

Терейковська Л.О.

Київський національний університет будівництва і архітектури

АНАЛІЗ ЗАСОБІВ РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ ЛЮДИНИ ЗА ГОЛОСОМ

Стаття присвячена проблемі впровадження засобів розпізнавання емоційного стану людини в інформаційні системи різноманітного призначення. Визначена обмеженість поширених засобів розпізнавання емоцій, що базуються на аналізі геометрії обличчя людини. Показано перспективність застосування засобів розпізнавання емоцій на основі аналізу інтонації та тембру голосу людини. На основі результатів аналізу відомих розробок сформовано базовий перелік характеристик, які можуть бути використані для оцінки відповідності зазначених засобів умовам поставленої задачі розпізнавання. Запропоновано співвіднести шляхи подальших досліджень із розробкою процедури оцінки ефективності засобів розпізнавання та формуванням переліку характеристик засобів розпізнавання емоцій на основі таких біометричних параметрів, як колір підшкірного візерунка кровоносних судин обличчя та малюнок сітківки ока.

Ключові слова: емоційний стан, голосовий сигнал, тембр голосу, інтонація голосу, розпізнавання емоцій, засоби розпізнавання.

Постановка проблеми. Однією з важливих тенденцій розвитку інформаційних систем загального та спеціального призначення є впровадження в них засобів розпізнавання емоційного стану (далі – ЗРЕС). Так, відомі спроби застосування ЗРЕС у системах розпізнавання впливу на підсвідомість у мультимедійних повідомленнях засобів масової інформації під час ведення інформаційних війн, для діагностування психічних захворювань, в індустрії розваг і системах управління розумним будинком. Викликає інтерес застосування ЗРЕС у системах дистанційної освіти для оптимізації навчального навантаження. Саме на ЗРЕС базується функціонування більшості відомих систем розпізнавання брехні (поліграфів). Крім того, впровадження ЗРЕС в інтерфейс інформаційних систем загального призначення є одним із сучасних напрямів їх вдосконалення. Нині найбільш поширеними є ЗРЕС на основі аналізу геометрії обличчя людини. Однак суттєвою перешкодою використання таких ЗРЕС є необхідність отримання якісного зображення обличчя. Тому одним з основних напрямів розвитку ЗРЕС є застосування в них аналізаторів інших біометричних параметрів. Перспективним вважається застосування аналізатора голосового сигналу (далі – ГС) людини [6; 10]. Це пояснюється тим, що сьогодні засоби реєстрації ГС є невід’ємною частиною більшості відомих інформаційних систем. Водночас практичний досвід і результати робіт [2; 9] вказують на необхідність модернізації таких ЗРЕС у напрямі зменшення ресурсоемності, збільшення точності розпізнавання, скорочення

терміну розробки та підвищення адаптованості до інших умов застосування, що і зумовлює актуальність досліджень у цьому напрямі.

Аналіз останніх досліджень і публікацій. Загалом під поняттям «емоції» розуміють особливий клас психічних станів, які відображають ставлення людини до навколишнього світу, до інших людей, до самої себе та до результатів своєї діяльності [1; 3; 8]. У сучасній психології виділяють сім базових емоцій, прояви яких не залежать від раси і культури людини. Зазначимо, що базові емоції – це елементарні емоції, які більше ні на що не розщеплюються і самі можуть бути складниками інших складних емоцій. До базових емоцій належать радість, гнів, відраза, смуток, страх, подив, зневага. Слід зауважити, що в деяких роботах [4; 7] виділяють не сім, а чотири базові емоції: щастя (радість), смуток (сум), переляк (страх), здивування і гнів / відраза (огиду). Кількість і номенклатура складних емоцій досі чітко не визначена, а відповідна теорія далека від досконалості. Всі емоції поділяються на позитивні і негативні, а також на стенічні й астенічні. Позитивні емоції пов’язані із задоволенням потреб, а негативні викликають почуття незадоволення і вимагають зміни ситуації. Стенічні емоції активізують організм і піднімають настрій (гнів, лют, захват), астенічні (туга, печаль, смуток, сором) – розслаблюють людину і пригнічують активність організму. Завдання розпізнавання емоцій є слабоформалізованим, а його рішення ускладнюється невизначеністю характеристик проявів емоцій у ГС. Зазначимо, що

