УДК 004.934.2+621.395 PACS 02.30.Nw, 02.60.Ed, 02.70.Hm, 05.45.Tp, 07.05.Tp

А. Ю. Небилиця

ЧАСТОТНА НОРМАЛІЗАЦІЯ СИГНАЛУ ПРИРОДНЬОЇ МОВИ В ЧАСОВІЙ ОБЛАСТІ

Запропоновано приведення фрагменту вхідного сигналу, який відповідає інтервалу основного тону, до вектора фіксованої довжини шляхом частотного еспандування. Обґрунтовано доцільність, встановлено спосіб і параметри транспозиції акустичного сигналу. Представлено алгоритм нормалізації сигналу та аналіз результатів досліджень. Встановлено обчислювальну складність методу та рівень його ефективності при виділенні інформаційних ознак природньої мови.

Ключові слова: *мовний людино-машинний інтерфейс, розпізнавання мови, мовний потік, сегмент основного тону, транспозиція акустичного сигналу.*

Вступ

Природня мова характеризується значною варіативністю акустичного сигналу, причому відмінності зумовлені як індивідуальним особливостями мовного тракту диктора, так і його емоційним станом, інтонаційними ознаками вимови. В кількісному вимірі такі відмінності проявляються у значній величині амплітудного діапазону 60 дБ та частотного діапазону основного тону 80..320 Гц , що становить чотири октави. Крім цього, зазвичай, на акустичний канал накладаються завади у вигляді респіраційних і реверберацийних звуків, фонових шумів. У сукупності такі чинники суттєво ускладнюють вирішення задачі розпізнавання мови, внаслідок чого достовірність ідентифікації мовних образів і на сьогодні не перевищує 80% [1]. Такий рівень надійності інтерфейсу людино-машинної взаємодії є недостатнім для його широкого використання [2].

Іншим стримуючим чинником поширення голосового управління є складність та громіздкість систем розпізнавання мови [3]. Складність проявляється у значних витратах обчислювальної потужності як на попередню обробку мовного сигналу, так і на аналіз мовного потоку та прийняття рішень щодо класифікації інформаційних ознак. Громіздкість проявляється в значних потребах оперативної та програмної пам’яті.

Часто можливо почути аргумент, що сучасна обчислювальна техніка наскільки потужна, що такий критерій як складність алгоритму не є критичним, а проведена оптимізація несуттєво впливає на функціональну ефективність програмного чи програмно-апаратного додатка. Хибність такого аргументу проявляється в тому, що в реаліях експлуатації важливо забезпечити не лише зручність використання, надійність та стабільність роботи додатка, а також досягти належного рівня опосередкованих показників якості, наприклад: автономності, компактності або мобільності. Такі показники тісно пов’язані із рівнем енергоспоживання, суттєве підвищення якого може бути зумовлено необхідністю витрат обчислювальної потужності на супутні процеси. Не варто забувати, що мовний людино-машинний інтерфейс є лише сервісом, а ніяк не основним функціоналом прикладної системи.

Таким чином, очевидною є потреба комплексного вирішення задач підвищення достовірності та спрощення алгоритмів розпізнавання природньої мови. В подальших викладках буде доведено, що таке вирішення можливе за рахунок обробки та аналізу сигналу мовного потоку на інтервалі основного тону.

Мета, актуальність та задачі досліджень

Метою досліджень є отримання способу нормалізації сигналу, який забезпечить мінімізацію спотворення інформаційних ознак природньої мови на етапі первинної обробки акустичних даних.

Актуальність роботи полягає у виявленні нових підходів попередньої обробки сигналу мовного потоку систем розпізнавання природньої мови.

Практична цінність роботи полягає у підвищенні ефективності використання людино-машинного інтерфейсу на базі природньої мови для автономних пристроїв, утилітарних роботів, систем та комплексів виробничого призначення.

Для досягнення мети роботи були вирішені наступні задачі:

- вибір інтервалу аналізу сигналу, який мінімізує спотворення параметрів мовного сигналу;
- визначення способу нормалізації сигналу в часовій області;
- розробка методології нормалізації сигналу для підсистем розпізнавання мови природнього людино-машинного інтерфейсу;
- синтез алгоритму частотного еспандування сигналу;
- встановлення рівня ефективності отриманого методу нормалізації.

Подальші викладки будуть проводитись у відповідності до зазначеної послідовності задач досліджень.

