

## МЕТОДИКИ РОЗПІЗНАВАННЯ МОВЛЕННЯ: ДОСЛІДЖЕННЯ ВІДБИТКУ ГОЛОСУ

*Стаття знайомить з деякими аспектами дослідження відбитків голосу (спектрограм) як однією з методологій розпізнавання мовлення. Після короткого історичного нариску описується сучасний стан справ у цій галузі. Серійні і проектні розробки методологій розпізнавання мовлення вже використовуються як у цивільній, так і у військовій сферах. У статті наводяться деякі найперспективніші напрями використання методологій розпізнавання мовлення.*

**Ключові слова:** відбиток голосу, спектрограма, розпізнавання мовлення і мови, цивільне застосування, військова сфера.

*В статті описуються такі методи розпізнавання речі як дослідження відбитку голосу (спектрограм). Після короткого історичного екскурсу слідують описи нинішньої ситуації в даній області. Приводяться приклади застосування серійних і ще не розроблених методик розпізнавання речі в цивільній і військовій сферах. Також представлені деякі перспективні напрями використання методик розпізнавання речі.*

**Ключевые слова:** отпечаток голоса, спектрограмма, распознавание речи и языка, гражданское использование, военная сфера.

*This article provides some insight into voiceprint (spectrogram) studies as one of speech recognition technologies. A brief historic outline is followed by the description of the present-day state of affairs in this area. Pre-commercial and commercial off-the-shelf speech recognition technologies find their way into both civil and military applications. The article also offers some ideas as to the most advanced and novelty ways of using speech recognition methods.*

**Key words:** voiceprint, spectrogram, speech and language recognition, civilian application, military domain.

### ***Відбиток голосу (спектрограма): історичний екскурс і трохи теорії***

Відбиток голосу виступає найбільш унікальною формою розпізнавання людини серед усіх інших, бо така форма більш складна і неповторна, ніж відбиток пальця чи сітківка ока. Відбиток пальця представляє собою лише точне відтворення ліній шкіри, в той час як відбиток голосу – це поєднання вимовного акценту, специфіки промовляння ненаголошених складів і ритміки висловлення, закінчень слів і фраз із такими фізичними факторами як розмір і форма голосових зв'язок мовця.

Саме через таку складну комбінаторику характеристик відбиток голосу практично неможливо підробити. Ось чому біометрична перевірка/верифікація голосу особистості сьогодні вважається однією з найнеобхідніших безпекових технологій як для військових, так і для цивільних цілей і завдань.

На перший погляд може скластися враження, що дослідження відбитків голосу почалися лише нещодавно. Насправді, підвалини дослідженням біометрики голосу було закладено ще наприкінці ХІХ століття Олександром

Мелвіллом Беллом, батьком винахідника телефону. Методика полягає у транскрибуванні (спеціальному записі мовленнєвих малюнків) не лише того, що було вимовлено, але і того, як це було вимовлено. І, як це часто буває, новий винахід спочатку здебільшого використовувався для специфічних цілей представниками спеціальних органів: слідчими, поліцейськими, військовими тощо. Перший прототип «Поліграфа», або «детектора брехні», технічного приладу, який через низку технічних характеристик «надає оцінку» мовленню, було розроблено приблизно у 1917 році. Проте, як тоді, так і зараз проблема із використанням такого пристрою полягає у тому, як інтерпретувати отримані за його допомогою дані.

Записи мовлення, про які згадувалося, спочатку носили назву «спектрограми», і лише у 1944 році у роботі С. Грея і Г. Коппа (Gray, Корр, 1944) замість терміну «спектрограма» з'явився «відбиток голосу» як метафоричне посилання на відбиток пальця.

Активні дослідження біометрики голосу з метою перехоплення голосових повідомлень і відстеження супротивника проводилися під час Другої Світової війни, зокрема, у СРСР (lawbook.online, 2017; nauka.x-pdf.ru, 2017) і США (voicevault.com, 2015).

Безсумнівно, що і в інших країнах – основних учасниках Другої Світової війни такі дослідження велися, проте, документальні підтвердження ще і досі здебільшого залишаються під грифом «секретно».

Напочатку 1980х років дослідження відбитків голосу як методу ідентифікації мовця використовувалося у судах 25 штатів у США, а також двома військовими судами і двома судами загальної юрисдикції Канади. Існують також дані про те, що «відбиток голосу» як свідчення приймався до уваги апеляційними і, навіть, деякими верховними судами у США та низці країн Європи (voicevault.com, 2015). Проте, в науково-дослідницькому й академічному середовищі немає єдиної точки зору щодо припустимості використання відбитку голосу в якості свідчення чи доказу у судовій справі.

