

УДК 004.934

А.М. Продеус

МОВЛЕННЄВІ КОРПУСИ ЯК ЗАСОБИ СТВОРЕННЯ ТА ЗБЕРІГАННЯ ЗРАЗКОВИХ МОВЛЕННЄВИХ СИГНАЛІВ

Speech corpora are an important constituent of modern investigators' toolkit in such areas as speech correction, designing and testing elements of telecommunication systems and systems of automatic speech recognition. In this paper, we search for elements of construction technology of the sound part of noisy Ukrainian speech corpora. To this end, we consider characteristics of the most widely used modern noisy speech corpora which allow formulating principles of such corpora design. The regularity of formulated principles is shown by an example of known modern program toolkit FaNT which permits constructing quickly speech corpora with required properties. The guidelines on constructing similar program toolkit in Matlab environment are developed. Such toolkit will allow not only to work out by joint efforts the best version of a Ukrainian noisy speech corpus, but to compare algorithms of noise reduction and algorithms of automatic speech recognition elaborated by various scientists with one another in the future.

Вступ

Відомо, що ефективність функціонування систем мовленнєвої комунікації та автоматичного розпізнавання мови (АРМ) залежить не тільки від виду мовлення (команди, читання, спонтанне мовлення), але й від акустичного оточення (шум різної природи, реверберація) і характеристик каналів комунікації (фільтри, кодеки) [1–3]. Тому в останнє десятиліття особливу увагу дослідників привертають методи моделювання мовлення в умовах підвищеного рівня шумів оточення та спотворень сигналів, притаманних лініям комунікації. Одним із таких методів є створення мовленнєвих корпусів [4].

Мовленнєві корпуси являють собою бази даних у вигляді звукових і пов'язаних з ними текстових файлів [4]. Звукові файли містять зразки елементів мовлення (звуки, склади, слова, фрази), у текстових файлах розміщені відповідні транскрипції. Зазвичай до складу мовленнєвих корпусів включають допоміжний програмний інструментарій, який забезпечує редагування файлів, автоматизацію транскрибування звукових файлів і т.п., а також файли з різноманітною службовою інформацією (інструкції з експлуатації, вихідні тексти програм тощо).

У далекому зарубіжжі найбільшим попитом користуються мовленнєві корпуси NTIMIT, 22 Language, Isolet, Aurora, NOIZEUS [5–8]. У Росії також розроблено кілька мовленнєвих корпусів зашумленої мови [9–12]. Разом з тим українські фахівці дещо відстають у створенні корпусів мовлення, записаного в умовах впливу реальних завод [13, 14]. Оскільки відсутність українських мовленнєвих корпусів істотно пе-

решкоджає подальшому розвитку цифрових технологій, пов'язаних із використанням вітчизняних систем цифрового телефонного зв'язку та систем АРМ, необхідно шукати шляхи усунення цього недоліку. Зважаючи на принципову подібність урбаністичних шумів у різних містах світу, можна скористатися записами шумів, задіяними при створенні відомих мовленнєвих корпусів. Так зробили, наприклад, розробники мовленнєвого корпусу NOIZEUS, які скористалися шумами мовленнєвого корпусу Aurora [8].

На жаль, такий шлях не завжди може задовольнити вимоги дослідників. По-перше, навіть за умов наявності "готових до використання" записів шумів акустичного оточення залишається актуальним завдання самостійного формування "чистих" мовленнєвих сигналів, тобто сигналів, не спотворених різноманітними завадами. По-друге, характеристики наявних записів шумів можуть не узгоджуватися із вимогами дослідників. Таким чином, дослідники мають вміти самостійно й доволі швидко створювати мовленнєві корпуси із потрібними характеристиками.