Вибір інтервалу аналізу сигналу мовного потоку

Найбільш змістовними інформаційними ознаками природньої мови для ідентифікації фонем є спектральні параметри [4]. Для їх отримання проводять перенесення сигналу із часової області в частотну шляхом застосування дискретного перетворення Фур'є (ДПФ). У випадку використання ДПФ для обробки сигналу на кінцевому інтервалі часу мають місце значні спотворення спектра, причому не має суттєвого значення факт накладання чи вид зважувального вікна. Так, прямокутне вікно мінімізує розширення інформаційних спектральних компонент, але вносить значний рівень шумових. Натомість, накладання на сигнал віконних функцій розширює інформаційні та мінімізує шумові компоненти спектра. Якщо величину відносного розширення головної пелюстки амплітудно-частотної характеристики (АЧХ) позначити як K, а відносний рівень висоти бокових пелюсток як γ, тоді властивості віконних функцій високої роздільної здатності за даними [5] можливо представити так: прямокутне вікно (без зважування) – $K = 1$, $\gamma = -13$ дБ; Бартлета (трикутне вікно) – $K = 2$, $\gamma = -26{,}5$ дБ; Ханна – $K = 2$, $\gamma = -31{,}5$ дБ; Хемінга – $K = 2$, $\gamma = -42$ дБ; Блекмана – $K = 2$, $\gamma = -58$ дБ. Слід відмітити, що використання зважувальних вікон, крім недоліку – значного розширення головної пелюстки, вносять суттєве ускладнення обчислень, наприклад, функції Ханна і Хемінга включають по одній операції множення та одній віднімання, а от функція Блекмана по дві операції множення та додавання, причому обрахунки в зазначених функціях проводяться з раціональними числами розширеної точності та потребують використання таблиць синусів і косинусів.

В роботі [4] детально викладено результати досліджень стосовно вибору віконної функції та інтервалу аналізу сигналу природньої мови. Однак, не з усіма висновками автора цієї роботи можливо цілковито погодитись. Так, мовний сигнал є квазіперіодичним лише на значних інтервалах спостереження, але в межах сегменту основного тону (ОТ) він затухаючий, тому накладання віконних функцій Бартлета, Ханна, Хемінга чи Блекмана спричинить ризики обрізання високочастотних компонент спектра. Найбільш спотворення такого виду будуть мати місце у випадку накладання вікна Бартлета [6]. Внаслідок зазначеного та з точки зору спрощення аналізу мовного сигналу бажано відмовитись від використання зважувальних вікон, а мінімізацію спотворення мовного спектру досягти в інший спосіб.

В даних дослідженнях мінімізація спектральні спотворення сигналу досягалась шляхом синхронізації процесу аналізу із слідуванням сегментів ОТ. Так, відомо, що зменшити спотворення спектру можливо у випадку, коли початок і кінець вибірки відповідають моменту часу проходження сигналу через ізолінію. У випадку акустичного сигналу мови найкраще цій умові відповідають сегменти ОТ мовного потоку. Рисунок 1 відображує два можливі підходи у виборі інтервалу для ДПФ сигналу.

Перший підхід полягає у виділення періоду T_{OT_A}, який знаходиться між моментами часу перетину ізолінії п.2 передніми фронтами перших пелюсток сусідніх сегментів основного тону сигналу п.1. Другий – позначений на рисунку 1 як $T_{OT\_Б}$ – являє собою проміжок між моментами часу перетину ізолінії сигналу спадаючими фронтами останніх в сегментах пелюсток. Оскільки швидкість наростання сигналу в другому підході значно менша, то і спотворення спектру сигналу на обмеженому інтервалі часу $T_{OT\_Б}$ будуть несуттєвими, а значить, прийнятними. З позицій цифрової обробки, зазначений підхід еквівалентний застосуванню синус-вікна без проведення математичних маніпуляцій. Такий висновок слідує з того, що компонента сигналу, яка відповідає частоті ОТ, сама низькочастотна, і тому має найменший рівень затухання і з цієї причини більшою мірою відповідає ознаці гармонічного сигналу. Вибір за початок відліку фазу спаду мовного сигналу надає можливість досягти зменшення розтікання спектру з величини -13дБ до -23дБ. В подальших викладках будуть приводитись результати спектральної обробки, отримані саме за другим підходом.

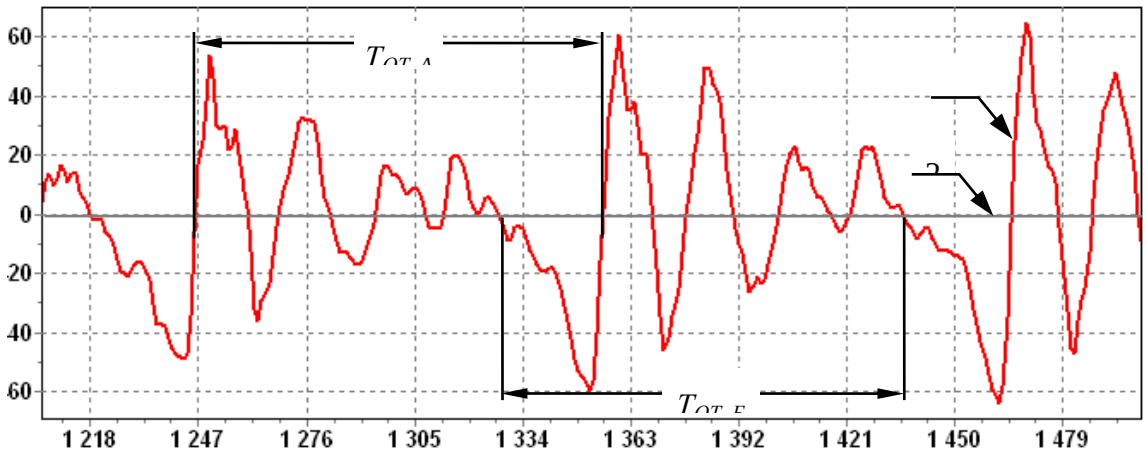

Рис. 1. Схема вибору меж інтервалу для ДПФ сигналу природньої мови.