ГС – це складний акустичний сигнал, джерелом якого є голос людини. Оскільки близько 40% інформації людина сприймає завдяки інтонації голосу та тембру ГС, розробку ЗРЕС пов'язують з аналізом вказаних характеристик. Зазначимо, що інтонація ГС є збірним поняттям, що включає тональність, силу звучання, темп мовлення, наголоси і паузи. З погляду розпізнавання емоцій найбільш важливими складниками є тональність ГС і темп мовлення [3; 9]. Тональність ГС відображає в ньому емоційно-вольову установку людини у досягненні конкретної мети. За типом впливу розрізняють три види тональності: мажорну, мінорну, нейтральну. Виділяють такі характеристики тональності, як спокій і напруженість. Спокій – відносно рівний стан, а напруженість – нестійкий стан тональності, що характеризується накопиченням певної кількості сигналів іншої тональності, подальше збільшення якої здатне привести до якісного стрибка і переходу в іншу тональність. Темп мовлення відображає швидкість проголошення різних елементів мови (звуків, складів, слів). Тембр голосу – психоакустична характеристика, зумовлена формантною структурою спектра ГС, ступенем вираженості шумових компонентів, негармонійних обертонів, модуляційними процесами, зростанням і згасанням ГС. Як і тональність ГС, тембр голосу відображає яскравість звуку, його індивідуальність і багато в чому визначається поточним психологічним станом людини [10; 11].

У більшості відомих ЗРЕС тональність і тембр ГС оцінюють на основі висоти основного тону й обертонів. Типова технологія оцінки полягає у фільтрації ГС, визначенні квазістаціонарних фрагментів, спектральному аналізі кожного із вказаних квазістаціонарних фрагментів, розрахунку мел-кепстральних коефіцієнтів (коефіцієнтів лінійного передбачення) та власне у розпізнаванні поточних емоцій. Експлуатаційні характеристики конкретних ЗРЕС визначаються реалізацією кожного з означених етапів. Наприклад, спектральний аналіз квазістаціонарних фрагментів може бути реалізований за допомогою методу Фур'є або за допомогою вейвлет-перетворень. У проаналізованій літературі [1–11] не знайдено обґрунтованого переліку характеристик ЗРЕС, адаптація яких до поставленого завдання розпізнавання зумовлює ефективність використання зазначених засобів.

Постановка завдання. Основною метою публікації є формування переліку характеристик, що можуть бути використані для оцінки відпо-

відності засобів розпізнавання емоційного стану людини на основі ГС умовам поставленої задачі розпізнавання. Для цього слід провести аналіз відомих рішень у сфері створення таких засобів розпізнавання.

Виклад основного матеріалу дослідження. Сьогодні відома достатньо велика кількість ЗРЕС на основі голосового сигналу. Однак обмежений обсяг публікації й обмежена доступність деяких ЗРЕС зумовила можливість аналізу тільки найбільш відомих та апробованих рішень.

Комп'ютерна система Вищої школи економіки (Росія) (perm.hse.ru) дозволяє на основі аналізу ГС розпізнавати вісім базових емоцій – нейтральність, спокій, щастя, сум, гнів, страх, відразу та здивування [8]. Заявлена середня точність розпізнавання сягає 71%. Точність розпізнавання нейтральності та спокою є більш високою, а точність розпізнавання страху, суму, здивування та відрази нижча за середню. Система має експериментальний характер. Основою блоку дикторонезалежного розпізнавання є рекурентна нейронна мережа типу LSTM, що застосовується для аналізу ГС, поданого у вигляді набору мел-кепстральних коефіцієнтів. У системі передбачено фільтрацію білого шуму та вирівнювання гучності ГС.