Постановка задачі

Мета роботи полягає у виявленні елементів технології порівняно швидкої побудови корпусів зашумленого мовлення із потрібними характеристиками, що дасть можливість прискорити й підвищити якість досліджень ефективності математичного та програмного забезпечення систем зв'язку й АРМ. Для досягнення цієї мети доцільно виконати огляд характеристик найбільш вживаних сучасних корпусів зашумленого мовлення та сформулювати прин-

ципи розроблення таких корпусів. Після перевірки правильності сформульованих принципів, доцільно розробити практичні рекомендації із власноручної побудови зручного програмного інструментарію для швидкого створення потрібних мовленнєвих корпусів.

Характеристики сучасних корпусів зашумленої мови

У цій статті обмежимося оглядом характеристик найбільш вживаних мовленнєвих корпусів, що використовуються для налагодження і тестування систем мовного зв'язку й АРМ.

Корпуси англійського мовлення. У далекому зарубіжжі, при дослідженнях і розробленні алгоритмів АРМ, найбільшим попитом користуються мовленнєві корпуси ТІМІТ і NTІМІТ [5, 16]. Акустико-фонетичний корпус ТІМІТ призначений як для фонетичних досліджень, так і для розроблення й тестування автоматичних систем розпізнавання злитого мовлення в межах американського варіанту англійської мови. Це дуже великий за обсягом корпус: текстовий матеріал ТІМІТ включає 6300 речень, по десять речень для кожного із 630 дикторів з восьми діалектних зон США. Недоліком частини корпусу ТІМІТ, призначеної для тестування систем АРМ, є те, що вона не враховує спотворення мовленнєвих сигналів, викликаного шумами акустичного оточення та нерівномірністю частотних характеристик каналів зв'язку.

Мовленнєвий корпус NTІМІТ (Network ТІМІТ) являє собою надбудову до корпусу ТІМІТ у вигляді мовленнєвих сигналів у смузі частот, що використовуються в телефонії. NTІМІТ призначений для проведення досліджень впливу телефонних ліній на якість і розбірливість мовлення, на ефективність систем АРМ. Мовленнєвий корпус NTІМІТ сформовано за допомогою запису й оцифрування всіх 6300 оригінальних сигналів корпусу ТІМІТ, які було піддано спеціальним спотворенням. Так, мовні сигнали корпусу ТІМІТ передавалися по телефонних мережах десяти зон, половина з яких були зонами дальнього зв'язку. Крім того, передані сигнали піддавалися спотворенням за допомогою системи "штучний рот" і системи, яка моделює слухавку. Дослідження систем АРМ із застосуванням мовного корпусу NTІМІТ дали змогу встановити, зокрема, що практично завжди в телефонних трактах наявні шумова й імпульсна перешкоди, а також спотворення амплітудного

й фазового спектрів переданого сигналу істотно, на 10–15 %, знижують якість автоматичного розпізнавання мовлення [1, 3].

Перелік двадцяти мовленнєвих корпусів, розроблених у центрі вивчення усного мовлення аспірантури при Орегонському медичному університеті (Center for Spoken Language Understanding of Oregon Health & Science University) і призначених для досліджень алгоритмів систем шумозаглушення й АРМ, наведено в [6]. Більшість корпусів являють собою записи телефонного мовлення. Зокрема, розроблений у 2000 р. корпус 22 Language цікавий тим, що до його складу включено зразки "телефонного" мовлення 22 мов світу, у т.ч. російського. Мовленнєві сигнали записувалися для цифрових телефонних ліній з такими параметрами кодування: 8 кГц, 8 біт, нелінійне компандування за μ -законом.

У згаданій вище перелік двадцяти корпусів входить і мовний корпус ISOLET, розроблений у 1990–1996 рр. і модифікований у 2000–2002 рр. Він містить записи окремих звуків (26 звуків) англійського мовлення. В цілому у базі є 7800 записів довжиною близько 1,25 год. Параметри оцифрування звуку такі: 16 кГц, 16 біт, моно.

