ширення його можливостей для підвищення ефективності роботи програми.

3. Дослідження залежності ефективності формального алгоритму від розкиду значень ваг та тривалостей виконання робіт вихідного графа показали незначне зниження ефективності із зростанням максимуму значень w та l, що задаються випадково (рис. 3,в). Це пояснюється ускладненням структури множин максимального пріоритету, які формуються під час оптимізації розкладу.

Діючий варіант алгоритму можна застосовувати для графів, які мають довільні надлишкові відношення передування, які припустимо ігнорувати, якщо вони не порушують хід алгоритму, і при видаленні яких формується еквівалентний послідовно-паралельний підграф, оптимальний розклад для якого є також оптимальним і для вихідного графа [2].

Дані статистичних досліджень дають змогу зробити висновок про те, що найкращих результатів застосування цього програмного продукту досягаємо на графах, що містять до 50 робіт, максимальні значення ваг та тривалостей яких не перевищують 60. Оптимальний розклад на графах, що генеруються випадково, з кількістю робіт близько 100 буде формуватись, головним чином, за рахунок вдосконаленого алгоритму, який реалізує "перестановки погіршення" та рекурсивний аналіз конструкцій.

1. *Танаев В.С., Гордон В.С., Шафранский Я.М.* Теория расписаний. Одностадийные системы. М., 1984.
2. Конструктивные полиномиальные алгоритмы решения индивидуальных задач из класса NP / А.А. *Павлов, А.Б. Литвин, Е.Б. Мисюра* и др. К., 1993.
3. *Pavlov A.A., Pavlova L.A.* About one subclass of polynomially solvable problems from class "Sequencing jobs to minimize total weighted completion time subject to precedence constraints"//Вестник международного Соломоновского университета. № 1. С. 109-116.

УДК 681. 84. 087. 4

МЕТОДИ МОДИФІКАЦІЇ ЧАСОВОГО МАСШТАБУ МОВНОГО СИГНАЛУ В СИСТЕМАХ АНАЛІЗУ-СИНТЕЗУ МОВИ

© Роман Марцишин, Юрій Рашкевич
НУ "Львівська політехніка", м. Львів, вул.С. Бандери, 12

В статті наведено огляд моделей представлення мови та методів модифікації часового масштабу мовного сигналу в системах аналізу-синтезу мови.

In this paper the browse of the models of speech and methods of time-scale modification of a speech in systems of analysis -synthesis of speech is presented.

Вступ

Методи часової модифікації мови використовуються в різноманітних випадках мовного спілкування: для нормалізації за темпом на етапі попередньої обробки при розпізнаванні, при забезпеченні необхідної швидкості відтворення мови в синтезаторах, для верифікації та ідентифікації дикторів, в системах кодування, при передаванні мовної інформації каналами зв'язку, для швидкого прослуховування та передачі мовних повідомлень в системах голосової пошти [1]. Методи перетворення часового масштабу мови також широко використовуються в багатьох задачах аналізу-синтезу мови та в інтерактивних системах керування, де вимагається постійний аналіз психофізіологічних параметрів людини, для забезпечення їй оптимальних умов роботи та надійного функціонування системи загалом. В статті коротко розглянуто перетворення часового масштабу мови в системах аналізу-синтезу. Детальніше описання деяких методів та алгоритмів наведено в монографії [1].

Моделі представлення мови

Відомо, що під час формування звуків мови утворюється сигнал збурення у вигляді потоку повітря, який проходить через голосові зв'язки, що коливаються (вокалізований звук) або ні (невокалізований звук). Потім сигнал збурення надходить через певної форми голосовий тракт на голосову щілину. Загальновизнана цифрова модель мовотворення, запропонована Рабінером та Шафером [2], де сигнал збурення для вокалізованих звуків - квазіперіодична послідовність імпульсів, сформованих генератором з періодом основного тону (ОТ), які проходять через фільтр з передатною характеристикою $G(z)$, - модель голосової щілини. Сигнал збурення для невокалізованих звуків - шум, який формується генератором випадкових чисел так, щоб отримати послідовність з рівномірним спектром. Голосовий тракт подано у вигляді лінійної системи з постійними параметрами (цифровий багатополюсний фільтр з передатною функцією $V(z)$) та моделі випромінювання (цифровий фільтр з передатною функцією $R(z)$), що описує поведінку потоку повітря біля губ. В більшості задач аналізу-синтезу (наприклад, лінійного передбачення) моделі голосового збурення тракту та випромінювання спрощуються до однієї з передатною функцією:

$$H(z) = G(z)V(z)R(z). \qquad (1)$$

Така модель мовотворення також є основою цілого ряду широко вживаних моделей мовних сигналів, описання та аналіз яких наведено далі.

Вищеописана модель має ряд обмежень, основними з яких є:

- параметри моделі постійні на інтервалі 10-20 мс (модель неточна щодо вибухових звуків),
- передатна функція $V(z)$ не має нулів, що важливо для представлення носових та фрикативних звуків,
- спрощений поділ збурення на невокалізоване (шум) та вокалізоване (тон), що не відповідає вокалізованим фрикативним звукам.

