Проблематика фонетичної ідентифікації емоційних станів мовця за допомогою лінгвістичних та паралінгвістичних засобів

К. В. Лисенко

Київський національний університет імені Тараса Шевченка, місто Київ, Україна
Corresponding author. E-mail: lysenkokath@gmail.com

Paper received 11.11.20; Accepted for publication 25.11.20.

https://doi.org/10.31174/SEND-Ph2020-240VIII71-09

Анотація. У статті розглядаються проблеми визначення емоційних станів мовця за допомогою лінгвістичних та паралінгвістичних засобів. Розмовляючи, ми передаємо нашим слухачам інформацію про наш емоційний стан та ставлення як до наших слухачів, так і до того, що ми говоримо. Хоча в цілому специфічна здатність до декодування емоційної складової висловлювання залежить від окремо взятої емоції. Відомо, що носії мови, при прослуховуванні акторів, які читають емоційно нейтральний текст, водночас зображуючи емоції, перцептивно не завжди здатні правильно їх ідентифікувати. Навіть якщо емоційне вираження через просодію не завжди може бути усвідомлено розпізнане учасниками розмови, на неї може підсвідомо впливати інші показники, такі як тональність голосу, наприклад, або міміка. Цей вид вираження випливає не лише з лінгвістичних чи семантичних ефектів. Вміння пересічної людини розшифрувати за просодичними малюнками емоційні стани протягом розмовн розвинене меньше, ніж традиційна здатність до ідентифікації виразів обличчя. Чи сприймає людина просодію як позитивну, негативну чи нейтральну, завжди координується з мімікою. Якщо вона стає ближчою до нейтральної, на інтерпретацію емоційного стану впливає просодична інтерпретація. Для прийняття рішення щодо рівня емоційності слухачі часто змушені вслуховуватись в просодію висловлювання. Щоб мати можливість визначити рівень афективності висловлювання, слухачам необхідно не меньше 600 мілісекунд просодичної інформації, нижче яких буває недостатньо часу для обробки емоційного контексту висловлювання. Оскільки дослідження попередніх часів зосереджувались на таких просодичних особливостях мовлення, як основна середня частота та діапазон, темп та інтенсивність, в наш час стає необхідним проведення аналізу даних з використанням ширшого спектру голосових ефектів, включаючи паралінгвістичні. Подібні дослідження не часто проводились в минулому, хоча вони виконують важливу роль у вираженні емоцій. Розробивши градієнтне розмежування просодичних та паралінгвістичних особливостей у мовленні, вчені-дослідники на просодичному кінці шкали розташували інтонацію з її показниками, що свідчить про його лінгвістичність, а на іншому кінці шкали - такі риси, як якість голосу, або шуми та всілякі клацання, що є паралінгвістичними ознаками. Нелінгвістичні або паралінгвістичні особливості класифікують на два види: індивідуальні варіації та рефлекси. Індивідуальна варіація включає ефекти фонації та резонансу через фізіологію мовця та особливості голосового тракту. Слід брати до уваги також рефлекторні стани, оскільки вони часто є мимовільною ознакою справжнього емоційного напруження. В цілому слід зазначити, що емоційна або афективна просодія проявляється як у лінгвістичних, так і в паралінгвістичних аспектах мовлення, які дозволяють та допомагають як передавати, так і розуміти емоції. Вона включає індивідуальний тон голосу, який передається через зміну висоти, гучності, тембру, темпу мовлення та пауз.

Ключові слова: *просодія, емоційна реакція, лінгвістичні та паралінгвістичні особливості, сегментарні та супрасегментні функції та ознаки, голосова модифікація*.

