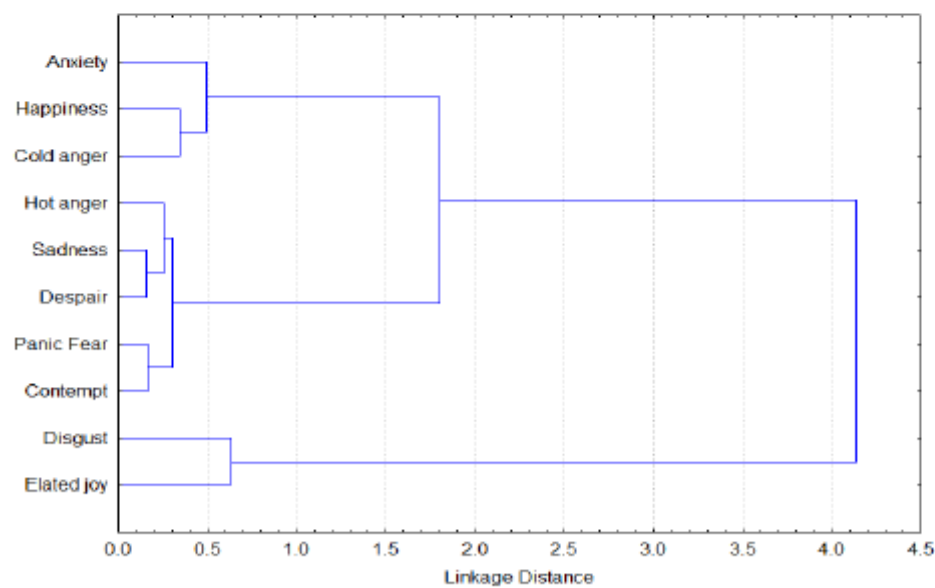


Figure 2. Results of cluster analysis of the matrix of correlations across confusion patterns.