Практично всі розглянуті далі моделі аналізу чи синтезу мовного сигналу базуються на цифровій моделі мовотворення. Гнучке квазістаціонарне представлення вибірок мовної хвильової форми запропоноване в 1981р. в роботах Портноффа [3,4] де він використовує короткочасове перетворення Фур'є (КПФ) для модифікацій мовного сигналу. Згідно з ними вибірка мовної хвильової форми моделюється, як вихід змінного в часі лінійного фільтра, керованого сигналом збурення, що є або сумою вузькосмугових сигналів з гармонійно зв'язаними миттєвими частотами (вокалізована мова), або стаціонарною випадковою послідовністю з плоским спектром потужності (невокалізована мова).

Модифікація збурення загальної цифрової моделі мовотворення - змішана модель збурення, розроблена Гріффіном і Лімом [5], де спектр сигналу ділиться на смуги, кожна з яких оголошується вокалізованою чи невокалізованою. В роботі Мулінеса [6] пропонується змішана модель, в якій сигнал збурення для невокалізованих звуків утворюється проходженням шуму через фільтр із змінними в часі параметрами, а для вокалізованих - сигнал є сумою невокалізованих та вокалізованих компонент.

Кватієрі і Мак-Аулай запропонували в [7] систему аналізу-синтезу мови, де використовується математична модель мовного сигналу у вигляді лінійної комбінації K відповідно зважених синусоїдальних хвильових компонент, частоти яких не є гармонічно залежними. Для побудови такої моделі необхідне точне оцінювання амплітуди і фази сигналу збурення та моделі голосового тракту для кожної синусоїдальної складової. Оцінки отримуються використанням короткочасного перетворення Фур'є (КПФ), значення якого вимірюються на частотах локальних піків спектра. Оскільки частоти цих складових є гармонічно незалежними, то не потрібно визначати частоту основного тону (ОТ). Оцінки амплітуди і фази використовуються при синтезі для отримання залежностей зміни в часі кожного із параметрів. Згідно із поданою вище цифровою моделлю мовотворення мовний сигнал формується в результаті проходження сигналу збурення через голосовий тракт. Для побудови даної моделі в процесі аналізу спочатку на основі КПФ визначаються значення частот, амплітуди та фази, потім - розділяють компоненти. Дана синусоїдальна модель може бути використана і для описання невокалізованого мовного сигналу введенням випадкової фази [4]. Тоді цифрова модель мовотворення модифікується в більш загальну: сигнал збурення у вигляді імпульсів основного тону, або відліків білого шуму, представляється в загальнішому вигляді - K штук спеціально визначених амплітуд, частот і фаз. Треба відзначити, що хоч дана модель і не потребує точного визначення значення періоду основного тону, наближене значення ОТ використовується для визначення тривалості вікна аналізу вокалізованого сигналу.

Модель мовотворення з використанням елементарних хвильових форм (*ЕХФ*) подібна до паралельної формантної моделі та до синусоїдальної моделі. Модель оперує двома основними типами елементарних хвильових форм: формантними (у високочастотному регіоні) та синусоїдальними (в низькочастотному регіоні, який охоплює смугу до другої форманти). У формантних типах *ЕХФ* розглядаються як незалежні спектротемпоральні акустичні категорії, інакше обробляється глотальний потік (збурення) та використовується часова область перетворень. Формантні типи *ЕХФ* використовуються для представлення невокалізованої або змішаної мови через визначення окремого збурення для кожної форманти (тобто можливо синтезувати періодичні, випадкові та

змішані сигнали). Синусоїдальні типи *ЕХФ* відрізняються від стандартної синусоїдальної моделі тим, що оперують синусоїдальним представленням тільки в низькочастотному регіоні спектра та для них не важлива поведінка синусоїдальних компонент частотних каналів, для забезпечення параметра інтерполяції фреймів [8].

Використовується також представлення мовного сигналу у вигляді суми періодичної та аперіодичної складових [9, 10]. Аперіодична складова (наявна навіть у дуже виражених вокалізованих звуках) описується сумою елементарних хвиль, точно локалізованих у певних частинах спектральної області.

Останнім часом з'явилася робота [11], де мовний сигнал представлений сумою синусоїдальних складових і залишку, в якості якого є надалі розділені перепади (піки) та шумові компоненти. Залишок розраховується методом аналіз-через-синтез та подальше розділення відбувається через дискретне хвильове представлення (ДХП). Оскільки використовується процедура гармонічного аналізу Томпсона та перевірка на статистичну значимість кожної синусоїди, то автори підкреслюють позитивний ефект такої моделі, особливо в комплексних та багатотонових сигналах.

Всі методи аналізу-синтезу можна розділити на дві групи щодо моделей представлення мови. Методи, де параметри моделі неявно оцінені, називаються непараметричними. Параметричні методи вимагають явно оцінити параметри моделі та використати їх на етапах аналізу та синтезу.

