РОЗПОДІЛЕНЕ КОМП’ЮТЕРНЕ ДОКУМЕНТУВАННЯ ГОЛОСОВИХ МОВНИХ ФОНОГРАМ

О.С. Загваздін

Аналіз предметної області

- Можливі сфери застосування включають: стенографування засідань представницьких органів, органів виконавчої влади, судів, інших засідань
- Користувачі системи мають обмежені навички користування комп’ютером і вимагають простого інтерфейсу
- Експлуатація системи має бути простою і не вимагати адміністрування
- Якість звукового сигналу, який подається на вхід може бути досить невисокою

Аналіз існуючих рішень і систем

- Система “Нестор” Центра Речевых Технологий, Москва
- Комплекс оперативного стенографування «SRS Report 2000», команія SRS, Москва
- Проект системи стенографування засідань університету Berkeley, США
- Система стенографувань засідань ILS, Німеччина

Постановка задачі

- Отримання звукового та відео сигналу для широко вживаних типів форматів (wav, mp3, wma, avi, mpeg тощо)
- Розбиття сигналу на рівноцінні сегменти з автоматизованою фільтрацією від сторонніх шумів
- Створення багатористувацької системи з автоматичним розподіленням сегментів, яка б не вимагала адімінстрування
- Створення простого і інтуїтивно зрозумілого інтерфейсу користувача
- Визначення позиції зміни диктора у голосовому сигналі для більш інтелектуальної сегментації
- Зміна швидкості відтворення сигналу без зміни його основних акустичних характеристик

Сегментація звукового сигналу: алгоритм

- Пошук пауз: проходження вікном визначеної довжини по всьому сигналу і пошук інтервалів, в яких середньоквадратичне відхилення не перевищує заданої межі
- Межі сегментів визначаються по знайденим паузам
- Довжина сегменту є не меншою від деякої заданой величини

Визначення пауз: адаптивний підхід

- Середньоквадратичне відхиделння сигналу обраховується для кожного вікна
- На кожному кроці маємо *N* попередніх значень, де *N=10*(Ds / Lw).* Тут *Ds* – частота дискретизації, *Lw* – довжина вікна. Тобто припускаємо, що на кожному проміжку в 10 секунд є принаймні одна пауза
- Значення порогу покладаємо *T=2*min(D),* тут *D* – множина значень, отриманих на *N* попередніх кроках

Розподілення сегментів між операторами

- Серед комп’ютерів робочої групи один визначається як головний – на ньому відбувається завантаження звукового сигналу, сегментація і попередня цифрова обробка сигналу
- Решта комп’ютерів по мережі отримують наступний сегмент, що не оброблено і не надано жодному іншому оператору
- По завершенню обробки, результат обробки надсилається на головний комп’ютер з поміткою чи вдалося коректно обробити сегмент
- Остаточний документ зі стенограмою створюється на головному комп’ютері

Видалення сторонніх шумів

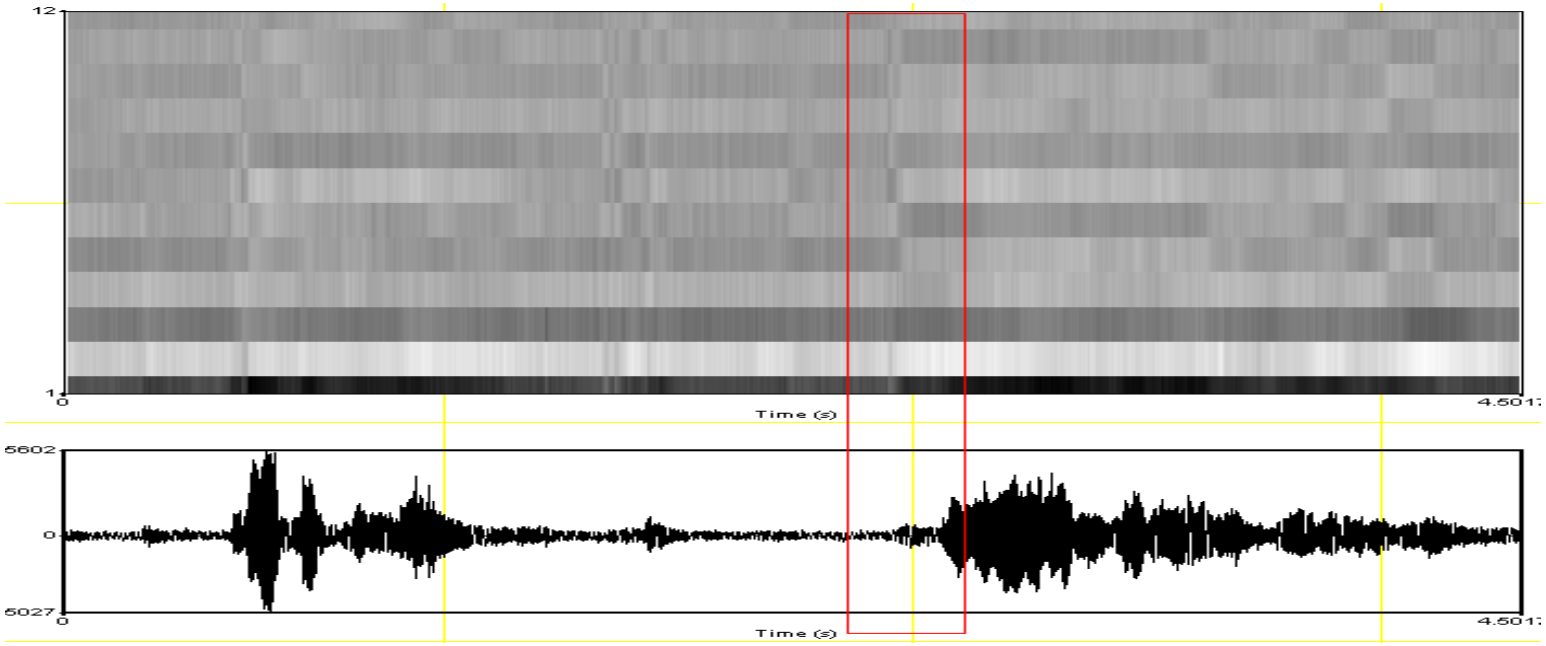
- Використовуються паузи, що були знайдені в процесі сегментації сигналу: ділянка сигналу, яку визначено як паузу містить лише шум
- Використовується видалення адитивного шуму за методом спектрального віднімання