Тим не менш, вагомим і цікавим фактором на користь використання відбитків голосу (спектрограм) можна вважати книгу Є. Блока «Відбиток голосу: як закон може зрозуміти голос злочину» (Block, 1975). Автор детально описує необхідні параметри для роботи із відбитками голосу (спектрограмами): якість і тривалість запису, наявність чи відсутність фонових перешкод (шумів), підготовленість слухачів-оцінювачів.

Недивно, що лінгвісти та люди зі знанням іноземних мов («просунутий/ майже рідний») рівень володіння мовою) краще розпізнають не лише голоси, але і діалекти чи акценти, які можуть «вдягати на себе» мовці, аби не бути упізнаними. Фонетика і вимова рідної мови дуже часто «проступають» крізь імітований акцент чи діалект мовця. Також удаваний акцент чи мовлення нерідною мовою можуть бути насичені такими помилками, як емоційна удаваність, плутаність, невідповідності природному невимушеному мовленню. З іншого боку, коли мовець демонструє стопроцентну прихильність усім визначним мовним правилам, таке мовлення сприймається як удаване і, відповідно, нерідне.

Ступінь знайомства слухача із мовцем також підвищує здатність коректного розпізнавання голосу, навіть не зважаючи на різні «маскувальні» технології. Серед найбільш уживаних так званих «фізичних» способів замаскувати голос використовуються шепотіння, зміна діапазону голосу (зависокий чи наднизький), артикуляційне маніпулювання (набуття акценту/діалектної вимови, або назалізація вимови).

Окрім фізичних способів маскування голосу, існують і широко використовуються електронні маніпулювання відтвореним мовленням і синтез мовленнєвого сигналу.

### ***Практичні аспекти застосування технологій розпізнавання мовлення і мови сьогодні: ким і для чого***

Відбиток голосу (спектрограми) сьогодні широко використовують приватні детективи та інші експерти поза межами академічного товариства. Зазвичай, такі дослідження проводяться у лабораторних умовах. Тут мова йде про вивчення відбитків голосу (спектрограм) тривалістю від одного слова до закінченого речення, а також окремо вимовлених слів («ізольоване слово») та відокремлених ключових («наповнених») слів у контексті. Значення також має тривалість записаного уривку та акустична якість запису, присутність чи відсутність шумового фону і діапазон частот.

Однією із найбільш уживаних вважається методика так званого «параду голосів» для розпізнавання безпосереднім свідком. Це – аудиторський еквівалент «лінійки на упізнання», що використовується поліцією. Так звані «контрастуючі голоси» повинні водночас бути доволі подібними до досліджуваного відбитку голосу, але і містити певний ступінь варіативності.

З метою більш точного виявлення і розпізнавання відбитку голосу використовують також методики сканування мозку, аби відокремити правдиві заяви від удаваних чи брехливих стверджень. В рамках лабораторних досліджень також часто проводять функціональні магнітно-резонансні томографії, проте, така методика вважається доволі складною, дорогою і такою, що забирає багато часу. Отже, при розслідуванні приватних справ, такі методики практично не застосовуються.

Сьогодні дослідження відбитків голосу проводяться не лише лабораторно-аудитивними методами, як згадувалося вище, а також із використанням доробків криміналістичної акустики, фонетики і фонології у галузі технологій розпізнавання мовлення широким колом зацікавлених сторін, перш за все, із військовими цілями. Визначення мовця через використання відбитку голосу (спектрограму) вважається однією з прикладних галузей, що поєднують фонетичні науки та інженерію. Проте, дуже часто різноманітні конгреси з питань фонетичних наук, наприклад, збори Акустичного Товариства Америки (ASA, 2017) займаються здебільшого лише обговоренням теоретичної сторони питання і дослідженням фактів, які часто протирічають одне одному. Як зазначив один із учасників останніх 2016 зборів ASA 2016 року: «Ми

обговорюємо різні точки зору і це – нормальний стан справ для науки. Говоримо про теорію, інтерпретацію даних і взаємини між подіями, тобто, ретельно проробляємо кожну сторону наукового процесу. І в якості відповіді на поставлені питання ми пропонуємо ще один раунд обговорення, додаткові перевірки конкуруючих ідей, їх подальші, наступні, майбутні перевірки. І лише згодом, після з'ясування фундаментальних залежностей, напрацьовується певне рішення. Однак, часто-густо, це рішення взагалі не буде застосовано на практиці. Технології, зокрема, військові чи цивільно-безпекові, не можуть так довго чекати, бо, на жаль, між надійними результатами дослідження та їх можливістю практичного застосування у польових умовах часто простягається вічність...» (ASA, 2016).

