

УДК 004.8+004.5+004.65

Двоступенева система аналізу голосового сигналу для задачі контролю стану водія під час керування автомобілем

Мєшков О.Ю., Новіков О.О.

UDC 004.8+004.5+004.65

A two-step system of voice signal analysis for the task of driver condition monitoring while driving a vehicle

Mieshkov O.Yu., Novikov O.O.

Запропоновано двоступеневу біотехнічну систему аналізу голосу водія. На першому етапі система проводить голосову ідентифікацію водія. За результатами ідентифікації система приймає рішення про допуск людини до керування автомобілем. На другому етапі система періодично аналізує стан водія під час руху. Аналіз проводиться за аналогічним алгоритмом на основі аналізу голосового сигналу людини. У випадку певних відхилень у стані голосу система на першому етапі може не допустити водія до керування автомобілем, а у другому – рекомендувати припинити рух задля уникнення можливих небезпек чи дорожньо-транспортних пригод.

Ключові слова: голосовий сигнал, аналіз сигналів, контроль стану людини.

A two-step biotechnical system for driver voice analysis is proposed. At the first step the system performs driver voice identification. According to the results of identification the systems decides the admission of the person to drive. At the second step the system periodically analyze the condition of the driver while driving. This analysis is performed by the same algorithm on the basis of human voice signal. In the case of some deviations in voice signal the developed system at the first stage can prevent the driver to operate the vehicle, and at the second one – recommend to stop driving in order to avoid possible dangers or accidents.

Key words: voice signal, signal analysis, human condition monitoring.

Вступ. Однією із важливих задач на сьогоднішній день є контроль стану водія під час виконання безпосередніх обов'язків – керування автомобілем. На сьогоднішній день багатьма автомобільними компаніями створено ряд технічних засобів, що контролюють різні аспекти поведінки водія за кермом. Більшість із них стосуються попередження водія від засинання і орієнтовані на моніторинг положення очей чи повік водія відносно дороги [1]. У той же час дослідниками розроблено новий тип автомобільних крісел із вбудованими датчиками, які відслідковують частоту серцевого ритму у водія [2]. В обох випадках система надсилає попередження у вигляді різного роду сигналів, а у випадку ігнорування – втручається у керування автомобілем.

У той же час відомо, що стан водія можна відслідковувати за рахунок аналізу голосового сигналу людини. На голосову функцію людини впливає практично кожен орган чи система людського організму. При певних змінах у стані людини вони автоматично будуть відображатися на стані її голосу.

З іншого боку, голосовий сигнал людини є індивідуальним, що також можна використати з метою ідентифікації водія перед початком руху. Це дозволить значно зменшити ймовірність викрадення автомобіля – система просто не дозволить завести двигун автомобіля у випадку, якщо це намагається зробити не власник автомобіля (або один із власників). У той же час така система робитиме попередній аналіз стану водія і рекомендуватиме йому, чи варто сідати за кермо у його поточному стані.

Мета та завдання дослідження. Метою роботи є розробка системи, яка підвищить безпеку дорожнього руху завдяки ідентифікації водія перед початком руху та контролю стану водія безпосередньо під час руху за аналізом голосового сигналу. Аналіз голосу водія у режимі реального часу дозволить виявляти зміни у його стані та сигналізувати про можливі небезпеки.

Для досягнення поставленої мети необхідно виконати наступні завдання:

- розглянути існуючі системи контролю водія під час керування автомобілем;
- визначити можливості ідентифікації людини за голосовим сигналом;

- розробити систему аналізу стану людини за голосовим сигналом в режимі реального часу;
- визначити критерії аналізу стану водія під час руху за голосовим сигналом.

Матеріали та методи дослідження. Для задачі ідентифікації та аналізу голосового сигналу цей сигнал необхідно подати у вигляді малої кількості інформаційно-значущих параметрів, тобто параметризувати. Авторами пропонується використовувати з даною метою такі критерії як основна частота голосу та структура розподілу амплітуди сигналу у часовому просторі. З метою ідентифікації особистості пропонується виділяти з голосового сигналу фонему «А» та проводити аналіз її характеристик. Вибір саме цієї фонему пов'язаний з тим, що цей звук є найбільш інформативним – на основі аналізу цього звуку на сьогоднішній день проводять аналіз багатьох захворювань легенів, серцево-судинної, нервової системи тощо [3].

Під час досліджень було виявлено, що кожна фонема в одиничній вимові складається з набору квазіперіодичних коливань, які в літературі називають фреймами [4]. При цьому на початку та в кінці сигналу спостерігаються перехідні процеси, відомі як атака звуку та гасіння звуку (див. рис. 1).