Непараметричні методи

Підхід до перетворення часової структури мовного сигналу на основі систем аналізу-синтезу відомий вже з 1939 року та запропонований Дадлі [12]. Підхід неявно використовував цифрову модель мовотворення та ряд спрощень при аналізі-синтезі. Зміну частотної шкали можна було досягнути "простим масштабуванням центральних частот смугових фільтрів". Фазовий вокодер - інакший підхід до аналізу-синтезу, що базується на моделі КПФ. Представимо комплексні операції через дійсні і для одного каналу отримаємо:

$$\mathrm{Re}\left(P_k y_k(n)\right) = \left|P_k\right| \left|X_n\left(\exp\left(j\omega_k\right)\right)\right| \cos\left(\omega_k n + \theta_n\left(\omega_k\right) + \gamma_k\right), \qquad (2)$$

де P_k - множник, що дорівнює 1 якщо канал входить в підсумовування, w_k - частота сигналу, $\theta_n(w_k)$ - фаза сигналу, γ_k - фазова константа (для досягнення максимально плоскої загальної характеристики). Недолік канального вокодера - значні спотворення та низька розбірливість, що викликаються зміною формантних частот та інтерференцією суміжних частотних смуг.

В 1966 році ускладнений підхід до перетворення часу звучання мови також з використанням модифікацій в частотній області запропонований Фланаганом та Голденом [13] як варіант використання можливостей фазового вокодера. Підхід передбачав використання лінійного стиснення-розтягнення частот та фазових похідних спектра з відновленням спектральних спотворень, використанням відповідної зміни частоти дискретизації під час аналізу-синтезу чи зміни швидкості запису-відтворення мовного сигналу.

Однак було визнано, що модифікації масштабу часу (МЧ) мовного сигналу, які базуються на фазовому вокодері, дали хороші суб'єктивні результати, за винятком

ефекту змінного рівня реверберації або ефекту хорус у вихідному сигналі, особливо для великих коефіцієнтів модифікації (більше за 2).

Часова модифікація фазовим вокодером досягається через часове масштабування контуру тону мови і системної амплітудної функції, та за умови, що фази гармонік тону вільно змінюються. В результаті цього темпоральний аспект модифікованого сигналу значно відрізняється від початкового сигналу, бо не зберігаються локальні фазові відношення між тоновими гармоніками. У методі Силвестре [14] фаза у кожному КПФ-каналі не забезпечує вільної зміни, за винятком переустановлення в кожному часовому моменті синтезу. В результаті фазова неперервність між двома послідовними короткочасними сигналами синтезу гарантується недовго. Для відновлення неперервності фази до кожного каналу додається фіксоване фазове зміщення і залишок перерваної фази точно відновлюється через невелику модифікацією миттєвої частоти в кожному КПФ-каналі.

Цей алгоритм гарантує, що фазові відношення між гармоніками тону в околі часових моментів синтезу такі самі, як і в початковому сигналі; в околі часових моментів аналізу - аж до лінійного зсуву фази. Метод вносить деяке покращання якості сигналу порівняно з фазовим вокодером щодо проблеми небажаного ефекту реверберації-хорусу, однак були деякі спотворення для низькотонових голосів (чоловічих) та при прискоренні була тенденція до їх маскування, а при сповільненні спостерігались вібрації і вокалізованих порціях та розмитість невокалізованих порцій. При $0.5 < \beta < 2.0$ автори відзначали добру якість сигналу для чоловічого голосу.

Методи, що базуються на фазовому вокодері, не використовують кусково-постійної моделі представлення мови, а просто оперують з сигналом, який складається з синусоїд із частотами, достатньо далеко віддаленими, щоб бути розчиненими через вікно аналізу. Як результат - вони не відмовляють навіть у випадку одночасної розмови декількох дикторів чи інших дуже складних сигналів (наприклад, музика) [15] .

В кінці 70-х та на початку 80-х років як результат розвитку теорії КПФ з'являється ряд методів та алгоритмів перетворення часової структури мови. Один з перших алгоритмів розроблений Малахом [16] для передачі мови каналами зв'язку з використанням вокодерних систем аналізу-синтезу та часової нормалізації мови в системах розпізнавання. Алгоритм детально розглянуто в [1]. Алгоритм - лінійний та за своєю суттю близький до канального вокодера, однак не параметричний. Отже, є підстави вважати, що йому властиві такі ж межі зміни коефіцієнта трансформації часу та такі самі спотворення.

Найповніше теорія КПФ описана в роботах Портноффа [3,4], де використовуються модель гармонічного представлення мови на базі КПФ та цифрова модель мовотворення.

Послідовність операцій для виконання перетворення часового масштабу розділиться на три етапи:

- аналіз відрізку мовного сигналу на інтервалі часу T_1, та визначення параметрів,
- модифікація параметрів сигналу,
- синтез вихідного сигналу на інтервалі T_2.

Реалізація формальних перетворень під час часового масштабування мови вимагає надзвичайно великої кількості обчислень, які необхідно виконувати в реальному

часі. Крім того, розділення фазових компонент є дуже чутливим до похибок обчислення, і роздільна модифікація фазових складових часто є причиною спотворень у мовному сигналі.

Зміна темпу мови з використанням КПФ в нашій країні вперше розглядалася Гнатівим та Рашкевичем в [17].