Створена в Ізраїлі **технологія відстеження емоцій у реальному часі Beyond Verbal** (www.beyondverbal.com) передбачає можливість визначення близько 400 варіантів емоційного стану диктора. Технологія використовується рядом компаній і передбачає використання різноманітних механізмів обробки ГС. Остаточна класифікація емоційного стану реалізується за допомогою спеціальних шаблонів, для створення яких було проаналізовано 2 300 000 аудіозаписів ГС зі 170 країн світу. Технологія передбачає можливість як дикторозалежного, так і дикторонезалежного розпізнавання емоцій.

Мобільний додаток Xpression розроблений компанією EI Technologies із Великої Британії (www.ei-technologies.com) і здатен розпізнавати 4 емоційні стани людини: спокій, щастя, сум, гнів і тривогу. Додаток має клієнт-серверну архітектуру. Клієнтський модуль використовується для реєстрації ГС та його попередньої обробки. На сервері аналізуються фрагменти ГС тривалістю 200 мілісекунд. Для цього використовується нейронна мережа, на вхід якої подаються параметри, які описують гучність ГС, його тональність і темп мовлення. Процес розпізнавання має дикторозалежний характер,

Характеристики засобів розпізнавання емоцій на основі голосу

№	Опис характеристик
Особливості розпізнавання	
N _{1,1}	Повнота розпізнавання базових емоцій
N _{1,2}	Можливість розпізнавання складних емоцій
N _{1,3}	Адаптованість до мови диктора
N _{1,4}	Адаптованість до статі диктора
N _{1,5}	Адаптованість до віку диктора
N _{1,6}	Можливість фільтрації шуму
N _{1,7}	Можливість нормалізації гучності ГС
N _{1,8}	Можливість розпізнавання в умовах натовпу
N _{1,9}	Дикторозалежність процесу розпізнавання
N _{1,10}	Можливість числової оцінки емоційного стану
N _{1,11}	Можливість аналізу динаміки емоцій
N _{1,12}	Можливість аналізу спонтанного мовлення
N _{1,13}	Необхідність використання навчальних баз даних, що містять приклади ГС із різними емоціями
N _{1,14}	Точність розпізнавання
N _{1,15}	Обчислювальна складність розпізнавання
N _{1,16}	Часова складність розпізнавання
Технічна реалізація	
N _{2,1}	Можливість інтеграції з системами управління базами даних
N _{2,2}	Можливість інтеграції з різноманітними системами реєстрації аудіопотоку
N _{2,3}	Можливість аналізу різноманітних форматів аудіофайлів
Сервісні можливості	
N _{3,1}	Верифікація користувача
N _{3,2}	Визначення віку
N _{3,3}	Визначення раси
N _{3,4}	Визначення статі
N _{3,5}	Можливість інтеграції з ЗРЕС, що базуються на інших біометричних ознаках

що передбачає створення шаблону конкретного користувача. Надалі додаток стежить за відхиленнями тембру голосу, його гучності, зміною проміжків часу між звуками і словами та порівнює все це з означеним шаблоном. Класифікація здійснюється на підставі розрахованих відхилень. У процесі розробки визначено, що представники різних народів говорять із різною швидкістю та гучністю, а люди різного віку та статі мають різні інтонації.

У роботі [4] наведено опис експериментальної системи розпізнавання емоцій у людській мові з використанням вейвлетів. Система призначена для розпізнавання в аудіозаписах 4 базових емоцій: гніву, страху, радості та норми. Наведено детальний опис математичного забезпечення системи. Розпізнавання реалізується на основі величин коефіцієнтів кореляції, для розрахунку яких використовується спектральний аналіз ГС на основі вейвлету Морле.

У роботі [7] описана програма для вимірювання емоцій абонента мобільного телефону. Результатом роботи програми є графік емоційної кривої абонента, що може бути використаний для розпізнавання його емоцій шляхом порівняння з емоційними кривими базових і складних емоцій.

Програма *EmoWatch* від японської компанії Smartmedical (eng.smartmedical.jp/) призначена для ідентифікації та розпізнавання емоцій користувачів мобільних пристроїв компанії Apple. Програма базується на технології Emrpath, яка аналізує основний тон, обсяг і швидкість ГС. Завдяки цій інформації програма дозволяє аналізувати енергетичні рівні користувача і розпізнавати гнів, спокій, радість, сум. Програма є дикторозалежною та має можливість розпізнавання незалежно від мови користувача.