Окремо слід пояснити вибір обробки сигналу за інтервалом сегменту ОТ, оскільки аргументація відрізняється від поданої в [4]. Для повноти викладок необхідно зазначити, що більшість підсистем розпізнавання мови базуються на аналізі значних обсягів даних, які включають десятки сегментів ОТ. Мотивація такого способу реалізації обробки полягає у бажанні уникнути операції виділення сегменту ОТ та отримати більшу роздільність спектральних ліній. Однак, такий спосіб обробки мовного сигналу вносить спотворення, які проявляються у вигляді розщеплення спектральних пелюсток, що зумовлено відхиленнями реального сигналу від ознак стаціонарності та ергодичності внаслідок поліфазності сегментів ОТ, яка спричинена такими акустичними ефектами вимови як вібрато та джитер-ефект. Поява у спектрі близько розміщених спектральних максимумів суттєво ускладнює процес ідентифікації фонем і тому робить спосіб аналізу мовного потоку шляхом розширення та перекриття вибірок малоефективним.

Транспозиція сигналу

У рамках даного дослідження основною із задач було визначення способу та параметрів частотної нормалізації сигналу в часовій області. Під терміном «нормалізація» розуміють приведення основних параметрів вхідних даних до еталонної величини. Для мовного сигналу такими параметрами є амплітуда (потужність) сигналу та частота основного тону. Нормалізація за амплітудою в підсистемах розпізнавання мови використовується частіше, прикладом реалізації може бути [7]. Нормалізацію за частотою в підсистемах розпізнавання мови, зазвичай, як операцію окремо не проводять, оскільки в частотній області її реалізація є тривіальною і відповідає операції зміщення початку частотної осі. Однак, як буде доведено нижче, така маніпуляція спектром не може бути ефективною в силу його спотворення внаслідок ДПФ. Вихід вбачається у здійсненні трансформації сигналу таким чином, щоб в результаті проведення ДПФ було отримано вже нормалізований спектр. Операції на кшталт перенесення довільного частотної смуги акустичного сигналу в іншу ділянку діапазону називають транспозицією сигналу. В підсистемах розпізнавання мови частотні маніпуляції, зазвичай, не проводять, але в програмних продуктах студійного запису така опція є типовою. Транспозицію можливо здійснювати у два способи.

Перший спосіб являє собою гетеродинний метод перетворення частоти, який реалізується перемноженням акустичного сигналу на косинус опорної частоти та виділенням необхідної смуги частот шляхом цифрової фільтрації. Проте, для транспозиції мовного сигналу даний спосіб використовується рідко. Причина полягає в тому, що відразу перенести спектр неможливо, оскільки основна частка спектру сигналу природньої мови має широку смугу 80 .. 8000 Гц, а виконуване зміщення частоти незначне – 10 .. 240 Гц. За таких умов буде мати місце накладання спектрів бокових смуг, внаслідок чого стає неможливою фільтрація результуючого сигналу. На практиці гетеродинне перетворення мовного сигналу можливе лише в двохстадійній реалізації. На першій стадії сигнал переноситься з основної на проміжну смугу частот, причому остання знаходиться вище за спектром природньої мови, а також проводиться ослаблення сигналу в межах основної смуги. На другій стадії повертають сигнал в основну смугу частот таким чином, що спектр сигналу є зміщеним відносно вхідного сигналу на встановлену величину. Опис гетеродинного методу перетворення частоти для випадку обробки мовного сигналу засвідчує його складність, а тому і неприйнятність.

Другий спосіб транспозиції реалізується шляхом передискретизації сигналу таким чином, щоб відтворення голосу на базовій частоті дискретизації відповідало необхідній тональності. В більшості випадків транспозиція голосу проводиться саме за другим способом.

Методи передискретизації добре розроблені [8] і надають можливість проводити як пониження частоти – децимацію, так і підвищення частоти – еспандування (інтерполяцію). Можливо здійснити передискретизацію і за допомогою прямого та зворотного перетворення Фур'є [9]. Однак, на практиці, як правило, передискретизацію виконують шляхом згортки сигналу з імпульсною характеристикою відновлювальних фільтрів, а саме: віконного sinc-фільтра, рівнохвильового фільтра Чебишева, лінійних та поліфазних інтерполюючих фільтрів, CIC-фільтрів. За типом коефіцієнта розділяють цілочисельне та дробове масштабування частоти квантування. Передискретизація із дробовим коефіцієнтом більш складна, але у випадку нормування сигналу природньої мови, в якому частота основного тону змінюється у відповідності інтонації на малу величину, є єдино можливою.

Частотне еспандування мовного сигналу

Як вже було зазначено, передискретизацію сигналу можливо здійснювати як із пониженням, так і з підвищенням частоти. У зв'язку з чим постає питання: а який підхід слід вибрати? Щоб аргументовано відповісти на дане питання слід виділити ряд обставин, які пов'язані із особливостями мовного потоку та специфікою цифрової обробки.