Мовленнєві корпуси Aurora, які були розроблені в період з 1997 по 2002 р. однойменною робочою групою Європейського інституту телекомунікаційних стандартів (European Telecommunications Standard Institute – ETSI), містять записи у вигляді послідовності цифр (від 0 до 9), вимовлених так, як це роблять користувачі мобільних телефонів при голосовому наборі номера [7]. Різні версії корпусу Aurora відрізняються прототипами й способами додавання шумового фону. Так, у корпусі Aurora-2, який базується на прототипі у вигляді корпусу TIDigits, шум додавався штучно. У корпусі Aurora-3 записи зроблено на тлі природного шуму в салоні автомобіля. У корпусі Aurora-4 як прототип використано корпус Wall Street Journal [16], при цьому до мовленнєвих сигналів штучним шляхом були додані шуми різної природи. Корпус Aurora-5 було створено в 2006 р. Він базується на прототипі у вигляді корпусу TIDigits і призначений для відпрацювання програмного забезпечення мобільних телефонів, що експлуатуються у режимі handsfree у приміщеннях або салонах автомобілів. Параметри оцифрування мовленнєвих сигналів такі: 8 або 16 кГц, 16 біт, моно.

Англомовний (американський варіант англійської мови) мовленнєвий корпус NOIZEUS, розроблений у 2007 р. в лабораторії обробки мовлення університету Далласа (Speech Processing Lab of the University at Texas at Dallas), доступний для вільного використання [8]. Він містить усього 30 речень, вимовлених шістьма дикторами (три жінки й три чоловіки), спотворених вісьмома видами реальних шумів (запозичених з корпусу Aurora) для різних відношень сигнал/шум (0, 5, 10 і 15 дБ). Усі файли збережені у форматі WAV (8 кГц, 16 біт, моно).

Корпуси російського мовлення. Санкт-Петербурзькою компанією ЦРТ [9] створено мовну телефонну базу RUSTEN (250 фонограм телефонних переговорів між двома персонами), яку, зокрема, було використано при розробленні автоматизованого комплексу розпізнавання дикторів у телефонному каналі "Трал".

Санкт-Петербурзькою комп'ютерною фірмою "Одитек" розроблені два спеціалізованих корпуси російського телефонного мовлення для завдань верифікації диктора (65 дикторів) і автоматичного розпізнавання мовлення (1350 дикторів) [10, 11]. Іншим прикладом російського мовленнєвого корпусу для стаціонарних телефонних мереж є база даних Russian Speech-Dat(E) Database, розроблена співробітниками тієї ж фірми "Одитек". Вона містить записи мовлення 2500 російських дикторів (1242 чоловіків, 1258 жінок) [12]. Параметри оцифрування мовних сигналів такі: 8 біт, 8 кГц, компресія по А-закону. Нарешті, фірмою "Одитек" створено мовленнєвий корпус російської мови в межах європейського проекту SPEECON, спрямованого на створення мовленнєвих корпусів більшості європейських і деяких азіатських мов. Записи виконані в різних акустичних середовищах: офіс, автомобіль, громадське місце (у т.ч. вулиця), будинок. Усього записано 550 дорослих дикторів, 50 дітей. При організації записів враховано такі характеристики, як вік, стать, особливості діалекту. Запис провадився по чотирьох каналах одночасно (для чотирьох різних мікрофонів), параметри оцифрування такі: 16 кГц, 16 біт. Для кожного мовленнєвого файлу створено файл анотації, у якому, крім орфографічної анотації мовлення, міститься інформація про звукові дані, дикторів, сценарії запису, мікрофони.

Корпуси українського мовлення. Лідерами в Україні в цій галузі є відділ розпізнавання й синтезу звукових образів Київського МННЦІТС, а також Українська асоціація з обробки інформації й розпізнавання образів [13]. Цими організа-

ціями для досліджень у галузі автоматичного розпізнавання українського мовлення розроблено три мовленнєвих корпуси: багатодикторний мовленнєвий корпус UkReco, мовленнєвий корпус виступів депутатів Верховної Ради України, а нині триває розроблення акустичного корпусу українського ефірного мовлення [17].