$$y(m) = x(m) + n(m)$$

$$X_w(e^{i\omega}) = Y_w(e^{i\omega}) - N_w(e^{i\omega})$$

Визначення позиції зміни диктора

- В якості характеристичного ветору використовується вектор з коефіцієнтів мел кепстр



Визначення позиції зміни диктора

- Припускаємо, що зміна диктора відбувається в околі паузи
- Порівнюємо множини характеристичних векторів до і після паузи
- Якщо міра відмінності перевищує поріг – в околі паузи є зміна диктора

$$d(X_1, X_2) > \delta_2$$

$$d(X_1, X_2) = \mu_{1/2}(d(x_{1i}, x_{2j}))$$

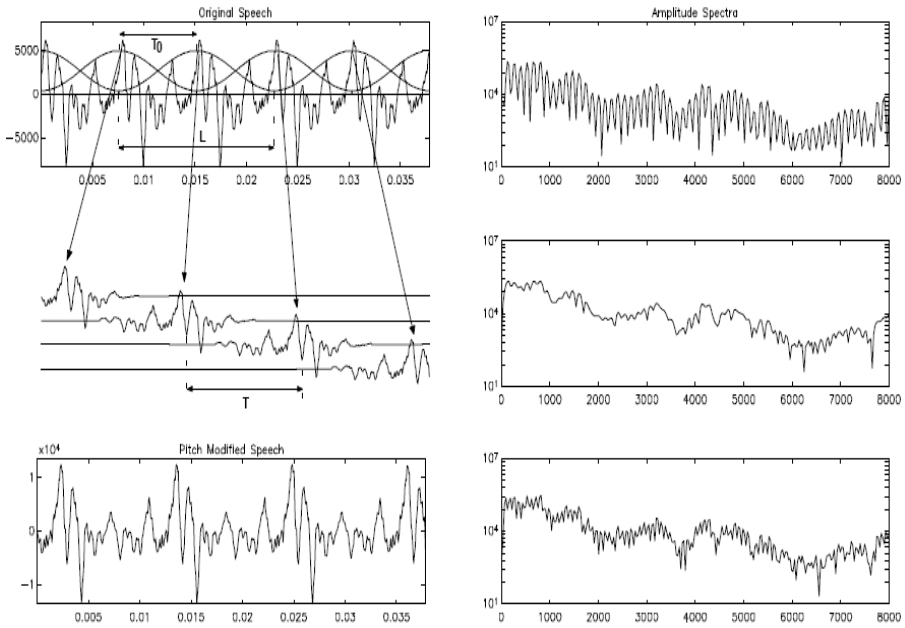
$$\forall x_{1i} \in X_1, x_{2j} \in X_2$$