По-перше, апріорно частота ОТ, яка буде мати місце в акустичному сигналі мови в наступний момент часу, достеменно не відома, а визначається лише на момент локалізації сегменту ОТ. По-друге, передискретизація шляхом пониження та підвищення частоти виконуються по різному. Так, для пониження частоти необхідно спочатку провести низькочастотну фільтрацію, а лише потім здійснити проріджування сигналу, що зумовлено умовою уникнення аліасингу. Для операції еспандування сигналу, навпаки, спочатку проводять вставку необхідної кількості відліків, а потім проводять низькочастотну фільтрацію. При застосування поліфазних фільтрів вид передискретизації не є критичним, однак, в такому підході алгоритм значно ускладнюється і не підлягає оптимізації.

По-третє, як відомо, важливим показником попередньої обробки сигналу є мінімізація втрати його інформативності, зокрема – частотної роздільної здатності. Із теорії спектральної обробки відомо, що в результаті ДПФ спектр сигналу є дискретним, а ширина окремого спектрального відліку відповідає $\Delta f = f_\partial / N$, де f_∂ – частота дискретизації, а N – довжина вибірки. В підсистемах розпізнавання та кодування мови частота дискретизації визначається шириною спектру природньої мови та теоремою Котельнікова, і, як правило, відповідає діапазону 8 .. 16 кГц. Діапазон довжин вибірки також обмежений з огляду значних спектральних спотворень, зумовлених усередненням, відхилення реального мовного сигналу від умов стаціонарності та ергодичності. У випадку обмеження довжини вибірки вхідних даних інтервалом ОТ, в кількісному вимірі вона буде становити для чоловічих голосів 100 .. 256, а для жіночих – 40 .. 120 семплів. Величина частотної селективності прямо пропорційна N, тому для сигналу жіночого голосу вона буде значно нижчою.

Виходячи із зазначених обставин, пропонується нормалізацію проводити в часовій області шляхом експандування сигналу природньої мови таким чином, щоб довжина вхідного вектора була постійною величиною незалежно від тривалості сегменту ОТ. Оскільки довжина вибірки визначається виходячи із f_∂ та тривалості найбільшого із періодів сигналу, тому можливо вважати зазначений вид частотної нормалізації як операцію приведення тривалість сегменту ОТ до тривалості самого низькочастотного чоловічого голосу типу бас. Конкретне значення тривалості вибирається розробником підсистем розпізнавання голосу виходячи із частоти дискретизації та частотної роздільної здатності спектра (ширини вікна), так для $f_\partial = 16$ кГц, та N=256 частота нормування складе $f_н = 62{,}5$ Гц. В своїй сутті зазначений підхід не новий і представлений у роботах [10; 11], але результати цих досліджень стосувалися області телефонії і не є оптимальними для підсистем розпізнавання мови. З цієї причини пропонується проводити трансформацію часової осі шляхом використання поліноміальних фільтрів.

Приймемо, що із вектора $\boldsymbol{y}$ довжиною N_{in}, який отримано з частотою дискретизації f_{in}, необхідно отримати вектор $\boldsymbol{Y}$ довжиною N_{out}. Враховуючи, що обчислення коефіцієнтів інтерполяції a_k можливо здійснити на всіх ділянках інтервалу за виключенням першого та останнього, ефективною буде довжина $N_{in} - 2$, а вихідна частота $f_{out} = f_{in} \cdot N_{out} / (N_{in} - 2)$. Виходячи із того, що у випадку кубічної інтерполяції мінімізація похибки обчислень забезпечується лише на середній ділянці області, знаходження інтерполяційних значень $Y(t)$ буде проводиться ітераційно на поточному інтервалі $t \in \left[t_{j-1}, t_j \right]$ згідно схеми рис.2 за поліномом (1), який представляє собою поліноміальний фільтр третього порядку і є модифікованим фільтром Фарроу [12]:

$$Y(t) = a_0 + y(t) \cdot \left[a_1 + y(t) \cdot \left[a_2 + a_3 \cdot y(t) \right] \right] \qquad (1)$$

Коефіцієнти фільтра a_k знаходяться для кожної із поточно-активних ділянок інтервалу N_{in} за виразами:

$$\left.\begin{aligned} a_3 &= 1/2 \cdot \left[3(y(t_{i+1}) - y(t_{i-2})) + y(t_{i-1}) - y(t_i)\right]; \\ a_1 &= 1/2 \cdot \left[y(t_{i+1}) - y(t_{i-1})\right] + a_3 ; \\ a_2 &= y(t_{i+1}) - y(t_i) - a_1 - a_3 ;\ a_0 = y(t_i) . \end{aligned}\right\} \qquad (2)$$