В [1] розроблено спрощений підхід та алгоритм модифікації КПФ, які використовуються при розробленні систем аналізу-синтезу мови на основі КПФ для використання її в задачах регулювання темпу мови :

- Аналіз, перетворення та синтез невокалізованих ділянок мовного сигналу здійснювалися аналогічно до перетворень вокалізованої мови. Це дало б змогу з одного боку, уникнути необхідності розділення початкового сигналу без втрати якості синтезованого.
- Розділення фазових компонент, як в [3], не проводилося, а модифікувалося лише основне значення фази, обчислене на основі дійсної та уявної частин КПФ.
- Як аналізуючий і як синтезуючий фільтр використовувалась вагова функція Хеммінга, що викликало необхідність чотирикратного перекривання відрізків мовного сигналу в процесах аналізу та синтезу.

В результаті у алгоритм регулювання темпу мови [1] входять такі процедури:

1. Виділення в момент часу n відрізку мовного сигналу $x(m)$ за допомогою вагової функції Хеммінга $h(m)$ довжина якої N відліків. Відстань між послідовними моментами часу n_i дорівнює $N/4$.
2. Обчислення короткочасного перетворення Фур'є $X(n,k)$ та виділення його модуля $A(n,k)$ та фази $\theta(n,k)$.
3. Модифікація КПФ зміною кількості відліків $A(n,k)$ обернено пропорційно до коефіцієнта регулювання темпу мови β та модифікація $\theta(n,k)$, діленням кожної частотної складової на β. Спосіб виконання модифікації $A(n,k)$ залежить від значення β:
   - якщо β є цілим числом, то кількість відліків КПФ збільшується за допомогою інтерполяції, а зменшується - проріджанням;
   - якщо β - дробове число виду $\beta = \tau/u$, де τ і u - цілі числа, то спочатку виконується інтерполяція в u разів, а пізніше - проріджання в τ разів;
   - якщо β - довільне раціональне число, то можливою є апроксимація обвідної спектра поліномом високого порядку з наступним квантуванням його з відповідно зміненою частотою.
2. Синтез на основі модифікованого КПФ відрізку вихідного сигналу з кількістю відліків N/β.
4. Перекриття з підсумовуванням $3N/4\beta$ отриманих відрізків і формування вихідного сигналу.

Метод та алгоритм може бути використаний для коефіцієнтів регулювання темпу мови близько 2,5. Недоліком підходу є великі обчислювальні затрати і чутливість до похибок, особливо при обчисленні фази.

Відомий також алгоритм модифікації на основі спрощеного підходу до модифікації КПФ Мулінеса [18], який базується на відсутності розділення фази та моделі гармонічного представлення КПФ. Аналіз КПФ здійснюється аналогічно [3,4] та модифікація

і синтез за таким алгоритмом:

1. Встановлення початкових миттєвих фаз. Обчислення КПФ в наступний момент часу аналізу $t_a(u)$ і миттєвої частоти в кожному каналі.
2. Обчислення наступних модифікованих моментів синтезу $t_s(u) = [F(t_a(u)]$ і відповідної миттєвої фази.
3. Відновлення модифікованих КПФ у часі $t_s(u)$.
4. Обчислити u-й короткочасний модифікований сигнал, використовуючи модифіковану формулу синтезу перекриттям з підсумовуванням і повернення до кроку 2.

Відомо, що модифіковане КПФ (МКПФ) не обов'язково відповідає сигналу, що є в часовій області. Фазова модифікація, притаманна модифікаціям часової шкали, не зберігає фазової когерентності, що існує у послідовних оригінальних короткочасних спектрах. Для уникнення цієї проблеми було запропоновано [19] відкинути фазову інформацію у короткочасному спектрі, і щоб реконструювати МКПФ знаючи тільки модуль, використовуючи велику надлишковість даних в КПФ для компенсації втрат інформації. Ця ідея привела до ітеративних алгоритмів, що сходяться до локального мінімуму відстані. Незважаючи на те, що збіжність цих алгоритмів може бути доведена в деяких випадках, не завжди досягається глобальний мінімум. Ці ітеративні методи реконструкції застосовуються в алгоритмах модифікації часового масштабу з використанням ітеративної схеми синтезу та у покращеному алгоритмі [20], де вперше запропонована нова схема синтезу - синтез синхронізованим перекриттям з підсумовуванням (СПП). Алгоритми базуються на обчисленні модуля КПФ вхідного сигналу з перекриттям S_1 та ітеративним синтезом КПФ вихідного сигналу з перекриттям S_2 з використанням в першому випадку схеми ПП, а в другому СПП. Коефіцієнт трансформації часового масштабу для першого алгоритму визначався як $\beta = S_1/S_2$, а для другого β змінювався нелінійно з врахуванням максимального значення кореляції ділянок сигналу перед ПП-синтезом. Фаза в даному алгоритмі непрямо визначається з КПФ без модифікації. Тобто якість вихідного сигналу вища від алгоритму Портноффа [4], і особливо зменшується ефект реверберації-хорусу. Як вже відзначалось, однак, збіжність є звичайно досить повільною, що вимагає великої кількості ітерацій.