Недоліком зазначених вище корпусів українського мовлення є те, що їм притаманні доволі високі значення сигнал/шум (25 дБ і вище). Ця обставина не дає можливості використовувати вказані корпуси для налагодження та тестування математичного забезпечення систем АРМ, що експлуатуються в умовах, притаманних реальним лініям зв'язку. Тому Київським МННЦІТС, а також Українською асоціацією з обробки інформації й розпізнавання образів, розробляється телефонний мовний корпус. Його особливості: російська і українська мови, реальні записи для різних моделей мобільних телефонів, частота дискретизації 8000 Гц, спонтанне мовлення. Об'єм корпусу сягає 5 Гб (формат GSM). На жаль, цей корпус поки що не анотовано.

Принципи розроблення корпусів зашумленого мовлення

За результатами наведеного вище огляду сформулюємо кілька загальних принципів формування корпусів зашумленого мовлення, ґрунтуючись на узагальненій схемі процедури тестування систем обробки мовленнєвих сигналів, показаній на рис. 1. До згаданих принципів належать врахування:

- різновидів мовлення, що моделюється;
- різновидів спотворень, яким піддаються мовленнєві сигнали в каналах зв'язку;
- різниці у призначенні та технічних характеристиках систем, що налагоджуються або тестуються із застосуванням мовленнєвого корпусу;
- значної різноманітності критеріїв і методів оцінювання якості систем АРМ та телекомунікаційних систем;
- потреби порівняння результатів, одержаних різними дослідниками.

Коротко прокоментуємо зазначені принципи. Врахування різновидів мовлення, що моделюється, означає, зокрема, врахування наявності доволі довгих пауз між словами при дискретному або спонтанному мовленні. На практиці це означає необхідність спеціальних заходів для коректного додавання шуму до первинного "чистого" мовленнєвого сигналу, який зазвичай записується в умовах заглушеного приміщення.



Рис. 1. Схема тестування системи обробки мовних сигналів

Врахування різновидів спотворень сигналів означає таке: за умов, коли відпрацьовуються алгоритми придушення шуму акустичного оточення, достатнім є моделювання мовного сигналу $y(t)$ у вигляді адитивної суміші чистого мовного сигналу $x(t)$ й шумового процесу $n(t)$:

$$y(t) = x(t) + n(t). \quad (1)$$

При врахуванні впливу характеристик приміщення, що характеризується імпульсною характеристикою $h_f(t)$, модель мовного сигналу має такий вигляд:

$$y(t) = x(t) \otimes h_f(t) + n(t), \quad (2)$$

де \otimes – символ згортки. Нарешті, враховуючи ще й вплив фільтруючих властивостей каналу зв'язку, що характеризуються імпульсною характеристикою $h_c(t)$, модель мовного сигналу має ще більш складний вираз:

$$y(t) = (x(t) \otimes h_f(t) + n(t)) \otimes h_c(t). \quad (3)$$

Наслідком різноманітності систем, які піддаються налагодженню або тестуванню, є різноманітність додаткових вимог до моделі мовного сигналу. До них належать такі:

- смуга частот (вузька – 30–3400 кГц; широка – 50–7000 Гц; суперширока – 50–14 000 Гц) [21];
- “априорне” відношення сигнал/шум $\xi = 10 \lg(D_x / D_n)$ (D_x і D_n – дисперсії сигналу $x(t)$ й шуму $n(t)$ відповідно);
- характер забарвленості шуму (при використанні штучних шумів);

- джерело походження шуму природного походження (шуми природних явищ, урбаністичного середовища, промислові шуми і т.п.);

- характер і ступінь нестационарності шумової завади;

- наявність додаткових видів завад і спотворень (реверберація, луна, особливості частотної характеристики тракту, спосіб кодування сигналу тощо);

- наявність міжнародних стандартів і рекомендацій щодо характеристик фільтрів та кодексів мовлення, які використовуються в мобільних лініях зв'язку.