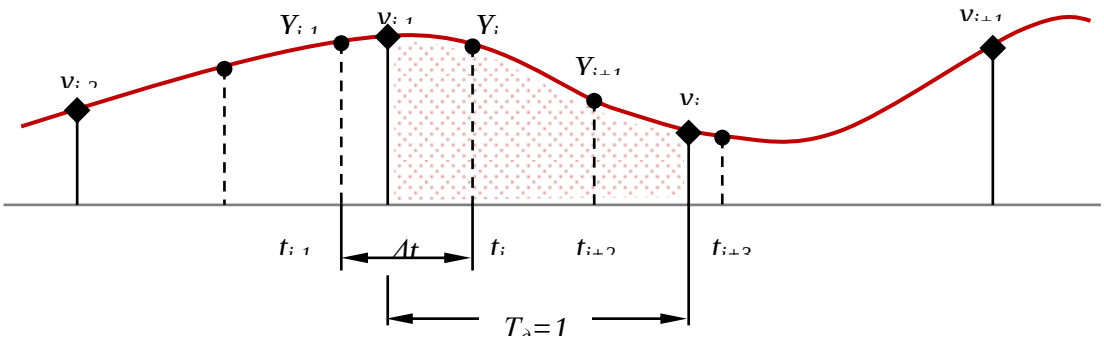

Рис. 2. Схема передискретизації шляхом застосування поліноміального фільтру.

Слід відмітити, що індексація у виразах знаходження a_k відрізняється від приведеної в [12], що зумовлено зміщенням системи координат задля оптимізації обчислень (див. рис.2).

Алгоритм передискретизації

Враховуючи, що моделювання проводилось в середовищі Mathcad 15, а графічне представлення коду засобами програмування даного середовища є достатньо наглядним, алгоритм передискретизації наведено функцією FarrowFilter(y, N) на рис.3. Першим аргументом функції є вказівник на вхідний вектор, а другим – довжина вихідного вектору.

Основною частиною алгоритму є цикл, в якому ітераційно обчислюються інтерполяційні значення за поліномом (1). У випадку, коли має місце вихід за поточний інтервал інтерполяції, проводиться обчислення коефіцієнтів КІХ-фільтра за формулами (2) та переноситься початок осі часу в наступну точку (дискретезація часу вхідного вектору нормована до 1). Вихід із інтерполяційного циклу здійснюється у випадку, коли поточною точкою стає остання точка вхідного вектора. Завершення виконання підпрограми передискретизації FarrowFilter(y, N) проводиться шляхом повернення вказівника на масив результату перетворення.

Ефективність методу нормалізації сигналу в часовій області

За критерії ефективності методу вибрано рівень спотворень спектру сигналу та обсяг обчислень. Дослідження з визначення величини спектральних спотворень проводилось у дві стадії – шляхом співставлення спектрів модельних сигналів та натурних сигналів звуків природньої мови. Синтез модельних сигналів проводився шляхом адитивного накладання декількох гармонічних сигналів. Реальні мовні сигнали фіксувалися стандартною програмою «Звукозапись» Windows XP SP3 шляхом використання мікрофону Canyon CNR-MIC2 та звукової карти Realtek HDA з режимом запису моно, 16 кГц, 8 біт. Для обох стадій обробка даних проводилась засобами середовища Mathcad 15.

```
FarrowFiltr(y,N) := | M ← last(y) − 1
                    | Δt ← M/N                                   – ініціалізація локальних змінних;
                    | j ← 1
                    | t ← 0
                    | Y_0 ← y_1
                    | for i ∈ 1..N − 1                           – цикл передискретизації;
                    |   t ← t + Δt
                    |   if t > 0                                 – перевірка виходу за межі інтервалу інтерполяції;
                    |     j ← j + 1
                    |     if j > M                               – перевірка виходу за межі діапазону;
                    |       Y_i ← a0 + t·[a1 + t·(a2 + a3·t)]
                    |       break
                    |     a3 ← 0.5·((y_{j+1} − y_{j−2})/3 + y_{j−1} − y_j)
                    |     a1 ← 0.5·(y_{j+1} − y_{j−1}) − a3      – обчислення коефіцієнтів інтерполяції;
                    |     a2 ← y_{j+1} − y_j − a1 − a3
                    |     a0 ← y_j                               – корекція на зміщення інтервалу;
                    |     t ← t − 1
                    |   Y_i ← a0 + t·[a1 + t·(a2 + a3·t)]        – поточне значення сигналу;
                    | Y                                          – повернення результату.
```

Сегменти ОТ визначались вручну з фонограм. В усіх дослідження ширина вікна аналізу складало 256 елементів, що відповідає ширині спектрального біна – $\Delta f = 62{,}5$ Гц.

В результаті проведених досліджень очікування зменшення спектральних спотворень знайшли підтвердження. Демонстрацією такого висновку може слугувати рисунок 4, на якому приведено спектр необробленого сигналу – фрагмент *а*, та спектр нормалізованого сигналу – фрагмент *б*. Враховуючи, що в обох випадках ширина вікна однакова і початковий сигнал той самий, то значні флуктації спектра фрагменту а порівняно із фрагментом *б* слід пов'язувати із впливом частини другого сегменту ОТ, оскільки в сигналі у випадку *а* на вибірку припадає 1,5 періоду ОТ. Близький до наведеного прикладу був стан спектрів і для

інших звуків. Для модельних сигналів відмінності були характерні меншою мірою, що пов'язано, на думку автора, з поліфазністю сегментів ОТ мовного сигналу.