Для короткочасового кепстрального перетворення (ККП) розглянута методика [21,22] синтезу мовного сигналу з модифікованого ККП на основі ітераційної процедури та алгоритм модифікації, який використовує розроблену в [23] методику та алгоритм модифікації кепстру. Непараметрична методика перетворення часової структури на базі короткочасового кепстрального перетворення забезпечує ітераційне обчислення синтезованого сигналу через мінімізацію кепстральної відстані між аналізованим та модифікованим короткочасовим кепстральним перетворенням, без обчислень фази та основного тону сигналу. Розроблений алгоритм модифікації з використанням модифікованого короткочасового кепстрального перетворення дає змогу при кількості ітерацій синтезу, вдвічі меншій від відомих алгоритмів, виконувати перетворення часового масштабу мовного сигналу з коефіцієнтом 2 при розбірливості, не меншій за 88%. Синхронізація за вхідним сигналом дає змогу за рахунок знаходження подібної ділянки на розширеному інтервалі пошуку забезпечити мінімізацію спотворень та підвищити якість перетвореного сигналу.

Параметричні методи

Яскравими представниками параметричних методів є вокодери, які базуються на цифровій моделі мовотворення та синтезують вихідний мовний сигнал з використанням її параметрів. Розглянемо основні види параметричних вокодерів, оскільки решта вокодерів - їх різновиди та модифікації, запропоновані для покращання розбірливості мови та зменшення кількості інформації, яка передається синтезатору для утворення мови.

Найбільш відкритим для параметричного підходу є лінійний передбачувальний синтезатор [24], в якому вокалізована мова модифікується згорткою серії періодичних імпульсів із змінним в часі фільтром. Метод лінійного передбачувального кодування (ЛПК) - можливо, найрозповсюдженіша модель, що використовується для оброблення сигналу мови. В ЛПК-структурі повільна змінна огинаюча спектра мовного сигналу оцінюється обчисленням коефіцієнтів всеполюсного лінійного фільтра в короткочасних фреймах, сконцентрованих навколо часових моментів аналізу (типовий розмір фрейму 20-30 мс, використовуються 10-20 коефіцієнтів, що залежать від ширини смуги мови). Декілька методів оцінки описані в літературі: стандартний автокореляційний метод, що частіше використовується через простоту побудови (не перевіряється стійкість синтезуючого фільтра на противагу коваріаційній процедурі оцінювання). При використанні цього напрямку для ЛПК-моделей вхідний сигнал визначений зворотною фільтрацією. Для уникнення проблем в сегментах, де коефіцієнти фільтра швидко змінюються, стандартно використовується інтерполяція коефіцієнтів фільтра. Ця інтерполяція може виконуватися або безпосередньо на коефіцієнтах передбачення (треба проявляти деяку обережність для уникнення нестабільності), або на коефіцієнтах "трансформації", таких, як коефіцієнти відбиття сигналу чи коефіцієнти площі голосового тракту.

Синтез в ЛПК-синтезаторі здійснюється згідно з формулою:

$$s(n) = \sum_{k=1}^{M} \alpha_k s(n-k) + Gu(n), \qquad (3)$$

де M - кількість коефіцієнтів фільтра; α_k - значення коефіцієнтів фільтра; $u(n)$ - сигнал збурення (тон чи шум); G - енергія сигналу збурення.

В [25] показано спрощений підхід до перетворення часового масштабу на базі ЛПК-вокодера. Перетворення здійснюють, аналізуючи параметри на одному інтервалі часу та синтезуючи на іншому інтервалі часу, без розгляду схеми синтезу, яка важлива для розбірливості та якості модифікованого сигналу.

Проблеми виникають із стандартним ЛПК при обробці жіночих голосів з високим тоном, тому що фільтр ЛПК має тенденцію до моделі індивідуальної гармоніки тону краще ніж до загальної огинаючої. В цьому випадку може використовуватися альтернативний метод оцінки параметра, як наприклад" дискретне всеполюсне моделювання", запропоноване Ельджаруді [26], з метою пристосування ЛПК огинаючої безпосередньо до гармоніки тону, при накладенні додаткових обмежень гладкості, щоб запобігти перемоделюванню.

В [27] також описується гомоморфна модель регулювання темпу мови, яка допускає зміну часового масштабу мови відповідно зміною кепстру та синтезом на

зміненому інтервалі часу. В моделі розглядається дійсний кепстр, але не вказується явно схема синтезу. Авторами відзначається краща якість сигналу відносно часового методу вибіркової сегментації та ширший діапазон зміни темпу. Відзначимо, що робота [27] - перший підхід до перетворення часового масштабу за допомогою операцій з кепстром сигналу.

Гомоморфний вокодер, на відміну від ЛПК, дає змогу оцінити, як параметри мовного тракту, так і параметри збурення. Для вокодера приймається, що мовний сигнал є згорткою (в часі) імпульсної характеристики голосового тракту та функції збурення. Оскільки згортка рівноцінна множенню в частотній області, то логарифм спектра мовного сигналу дорівнює сумі логарифмів спектрів збурення та вокального тракту. Кепстральні коефіцієнти сигналу обчислюються через зворотне перетворення Фур'є логарифма спектра сигналу мови.

Система аналізу-синтезу на основі дискретної згортки-розгортки [28] подана Сенефф, де перетворення часового масштабу виконується подібно до фазового вокодера, однак з використанням досить складної процедури "розгортування" фазового спектра. Дискретне згортування-розгортування - один з перших підходів до отримання сигналів збурення та спектральної огинаючої з використанням дискретного перетворення Фур'є (ДПФ), що зараховує його до параметричних методів. Система використовувалась з коефіцієнтом часової трансформації $\beta<2$.

