

РОЗПІЗНАВАННЯ МОВЛЕННЯ В СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ

Анотація:

Розглядається можливість використання періоду основного тону для визначення інтонації усного мовлення, що, в свою чергу, дозволить визначити зміст висловлення. Пропонується варіант застосування даного параметра мовного сигналу при розпізнаванні мови в системах зі штучним інтелектом. Приводиться приклад обчислення частоти основного тону за допомогою програмної реалізації алгоритму виділення основного тону.

Аннотация:

Рассматривается возможность использования периода основного тона для определения интонации устной речи, что, в свою очередь, позволит определить смысл высказывания. Предлагается вариант применения данного параметра речевого сигнала при распознавании речи в системах с искусственным интеллектом. Приводится пример вычисления частоты основного тона путем программной реализации алгоритма.

Abstract:

The possibility of using pitch period to determine the inflexion of speech, and thus to clarify the statement's sense is considered in this text. A choice for using this parameter of speech for speech recognition in artificial intelligence systems is purposed. An example of computing the pitch's frequency using the software implementation of pitch's extraction algorithm is also given.

Однією з невід'ємних задач систем штучного інтелекту є розпізнавання інтонації мовлення (питальної, стверджувальної, окличної, подиву). Важливість цієї задачі обумовлюється тим, що вона дозволяє точніше визначати зміст висловленого речення, що відрізняється від розгляду слів тільки як команд для комп'ютера. Відома система розпізнавання мовлення, лідер ринку, *Dragon Naturally Speaking*, розроблена для англійської мови, тому використання такої системи для розпізнавання інтонації української (російської) мови видається нелегким [1, 2, 3]. Крім того, дослідження параметрів мовних сигналів для різних мов має істотні відмінності. Наприклад, питання в англійській мові визначається порядком проходження слів у реченні, у той час як у російській мові інтонація може бути визначальною («Я був в інституті.», «Я був в інституті!», «Я був в інституті?»).

Існуючі зараз системи розпізнавання мови («Руссограф», *Dragon Naturally Speaking*, *ViaVoice*) ґрунтуються на порівнянні мовних сигналів, що визначаються, з великою кількістю еталонів. Даний підхід значно збільшує вимоги до необхідного обсягу пам'яті й продуктивності процесору пристрою розпізнавання. Крім того, для розробки таких систем необхідні потужні обчислювачі, що забезпечують порядку 10^{12} операцій у секунду [4].

Для вирішення подібних завдань, а в остаточному підсумку – побудови системи розпізнавання мовлення для української (російської) мови, пропонується алгоритм розпізнавання, основу якого становлять такі елементи:

1). Алгоритм виділення системи акустичних параметрів мови на основі параметрів, які виділяються смуговим вокодером. Такий вокодер забезпечує мінімізацію кількості чисельних характеристик мовлення і, у той же час, збереження розбірливості мовлення, синтезованого на основі виділюваних параметрів. Це свідчить про те, що обрана система параметрів достатньою мірою характеризує мовний сигнал. Виділяється дев'ятнадцять параметрів, які характеризують форму огинаючої спектра мовного сигналу (сімнадцять параметрів), частоту коливань голосових зв'язувань (період основного тону), вид спектра (дискретний або безперервний) [5].

2). Нормоване представлення мовного сигналу, необхідне для порівняння параметрів мовлення.

3). Програма розпізнавання, побудована в системі математичного моделювання *Matlab*, що дозволяє здійснювати роботу зі звуковими файлами із заданою частотою дискретизації, здійснювати розрахунок цифрових фільтрів, фільтрацію сигналів, і, в остаточному підсумку, перетворити безперервний мовний сигнал у кінцеву множину параметрів (дев'ятнадцять параметрів). Надалі ці параметри будуть використані для розпізнавання шляхом порівняння з еталонними наборами таких параметрів, що характеризують деякі фонетичні елементи мови (фонемі). Дана система дозволить також здійснювати обробку отриманих даних і виведення результатів у вигляді графіків або текстової інформації.