При розробленні мовленнєвих корпусів доцільно враховувати, за якими критеріями буде оцінюватися система, що налагоджується або тестується. Справа в тому, що об'єктивним (інструментальним) методам акустичної експертизи властиві обмеження [18]. Наприклад, при застосуванні формантного методу оцінювання розбірливості мови тестовий сигнал має бути неперервним. Це може бути або стаціонарний випадковий процес із спектром, подібним до довготривалого спектра неперервного мовлення, або зразок неперервного мовлення. Таким чином, для оцінювання розбірливості мови на виході системи, що тестується, потрібно мати корпус неперервного мовлення. Дещо складнішою є ситуація при застосуванні модуляційного або формантно-модуляційного [19] методів оцінювання розбірливості мови, де тестовий сигнал є принципово нестационарним. У цьому випадку корпус має містити зразки дискретного мовлення із певними вимогами до довжини пауз.

Нарешті, при розробленні нових алгоритмів заглушення завад актуальною є необхідність порівняння результатів, одержаних різними дослідниками. Мовленнєві корпуси як спосіб створення та зберігання зразкових мовленнєвих сигналів якнайкраще відповідають цій потребі.

Програмний інструментарій FaNT для створення зразків зашумленого мовлення

Правильність сформульованих вище принципів можна перевірити шляхом аналізу можливостей програмного інструментарію FaNT (Filtering and Noise Adding Tool), який свого часу було створено та використано для синтезу штучно спотворених мовленнєвих даних у вигляді корпусів Aurora-2 та Aurora-4 [7, 15]. Система FaNT дає змогу додавати шум до мовленнєвого сигналу таким чином, щоб забезпечити бажане відношення сигнал/шум, а також врахувати вплив частотних характеристик ліній зв'язку. При синтезі сумішей із заданими відношеннями сигнал/шум дає можливість врахувати особливості сприйняття звуку слуховою системою людини шляхом частотного А-зважування. Крім того, система FaNT містить детектор голосової активності (Voice Speech Detection – VAD), використання якого дає змогу коректно реалізувати бажане відношення сигнал/шум за умов наявності пауз у мовленнєвих сигналах.

Особливістю системи FaNT є те, що її програмні модулі написано на C++, тому після завантаження на ПК цих модулів, їх потрібно піддати компіляції та одержати файл `filter_add_noise.exe`. Обробку мовленнєвих і шумових сигналів виконують у режимі командного рядка, що не завжди зручно.

Дещо пізніше автори системи FaNT розвинули її, додавши можливість врахування впливу ревербераційної перешкоди (салон автомобіля, офісне чи житлове приміщення) на мовленнєвий сигнал, а також врахувавши спотворення мови кодеками мобільних ліній зв'язку. Для зручності користувачів керування новою версією системи FaNT можна здійснювати в режимі реального часу, через www-презентацію із простим та інтуїтивно зрозумілим графічним інтерфейсом [20].

Варто зазначити, що розроблення системи FaNT було здійснено із врахуванням глобального стандарту мобільного зв'язку GSM та європейського стандарту мобільного зв'язку UMTS (який інколи називають 3GSM з метою підкреслити приналежність технології до мереж 3G та його спадкоємність у розробках із мережами GSM),

використано бібліотеки програмних модулів, рекомендованих Міжнародним союзом телекомунікацій (International Telecommunication Union – ITU), а також програмні модулі мовленнєвого кодека AMR, стандартизовані Європейським інститутом телекомунікаційних стандартів та консорціумом 3rd Generation Partnership Project [21].

Програмний інструментарій для середовища Matlab

Для середовища Matlab написано доволі багато програмних модулів з обробки мовленнєвих сигналів й однією із найбільш вдалих бібліотек такого типу є VoiceBox [22]. Проте навіть у ній відсутні вказівки стосовно інструментарію, необхідного для швидкого розроблення потрібного мовленнєвого корпусу. Тому доцільно зробити кілька зауважень з приводу можливих шляхів здійснення такої розробки.