Останнім часом розшифровка емоцій у мовленні з використанням компьютерного інструментарію стає доволі популярним завданням. Вважається, що визначити та правильно усвідомити емоційний стан мовця неважко. У випадках яскравих емоційних реакцій людина говорить гучніше, швидше, та частота основного тону у неї є незвично високою, незвично низькою або зі значними перепадами висоти тону. Однак це занадто спрощений підхід, оскільки носії мови, слухаючи акторів, які читають емоційно нейтральний текст, водночас зображуючи емоції, не завжди здатні правильно їх ідентифікувати: так, «щастя» визначається лише в 62% випадків, «гнів» - у 95%, «здивування» у 91%, «смуток» в 81% та «нейтральний стан» у 76%. При обробці цієї самої бази даних за допомогою комп'ютера сегментарні функції дозволяли більше, ніж в 90% випадків визнати емоції «щастя» та «гніву», тоді як надсегментарні, просодичні особливості дозволили розпізнати емоції лише у від 44% до 49% випадків. Що стосувалось емоційного стану «здивування», то ця емоція була ідентифікована лише у 69% випадків за сегментарними ознаками, і у 96% випадків - за суперагментальною просодією.

У звичайній, типовій розмові (без участі акторської гри) можливості розпізнавання емоцій можуть бути ще нижчими, порядку 50%, що перешкоджає складній взаємопозв'язаній функції мовлення.

Однак, навіть якщо емоційне вираження через просодію не завжди може бути усвідомлено розпізнане учасниками розмови, на неї може підсвідомо впливати тональність голосу. Цей вид вираження випливає не з лінгвістичних чи семантичних ефектів. Вміння пересічної людини розшифрувати за просодичними малюнками емоційні стани протягом розмовн меньш розвинене, ніж традиційна здатність до ідентифікації виразів обличчя, хоча слід зауважити, що в цілому специфічна здатність до декодування залежить від окремо взятої емоції. Отже, точність ідентифікації найбільш універсальних емоцій не є однаковою. Так, "гнів" і "смуток" носіями мови ідентифікуються найшвидше, "страх" і "щастя" не так точно та швидко. Натомість "огида" є найбільш проблематичною емоцією для ідентифікації.

Іншими словами, найточнішим є визначення психічних станів «гніву» та «смутку», натомість стан «огиди» ідентифікується найважче. Скоріше за все для прийняття рішення щодо рівня емоційності слухачі змушені вслуховуватись в просодію висловлювання. Крім того, чи сприймає людина просодію як позитивну, негативну чи нейтральну, завжди координується з мімікою. Якщо міміка обличчя стає ближчою до нейтральної, просодична інтерпретація впливає на інтер-

претацію емоційного стану. Разом з тим дослідження Марка Д. Пелла [14] показало, що слухачам необхідні 600 мілісекунд просодичної інформації, щоб мати можливість визначити афективний тон висловлювання. Нижче цього показника слухачам буває недостатньо інформації для обробки емоційного контексту висловлювання. Розмовляючи, ми передаємо нашим слухачам інформацію про наш емоційний стан та ставлення як до наших слухачів, так і до того, що ми говоримо.

Збереження на комп'ютері та аналіз достатньо великого кола аудіозаписів емоційного мовлення зробив би можливим методологічно обґрунтовані узагальнення щодо вираження емоцій у мовленні.

Література про емоційно-позитивний вплив паралінгвістичних особливостей здебільшого ґрунтується на дослідженнях досить малих кількостей даних, як правило, імітованих професійними чи аматорськими акторами в лабораторії. В декількох дослідженнях вчені намагалися штучно за допомогою акторів зобразити емоції, чим викликати емоційну реакцію у суб'єктів [7], [8]. З іншого боку, деякі вчені ставлять під сумнів обґрунтованість подібних симуляційних досліджень. Так, у критичному огляді подібних робіт Крамер [9] вказує, що актори, зіткнувшись із завданням висловити велику кількість емоцій таким чином, щоб їх можна було диференціювати, можуть створити "стереотипні уявлення про емоції, які можуть не виникати природно", на що важко знайти адекватний коментар. Крім того, дослідженням навмисного збудження емоцій зазвичай перешкоджають етичні проблеми. Навіть якщо їх можна подолати, використовуючи більш м'які форми подразників, зв'язок між стимулом та реакцією аж ніяк не є простим. Реакції різних суб'єктів на один і той же стимул можуть не тільки бути різними, а й непередбачувано змінюватися залежно від їх досвіду та психотипу особистості [17]. Крім того, часто буває практично неможливо виміряти рівень реально пережитого стресу.