Для розпізнавання інтонації мовлення пропонується використовувати період основного тону. Основний тон – це найбільш сильна гармоніка в спектрі мовного сигналу, що визначається частотою коливань голосових зв'язувань людини при вимові голосних і дзвінких приголосних звуків. Основний тон (ОТ) людської мови лежить у діапазоні приблизно 60...150 Гц у чоловіків і 150...350 Гц у жінок.

Для виділення періоду ОТ використовується характерна ознака голосного звуку – при малому рівні шумів форма мовного сигналу майже точно повторюється в кожному наступному періоді ОТ. Тому задача визначення періоду ОТ зводиться до визначення моментів появи екстремальних значень, наприклад, найбільших, і виміру часових інтервалів між сусідніми значеннями.

Виділення сигналів основного тону здійснюється багатоканальним часовим визначником, алгоритм роботи й основні елементи якого аналогічні описаному в [5, 6]. Визначник ОТ здійснює попередню й статистичну обробку. При попередній обробці, здійснюється грубе визначення можливого періоду основного тону за допомогою шести елементарних вимірників основного тону (ЕВОТ) на основі часових інтервалів між екстремумами відфільтрованого сигналу. Точне визначення періоду основного тону здійснюється в результаті статистичної обробки результатів попереднього визначення періоду основного тону. Блок-схему алгоритму оцінки висоти тону наведено на рис. 1.

Алгоритм зручно розділити на чотири операції, пов'язані з обробкою або з ухваленням рішення:

- 1). Низькочастотна фільтрація мовного сигналу;
- 2). Утворення шести функцій по екстремумам відфільтрованого сигналу – так званих m -параметрів. Значення m_1 , m_2 , і m_3 , формуються для кожного максимуму, а m_4 ,

m_5 і m_6 – для кожного мінімуму;

3). Одержання на основі значень функцій шістьох оцінок висоти тону в шістьох однакових вимірниках (ЕВОТ);

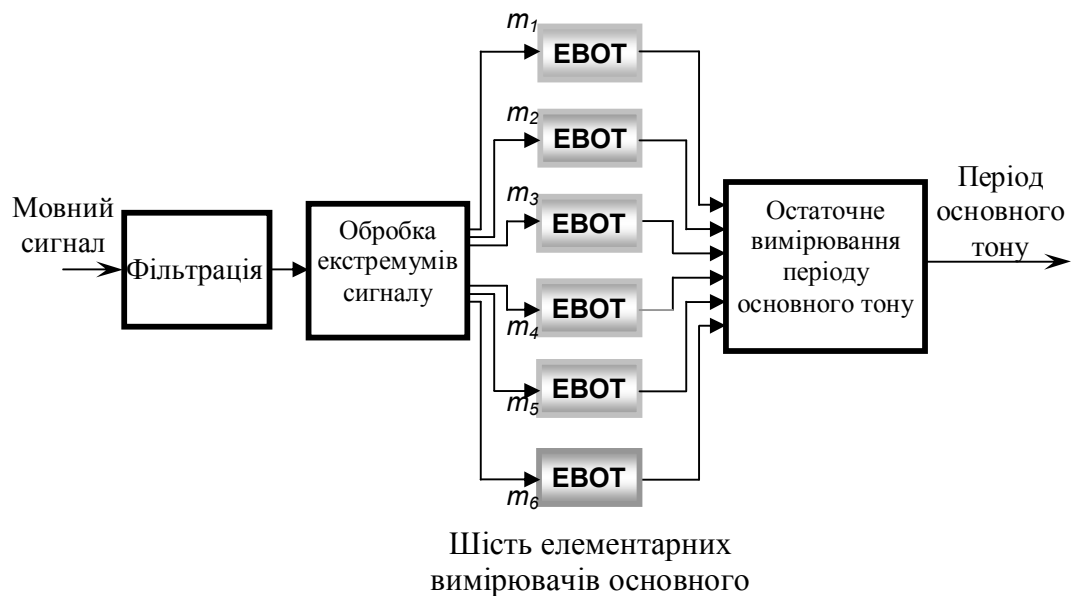


Рис. 1. Блок-схема визначника основного тону

4). Прийняття остаточного рішення на основі оцінок елементарних вимірників висоти тону.