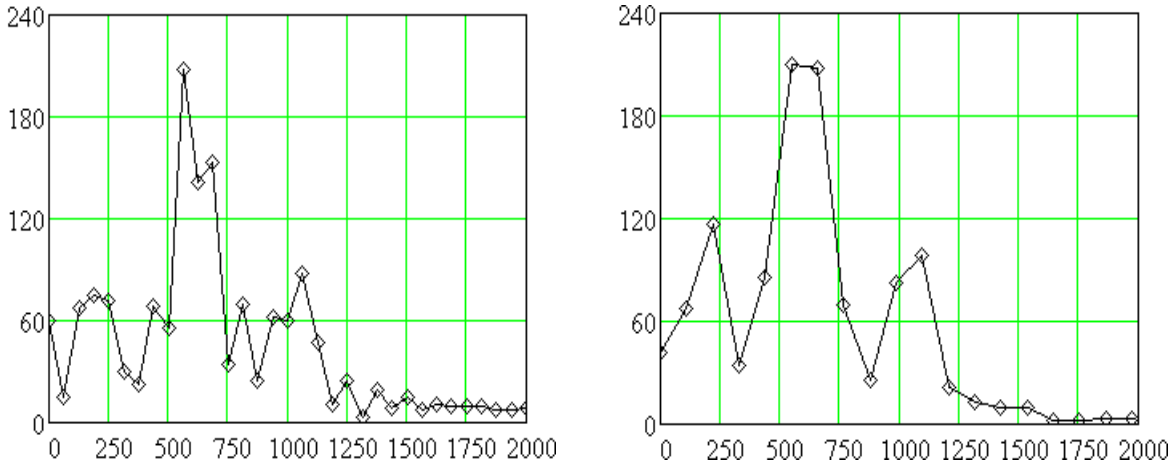

Рис. 4. Спектри мовного сигналу вимови звуку / а / – чоловічий голос.

Другим позитивом частотної нормалізації сигналу в часовій області стала закономірність, яка проявляється у значно меншому рівні варіативності форми спектру при зміщенні вікна аналізу. Виявлена закономірність пояснюється тим, що в межах сегментів нормований мовний сигнал є квазістаціонарним.

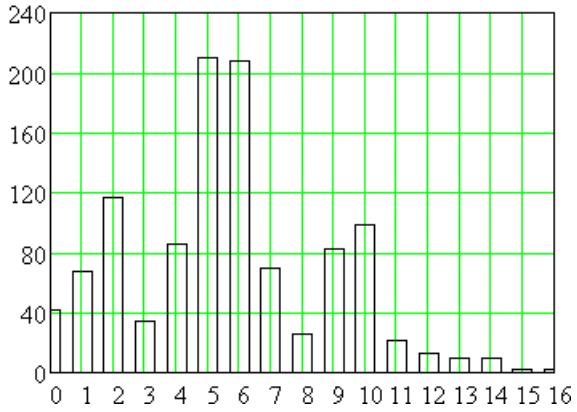

Рис. 5. Гармонічний склад сигналу вимови звуку / а /.

Третій позитив трансформації часової осі при спектральному аналізі мовних сигналів полягає у формі відображення результату обробки. Так, ДПФ «звичайного» сигналу дає спектр, в якому частотна вісь розбита в одиницях Гц. Частотне ж експандування зумовлює розбиття осі частот на гармоніки основного тону, тобто величина Δf приведена до параметрів мовного сигналу на поточний момент часу (див. рис. 5), а саме, $\Delta f = f_{OT}$. Така форма представлення спектру значно спрощує подальший аналіз даних та прийняття рішень на етапі ідентифікації фонем. Спрощення аналізу даних полягає в тому, що не має потреби проводити операцію згладжування спектру, для чого зазвичай використовують медіанний фільтр високих порядків. Використання таких фільтрів несе загрозу втрати інформативності даних, оскільки можуть мати місце маскування домінантних бінів, а це негативно позначається на достовірності розпізнавання мовних образів. З точки зору спрощення ідентифікації фонем слід відмітити таке. Викиди спектру в області першої та другої гармонік ОТ мають місце у всіх вокалізованих звуках мови, і тому не несуть корисну інформацію, однак можуть бути використані для приведення до норми амплітуд без усереднення по всьому спектру. З рисунку 4 видно, що звуку / а / голосу в області 5 та 6 гармонік зосереджують найбільшу енергію і є першою формантою даного звуку. Пелюстка спектру 10 гармоніки являють собою

другу форманту. Ніяких інших формант чи домінантних ознак у спектрі звуку / а /, який отримано шляхом ДПФ нормалізованого сигналу, немає, на відміну від даних [13], в якій голосні звуки представлені чотирьохформантною моделлю. З цієї причини на рис.5 спектр обмежено 16 гармонікою ОТ.

Стосовно обчислювальної складності варто зазначити наступне. Як відомо, мінімізація виконання швидкого ДПФ досягається у випадку, коли вибірка має фіксовану довжину і відповідає умові $N = 2^m$, де число два є основою базису перетворення. За такої умови досягається найкраща факторізація матриць і складність алгоритму становить $O(N \cdot \log_2 N)$. Така реалізація перетворення називається Radix2. У випадку, коли довжина вибірки має довільну довжину, необхідно використовувати алгоритм ДПФ з розчепленню основою $K = L \cdot M$, де $L = 2^n$, а M – довільне позитивне ціле число. Складність такого перетворення $O(L \cdot \log_2 L + K^2 / L)$, тобто тим менша, чим менше значення *M*, але значно більша за Radix2. Коли значення *K* несуттєво відрізняється від *N*, тоді можливо привести довжину вибірки шляхом додавання до вхідного вектора відліків з нульовими значеннями або обмежувати довжину. Обидва із зазначених підходів зумовлюють спотворення спектру вхідного сигналу, тому, коли відносна розбіжність *K* і *N* сягають більше 10%, необхідно виконувати ДПФ із розчепленню основою.

