

КОНТРОЛЬ ЕМОЦІЙНОГО СТАНУ ОПЕРАТОРА В СИСТЕМАХ МОВНОГО ВВЕДЕННЯ ІНФОРМАЦІЇ

MONITORING THE EMOTIONAL STATE OF THE OPERATOR IN LANGUAGE DATA ENTRY SYSTEMS

У статті представлені метод і пристрій, що дають можливість компенсувати відхилення зміщення складників мовного спектру. Суть такої компенсації полягає в тому, що у процесі мовного введення інформації оператор має можливість візуально контролювати власний стан. При цьому з'являється можливість корекції спектру самим оператором у реальному часі шляхом самоконтролю свого стану та подальшого апаратного усунення відхилення спектру від «еталону». Такий «еталон» представляє заздалегідь наговорений ряд слів (алфавіт), вживаних у процесі діяльності оператора. Візуалізація може бути організована у вигляді світлодіодної матриці, де кожній із трьох градацій функціонального стану – «велике відхилення», «середнє відхилення», «норма» – відповідатиме певний колір матриці. Точність розпізнавання є основним показником надійності таких систем. Точність розпізнавання знижується в разі збільшення кількості слів, використовуваних у діяльності оператора. Окрім цього, зниження точності розпізнавання пов'язане зі зміщенням спектру мовного сигналу діючого оператора. Таке зміщення спектру пов'язане зі зміною психологічного стану оператора. У кількісному вираженні точність розпізнавання знижується за кількості слів у словнику, що дорівнює десяти, до шести відсотків. Достовірність розпізнавання може становити дев'яносто відсотків за довжини словника, що дорівнює шести словам, за умови, що не буде «підгонки» словника під певного оператора. Якщо така «підгонка» буде здійснена, кількість слів, розпізнаваних системою, може збільшитися до п'ятисот. Пропонований спосіб і пристрій для його реалізації є автономним доповненням систем мовного введення інформації. Пристрій може застосовуватися як у процесі діяльності операторів, так і в лабораторних умовах для дослідження психологічного функціонального стану операторів. Крім того, можливе його застосування в системах контролю доступу. Пропонований спосіб підвищення надійності систем введення (розпізнавання) інформації заснований на психологічній саморефлексії, людини до саморегулювання.

Ключові слова: зміщення складників мовного спектру, компенсація, рефлексія, роз-

пізнавання мови, візуалізація, штучний інтелект.

The article presents a method and device that allow compensating for the deviation of the displacement of the components of the speech spectrum. The essence of such compensation is that in the process of speech input, the operator has the opportunity to visually monitor his own condition. At the same time, it becomes possible to correct the spectrum by the operator himself in real time through self-monitoring of his condition, and subsequent hardware elimination of the deviation of the spectrum from the "standard". Such a "standard" represents a pre-spoken series (alphabet) of words used in the process of the operator's activity. Visualization can be organized in the form of an LED matrix, where each of the three gradations of the functional state – "high deviation", "average deviation", "normal" – will correspond to a certain color of the matrix. Recognition accuracy is the main indicator of the reliability of such systems. Recognition accuracy decreases with an increase in the number of words used in the operator's activity. In addition, the decrease in recognition accuracy is associated with the shift in the spectrum of the speech signal of the active operator. Such a shift in the spectrum is associated with a change in the psychological state of the operator. In quantitative terms, the accuracy of recognition decreases when the number of words in the dictionary is equal to ten, to six percent. The recognition accuracy can be ninety percent with a dictionary length of six words, provided that the dictionary is not adapted to a specific operator. If such adjustment is carried out, the number of words recognized by the system can increase to five hundred. The proposed method and device for its implementation is an autonomous addition to speech input systems. The device can be used both during the activity of operators and in laboratory conditions to study the psychological functional state of operators. In addition, it can be used in access control systems. The proposed method of increasing the reliability of information input (recognition) systems is based on psychological self-reflection, a person's ability to self-regulate.

Key words: displacement of components of the speech spectrum, compensation, reflection, speech recognition, visualization, artificial intelligence.

УДК 159.9:62

DOI <https://doi.org/10.32782/2663-5208.2022.42.42>

Тиньков О.М.

к.психол.н.,
доцент кафедри психології
Національний аерокосмічний
університет імені М.Є. Жуковського
«Харківський авіаційний інститут»

Долгополова О.В.

к.психол.н.,
доцент кафедри психології
Національний аерокосмічний
університет імені М.Є. Жуковського
«Харківський авіаційний інститут»

Фаворова К.М.