Запропонований алгоритм програмно реалізовано у системі *Matlab* і проведено пробні обчислення частоти ОТ. У якості тестових мовних сигналів були записані звукові файли, в яких вимовлялися слова «ні» і «так» з різними інтонаціями (стверджувальною і питальною). На рис. 2 і рис. 3 показані значення частоти основного тону мовлення тестових сигналів протягом вимови слів. У сигналі не аналізувалися ділянки шумових звуків – глухих приголосних, енергія яких зосереджена у високочастотній частині спектра, тому що при вимові таких звуків голосові зв'язування не беруть участь. Для ділянок тональних звуків, характерних для голосних і дзвінких приголосних звуків, енергія яких зосереджена в низькочастотній частині спектра, а форма звичайно з великою точністю повторюється через період ОТ, проводилися обчислення основного тону.

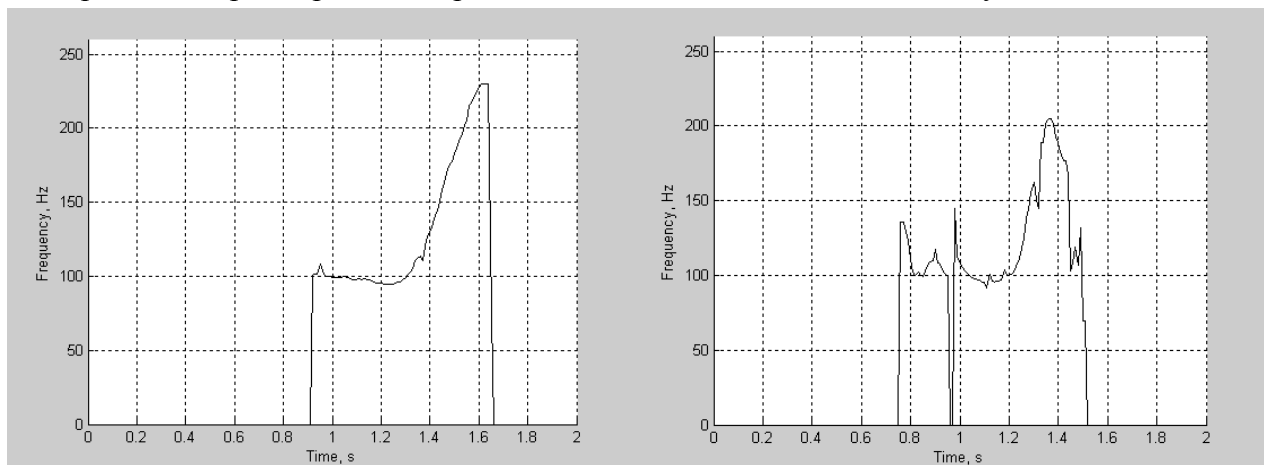


Рис. 2. Зміна частоти основного тону для висловлення з питальною інтонацією (для слова «ні» – ліворуч, «так» – праворуч)