Період ОТ природньої мови апріорно невідомий, більш того, постійно змінюється, а тому ймовірність випадку $K = N$ прямує до нуля. За цих обставин перед розробником постає питання: який підхід застосувати – ДПФ з розчепленню основою чи провести транспозицію сигналу та використати Radix2? Відповідь однозначна: другий підхід кращий, оскільки складність запропонованої частотної нормалізації складає всього $O(N + K - 2)$, причому завжди $N > K$.

Висновки

Нормалізація сигналу природної мови саме в часовій області потребує мінімальних витрат обчислювальної потужності та оперативних ресурсів, натомість, мінімізує спотворення вхідних даних, зменшує втрати інформативності мовного потоку, чим забезпечує суттєве підвищення ефективності подальших етапів обробки і аналізу даних. Однозначність операції частотної нормалізації сигналу мови досягається тим, що така маніпуляція здійснюється завжди шляхом експандування, що помітно спрощує алгоритм її реалізації.

Представлення спектральних даних у базисі гармонік основного тону надає можливість покращити параметризацію сигналу, підвищити рівень ідентифікації фонем та здійснити перехід до дикторонезалежних підсистем розпізнавання мови нового покоління. При цьому локалізацію основного тону доцільно проводи за методикою [14].

Окремо слід відмітити, що отримані результати досліджень будуть корисними у сфері цифрової телефонії, оскільки можуть оптимізувати процес кодування мовного сигналу.

Список використаної літератури:

1. Ализар А.. Незаметная смерть распознавания речи. [Електронний ресурс]. – Режим доступу: https://geektimes.ru/post/92771.
2. Автоматизированная система распознавания интонации, подсчета слов и устойчивых словосочетаний в речи человека. [Електронний ресурс]. – Режим доступу: http://it-claim.ru/Persons/Semenova_Yana/speech.htm.

3. Гаюха А. А. Способы и особенности организации систем распознавания речи. System Analysis and Information Technology 18-th International Conference SAIT 2016. – Kyiv, Ukraine. – 2016.
4. Огородников А. Н. Выбор интервалов анализа сигнала при распознавании речи. Вестник Томского государственного университета. – 2003. – № 280. – С. 285-304.
5. Спектральный анализ на ограниченном интервале времени. Оконные функции. [Електронний ресурс]. – Режим доступу: http://www.dsplib.ru/content/win/win.html.
6. Шрюфер Э. Обработка сигналов: цифровая обработка дискретизированных сигналов. / Учебник // Под ред. проф. В.П. Бабака. - К.: Либідь. – 1995. – 320 с.
7. Авторское свидетельство RU 555411 Способ амплитудной нормализации сигналов и устройство для его осуществления. G06 K 9/00 // Б. В. Болотов, А. Н. Пивоваров, А. Д. Рябинин, Н. Я. Искренко, С. Н. Сотсков, Г. И. Куц. Опубл.25.04.77. Бюллитень №15.
8. Передискретизація. [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/%D0%9F%D0%B5%D1%80%D0%B5%D0%B4%D0%B8%D1%81%D0%BA%D1%80%D0%B5%D1%82%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%8F.
9. United States Patent US6549884 B1 Phase-vocoder pitch-shifting. Jean Laroche, Mark Dolson. G10L 19/14 Date of Patent: 15 апр 2003.
10. Азаров И. С. Изменение основного тона речевого сигнала в реальном масштабе времени. // Международная научно-техническая конференция, приуроченная к 50-летию МРТИ-БГУИР (Минск, 18-19 марта 2014 года) : Материалы конф. В 2 ч. Ч 1. – Минск. – 2014. – С. 274-275.
11. Вашкевич М. И. Передискретизация речевого сигнала, согласованная с частотой основного тона. // Международная научно-техническая конференция, приуроченная к 50-летию МРТИ-БГУИР (Минск, 18-19 марта 2014 года) : Материалы конф. В 2 ч. Ч 1. – Минск. – 2014. – С. 308-309.
12. Фильтры Фарроу на примере фильтра третьего порядка. Ресэмплинг сигналов. [Електронний ресурс]. – Режим доступу: http://www.dsplib.ru/content/farrow/farrow.html
13. Устойчивость оценок формантных частот / В.Н.Сорокин, А.С.Леонов, И.С.Макаров // Речевые технологии. – 2009. – №1. – С. 3-21.
14. Небилиця А. Ю. Оптимізація методу виділення мінімальних сегментів мовного потоку / А. Ю. Небилиця. – Вісник Черкаського університету: Серія «Прикладна математика. Інформатика», № 18 (311). – 2014. – С. 59-67.