Слід зазначити, що випадки використання аудіо записів справжніх, не-імітованих чи спеціально викликаних емоцій для подальшого фонетичного дослідження є малочисельними та обмеженими. Серед найпоширеніших джерел подібних даних є записи бортових реєстраторів пілотів літаків або космонавтів під час небезпеки. Як от, наприклад, дослідження Williams & Stevens [18], Kuroda et al., [10] та Simonov & Frolov [16], а також дані аудіо записів розмов із різноманітних менш екстремальних ситуацій. Помітним джерелом є також робота Браун [2], який робив аудіо записи людей під час різних емоційних ситуацій, включаючи схвильовані реакції батьків до і після пологів, а також спортсменів під час та після вдалих або невдалих спортивних епізодів, а також перемог та програшів.

Оскільки дослідження попередніх часів зосереджувались на відносно обмеженому діапазоні особливостей мовлення, таких як основна середня частота та діапазон, темп та інтенсивність [14], [12], існує необхідність проведення аналізу даних з використанням більшої кількості природних показників, а також ширшого спектру голосових ефектів, включаючи паралінгвістичні. Подібні дослідження не часто проводились в минулому, хоча вони виконують важливу роль у вираженні емоцій.

Вивчення паралінгвістичних особливостей у мовленні не має чітко визначених рамок, що пояснює розбіжності в літературі стосовно меж між паралінгвістичними та просодичними ознаками (див., наприклад, Brown, [3]; Ladd, [11]; Cruttenden, [4]). Однак Crystal [5] проводить розмежування з одного боку між лінгвістичними показниками, для яких характерні зміни висоти тону, гучності, тривалості та пауз, та між паралінгвістичними рисами, з іншого, оскільки, хоча вони є голосовими, вони є незалежними від висоти, гучності та тривалості. Хоча просодичні ознаки є необхідною складовою всього мовлення, паралінгвістичні ознаки можуть бути відсутніми і дозволяти більше ідіосинкразії в їх реалізації. Crystal & Quirk [6] надають найбільш детальну класифікацію просодичних та паралінгвістичних особливостей англійською мовою, і саме на їх класифікації в основному базується нинішня система паралінгвістичних анотацій.

Розробивши градієнтне розмежування просодичних та паралінгвістичних особливостей у мовленні, Crystal & Quirk [6] на просодичному кінці шкали розташовують інтонацію з її показниками, чим однозначно демонструють риси лінгвістичності, а на іншому кінці шкали - такі риси, як якість голосу, або шуми та всілякі клацання, що є паралінгвістичними ознаками [6].

Паралінгвістичні особливості Crystal & Quirk [6] поділяють на два типи: голосові якості та голосову кваліфікацію. Голосові якості зумовлені різними режимами звучання, а саме нормальний голос, фальцет, шепіт, скрип, охриплість і придихання.

Тут слід зупинитись на понятті "мовний голос". Він в цілому являє собою соціокультурне явище, тобто є показником соціальної індексальної інформації, звучання якого дає уявлення про освіту, соціальний статус, професії мовця і національних рисах його голосоутворення. Мовний голос є базовим звуковим компонентом риторичного дискурсу і сприяє реалізації дискурсивного плану оратора, який залежить від жанрової приналежності публічного виступу і від індивідуальних переваг виступаючого. Таким чином, прояв компонентів мовного голосу в тексті що звучить залежить від виду риторичних стратегій мовця. Відзначаючи специфіку прояву окремих компонентів мовного голосу при реалізації риторичних стратегій мовця, необхідно підкреслити, що в першу чергу використовуються різні голосові модуляції, а також різні відтінки тону голосу, що супроводжуються варіюванням інтонаційних параметрів. У той же час природна якість голосу і тембр мовця надають додатковий звуковий вплив на його переваги щодо вибору голосових модуляцій, які є поліфункціональними за своєю природою. Таким чином, одна і та ж голосова модифікація сприяє реалізації різних комунікативних інтенцій мовця, і навпаки одна і та ж риторична стратегія може бути реалізована кількома модуляціями.