Синусоїдальна модель, яка були запропонована незалежно Альмейдою [29] та незалежно Мак-Аулау та Кватієрі [8], є надійнішим підходом до перетворень мовного сигналу. Параметри моделі отримуються використанням КПФ разом з відповідною процедурою пошуку та відслідковування піків. Рекомендованою процедурою розділення параметрів збурення та тракту є гомоморфна фільтрація. Причому визначається лише модуль частотної характеристики голосового тракту, а фаза визначається із властивості, за якою логарифм амплітуди і фаза - пара перетворень Гільберта. В результаті фаза згладжена та нерозривна, що є дуже важливим для подальших модифікацій.

Для синтезу складових хвиль спочатку необхідне визначення-розмічення вкладів збурення і тракту в кожну із хвиль, яке здійснюється на послідовних границях ділянок. Далі вздовж ділянки проводиться інтерполяція всіх параметрів збурення і тракту. Наприкінці процедури хвилі синтезуються на базі інтерпольованих компонент. Для додаткових складових амплітуд збурення і тракту а також фази тракту використовується лінійна інтерполяція при цьому при виконанні інтерполяції фази існує додаткове обмеження - фаза повинна бути гладкою і нерозривною.

Оскільки фаза збурення КПФ визначається за модулем 2π, то вона має розриви, тобто потрібно виконувати "вирівнювання фази". Автори пропонують кубічну інтерполяцію. Отримана фаза вирішує проблему вирівнювання та відповідає граничним вимогам. Далі відбувається декомпозиція фази збурення на сталу та змінну в часі складові.

Часове масштабування може здійснюватися як при постійному коефіцієнті β, так і змінне в часі, коли β є функцією часу. Для часового масштабування необхідна модифікація параметрів - амплітуди $M(t,\omega)$ та фази $F(t,\omega)$ голосового тракту, а також амплітуд $a_i(t)$ і частот $\omega_i(t)$ кожної із виділених хвильових компонент. Модифікація параметрів тракту відображає процеси прискорення-сповільнення роботи артикуляторів, модифікація параметрів збурення приводить до розширення-скорочення частотних

траєкторій при збереженні частоти ОТ. Часове масштабування при змінному (залежному від часу) коефіцієнті реально зводиться до випадку сталого коефіцієнта, оскільки приймається, що на конкретній ділянці коефіцієнт не міняється, а лише змінюється від ділянки до ділянки.

В 1992 р. цими ж авторами запропонований новий метод часового перетворення мови на базі синусоїдального представлення [7]. Його особливістю є збереження темпоральної структури початкового сигналу (інваріантність форми сигналу до перетворень) а також збереження початкових фазових співвідношень поміж хвильовими компонентами.

Синусоїдальне представлення початкового сигналу є аналогічним попередньому методу. Параметри хвиль оцінюються в моменти початків ділянок $m = 0, Q, 2Q$ ..., де Q - кількість відліків у ділянці. Рекомендується довжина ділянки 10 мс і тривалість вікна аналізу - 2.5 періоду ОТ. Вікно розташовується симетрично до центру ділянки, КПФ рахується на 1024 точках. Отримані амплітуди і фази хвиль в центрі ділянок повинні інтерполюватися (амплітуди - лінійно, фази - кубічною інтерполяцією) вздовж границь ділянок. Для цього спочатку виміряні частоти $\omega_k(m)$ асоціюються з частотами, отриманими на попередній ділянці. Допускається, що деякі частоти "помирають", а відповідно деякі - "народжуються".

При часовому масштабуванні початковий час t_0 трансформується в новий t'_0 за допомогою лінійного перетворення $t'_0 = \beta t_0$.

У даній спрощеній моделі зміна часу артикуляції відбувається масштабуванням системних амплітуд і фази. Параметри збурення повинні модифікуватися так, щоб змінити частотні траєкторії, не змінюючи частоту ОТ. Однак масштабування фази змінює ОТ. Перетворення фази зберігає ОТ, але приводить до появи хвильових дисперсій, оскільки змінює фазові співвідношення між хвилями. За новим підходом сигнал збурення з'являється в моменти розташування імпульсів ОТ, тобто моменти "запуску" сигналу збурення пересуваються в часі відносно початкової шкали часу. Це нововведення - принципова відмінність від запропонованого раніше методу, коли "запуск" збурення здійснювався на початку ділянки.

Масштабування часу відбувається аналогічно попередньому методу із змінним в часі коефіцієнтом β та полягає в зміні коефіцієнта від ділянки до ділянки, а синтез відбувається за звичайною схемою, також аналогічно попередньому методу.

У методі пропонується адаптивне до змісту сигналу масштабування, введенням ймовірності вокалізованості, що може набирати значення в інтервалі між 0 та 1 залежно від ступеня вокалізованості на ділянці мови. Перетворюються тільки вокалізовані ділянки, при цьому ступінь зміни тривалості ділянки залежить від значення ймовірності вокалізованості. Авторами відзначається спрощеність підходу, хоч навіть такий крок дає змогу підвищити розбірливість мови.