к.психол.н.,
доцент кафедри психології
Національний аерокосмічний
університет імені М.Є. Жуковського
«Харківський авіаційний інститут»

Постановка проблеми. Існує проблема здійснення незалежного від диктора (дикторів) розпізнавання мови. Складність такого розпізнавання пов'язана не лише з відмінностями в голосах дикторів, але і з емоційним станом того, хто говорить.

Проблема незалежного від диктора розпізнавання мови може бути частково розв'язана шляхом створення встановленого ряду слів, які приймають за еталони. Якщо йдеться про

декількох дикторів, то може бути застосований банк фраз, що відбиває особливості частотних складових мовних сигналів кожного окремо диктора. Проте такого рішення недостатньо, потрібний облік емоційного стану оператора.

Діяльність оператора проходить в певному доквіллі, такому, як фізичне і соціальне. Окрім цього є ще і внутрішнє психологічне середовище. Цей комплекс "оточення" можна представити у вигляді багатofакторної моделі.

Чинниками зовнішнього фізичного середовища є: атмосферний тиск, температурні дані довкілля, хімічний склад повітря, швидкість вітру, фази місяця, магнітні бурі. Чинниками зовнішнього соціального середовища є: згуртованість і сумісність колективу, конфліктність, психологічне здоров'я колективу, комунікабельність, корпоративна культура. Окрім цього на функціональний стан оператора впливає: інтенсивність і тривалість діяльності, виникаючі ситуації (аварії, збої в роботі техніки). Внутрішніми чинниками можуть бути: мотивація до праці, характер, темперамент, психічні процеси (увага, пам'ять, мислення, воля). Під впливом цих чинників відбувається зміна функціонального стану оператора, як наслідок, знижується надійність систем мовного введення інформації.

Аналіз останніх досліджень та публікацій. Технічні системи для розпізнавання мови застосовуються в основному для забезпечення діалогового режиму в системах з штучним інтелектом. Наприклад, формування текстових файлів; інформаційно-довідкові системи; мовний набір номерів, кодів в системах доступу; системи управління диспетчерського типу; системи голосових повідомлень; системи навчання і інші.

Точність розпізнавання є основним показником надійності таких систем. Точність розпізнавання знижується при збільшенні кількості слів, використовуваних в діяльності оператора. Окрім цього, зниження точності розпізнавання пов'язана зі зміщенням спектру мовного сигналу діючого оператора. Так, у своїх дослідженнях K.R. Scherer [10] відмічав той факт, що у операторів, що знаходяться в стані стресу спостерігалось розширення діапазону варіювання частоти основного тону, як у бік пониження, так і у бік підвищення.

При виникненні (в процесі діяльності) у операторів будь-якого з таких станів, як ейфорія, розгубленість, тривога, страх, стомлення, емоційній напруженості, нудьга, апатія, монотонія – цей показник істотно змінюється. Така зміна може бути спрямована, як на підвищення, так і на пониження частоти.

У кількісному вираженні, точність розпізнавання знижується при кількості слів в словнику рівним десяти, до шести відсотків [2]. Достовірність розпізнавання може скласти дев'яносто відсотків при довжині словника рівному шістдесят слів, за умови, що не буде підгонки словника під певного оператора. Якщо така підгонка буде здійснена, кількість слів, розпізнаваних системою може збільшитися до п'ятисот [3]. Нині накопичений величезний арсенал способів і технічних засобів розпізнавання мови людини. Зупинимося на деяких, на наш погляд найбільш перспективних підходів.

Так, був запропонований спосіб для пригнічення перешкод, що виникають в каналі мовного сигналу, заснований на аналізі в модуляційній області. Цей спосіб дозволяє значно понизити потужність перешкоди в порівнянні з широко вживаним для цих цілей алгоритму на базі RASTA (RelAtive SpecTrA).

Велика частина досліджень, присвячених розпізнаванню диктора використовують розрахунок коефіцієнта кепстра, який визначається по огибаючій спектру. Такий розрахунок припускає швидке перетворення Фур'є і реалізується шляхом застосування смугових фільтрів. Такий підхід забезпечує простоту обчислень і враховує індивідуальні характеристики голосу диктора. Окрім викладеного, набирають популярність методи сумішей (GMM) гауса, і опорних векторів (SVM). Також застосовуються способи розпізнавання із застосуванням штучних нейронних мереж і прихованих Марківських ланцюгів. Тембром мови людини, пов'язаний з його емоційним станом. У дослідженнях [1], показано, що люди здатні розрізняти емоції інших по зміні тембру мови. Було встановлено, що в порівнянні із спектром, виміряним у людини в стані спокою, спектр емоційної мови значно змінюється, форманти зрушуються по частоті, вони звужуються або розширюються, спостерігається зрушення енергії між областями високими і низькими частот [1].