Опис якості голосу Лавера [12] включає в себе не тільки вищезазначені режими звучання, але і всі зміни, навмисні або ненавмисні, на які здатний мовець, враховуючи фізіологічні обмеження голосових органів, з урахуванням поздовжніх артикуляційних параметрів (підвищений і опущений голос гортані, губне

випинання і лабіодентальний голос), широтних параметрів (губні, язикові, фокальні, глоткові та нижньощелепні), велофарингальних та фонаторних параметрів (включаючи комбіновані способи фонації).

Додаткові якості голосу - це нелінгвістичні голосові ефекти, що проходять через мовлення або переривають його, і включають: сміх, хихикання, трепетність, ридання та плач.

Просодичні особливості складаються з особливостей темпу, гучності, ритмічності, діапазону тону, напруженості, паузи та інтонації. Розрізняють особливості темпу, гучності та висоти між простими та складними варіативними системами, коли перша є варіацією по контрасту з нормальним мовленням, а друга описує внутрішньомовний контраст [6].

Темп це швидкість мовлення, який має два різні прояви залежно від того, чи він сприймається через багатоскладові розтяжки чи на окремих складах. На багатоскладових розтяжках «простий» темп, тобто швидка або повільна мова, поділяється на чотири позначені швидкості: allegrissimo; алегро; ленто; і lentissimo. "Складний" темп відноситься до прискорення та уповільнення швидкості мовлення, і називається в музичній термінології «прискоренням» або раленнтандо. На одному складі темп або урізаний, або розтягнутий.

Особливості гучності можуть бути простими, тихими або гучними, (піаніссімо, фортепіано, форте, або фортіссімо), або складними (crescendo) або (diminuendo).

Опис діапазону тону на окремих складах ускладнюється впливом гучності та інтонаційної системи, але на багатоскладових ділянках та на ядерному складі є або простими (низькими чи високими) або складними (монотонними, вузькими чи широкими).

Ритмічність залежить від трьох дискретних факторів: сприйнятої регулярності напружень, як ритмічних, так і аритмічних, різкості зміни тону та гучності, яка класифіковується як різка або гліссандо, і варіативності гучності без зміни тональності (стаккато або легато). Напруженість відноситься до точності артикуляційних жестів і може бути як розмитою, як у нетверезому мовленні, в'ялою, так і напруженою, або точною.

Паузи можуть бути або тихими, або озвученими і класифікуються за сприйнятим відхиленням від норми як такі, що мають чотири ступені тривалості: короткі; єдині; подвійні, або високі.

Моделі інтонації (напрямок руху тону) можуть бути простими: (низхідно-висхідний або рівний), складними (низхідно-висхідний або висхідно-низхідний) або складеними (падіння-плюс-підйом; підйом-плюс-падіння). Монотонність на ядерному складі реалізується як рівень ядерного тону. Ball та ін. [1] представляє систему транскрипції якості голосу, засновану Laver [12]. Ця система розроблена для використання як з нормальним, так і з невпорядкованим мовленням. Поділяють використання дужок і подвійних дужок для позначення ступеня голосових ефектів, а також скалярні позначення, які показують ступінь певних ефектів.

Нелінгвістичні особливості передають таку інформацію, як вік, стать, стан здоров'я та ін. Їх можна класифікувати на два види: індивідуальні варіації та рефлекси. Індивідуальна варіація включає такі ефекти фонації та резонансу через фізіологію мовця та гістологію голосового тракту. Слід брати до уваги рефлекторні стани, оскільки вони часто є мимовільною ознакою справжнього емоційного напруження. Вони створюють змінені форми дихання, ендокринної системи та обміну речовин в цілому, що також може спричинити чутні зміни у мовленні.

Дослідження виявили, що деякі емоції, такі як страх, радість і гнів, зображуються частіше, ніж такі емоції, як сум.