Тим не менш, прогрес зупинити не можна, і, переходячи від теоретичної стадії до практичної реалізації, результати всіх досліджень рано чи пізно стають на комерційні рейки. А далі, як би цинічно це не звучало, все залежить від ціни питання. Також, як ми всі знаємо, велика кількість наукових відкриттів не лише спонсорується, але і, насамперед, замовляється військовими. Найбільш відомим прикладом цього є народження Інтернету, який із вузько спеціалізованого військового проекту розвинувся у глобальну неосяжну мережу, захопивши усіх нас.

Незалежно від вартості розробок та їх втілення в практичне життя, безпекові технології завжди використовують найновітніші наукові доробки. І, мабуть, ми всі вже знайомі з деякими напрямками використання технологій розпізнавання мовлення у буденному житті. Інколи ми навіть не зважаємо на такі технології уваги, бо вони тишком-нишком вже доволі давно супроводжують нас, чи просто не замислюємося над тим, наскільки автоматизованим стало наше сьогоднішня. Отже, прикладами практичного застосування технологій розпізнавання мовлення є такі: (i) програвання записаної простої інформації; (ii) керування дзвінками; (iii) автоматичне розпізнавання чи ідентифікація (голосу) мовця, (iv) відмова від інтерактивних систем відповіді на телефонні дзвінки (можна вважати варіантом пункту (ii)), (v) розв'язання проблеми пікового навантаження дзвінків у мережах (callcentrehelp.com, 2008).

Якщо коротко, то (i) програвання записаної простої інформації використовується практично усюди, починаючи із компаній з надання житлово-комунальних послуг і до банків чи інформаційних служб аеропортів. Так само і пункт (ii) керування дзвінками набув широкого вжитку в службах клієнтської підтримки банків, в інформаційних меню «самообслуговування» пенсійних фондів, онлайн крамниць, тощо.

Подібно до двох вищевказаних прикладів, відмова від інтерактивних систем відповіді на телефонні дзвінки на користь так званого «розумного керування дзвінками» за допомогою систем розпізнавання природного мовлення допомагає різноманітним компаніям максимально знизити час очікування відповіді на дзвінок і, відповідно, зекономити свої ресурси.

*Автоматична ідентифікація (голосу) мовця* набула найширшого вжитку, насамперед, в сфері банківських послуг і в тих галузях, що пов'язані із

забезпеченням безпеки, охорони, посиленням захистом чи обмеженням доступу до певних даних. Зокрема, впровадження такої технології радо вітали в авіації. Відомо, що найнебезпечніший час польоту – це зліт і посадка, коли командир повітряного судна спілкується з наземними службами. І тут повне і чітке розуміння на обох сторонах комунікаційного ланцюга означає безпеку життя багатьох людей. Спілкування у повітряному просторі, наприклад, при перетині повітряних кордонів країн, коли керування літаком передається диспетчерами однієї країни колегам з іншої, – це багатомовна комунікація, тому коректне розпізнавання різноманітних мовленнєвих малюнків різними мовами набуває величезного значення. Не дивно, що передові технології автоматизованого розпізнавання мовлення в галузі авіації було запроваджено відносно давно порівняно з іншими галузями. Зрозуміло також, що такі технології безперервно оновлюються і покращуються.

І, нарешті, дієві способи *розв'язання проблеми пікових навантажень на телефонні мережі* пов'язані або із надзвичайними подіями (робота «гарячих ліній»), або із проведенням, наприклад, спортивних змагань високого міжнародного рівня, що цікавлять практично усіх. Типові телефонні запити, що надходять, скеровуються в автоматичному режимі, подібно до пунктів (ii) і (iv), а такі, що потребують індивідуального підходу і розлогої відповіді, передаються безпосередньо оператору.

Якщо говорити про військову сферу застосування технологій розпізнавання мовлення, то на основі відкритих даних можна зробити висновок про те, що лише за останні два десятиліття накопичена теоретична база нарешті перейшла до стадії практичного застосування. Відомо також, що найбільш широкого вжитку технологія розпізнавання мовлення набула при керуванні безпілотними апаратами (наземними, (під)водними, повітряними), а також при забезпеченні надійного стільникового зв'язку. Тим не менш, взаємодія акустичних і фонетичних наук із військовим сектором не обмежується лише цими напрямками.