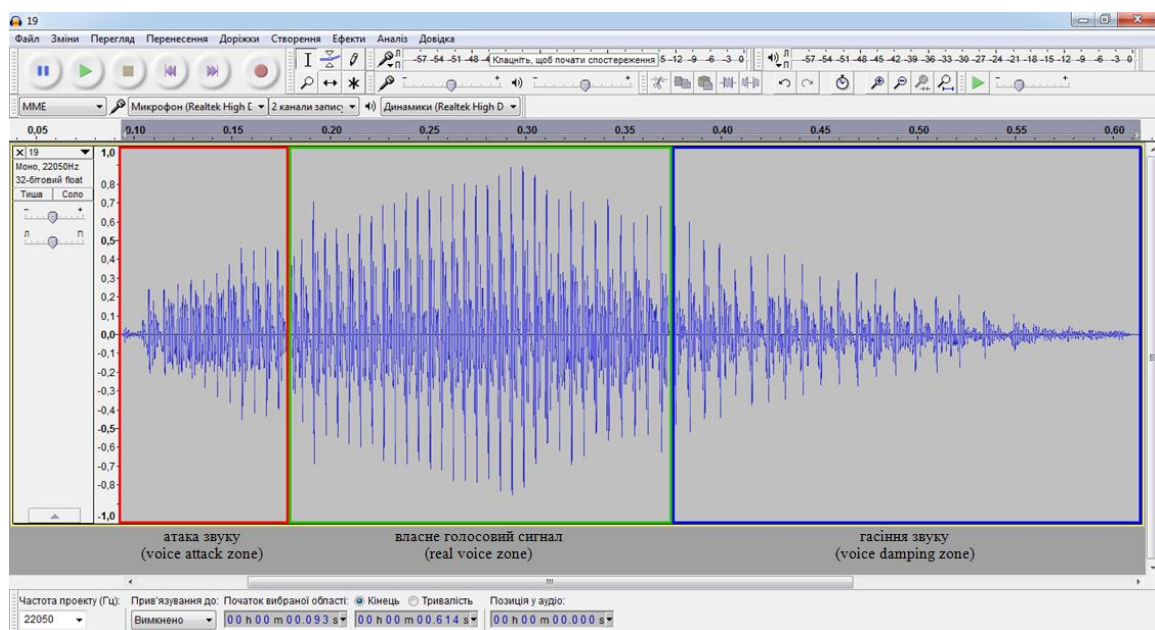


Рис. 1. Приклад розподілу амплітуди сигналу у часовому просторі (звук А, частота дискретизації – 22050 Гц)

Усі фрейми сигналу не можна використовувати з метою аналізу та для задачі ідентифікації людини. Тому авторами було розроблено алгоритм виділення необхідної кількості фреймів з потоку голосового сигналу. У результаті роботи даного алгоритму отримується сигнал, який буде складатися виключно з необхідних для аналізу фреймів. Для задачі параметризації даного сигналу пропонується визначити такі характеристики як основна частота голосу та розподіл амплітуди сигналу у часовому просторі. Визначити їх можна як для усього набору фреймів цілком, так і для кожного окремого фрейму. Останній варіант є найбільш актуальним для визначених характеристик, оскільки вони обидві змінюються у часі навіть у межах одиничного сигналу, від фрейму до фрейму. Тому у якості ще одного показника параметризації пропонується використовувати динаміку основної частоти голосу у часі та структуру розподілу амплітуди сигналу у часовому просторі.

Для визначення указаних характеристик необхідно розділити аналізований сигнал на окремі фрейми. З цією метою авторами було розроблено алгоритм пошуку точки розмежування фреймів. Приблизне положення цієї точки визначалось на основі значення середньої тривалості фрейму, а уточнювалося шляхом пошуку точки переходу сигналу з від'ємної у додатну область. У результаті роботи цього алгоритму отримується так звана хмара одиничних фреймів. Кожен фрейм має свою тривалість, а, відповідно, і частоту, а також структуру розподілу амплітуди у часовому просторі. Загалом, на наш погляд, кожен окремо взятий фрейм отриманої хмари може використовуватись для подальшого аналізу.

Оскільки кожен із отриманих фреймів має різну тривалість, то це може ускладнювати процес подальшого аналізу та, зокрема, ідентифікації людини. Тому на останньому етапі попередньої обробки сигналу кожен фрейм масштабується до чітко визначеної тривалості. У результаті цієї процедури структура сигналу зберігається незмінною. У результаті частота первинного фрейму та розподіл амплітуди масштабованого фрейму заносяться до спеціальної бази даних.