• **Гнів:** Гнів можна розділити на два типи: "гнів" і "гарячий гнів". У порівнянні з нейтральною мовою, гнів виробляється з нижчим тоном, більшою інтенсивністю, більшою енергією (500 Гц) протягом вокалізації, більшим першим формантом (перший виданий звук) і швидшим часом атаки при початку голосу (початок мовлення). "Гарячий гнів", навпаки, виробляється з більш високим, різноманітним тоном та ще більшою енергією (2000 Гц).

• **Огида:** У порівнянні з нейтральним мовленням, огида виробляється з меншим, спрямованим донизу тоном, з енергією (500 Гц), нижчим першим формантом і різким падінням тону, подібним до гніву. Менші варіативність та коротша тривалість є також характеристиками огиди.

• **Страх:** Страх можна розділити на два типи: «паніка» та «тривога». У порівнянні з нейтральним мовленням, емоції страху мають більшу висоту звучання, незначні варіації, меншу енергію та швидший темп мовлення з більшою кількістю пауз.

• **Смуток:** У порівнянні з нейтральним мовленням, сумні емоції виробляються з більшою висотою звуку, меншою інтенсивністю, але більшою голосовою енергією (2000 Гц), більшою тривалістю, більшою кількістю пауз і меншим першим формантом.

Розшифровка емоцій у мовленні включає три етапи: визначення акустичних особливостей, створення значущих зв’язків з цими ознаками та обробка встановлених акустичних типів. На етапі обробки зв’язки з базовими емоційними знаннями зберігаються окремо в мережі пам’яті, специфічній для асоціацій. Ці асоціації можуть бути використані для формування базової лінії емоційних проявів, що зустрінуться в майбутньому. Емоційні значення мовлення неявно та автоматично реєструються після аналізу обставин, важливості та інших оточуючих деталей події.

В цілому слід зазначити, що емоційна або афективна просодія проявляється як у лінгвістичних, так і в паралінгвістичних аспектах мовлення, які дозволяють та допомагають як передавати, так і розуміти емоції. Вона включає індивідуальний тон голосу, який передається через зміну висоти, гучності, тембру, темпу мовлення та пауз. Він може бути ізольований від семантичної інформації та взаємодіяти із словесним змістом. Емоційна просодія в мовленні сприймається або декодується дещо гірше, ніж міміка, проте точність сприйняття залежить власне від типу емоцій.

Хоча у середньому слухачі здатні сприймати передбачувані емоції із значною швидкістю, все ж рівень помилок є також високим, що частково пов’язано із спостереженням, що слухачі точніше

сприймають емоційний висновок певних голосів і сприймають одні емоції краще за інші. Вокальні вирази "гніву" і "смутку" сприймаються найлегше, "страх" і "щастя" сприймаються лише помірно, в той час як найнижчу сприйнятливість має "огида".