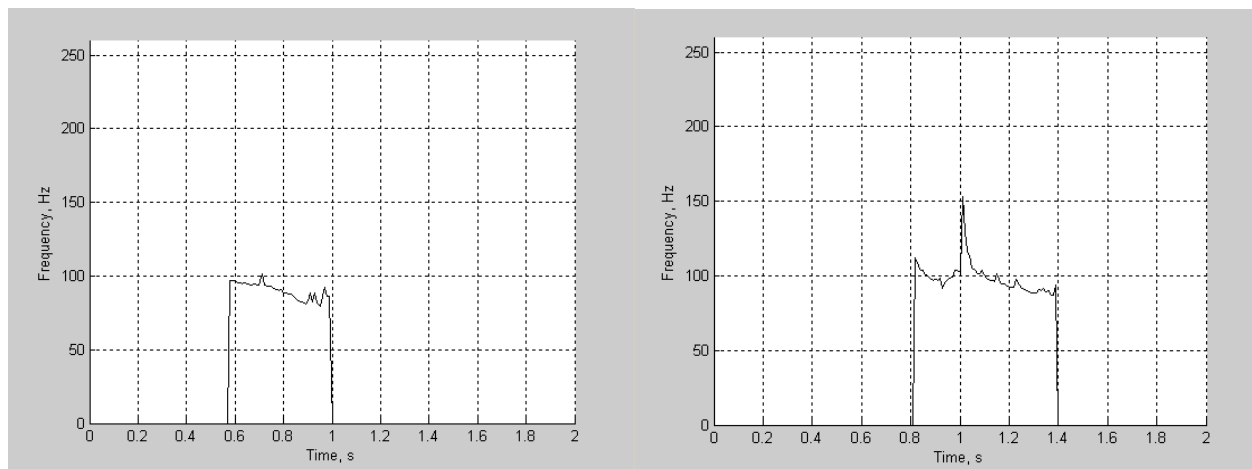


Рис. 3. Зміна частоти основного тону для висловлення зі стверджувальною інтонацією (для слова «ні» – ліворуч, «так» – праворуч)

В алгоритмі, що реалізовано в системі *Matlab*, визначення шумових і тональних ділянок сигналу здійснювалося за допомогою способу, описаному в [5]. З порівняння тестових сигналів мовлення можна укласти, що у варіанті вимови з питальною інтонацією частота ОТ досить сильно (в 2...2,5 рази) збільшується протягом аналізу, у той час як при вимові зі стверджувальною інтонацією частота ОТ – навпаки, незначно зменшується.

З метою визначення швидкодії обраного алгоритму, за допомогою можливостей системи *Matlab* був здійснений вимір часу, що затрачувався на обробку тестових сигналів. Дана величина залежить здебільшого від характеристик системи, на якій здійснювались обчислення, версій системи *Matlab* і операційної системи, а також від частоти дискретизації сигналу. При використанні в системі центрального процесора з тактовою частотою 3 ГГц на обчислення значень ОТ протягом однієї секунди вихідного сигналу затрачалося 0,035...0,1 секунди залежно від вищезгаданих факторів.

Отримані результати дозволяють стверджувати, що параметр “основний тон” можна застосовувати при смислому аналізі мовних сигналів, а використання методів аналізу періоду основного тону може сприяти зниженню обчислювальних вимог систем штучного інтелекту (розпізнавання мовлення).

Література:

1. Бердников О.М. Використання засобів MATLAB для рішення задач розпізнавання мови / Бердников О.М., Богуш К.Ю. // Зб. наук. пр. ВІТІ НТУУ “КПІ”. – 2004. – № 4. – С. 5-10.
2. Бердников О.М. Алгоритм розпізнавання мовних сигналів / Бердников О.М., Богуш К.Ю. // Зб. наук. пр. ВІТІ НТУУ “КПІ”. – 2007. – №2. – С. 4-10.
3. Бердников О.М. Модель пофонемного розпізнавання мови на основі акустичних параметрів смугового вокодеру / Бердников О.М., Богуш К.Ю. // Зб. наук. пр. ВІТІ НТУУ „КПІ”. – 2010. – №2. – С. 11-18.
4. <http://www.3dnews.ru/tags/Русскоязыф>

5. *Аппаратура АТ-3001М.* – К. : КВВИУС, 1988. – 96с.
6. *Рабинер Л.* Теория и применение цифровой обработки сигналов. / Рабинер Л., Гоулд Б. – М.: МИР, 1978. – 848 с.