$\mu_{1/2}$ - медіана

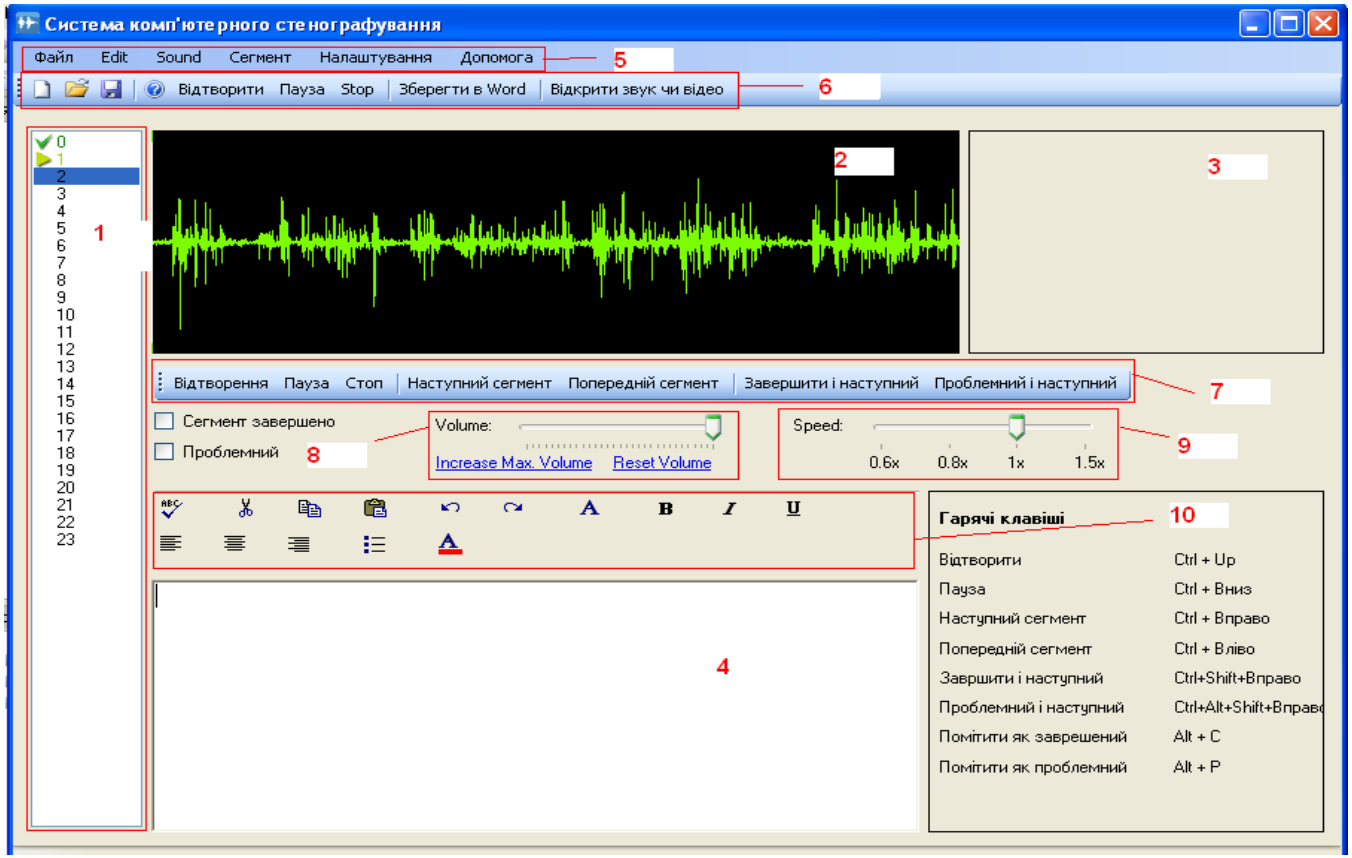
Зміна швидкості відтворення сигналу

- Використовується алгоритм типу PSOLA
- Для знаходження періодів псевдоперіодичності використовується фільтрація полосовим фільтром і медіанне згладження

$$x[n] = \sum_{i=-\infty}^{\infty} x_i[n - t_a[i]] \qquad y[n] = \sum_{j=-\infty}^{\infty} y_j[n - t_b[j]]$$



Інтерфейс користувача



Ергономіка робочого місйя оператора

- Відповідно до класичного дослідження Джорджа Міллера про короткочасну пам’ять людини (1956), людина здатна концентрувати увагу на 7+/-2 об’єктах
- Кожен сегмент для стенографування має містити 5-9 слів
- Кількість елементів керування в інтерфейсі користувача не повинна перевищувати 9

Результати експерименту (однокористувацький режим)

- Для створення стенограми засідання тривалістю 2 години при використанні системи одному непідготовленому користувачу потрібно близько 6 годин, проти 12-18 годин при використанні стандартних засобів.
- Користувач починає впевнено користуватися системою вже після перших 15-30 хвилин роботи