Базовими варто вважати записи мовленнєвих сигналів, виконані в заглушеному приміщенні із часом реверберації не більше 0,1 с. Запис має виконуватися за допомогою якісних мікрофона та звукової карти. Відношення сигнал/шум одержаного запису “чистого” мовленнєвого сигналу не має бути меншим за 40 дБ. Пікові значення записаного сигналу не мають перевищувати 12 дБ. Запис можна виконати як у середовищі Matlab за допомогою стандартної функції `wavrecord()`, так і з використанням звукових редакторів, які є в розпорядженні користувача. Довжина записів вибирається із врахуванням потреб дослідників (наприклад, в системі FaNT це 10 с, що є недостатнім при бажанні використати подовжені тестові сигнали).

Після запису дискретної мови доцільно створити звуковий файл із видаленими паузами для коректного визначення рівня мовленнєвого сигналу (крім того, файл із видаленими паузами знадобиться при оцінюванні розбірливості мови формантним методом, на виході системи, що тестується). Для процедури видалення пауз можна використати детектор голосової активності `vadsohn()` бібліотеки VoiceBox. Крім того, можна застосувати відповідну функцію звукових редакторів, яка є в розпорядженні користувача. Нарешті, видалення пауз можна виконати вручну в звукових редакторах, користуючись функціями виділення курсором і подальшого видалення виділених ділянок із паузами. Варто зауважити, що ручне видалення пауз якісніше, але й більш трудомістке, тому його слід використовувати при створенні невеликих за обсягом, проте, ду-

же якісних корпусів. Впевнитись у недостатньо високій якості автоматичного видалення пауз можна, аналізуючи рис. 2, де показано вихідний сигнал, що відповідає слову "листочок" (рис. 2, а), та результати видалення пауз ручним (рис. 2, б) і автоматичним (рис. 2, в) способами. Відносна різниця в довжині сигналів, одержаних після застосування процедур ручного й автоматичного видалення пауз, є близькою 20 %, а відносна різниця в оцінках ефективного рівня сигналу близька до 17 %. Одержані похибки значні, а їх причини доволі очевидні й полягають у помилковому видаленні при автоматичному способі пауз між складами. Зауважимо, що системи автоматичного видалення пауз зазвичай містять можливість регулювання параметрів процедури видалення пауз, тому наведений приклад можна

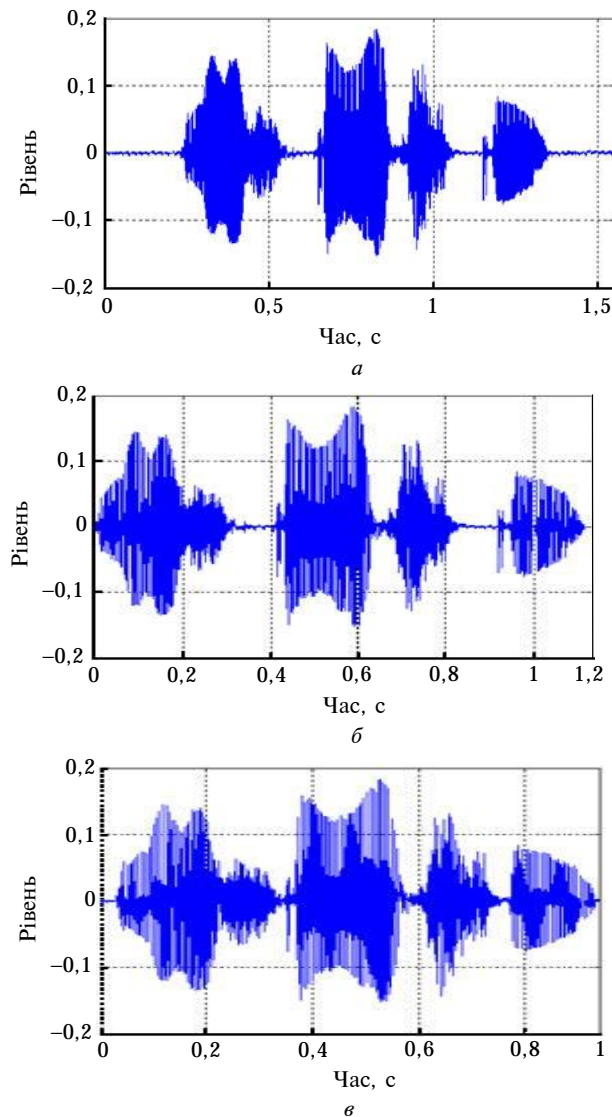


Рис. 2. Вихідний сигнал "листочок" (а) і результати видалення пауз ручним (б) та автоматичним (в) способами

тракувати як наслідки невдалого регулювання параметрів.