References:

1. Alizar A.. Nezametnaya smert raspoznavaniya rechi. URL: https://geektimes.ru/post/92771.
2. Avtomatizirovannaya sistema raspoznavaniya intonatsii, podscheta slov i ustoychivyih slovosochetaniy v rechi cheloveka. URL: http://it-claim.ru/Persons/Semenova_Yana/speech.htm.
3. Gayuha A. A. (2016) Sposobyi i osobennosti organizatsii sistem raspoznavaniya rechi. System Analysis and Information Technology 18-th International Conference SAIT 2016. Kyiv, Ukraine.
4. Ogorodnikov A. N. (2003) Vyibor intervalov analiza signala pri raspoznavanii rechi. Vestnik Tomskogo gosudarstvennogo universiteta. no. 280, pp. 285-304.
5. Spektralnyiy analiz na ogranichennom intervale vremeni. Okonnyie funktsii. URL: http://www.dsplib.ru/content/win/win.html.
6. Shryufer E. (1995) Obrabotka signalov: tsifrovaya obrabotka diskretizirovannyih signalov. Pod red. prof. V.P. Babaka. - K.: LibId, 320 p.

7. Avtorskoe svidetelstvo RU 555411 Sposob amplitudnoy normalizatsii signalov i ustroystvo dlya ego osuschestvleniya. G06 K 9/00 // B. V. Bolotov, A. N. Pivovarov, A. D. Ryabinin, N. Ya. Iskrenko, S. N. Sotskov, G. I. Kuts. Opubl.25.04.77. Byulliten no. 15.
8. PerediskretizatsIya. URL: https://uk.wikipedia.org/ wiki/%D0%9F%D0%B5%D1%80%D0%B5%D0%B4%D0%B8%D1%81%D0%BA%D1%80%D0%B5%D1%82%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%8F
9. United States Patent US6549884 B1 Phase-vocoder pitch-shifting. Jean Laroche, Mark Dolson. G10L 19/14 Date of Patent: 15 апр 2003.
10. Azarov I. S. (2014) Izmenenie osnovnogo tona rechevogo signala v realnom masshtabe vremeni. // Mezhdunarodnaya nauchno-tehnicheskaya konferentsiya, priurochennaya k 50-letiyu MRTI-BGUIR (Minsk): Materialyi konf. vol. 2, iss. 1, pp. 274-275.
11. Vashkevich, M. I. (2014) Perediskretizatsiya rechevogo signala, soglasovannaya s chastotoy osnovnogo tona. // Mezhdunarodnaya nauchno-tehnicheskaya konferentsiya, priurochennaya k 50-letiyu MRTI-BGUIR (Minsk): Materialyi konf. vol. 2, iss. 1, pp. 308-309.
12. Filtryi Farrou na primere filtra tretego poryadka. Resempling signalov. URL: http://www.dsplib.ru/content/farrow/farrow.html
13. Sorokin, V. N., Leonov, A. S., Makarov, I. S. (2009) Ustoychivost otsenok formantnyih chastot. Rechevyie tehnologii. no. 1, pp. 3-21.
14. Nebilitsya, A. Yu. (2014) Optimizatsiya metodu vidilennya minimalnih segmentIv movnogo potoku, Visnik Cherkaskogo universitetu: SerIya “Prikladna matematika. Informatika”, no. 18 (311), pp. 59-67, Cherkasy, Ukraine.

Summary

A. Yu. Nebylytsia

FREQUENCY NORMALIZATION OF SPEECH SIGNAL IN THE TIME DOMAIN

Introduction

Low fidelity level, high dependence on a speaker and complicated computations are the key factors that restrain wide use of machine speech recognition.

Purpose

The purpose of the research is to obtain a signal normalization method that will minimize distortions of informational attributes of a speech series at the stage of initial processing of acoustic data.

Method.

The efficiency of frequency normalization was tested by mathematical modeling of digital processing on synthetic and natural speech signals.

Results

The research results suggest that discrete short-time Fourier transformation for a selection of samples longer than one period of a pitch segment and without synchronization with it is not invariant to a window shift and results in significant distortions to a speech signal spectrum.

Originality

The research experimentally proved that a frequency normalization of signal can be carried out in a time frame by means of bringing a pitch segment length to a fixed value fold to exponential function of 2^n. It was established that utilization of a modified Farrow filter for frequency

normalization by resampling is expedient. The findings provide for optimization to the algorithm of polynomial resampling of a speech signal.

Conclusion

Normalization of a speech signal in the time domain requires minimum use of computation capacities and operational resources. It minimizes the input distortion, reduces losses in speech stream information, and thus provides a substantial increase of efficiency on the further stages of data processing and analysis. The unambiguity of an operation of speech signal frequency normalization is achieved due to conducting such a manipulation always by means of expansion, which makes its implementation algorithm tangibly simpler. Presentation of spectral data in harmonics basis of a pitch tone allows to improve signal parametrization and phonemes identification, and move to a new generation of speaker-independent speech recognition subsystems.

Keywords: *voice HMI, speech recognition, speech series, pitch segment, pitch-shifting, resampling, Farrow filters, frequency transposition of speech signal.*

Стаття надійшла 12.02.2016
Прийнято до друку 19.02.2016