ЛІТЕРАТУРА

1. Ball, M.J., Esling, J. & Dickson, C. The VoQS system for the transcription of voice quality//Journal of the International Phonetic Association, 1995. Is.25. P.71—80.
2. Brown, B.L. The detection of emotion in vocal qualities//Language: Social psychological Perspectives, Oxford: Pergamon. 1980. P.237-245.
3. Brown, G. Listening to Spoken English. (2nd Edition). London & New York: Longman, 1990.
4. Cruttenden, A. Intonation. (2nd Edition). Cambridge, UK: Cambridge University Press, 1997.
5. Crystal, D. Prosodic Systems and Intonation in English. Cambridge, UK: Cambridge University Press, 1969.
6. Crystal,D., Quirk,R. Systems of Prosodic and Paralinguistic Features in English. The Hague, Netherlands: Mouton, 1964.
7. Ekman, P., Friesen, W.Y., Scherer, K.R. Body movement and voice pitch in deceptive Interaction//Semiotica, 1976. Is.16. P.23—27
8. Johnstone, T. Emotional speech elicited using computer games. Swinburne University of Technology: Research Gate, 1996.
9. Kramer, E. Elimination of verbal cues in judgments of emotion from voice//The Journal of Abnormal and Social Psychology, 1964. Is. 68(4). P.390–396.
10. Kuroda, L., Fujiwara,O., Okamura,N., Utsuki. Method for determining pilot stress through analysis of voice communication//Aviation, Space, and Environmental Medicine, 1976. Is.47. P.528—533.
11. Ladd, R.D. Intonational Phonology. Cambridge, UK: Cambridge University Press, 1996.
12. Laver, J. The Phonetic Description of Voice Quality. Cambridge, UK: Cambridge University Press, 1980.
13. Murray, I.R., Arnott, J.L. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion//Journal of the Acoustical Society of America, 1993. Is. 93. P.1097—1108.
14. Pell, M.D. The look of (un)confidence: visual markers for inferring speaker confidence in speech//Frontiers in Communication, 2019. Is.4. P.63.
15. Scherer, K.R. Speech and emotional states//Speech Evaluation in Psychiatry, New York: Grune & Stratton, 1981. P. 189—220.
16. Simonov, P.V., Frolov,M.V. Utilisation of human voice for estimation of man' s emotional stress and state of attention//Aerospace Medicine, 1973. Is. 44. P.256—258.
17. Streeter, L.A., Krauss, R.M., Geller, V., Olson, C, Apple, W. Pitch changes during attempted deception//Journal of Personality and Social Psychology, 1977. Is. 35. P.345-350.
18. Williams, C.E., Stevens, K.N. On determining the emotional state of pilots during flight: an exploratory study//Aerospace Medicine, 1969. Is. 40, P.1369—1372.

Problems of phonetic identification of emotional states of the speaker with the help of linguistic and paralinguistic means

K. V. Lysenko

Annotation. The article considers the problems of determining the emotional states of the speaker with the help of linguistic and paralinguistic means. As we speak, we convey to our listeners information about our emotional state and attitudes toward both our listeners and what we say. Although in general the specific ability to decode the emotional component of the utterance depends on the individual emotion.It is known that native speakers, when listening to actors who read an emotionally neutral text, while depicting emotions, perceptually are not always able to correctly identify them. Even if the emotional expression through prosody may not always be consciously recognized by the participants in the conversation, it may be subconsciously influenced by other indicators, such as the tone of voice, for example, or facial expressions. This type of expression follows not only from linguistic or semantic effects. The ability of an average person to decipher emotional states during conversations is less developed than the traditional ability to identify facial expressions. Whether a person perceives prosody as positive, negative or neutral, he always coordinates with facial expressions. If it becomes closer to neutral, the interpretation of the emotional state is influenced by prosodic interpretation. To decide on the level of emotionality, listeners are often forced to listen to the prosody of the statement. To be able to determine the level of affectivity of the statement, listeners need at least 600 milliseconds of prosodic information, below which there is not enough time to process the emotional context of the statement. As previous research has focused on such prosodic features of speech as basic average frequency and range, tempo and intensity, it is now necessary to analyze data using a wider range of voice effects, including paralinguistic. Such studies have not been conducted in the past, although they play an important role in the depiction of emotions. Having developed a scale with a gradient distinction between prosodic and paralinguistic features in speech, researchers placed intonation with its indicators, which indicates its linguistic character, at the prosodic end of the scale, and at the other end of the scale - such features as voice quality, or noise and all sorts of clicks as paralinguistic features. Nonlinguistic or paralinguistic features are classified into two types: individual variations and reflexes. Individual variation includes the effects of phonation and resonance through the physiology of the speaker and the features of his vocal tract. Reflex states should also be considered, as they are often an involuntary sign of true emotional stress. In general, it should be noted that emotional or affective prosody is manifested in both linguistic and paralinguistic aspects of speech, which allow and help both to convey and understand emotions. It includes an individual tone of voice, which is transmitted through changes in pitch, volume, timbre, speech rate and pauses.

Keywords: *prosody, emotional reaction, linguistic and paralinguistic features, segmental and suprasegmental functions and features, voice modification.*