До параметричних методів також належить ряд методів часового масштабування мови на основі моделі елементарних хвильових форм (*ЕХФ*) Лінарда та Алессандро [30]. Метод часо- частотної модифікації мовного сигналу був запропонований Алессандро [31]. Перетворення часового масштабу розглядається як додаток до використання системи аналізу-синтезу. В системі розділяються періодичні та неперіодичні компоненти через ітеративну процедуру відновлення сигналу. Головними кроками аналізу є:

- виділення компонент збурення та огинаючої спектра, використанням лінійного передбачувального (ЛП) аналізу на коротких фреймах аналізу;
- ідентифікація періодичних та неперіодичних компонент в частотній області для кожного фрейму, причому періодичні компоненти шукаються як в спектрі, так і в кепстрі, а спектр розділяється на періодичну область (область гармонік) та неперіодичну (область формант);
- далі ітеративним алгоритмом будуються обидві компоненти через використання екстраполяційного алгоритму на базі ДПФ;
- з обох компонент визначаються *ЕХФ*.

Для обидвох компонент отримуються параметри елементарних хвильових форм: центри формантних частот, амплітуди, ширини смуг, початкові фази та періоди збурення (p/β). Для них *ЕХФ* мають одинакові довжини та різні параметри.

Далі відповідні періодичні компоненти трактуються як збурення моделі мовотворення. Для них алгоритм синтезу такий:

- встановлюються випадкові точки початків генерації;
- *ЕХФ* генеруються згідно із своїми детермінованими акустичними параметрами та згідно з випадковими точками початків (які визначаються часом надходження або моментами утворення);
- синтезований сигнал обчислюється для періодичних та неперіодичних компонент з використанням схеми перекриття з підсумовуванням.

Для неперіодичних компонент синтез базується на квазістаціонарності сигналу, тобто параметри *ЕХФ* (центри частот, фази, параметри огинаючих), знайдені фільтруванням в шести смугах, синтезуються для періодичних та неперіодичних компонент з використанням перекриття з підсумовуванням.

Модифікують часовий масштаб за визначенням авторів, простою модифікацією параметрів *ЕХФ*, які належать в часі до моментів утворення. Через просту модифікацію моментів утворення забезпечується якісне перетворення часового масштабу з $0.5<\beta<2$. Пришвидшення вдвічі досягається авторами простим діленням всіх моментів утворення *ЕХФ* на 2 перед синтезом.

Однак для зміни темпу більше ніж вдвічі вимагається складніша процедура: кожна *ЕХФ* повинна дублюватися в часі з деякою долею випадковості для запобігання тональної властивості в розтягнутих ділянках. Цей тип результатів масштабування часу в глобальному стисненні-розширенні без впливів періодичностей, що, можливо, лежать в основі шумової модуляції.

Процедура модифікації простіша від методу перекриття з підсумовуванням синхронно з ОТ (ОТСПП) та не вимагає міток ОТ. Подібність до ОТСПП є у зміні часів надходження сигналів короткої довжини, вибраних з вхідного сигналу мови, а відмінність в параметричному описанні коротких довжин сигналів та використання тільки вузькосмугових сигналів.

Метод дає можливість використовувати змінний коефіцієнт перетворення часового масштабу.

Висновки

Розглянуті моделі представлення мови базуються на цифровій моделі мовотворення. Аналіз існуючих моделей представлення мови на основі синусоїдальної, гармо-

нічної на базі КПФ та моделі елементарних хвильових форм показує, що вони вимагають складних процедур обчислення фаз чи параметрів хвильових форм. Огляд методів модифікації часового масштабу в системах аналізу-синтезу показує, що сьогодні найбільш розвинені дістали параметричні методи, незважаючи на їх обчислювальну складність. Дані методи можуть бути використані для стиснення мовної інформації при передачі її каналами зв'язку та для збільшення ступеня стиснення мови в параметричних кодеках, які застосовують і у IP-телефонії.