Крім розглянутих систем розпізнавання емоцій, відомий цілий ряд комплексів, функціональність яких також базується на аналізі ГС. Ці комплекси призначені для вирішення таких завдань: верифікації користувачів, визначення функціонального стану людини, розпізнавання брехні, створення різноманітних голосових ефектів, визначення раси, віку та статі.

Проведений аналіз дозволив виділити три групи характеристик ЗРЕС. Перша група характеризує особливості розпізнавання ЗРЕС, друга група – особливості технічної реалізації, а третя – сервісні (додаткові) можливості ЗРЕС. Короткий опис характеристик представлено в табл. 1. Надалі цей перелік може бути уточнений і доповнений.

За період 2014–2018 рр. опубліковано декілька баз даних (далі – БД), у яких містяться аудіозаписи голосів людей із різними емоціями. Найбільш відомими є Emo-DB, RAVDESS та UMSSSED. Наприклад, БД Emo-DB (emodb.bilderbar.info) містить записи голосу професійних акторів німецькою мовою. БД включає 535 записів ГС 10 дикторів (5 чоловіків, 5 жінок), що відтворюють набір дискретних 7 базових емоційних станів. Наявність цих БД, з одного боку, сигналізує про прогрес у сфері ЗРЕС на основі ГС, а з іншого – уможлиблює вдосконалення таких засобів.

Результати проведеного аналізу дозволили за допомогою виділених характеристик оцінити можливість проаналізованих засобів розпізнавання

емоцій. У базовому варіанті оцінки виставлені за бінарної шкалою (0/1), що відповідає відсутності / наявності зазначених можливостей. Отримані оцінки для першої групи характеристик частково показані в табл. 2. Акцентування уваги на першій групі характеристик продиктовано тим, що тематика подальших досліджень пов'язана з розробкою ефективних моделей і методів розпізнавання емоцій, виражених у ГС.

Таблиця 2

Оцінки першої групи характеристик засобів розпізнавання.

Назва ЗРЕС	Характеристика ЗРЕС				
	N _{1,1}	N _{1,2}	N _{1,3}	N _{1,4}	N _{1,5}
Комп'ютерна система Вищої школи економіки	1	0	0	0	0
Beyond Verbal	1	1	1	1	1
Xpression	1	0	1	1	1
Система розпізнавання емоцій у людській мові	1	0	0	0	0
Програма для вимірювання емоцій абонента мобільного телефону	1	0	0	0	0
EmoWatch	1	0	1	0	0

Як впливає з аналізу даних табл. 2, більшість проаналізованих ЗРЕС погано пристосовані до

зміни статі та віку диктора. Більшість відомих ЗРЕС не розпізнає складні емоції.

Формування переліку характеристик ЗРЕС дозволяє окреслити напрям наступного етапу досліджень – розробки процедури визначення шляхів забезпечення ефективності зазначених засобів в ІС різного призначення. За аналогією з відомими рішеннями у сфері розробки багаторівневих систем розпізнавання ГС [6] в основі процедури можливо використовувати такий принцип: ефективність ЗРЕС визначається тим, наскільки повно характеристики зазначених засобів відповідають особливостям ІС.

Висновки. Проведене дослідження дозволило обґрунтувати базовий перелік характеристик засобів автоматизованого розпізнавання емоційного стану людини за тембром та інтонацією голосу, який дозволяє оцінити відповідність зазначених засобів умовам поставленої задачі розпізнавання. Визначено, що напрями подальших досліджень полягають у розробці процедури визначення шляхів забезпечення ефективності засобів розпізнавання емоцій в інформаційних системах різного призначення. Також доцільно сформулювати перелік характеристик засобів розпізнавання емоцій на підставі аналізу таких біометричних ознак, як колір підшкірного візерунка кровеносних судин особи та малюнок сітківки ока.