На думку авторів, для задачі ідентифікації диктора достатньо порівняти вхідний акустичний сигнал з сигналами бази даних за вказаними характеристиками. З цією метою вхідний сигнал обробляється за алгоритмом, аналогічним алгоритму обробки базових сигналів, а отримана хмара фреймів усереднюється. Після цього будується двовимірний простір на базі виділених характеристик. Якщо для порівняння основних частот двох сигналів достатньо їх безпосередніх значень, то для порівняння розподілу амплітуди фрейму у часовому просторі необхідно обрати кількісний критерій. У якості цього критерію пропонується використовувати коефіцієнт середньоквадратичного відхилення структури сигналів, який розраховується за формулою:

$$K = \frac{\sqrt{\frac{1}{T} \sum_{i=1}^n (Y_{i,вх} - Y_{i,баз})^2 \Delta t}}{\sqrt{\frac{1}{T} \sum_{i=1}^n Y_{i,баз}^2 \Delta t}},$$

де $Y_{i,вх}$, $Y_{i,баз}$ - амплітуди і-го відліку вхідного та базового сигналів;

T – період (тривалість) фрейму;

Δt – інтервал часу між відліками [5].

У результаті будь-який вхідний сигнал, розбитий на фрейми, у запропонованому просторі ознак можна подати у вигляді точки, координатами якої будуть:

а) різниця основних частот фрейму вхідного та базового сигналів;

б) коефіцієнт середньоквадратичного відхилення вхідного сигналу від базового.

Оскільки, на думку авторів, обидві ці характеристики повинні враховуватись рівноправно, то обидві вони нормуються перед формуванням простору характеристик. З урахуванням того, що у базі міститься велика кількість дикторів і, відповідно, голосових фреймів, то порівняння усередненого фрейму вхідного сигналу утворить у даному просторі хмару точок (див. рис. 2).

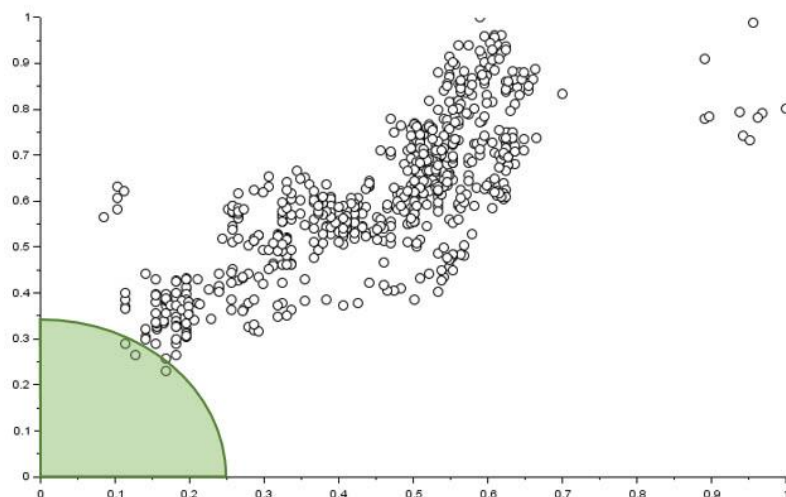


Рис. 2. Порівняння сигналів у двомірному просторі ознак «частота-структура»

Після цього визначається відстань від усередненого фрейму вхідного сигналу до кожного сигналу бази. Відстань між сигналами визначається як:

$$d_{\text{вх,баз}} = \sqrt{\omega_F (F_{0\text{вх}} - F_{0\text{баз}})^2 + \omega_K K_{\text{вх,баз}}^2},$$

де $d_{\text{вх,баз}}$ – відстань між вхідним та базовим сигналом (фреймом);

$F_{0\text{вх}}, F_{0\text{баз}}$ – нормовані значення частоти основного тону вхідного та базового сигналу;

$K_{\text{вх,баз}}$ – коефіцієнт середньоквадратичного відхилення вхідного сигналу від базового;

ω_F, ω_K – вагові коефіцієнти відповідних характеристик.

Сигнал з бази даних, відстань до якого буде найменшою, визначається як такий, що належить даному диктору. Отже вхідний сигнал визначатиме вхідного диктора як одного з дикторів бази даних. Відповідно, обов'язковою умовою для ідентифікації диктора є наявність декількох акустичних записів його голосу у базі даних.