На сприйняття звуків найбільший вплив має кількість формант і їх розташування в частотному спектрі. Формантами є частотні області з максимальною амплітудою, які можна спостерігати в спектрі мови. Розташування максимумів спектру відповідає резонансам мовного апарату людини. Анатомічно, на зміну спектру мовного сигналу людини, впливає положення (робота) м'язів глотки і гортані [1].

Автори ряду робіт [2] пропонують розділити увесь спектру мовного сигналу людини на області, зрушення в яких несуть інформацію про характеристики емоційного стану. Такими областями є: 250-980 Гц (діапазон частот основного тону і першої форманти), 980-2180 Гц (діапазон частот другої форманти), 2180-3880 Гц (діапазон частот третьої форманти). При дослідженні спектру розмови використовуються: значення частот перших трьох формант, спектральні моменти, ширина спектру.

Найбільш близьким до запропонованого є метод розпізнавання голосів, що визначає положення центру частотного спектру мовного сигналу людини [3]. Технічною реалізацією такого методу стала система для періодичного визначення середньої частоти і амплітуди мовного сигналу. Додатково система здатна визначати положення центру спектру мовного сигналу. При цьому в оперативну пам'ять

системи вводяться дані про локалізацію центру спектру мовного сигналу. Якщо на вході системи з'явиться сигнал, що належить іншій людині, або сигнал належить цій же людині, але зі зміненим емоційним станом, то дані про локалізацію центру спектру мовного сигналу поступає на обчислювальний пристрій, що визначає різницю локалізацій. Різницевий сигнал виходить внаслідок порівняння частотних центрів. Сигнал поступає на вхід системи порівнюється з сигналом, що зберігається в пам'яті системи. Якщо різниця між цими сигналами (центрами спектру) менше встановленої величини, робиться висновок про те, що обидва центри спектру належать одній і тій же людині.

Також був розроблений метод розпізнавання голосу диктора [1], заснований на порівнянні послідовності еталонів мовного сигналу на певному відрізку часу. Такими еталонами можуть бути виміри ряду характеристик спектру сигналу по кожній фонемі.

Мета статті – запропонувати метод і пристрій, що дозволяють компенсувати відхилення зміщення складових мовного спектру.

Виклад основного матеріалу. У статті представлений метод і пристрій, що дозволяють компенсувати відхилення зміщення складових мовного спектру. Суть такої компенсації полягає в тому, що в процесі мовного введення інформації оператор має можливість візуально контролювати власний стан. При цьому, з'являється можливість корекції спектру самим операторів в реальному масштабі часу шляхом самоконтролю свого стану, і подальшого апаратного усунення відхилення спектру від «еталону». Такий «еталон» представляє заздалегідь наговорений ряд (алфавіт) слів, вживаних в процесі діяльності оператора. Візуалізація може бути організована у вигляді світлодіодної матриці, де кожній з трьох градацій функціонального стану – «велике відхилення», «середнє відхилення», «норма», відповідатиме певний колір матриці.

Реалізація пропонованого методу полягає в тому, що відбувається реєстрація періоду фрагмента мовного сигналу. Період реєстрації дорівнює 2 с. (за часом). Впродовж цього періоду робиться вимір 100 сегментів сигналу. Далі обчислюється середня частота першої форманти і дисперсії основного тону мовного сигналу. Потім обчислюється різниця між частотою основного тону, яка зберігається в пам'яті пристрою і її поточного значення. При цьому дані про частоту, зберігаються в пам'яті й оцінюються у оператора, що знаходився в стані функціонального спокою (так званий – «фоновий» стан). Під поточним, розуміється значення частоти основного тону, отриманої в процесі діяльності оператора. Надалі, обчислюється величина розбіжності

частоти основного тону за критерієм t – Стьюдента.

Якщо виявиться, що різниця істотна, то з урахуванням достовірності статистичного критерію, робиться висновок про те, що оператор знаходиться в стані, що відрізняється від "фонового".

Прототипом розробленого методу, являється метод описаний в [3]. Автор цього методу пропонує на слуховий аналізатор людини (через телефон) подавати синхронний запис його голосу, який перевищує природну силу голосу на 40 dB із затримкою відтворення близько 0,15-0,3 с. Як наслідок – у людини відзначається стан психічної напруженості. Такий "зворотний акустичний зв'язок" дозволяє змінювати функціональний стан людини. Проте зміст такого стану носить негативний характер. Стан психічної напруженості негативно впливає на ефективність діяльності оператора. У нього неминуче з'являться помилки і збільшиться час рішення професійних завдань. Говорячи інакше, знизиться надійність діяльності. На відміну від розробленого, в пропонованому методі такий зв'язок має візуальний характер, іншу форму здійснення. Пропонований спосіб дозволяє операторові самостійно (за допомогою пропонованого пристрою) перейти з негативного поточного стану, в позитивний.