Генерування "первинного" штучного нормального білого шуму $n_1(t)$ виконується за допомогою стандартної функції `randn()`. Для одержання забарвлених шумів доведеться скористатися відповідними фільтрами. Зокрема, в праці [18] наведено *m*-коди програм для формування таких шумів із первинного білого шуму за допомогою гребінки октавних фільтрів. У бібліотеці VoiceBox також міститься різновид програми формування гребінки фільтрів (функція `gammabank()`), проте, вона менш зручна при формуванні забарвлених шумів.

Після обчислення дисперсії D_{n1} штучного або природного шуму $n_1(t)$ виконується розрахунок дисперсії D_x сигналу без пауз, після чого обчислюється наявне відношення сигнал/шум $\xi_1 = 10 \lg(D_x / D_{n1})$. Оскільки значення ξ_1 , як правило, не відповідає бажаному відношенню ξ_0 , варто обчислити коефіцієнт *k* корекції миттєвих значень шуму: $k = 10^{0.05(\xi_1 - \xi_0)}$.

Формування суміші із бажаним відношенням сигнал/шум ξ_0 виконується за формулою

$$y(t) = x(t) + kn_1(t).$$

Формування сумішей за формулами (2)–(3) вимагає виконання процедури згортки, для реалізації якої в бібліотеці Matlab є кілька стандартних функцій (`conv()`, `filter()`, `xcorr()`, `fftfilt()` тощо).

Висновки

Виконаний у статті огляд характеристик найбільш вживаних сучасних корпусів зашумленого мовлення дає можливість не тільки одержати уявлення про загальноприйняті у світовій практиці підходи до створення таких корпусів, але й зробити висновок про стан справ у створенні українських корпусів зашумленого мовлення. Хоч українські дослідники досягли значних успіхів у створенні мовленнєвих корпусів, робота із корпусом телефонної мови незакінчена. Крім того, що цей корпус не анотовано, після аналізу наявної інформації складається враження, що його не впорядковано за значеннями низки важливих параметрів, зокрема, за значеннями відношення сигнал/шум. Такий стан речей спонукає дослідників до пошуку шляхів самостійного формування корпусів зашумленого українського мовлення.

Для полегшення цієї справи в статті сформульовано принципи формування корпусів за-

шумленого мовлення із використанням узагальненої схеми процедури тестування систем обробки мовленнєвих сигналів. Правильність сформульованих принципів продемонстровано на прикладі програмного інструментарію FaNT, створеного та використаного для синтезу мовленнєвих корпусів Aurora-2 та Aurora-4.

Показано принципову можливість створення аналогічного (до інструментарію FaNT) власного інструментарію для роботи в середовищі Matlab, із застосуванням невеликої кількості програмних модулів сторонніх розробників.

Одержані результати будуть корисними для розробників математичного та програмного забезпечення систем корекції мовних сигналів, телекомунікаційних систем і систем АРМ, а також для студентів відповідних спеціальностей. Так, після створення кількох пробних корпусів зашумленого українського мовлення можна спільними зусиллями вибрати кращий варіант як стандартний.

У подальшому це дасть можливість порівнювати між собою алгоритми придушення завад та алгоритми автоматичного розпізнавання мови, запропоновані різними розробниками.