1. *Рашкевич Ю.М.* Перетворення часового масштабу мовних сигналів. Львів. 1997.
2. *Рабинер Л. , Шафер Р.* Цифровая обработка речевых сигналов. М. 1981.
3. *Portnoff M.* Short-Time Fourier Analysis of Sampled Speech// IEEE Trans. on Acoust. , Speech and Signal Proc. 1981. № 29. P. 364-373.
4. *Portnoff M.* Time-scale modification of speech based on short-time Fourier analysis// IEEE Trans. Acoust. , Speech and Signal Proc. 1981. № 30. P. 374-390.
5. *Griffin D. , Lim J.* Multiband exitation vocoder// IEEE Trans. Acoust. , Speech, Signal Proc. 1988. №30. P. 1223-1238.
6. *Moulines E. , Laroche J.* Non-parametric techniques for pitch-scale and time-scale modification of speech// Speech Communication. 1995. № 16. P. 175-205.
7. *Quatieri T. , McAulay R.* Shape Invariant Time-Scale and Pitch Modification of Speech. IEEE Trans. on Signal Processing. 1992. № 40. P. 168-175
8. *McAulay R. , Quatieri T.* Speech analysis-synthesis based on a sinusoidal representation// IEEE Trans. Acoust. , Speech and Signal Proc. 1986. № 4. P. 744-754.
9. *Richard G. , Alessandro C.* Analysis/synthesis and modification of the speech aperiodic component// Speech Communication. 1996. № 19. P. 224-244.
10. *Laroche J, Stylianou Y. , Moulines E.* HNS: Speech Modification Based on a Harmonic+Noise Model// IEEE Trans.
11. *K. Hamdy, M. Ali, and A. Tewlik.* High quality audio coding of audio signals with a combined harmonic and wavelet representation// ICASSP-96, Atlanta, GA. on Acoustic, Speech and Signal Processing. 1993. № 4. P. II-550-II-553.
12. *Dudley H.* The Vocoder. Bell Labs Record. 1939. Vol. 17. P. 122-126
13. *Flanagan J. , Golden R.* Phase vocoder// Bell Syst. Tech. J. 1967. № 45. P. 1493-1509.
14. *Sylvestre B. And Kabal P.* Time-scale modification of speech using an incremental time-frequency approach with waveform structure compensation// Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. 1992. P. 81-84.
15. *Dolson M.* The phase vocoder: A tutorial//Computer Music J. ,1986. Vol. 10, №4. P. 14-27.
16 Malah D. Time-Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals/ /IEEE Trans. Acoust. , Speech and Signal Proc. 1979. № 27. P. 121-133.
17. *Гнатив Я. Н. , Рашкевич Ю. М.* Регулирование темпа сообщений в речевом канале системы "машина-оператор"// Тез. докл. 2-й Республ. НТК "Автоматизация научных исследований". К. 1981. С. 89-91.
18. *Moulines E. , Laroche J.* Non-parametric techniques for pitch-scale and time-scale modification of speech// Speech Communication. 1995. № 16. P. 175-205.
19. *Nawab S. , Quatieri T.* Short-time Fourier transform// Lim J. , Oppenheim A. Advanced Topics in Signal Processing. Prentice-Hall: Englewood Cliffs, 1988. 193-211.
20. *Roucos D, Wilgus A.* High Quality Time-Scale Modification for Speech// Proc. International Conf. ICASSP-85. Tampa (USA). 1985. № 3. P. 493-496.
21. *Марцишин Р. С.* Модифікація часового масштабу мови через кепстральний аналіз-синтез // Вісн. ДУ "Львівська політехніка". 1998. №351. С. 190-196.
22. *Марцишин Р. С.* Нелінійне перетворення часового масштабу мови в системах керування // Вісн. ДУ "Львівська політехніка". 1999. №370. С. 50-57.
23. *Р. Марцишин, Ю. Рашкевич.* Методика модифікації кепстру для задач перетворення часового масштабу мови в системах аналізу-синтезу //Мат. міжн. НТК "Інформаційні системи та технології". Львів. 1999. С. 30-33.
24. *Маркел Дж. Д. , Грэй А. Х.* Линейное предсказание речи/ Под ред. Ю. Н. Прохорова, В. С. Звездина. М.,

1980.
25. *Шпак З. Я. , Рашкевич Ю. М.* Регулирование темпа подачи речевой информации//Распознавание и синтез звуковых сигналов. К: ИК АН УССР. 1987. С. 70-77.
26. *ElJaroudi . A.* Discrete all pole modeling//IEEE Trans. Acoust. Speech Signal Process. 1991. Vol. 39. - № 2. P. 411-423.
27. *Гнатив Я. Н. , Рашкевич Ю. М.* Гомоморфная модель регулирования темпа речи//Контрольно-измерительная техника. 1984. № 26. С. 84-87.
28. *Seneff S.* System to Independently Modify Excitation and/or Spectrum of Speech Waveform Without Explicit Pitch Extraction// IEEE Trans. on Acoust. Speech and Signal Proc. 1982. № 4. P. 566-578.
29. *Marques J. S. and Almeida L. B.* Frequency-varying sinusoidal modeling of speech// IEEE Trans. Acoust. Speech Signal Process. 1989. Vol. 37. № 5. P. 763-765.
30. *Lienard J. S. and d'Alessandro C.* Wavelets and granular analysis of speech. In:"Wavelets, Time-Frequency Methods and Phase Space, ed. by Combes J. M. Berlin,Springer. 1989. P. 744-754.
31. *Alessandro C.* Time-frequency speech transformation based on an elementary waveform representation// Speech Communication. 1990. № 9. P. 419-431.

УДК 621.391.19

ПРО ПІДХОДИ ДО ОРГАНІЗАЦІЇ СИСТЕМ РОЗПІЗНАВАННЯ МОВИ

© Роман Попович
НУ "Львівська політехніка", м.Львів, вул.С.Бандери, 12

Розглянуто підходи до організації сучасних систем розпізнавання суцільної мови з великим словником. Оцінено кількість трифонів для словника найуживаніших слів української мови та запропоновано питання для фонетичних вирішуючих дерев.

Approaches to organization of current large vocabulary continuous speech recognition systems have been considered. Number of triphones for Ukrainian mostly used words dictionary has been estimated and phonetic decision trees questions have been proposed.

Сучасні системи для розпізнавання суцільної мови з великим словником грунтуються на принципах статистичного розпізнавання образів [1].

На першому етапі мовний зразок перетворюється акустичним процесором на послідовність акустичних векторів $Y = y_1, y_2, ..., y_T$. Кожен вектор є стислим поданням короткочасного мовного спектра на інтервалі, як правило, близько 25 мс зі зсувом ін-