Пропонований пристрій з можливістю візуального контролю власного стану містить: мікрофон, підсилювач низької частоти, компресор, шість смугових фільтрів, блок управління, шість аналого-цифрових перетворювачів (АЦП), блок обчислювача з оперативним пристроєм (ОЗУ), що запам'ятовує, блок порівняння, підсилювач різницевої частоти, блок порогів і блок індикації (візуалізація). Блок візуалізації служить для графічного відображення різниці (рівня) між функціональним станом «спокою» і «поточним» станом діючого оператора. Формула роботи пристрою може бути представлена таким чином. Пристрій мовного введення інформації, трансформує «поточний» функціональний стан оператора в заданий стан – функціонального спокою. Така трансформація (управління) відбувається за рахунок внутрішнього психологічного ресурсу людини, його здібності до саморегуляції. У цьому випадку саморегуляція робиться таким чином, що той, хто говорить, орієнтується на колірні зміни у блоці індикації (системі відображення інформації про стан людини). При цьому метою того, хто говорить є не лише мовна інформація, необхідна для вирішення професійного завдання, але і самостійна перебудова голосу. Індикатором норми такої перебудови в цьому випадку є колір світіння світлодіодних індикаторів блоку індикації.

Блок-схема пропонованого пристрою наведено на рис. 1.

Блок 1 перетворює акустичні сигнали мови оператора в електричні імпульси.

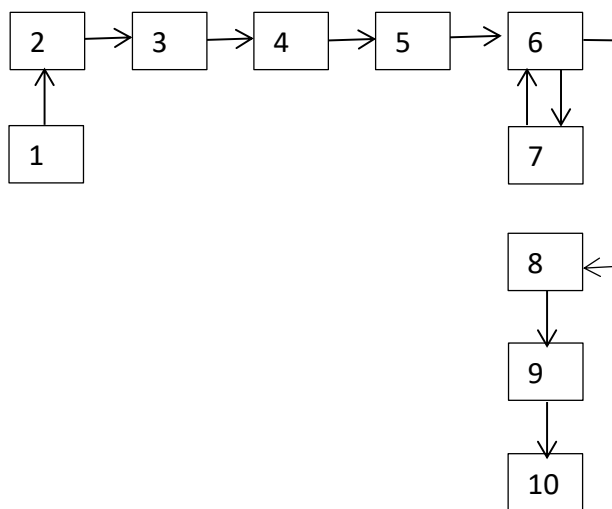


Рис. 1. Блок-схема пристрою

1. Мікрофон.
2. Компресор.
3. Підсилювач низької частоти.
4. Блок смугових фільтрів.
5. Блок аналого-цифрових перетворювачів (АЦП).
6. Блок управління.
7. Блок пам'яті (ОЗУ).
8. Обчислювач різницевої частоти.
9. Блок порогів.
10. Блок індикації (візуалізація).

Пристрій працює таким чином.

Компресор (блок 2) має смугу пропускання від 250 Гц до 4500 Гц. Його застосування обумовлене необхідністю стабілізувати амплітудні зміни сигналу, що поступають на його вхід.

Підсилювач низької частоти (блок 3) здійснює посилення сигналу до величини, достатньої для нормальної роботи АЦП (блок 5).

Блок 4 є смуговими пасивними фільтрами типу RC, з центральними частотами пропускання відповідно до 300 Гц, 700 Гц, 1200 Гц, 1800 Гц, 2300 Гц, 3000 Гц, 3800 Гц,

Блок управління 6 забезпечує:

- синхронізацію роботи АЦП з наявністю (відсутністю) мовного сигналу на вході пристрою (з мікрофону);
- розбиття мовного сигналу на сегменти фіксованої довжини – 2 с. Період реалізації мовного фрагмента у свою чергу розбивається на сегменти фіксованої довжини (20 мс.);
- вироблення сигналу зупинки рахунку сегментів після підрахунку 100 сегментів;
- запис інформації (у цифровому коді) про мовний сигнал в ОЗУ;
- дозвіл обчислень у блоці 8.

Блок 5 – АЦП, перетворює мовний сигнал в цифровий код, відповідно у визначеній смуговими фільтрами смуги частот.