Експериментальні дані та їх обробка. У ході експериментального дослідження для навчальної бази дикторів (75 жіночих, 75 чоловічих голосів) було визначено значення вагових коефіцієнтів характеристик ($\omega_F = 0,67, \omega_K = 0,33$), та граничного значення відстані ($d_{\text{max}} = 0,2$) оптимальних для

ідентифікації дикторів. При цьому досить поширеною була ситуація, коли в область допустимої ідентифікації потрапляли голосові сигнали декількох дикторів. У такій ситуації система приймала рішення, ідентифікуючи вхідного диктора як того, відсоток базових фреймів якого у даній області був найвищим.

Робота системи також була перевірена на тестовій вибірці дикторів (30 жіночих, 30 чоловічих голосів) і показала абсолютну точність ідентифікації дикторів у короткостроковому періоді дослідження (1 тиждень). Однак дослідження на довгостроковому періоді (2-3 тижні) показали, що для більшості дикторів сигнал бази перестає бути актуальним, отже його треба оновлювати.

На даному етапі досліджень автори використовували розроблену систему переважно з метою ідентифікації. Для задачі аналізу голосового сигналу пропонується використовувати вже описані параметри – основну частоту голосу та розподіл амплітуди у часовому просторі. При цьому їх можна доповнити такими допоміжними параметрами як загальна тривалість звуку та динаміка зміни основної частоти протягом вимови звуку. У роботі [5] авторами також розроблено ряд алгоритмів побудови еталонних голосових сигналів людини, з якими необхідно порівнювати голосові сигнали людини. Для аналізу стану людини за голосовим сигналом необхідно розробити алгоритм порівняння цих сигналів на основі виділених параметрів та визначити граничні норми відхилень вхідних параметрів від еталонних значень. Дослідження у даному напрямку проводяться авторами на даний момент. Також розглядається можливість апаратної та програмної реалізації даної системи для контролю стану водіїв перед початком керування автомобілем (етап ідентифікації) та безпосередньо під час руху (етап аналізу сигналу).

Висновки. Авторами розроблено ряд алгоритмів обробки первинного акустичного матеріалу, які дозволяють виділити із голосового сигналу характеристики, необхідні для аналізу голосового сигналу водія під час керування автомобілем. На основі цих параметрів будується двомірний простір характеристик, в якому проводиться процедура ідентифікації за методом

найближчого сусіда. З використанням аналогічних просторів характеристик є можливість формування так званих еталонних голосових сигналів людини на основі реальних голосових сигналів. Авторами також запропоновано дві моделі формування таких еталонних сигналів та доведено доцільність їх використання для поставленої задачі.

Література.

1. *Кравченко А.П.* Система контролю и предупреждения водителя от засыпания / Кравченко А.П., Локотош Б.Н., Морозов И.В. // Вісник Донецької академії автомобільного транспорту №4 2010 с. 22-27.

2. *Создана система контролю сна за рулем на основе анализа сердечного ритма* // Електронний ресурс // Режим доступу до ресурсу: <http://hi-news.ru/technology/sozdana-sistema-kontrolya-sna-za-rulem-na-osnove-analiza-serdechnogo-ritma.html>

3. *Чорна Л.Б.* Стохастична модель голосового сигналу для задачі діагностики ритміки серця людини: автореф. дис. канд. техн. наук: 01.05.02 / Л.Б. Чорна; Терноп. держ. техн. ун-т ім. І.Пулюя. – Т., 1999. – 19 с.

4. *Mieshkov O.Y.* Development of Universal Program Complex for Human Condition Analysis, Based on the Analysis of Human Voice / O.Y.Mieshkov, O.O.Novikov // Theoretical and Applied Aspects of Cybernetics. – Proceedings of the 4th International Scientific Conference of Students and Young Scientists – Kyiv: Bukrek, 2014. – P. 294-305.

5. *Mieshkov O.Y., Novikov O.O.* Automated system for identification and human condition diagnostics based on its voice signal analysis // System analysis and information technology: 18-th International conference SAIT 2016, Kyiv, Ukraine, May 30 – June 2, 2016. Proceedings. – ESC “IASA” NTUU “KPI”, 2016. – P.35–38.

Відомості про авторів:

Мешков Олександр Юрійович – аспірант 1-го року навчання кафедри Інформаційно-вимірювальних технологій, електроніки та інженерії Херсонського національного технічного університету; моб. тел. +380951757216, e-mail: alexander.meshkov@gmail.com.

Новіков Олександр Олександрович – д.х.н, проф., завідувач кафедри Інформаційно-вимірювальних технологій, електроніки та інженерії Херсонського національного технічного університету; моб. тел. +380504946106, e-mail: novikov.oleksandr@kntu.net.ua.