Військова галузь користується як такими технологіями, що вже знаходяться у комерційному обігу, так і тими, які ще не представлено на ринку. Основними напрямками, де використовуються технології розпізнавання мовлення і мови з військовими цілями є, наприклад, наступні: (i) командування і контроль; (ii) комунікації; (iii) комп'ютери і системи (інформаційного) доступу; (iii) розвідка; (iv) навчання; (v) багатонаціональні сили<sup>1</sup> (rta.nato.int, 2005–2017).

Командування і контроль. Системи розпізнавання мовлення і мови допомагають забезпечити голосову взаємодію між людиною і комп'ютером, різними видами озброєння і сенсорних систем. Передумовою такої взаємодії мусить бути високий рівень технології розпізнавання мовлення і мови в режимі реального часу, дуже часто у несприятливих умовах (фонові шуми, рухи, різноманітні наслідки стресових ситуацій, багатомовний ландшафт,

---

<sup>1</sup> Ми розглядаємо обмежений перелік напрямів відповідно до письмового зобов'язання автора про дотримання тасмніці.

тощо). Зокрема, для забезпечення якості командування і контролю рішучої важливості набувають: (де)кодування мовлення і мови; коригування мовленнєвих сигналів; синтез мовлення і мови; розпізнавання мовлення.

Важливими складовими також виступають технології забезпечення (взаєм)розуміння, інтерактивного діалогу і мульти-модальних комунікацій особового складу і різних родів військ.

Комунікації. Основна вимога до комунікацій – це їх надійність і безперервність. Окрім цього, комунікації повинні бути безпечними, зрозумілими і чіткими, не зважаючи на різноманітні перешкоди і заглушення сигналу. Самий звуковий сигнал може бути, наприклад, закодований/ зашифрований і повинен передаватися із високим ступенем надійності комунікаційного каналу, аби бути правильно декодованим/ дешифрованим і зрозумілим слухачами, які не обов'язково є носіями мовлення чи мови сигналу. Для цілей забезпечення надійних комунікацій наступні технології набувають основного значення: (де)кодування мовлення і мови; коригування мовленнєвих сигналів; переклад (усний і письмовий).

Комп'ютери і системи (інформаційного) доступу становлять сьогодні найнеобхіднішу складову сучасних військових операцій. Технології розпізнавання мовлення і мови використовують для того, аби військові надавали голосові команди комп'ютерам чи сенсорним пристроям, або запитували пристрої про надання інформації. При цьому всі сигнали повинні бути надійно захищені від: а) перехоплення і б) декодування або навмисного викривлення і подальшої передачі. Отже, збалансована і чітка взаємодія між людиною і технікою надзвичайно важлива, оскільки військовий персонал мусить постійно й уважно слідкувати за подіями в режимі реального часу і одночасно виконувати декілька невідкладних завдань.

При проведенні військових операцій із залученням коаліційних чи багатонаціональних сил ставляться додаткові вимоги до забезпечення безпечного, безперешкодного і надійного доступу до інформаційних систем. Крім того, в умовах багатонаціональних сил і різних вертикалей командування, для забезпечення конфіденційності та безпеки інформація мусить бути класифікованою за певними відповідними рівнями доступу. З точки зору технічного забезпечення необхідно проводити ідентифікацію (розпізнавання) мовців, захистити аудіо-дані цифровими водяними знаками та/ або підписами, здійснювати перевірку і правильність вхідного багатомовного сигналу, а також забезпечувати стрімкий переклад (як письмовий, так і усний) чи стислий підсумковий виклад інформації з однієї мови іншою. Тому технології синтезу мовлення і мови, а також виявлення і розпізнавання звукових сигналів, що надходять, їх переклад і можливість застосування в інтерактивному діалозі набувають критичного значення.

Розвідка. Обробка інформації та її поширення є складовими елементами і щаблями розвідки і (контр)розвідки. Успіх (контр)розвідки забезпечується обробкою такої інформації, яка має високу якість і цінність, точність і вчасність. Через бурхливе зростання кількості комунікаційних каналів і, зокрема,

Інтернету, а також інших відкритих інформаційних джерел, в геометричній прогресії зростає й обсяг аудіо- і різноформатних текстових даних.

Зрозуміло, що не вся інформація однаково цікава для цілей (контр)розвідки, тому їй треба ефективно відфільтрувати. І тут на допомогу так само приходять технології розпізнавання мовлення і мови, коли спеціально навчені експерти й автоматизовані пристрої відфільтровують дані, аналізуючи відбитки голосу/ мовлення, що звучить. Найбільш необхідними і вживаними стають такі технології: кодування і декодування мовлення і мови; коригування мовленнєвих сигналів; відстеження власного сигналу і супровід ворожого, розпізнавання мовлення і конкретного мовця, а також коректне визначення мови сигналу. З'ясування теми і визначення ключових слів (перехопленого) сигналу, навіть за умови його неповного відтворення, переклад тексту (письмового чи такого, що звучить) або його частини, розуміння і можливість подальшого використання в мульти-модальному комунікаційному просторі також надзвичайно важливі для (контр)розвідки.