Блок 8 – обчислювач, робить контроль за зрушеннями спектру сигналу. Обчислювач «порівнює» інформацію, що зберігається в ОЗУ.

Блок індикації 10 служить для відображення інформації про зміщення частотного спектру мовного сигналу. Величина такого зміщення пов'язана з певними кольорами світіння екрану, за яким спостерігає людина. Різниця в кольорі екрану пов'язана із заданими порогоми, встановленими у блоці Блок 9. Цей блок містить порогові значення характеристик спектру мовного сигналу для кожного з шести фільтрів і АЦП.

Необхідно відмітити, що пропонований спосіб і пристрій для його реалізації є автономним доповненням систем мовного введення інформації. Пристрій може застосовуватися як в процесі діяльності операторів, так і в лабораторних умовах, для дослідження психологічного функціонального стану операторів. Крім того, можливе його застосування в системах контролю доступу. Проте існують обмеження з практичного застосування у разі безпосереднього виконання професійних дій оператором. Чим менше ліміти часу на ухвалення рішень у операторів, тим складніше буде проходити процес «самонастрою» оператора на «заданий» (еталонне, стан спокою) функціональний стан. Йому може не вистачити час. Проте з іншого боку – спроба введення мовної команди оператором, що знаходиться в негативному стані (стомлення, емоційна напруженість) неминуче призведе до збільшення помилок введення. Таким чином, з одного боку – зниження швидкодії, а з іншої – ріст кількості помилок. І це можна розглядати як взаємовиключні чинники. Надійність розпізнавання (введення) буде однаковою. Практичний висновок із сказаного – додаткові тимчасові витрати при введенні голосової команди не зроблять істотного зниження надійності систем розпізнавання. У разі застосування пропонованого пристрою в системах контролю доступу, коли час введення мовної команди не має жорсткого обмеження, надійність систем розпізнавання буде високою. Необхідно відмітити, що пропонований спосіб і пристрій для його реалізації є автономним доповненням систем розпізнавання мовного сигналу.

Висновки з проведеного дослідження.

За допомогою пропонованого способу можна контролювати психологічний стан оператора по його мовному сигналу.

Розроблений пристрій забезпечує безперервний, дистанційний контроль психологічного стану операторів і є автономним доповненням систем мовного введення інформації.

Візуалізація функціонального стану операторів може бути організована у вигляді світ-

лодіодної матриці, де кожній з трьох градацій функціонального стану – «велике відхилення», «середнє відхилення», «норма», відповідатиме певний колір світлодіодної матриці.

Пропонований спосіб підвищення надійності систем введення (розпізнавання) інформації заснований на психологічній саморефлексії, здібності людини до саморегулювання.

ЛІТЕРАТУРА:

1. Галунов В.И., Пиктурна В.В., Янушавичюс В.Й. Акустические корреляты эмоциональной речи. Матер. докл. и сообщ. 5-го Всес. совещ. – симп. цикла «Акуст. речи и слуха»: Эмоции и автомат. распознавание речи. Одесса. 1989. С. 16-25.
2. Гомина Т.Г. Влияние различных эмоциональных состояний на изменение спектра английских гласных. Матер. и сообщ. Всес. симпозиума «Речь, эмоции и личность». Л.: Наука, 1978. С. 94-96.
3. Кучерявый А.А. Бортовые информационные системы: Курс лекций / под ред. В.А. Мишина и Г.И. Клюева. Ульяновск; УлГТУ, 2004. 504 с.
4. 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.
5. Леонова, А. Б. Психодиагностика функциональных состояний человека. Москва: МГУ, 1984. 200 с.
6. Носенко, Э.Л. Изменения характеристик речи при эмоциональной напряженности. *Вопросы психологии*. 1978. № 6. С. 76-85.
7. Сидоров К.В., Филатова Н.Н. Анализ признаков эмоционально-окрашенной речи. *Вестник ТвГТУ*. 2012. № 20. С. 26 – 32.
8. Смирнов, Б. А., Тиньков А.М. Методы инженерной психологии. Харьков: Гуманит. центр, 2008. 528 с.
9. Фролов М.В., Милованова Г.Б. Особенности контроля состояния человека-оператора по показателям основного тона и спектра его речи. *Физиология человека*. 2009. Т. 35. № 2. С. 136 – 138.
10. Scherer K.R. Emotion effects on voice and speech: Paradigms and approaches to evaluation. ISCA Workshop on Speech and Emotion, Belfast 2000. <http://www.qub.ac.uk/en/isca/proceedings/pdfs/scherer.pdf>