1. *Ладошко О.М.* Дослідження впливу характеристик телефонного каналу зв'язку на надійність розпізнавання фонем // Інфор. сист. управ. і комп. моніторинг: Зб. пр. міжнар. наук.-техн. конф. – Київ, 2012. – С. 308–313.
2. *S. Moller*, Quality of Telephone-Based Spoken Dialogue Systems. Boston: Springer Science+Business Media, Inc., 2005, 490 p.
3. *P. Moreno and R. Stern*, "Sources of Degradation of Speech Recognition in the Telephone Network", Proc. of the IEEE International Conf. on Acoustics, Speech, and Signal Processing, Adelaide, Australia, vol. I, pp. 109–112, April 1994.
4. *Кривнова О.Ф.* Речевые корпусы на новом технологическом витке // Речевые технол. – 2008. – № 2. – С. 13–23.
5. *Jankowski C. et al.*, "NTIMIT: A Phonetically Balanced, Continuous Speech, Telephone Bandwidth Speech Database", Proc. ICASSP-90, vol. 1, pp. 109–112, 1990.
6. *Corpora Group at CSLU* [Online]. Available: <http://www.cslu.ogi.edu/corpora/corpCurrent.html>
7. *H.-G. Hirsch*, The Aurora-5 Experimental Framework for the Performance Evaluation of Speech Recognition in Case of a Hands-free Speech Input in Noisy Environments [Online]. Available: <http://aurora.hsnr.de/background.html>
8. *The University at Texas at Dallas*. Speech Processing Lab. Noisy Speech Corpus [Online]. Available: <http://www.utdallas.edu/~loizou/speech/noizeus/>
9. *Центр Речевих Технологій* [Електронний ресурс]. – Режим доступу: <http://speechpro.ru/>
10. *Викторов А.Б., Викторова К.О., Воронцова А.В.* Речевые базы данных для задач автоматического распознавания речи и верификации говорящего // Сов. речевые технол.: Сб. тр. IX сессии Рос. акустич. общества. – 1999. – С. 5–15.
11. *Wideband Speech Database for Russian* [Online]. Available: <http://www.auditech.ru/page/widerband.html>
12. *SpeechDat-Car data base* [Online]. Available: <http://www.fee.vutbr.cz/SPEECHDAT-E/sample/russian.html>
13. *Сайт з розпізнавання та синтезу мовлення в Україні* [Електронний ресурс]. – Режим доступу: <http://www.speech.com.ua/developers.html>
14. *Центр Глобальних Повідомлень Україна – Global Message Services* [Електронний ресурс]. – Режим доступу: <http://www.gmsu.ua/>
15. *H.G. Hirsch and H. Finster*, "The Simulation of Realistic Acoustic Input Scenarios for Speech Recognition Systems", in 9th European Conf. on Speech Commun. and Technol., Lisboa, Portugal, September 2005, pp. 1–4.
16. *LDC Top Ten Corpora* [Online]. Available: <http://www.ldc.upenn.edu/Catalog/topten.jsp>
17. *Васильева Н.Б., Пилипенко В.В., Радуцкий А.М. и др.* Корпус украинской эфирной речи // Речевые технол. – 2012. – № 2. – С. 12–21.
18. *Продеус А.Н., Дидковский В.С., Дидковская М.В.* Акустическая экспертиза каналов речевой коммуникации: Монография. – К.: Имэкс-ЛТД, 2008. – 420 с.
19. *Продеус А.Н.* О некоторых особенностях развития объективных методов измерений разборчивости речи // Электрон. и связь. Тем. вып. Электрон. и нанотехнол. – 2010. – № 2. – С. 217–223.
20. *FaNT and the Calculation of the Signal-to-Noise-Ratio (SNR)* [Online]. Available: http://dnt.kr.hsnr.de/download/snr_comments.html
21. *Recommendation ITU-T P.56*. Series P: Terminals and Subjective and Objective Assessment Methods Objective measuring apparatus. Objective Measurement of Active Speech Level, Telecommun. Standartisation Sector of ITU, vol. 12, 24 p., 2011.
22. *VoiceBox: Speech Processing Toolbox for MATLAB* [Online]. Available: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html#file>