Навчання особового складу, а також машин і механізмів, якими користуються військові, іноземним мовам – це необхідність, хоч як банально це не звучить. Сьогодні навчання іноземним мовам за допомогою комп'ютерів і технологій – це, передусім, опанування мовами союзників і супротивників. Ключовими навичками є розуміння мови і мовлення, а також розпізнавання мовця. Крім того, акустичні і фонетичні навички потрібні для здійснення інтерактивного діалогу і залучення співрозмовника до мульти-модальної комунікації.

Багатонаціональні сили. За визначенням, багатонаціональні сили вимагають чіткого координування зусиль усіх військових підрозділів, які спілкуються різними мовами. Тому, розуміння мовлення і мови і переклад (усний і письмовий) за допомогою технологій розпізнавання мовленнєвих і мовних сигналів набувають величезного значення і на практиці сприяють підвищенню ефективності і досягненню успіху військових операцій. З цією метою необхідно використовувати цілий спектр різноманітних технологій. На початковому етапі можна обмежитися лише деякими із них, зокрема: перекладом і багатомовним обміном усною і письмовою інформацією на основі стандартизованих термінологічних і фразеологічних словників. Це означає, що військові повинні вміти (де)кодувати мовлення і мови; коригувати мовленнєві сигнали; здійснювати усний і письмовий переклад повідомлень і забезпечувати їх розуміння колегами-мовцями інших мов.

Серед найбільш запитаних для використання у найближчому майбутньому методик є ті, що можна охарактеризувати одним словом «автоматизовані». Наприклад:

- 1) автоматизований переклад наказів і повідомлень, що скорочує час відтермінування військової реакції;
- 2) переведення голосових розвідданих, переданих людиною, до формату «мово-незалежних» даних;

3) автоматизоване генерування специфічних військових глосаріїв і багатомовних словників для використання на штабному й операційному рівнях відповідно до стану справ на театрі, що постійно видозмінюється;

4) висока розбірливість мовленнєвого сигналу, навіть за умови його передачі гетерогенними засобами зв'язку;

5) можливість з'єднання різних гетерогенних терміналів (штаби, мобільні штаб-квартири, морські і повітряні кораблі);

6) захищена передача зашифрованого (кодованого) мовленнєвого сигналу із можливістю негайного перекладу на обох сторонах комунікаційного каналу.

Отже, підсумовуючи, можна стверджувати, що у технології розпізнавання мовлення і мови не лише вражаюча історія, але і цікаве, повне захоплюючих відкриттів, майбутнє.

Автор вдячна колегам-перекладачам і учасникам цьогорічної Парламентської Асамблеї НАТО за практичну допомогу і мудрі поради при підготовці цієї статті.

### *Література*

1. Asa.com. (2016). *Acoustic Analysis: New Ideas*. Retrieved from <http://asa.scitation.org/journal/publications>.
2. Asa.com. (2017). *News and Education/Outreach*. Retrieved from <http://acousticalsociety.org/news/outreach>.
3. Block, E. B. (1975). *Voiceprinting: how the law can read the voice of crime*, 43–57. New York: D. McKay Co. 244 p.
4. Callcentrehelper.com. (2008). *Speech Recognition Methods*. Retrieved from [https://www.callcentrehelper.com/speech\\_recognition](https://www.callcentrehelper.com/speech_recognition).
5. Gray, C. and Kopp, G. (1944). *Voiceprint Identification*. Report Presented to the Bell Telephone Labs (P. 1–14). Bell Laboratories Ltd.
6. Lawbook.online. (2017). *Phonoscopy: Now and Tomorrow*. Retrieved from <https://lawbook.online/phonoscopy/history/studies.html>.
7. Nauka.x-pdf.ru. (2016). *Forensic Phonetics: Means and Ways*. Retrieved from <https://nauka.x-pdf.ru/17story/forensicphonetics/428848.php>.
8. Rta.nato.int. (2005-2017). *Speech Recognition in Military Environments*. Retrieved from [www.rta.nato.int/Docs/IT-071-ALL.pdf](http://www.rta.nato.int/Docs/IT-071-ALL.pdf).
9. Voicevault.com. (2015). *Voice Biometrics: Brief History*. Retrieved from [https://voicevault.com/voice\\_biometrics\\_in\\_brief](https://voicevault.com/voice_biometrics_in_brief).