Список літератури:

1. Вартанов А.В. Антропоморфный метод распознавания эмоций в звучащей речи. *Национальный психологический журнал*. Москва, 2013. № 2 (10). С. 69–79.
2. Гришук Т.В., Биков М.М. Моделирование процесса анализа и классификации голосовых команд: монография. Вінниця: УНІВЕРСУМ-Вінниця, 2008. 146 с.
3. Заболеева-Зотова А.В., Орлова Ю.А., Розалиев В.Л., Бобков А.С. Задача создания системы автоматизированного распознавания эмоций. *Материалы междунар. науч.-техн. конф. «OSTIS-2012»*. Минск, 2012. С. 347–350.
4. Иванович А.В., Бредихин А.И. Распознавание эмоций в человеческой речи с использованием вейвлетов. *Материалы LXII междунар. науч.-практ. конф.* Новосибирск, 2018. С. 30–42.
5. Лебедева Н.Н., Каримова Е.Д. Акустические характеристики речевого сигнала как показатель функционального состояния человека. *Успехи физиологических наук*. 2014. Т. 45. № 1. С. 57–95.
6. Михайленко В.М., Терейковська Л.О., Терейковський І.А., Ахметов Б.Б. Нейромережеві моделі та методи розпізнавання фонем в голосовому сигналі в системі дистанційного навчання: монографія. Київ: ЦП «Компринт», 2017. 252 с.
7. Михайлов В.О. Программная реализация измерения эмоций абонента мобильного телефона. *Современные проблемы науки и образования*. 2012. № 5. С. 45–62.
8. Попова А.С., Рассадин А.Г., Пономаренко А.А. Детектирование эмоций в мультимедиа контенте. *Материалы междунар. науч.-техн. конф. «ИСТ – 2017»*. Москва, 2017. С. 852–857.
9. Ребрун І.А., Сидоров К.В., Филатова Н.Н., Ханеев Д.М. Модель проявления эмоций в естественной речи. *Материалы четырнадцатой национальной конф. по искусственному интеллекту «КИИ – 2014»*. Казань, 2014. С. 112–121.
10. Сидоров К.В. Биотехническая система мониторинга эмоций человека по речевым сигналам и электроэнцефалограммам: дис. ... канд. тех. наук: 05.11.17. Тверь, 2015. 182 с.
11. He L. Stress and Emotion Recognition in Natural Speech in the Work and Family Environments: Ph.D. Dissertation. *RMIT University*. 2010. 197 p.

АНАЛИЗ СРЕДСТВ РАСПОЗНАВАНИЯ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА ПО ГОЛОСУ

Статья посвящена проблеме внедрения средств распознавания эмоционального состояния человека в информационные системы различного назначения. Определена ограниченность распространенных средств распознавания эмоций, основанных на анализе геометрии лица человека. Показана перспективность применения средств распознавания эмоций на основе анализа интонации и тембра голоса человека. На основе результатов анализа известных разработок сформирован базовый перечень характеристик, которые могут быть использованы для оценки соответствия указанных средств условиям поставленной задачи распознавания. Предложено соотнести пути дальнейших исследований с разработкой процедуры оценки эффективности средств распознавания и формированием перечня характеристик средств распознавания эмоций на основе таких биометрических параметров, как цвет подкожного узора кровеносных сосудов лица и рисунок сетчатки глаза.

Ключевые слова: эмоциональное состояние, голосовой сигнал, тембр голоса, интонация голоса, распознавание эмоций, средства распознавания.

ANALYSIS OF MEANS OF RECOGNITION OF THE EMOTIONAL STATE OF HUMAN BY VOICE

The article is devoted to the problem of introducing means of recognition of the human emotional state into information systems of various purposes. The limited limitations of the common means for recognizing emotions based on the analysis of human geometry are determined. The perspective of the use of emotional recognition methods is shown on the basis of intonation analysis and human voice timbre. On the basis of the results of the analysis of known developments, a basic list of characteristics has been formed that can be used to assess the conformity of the indicated means with the conditions of the problem of recognition. It is proposed to correlate the ways of further research with the development of a procedure for evaluating the effectiveness of means of recognition and the formation of a list of characteristics of emotional recognition devices based on such biometric parameters as the color of the subcutaneous pattern of blood vessels of the face and the retinal eye.

Key words: emotional state, voice signal, timbre of voice, voice intonation, emotional recognition, recognition mean.