

# Система распределённого автоматизированного документирования речевых сигналов

*Крак Ю.В., доктор физико-математических наук, профессор*

*Загваздин А.С., научный сотрудник*

В статье рассматривается система автоматизированного компьютерного стенографирования, которая включает в себя механизмы предварительной обработки сигнала и сегментацию сигнала, основанную на определении голосовой активности в сигнале. Предварительная обработка сигнала включает в себя уменьшение уровня шума и возможность изменения скорости воспроизведения сигнала без изменения его акустических характеристик. Система также предоставляет возможность сегментации голосового сигнала на основании позиций изменения диктора. Система позволяет распространять сегменты голосового сигнала между операторами-стенографистами, что позволяет повысить продуктивность стенографирования в сравнении со стандартными способами стенографирования.

• стенографирование • распределённое документирование • обнаружение изменения диктора • определение голосовой активности • речевая сегментация сигнала • шумоподавление.

A system for distributed automated transcription is described. The system features automated speech signal pre-processing and segmentation based on voice activity detection. Signal pre-processing includes the noise reduction and the ability to change the speech rate of the signal without affecting its acoustical characteristics. The system also allows to segment the signal based on the speaker change. The system allows to automatically distribute the signal segments among the members of the group of transcriptionists that allows to increase the transcription performance compared to the standard transcription methods.

• distributed transcription • speaker change detection • voice activity detection • speech signal segmentation • noise reduction.

Текстовая стенограмма заседания — необходимая составляющая в работе многих организаций. Как правило, процесс создания и расшифровки стенограмм достаточно продолжителен и попытки его ускорения путём расширения персонала, вовлечённого в процесс, представляются неэффективными. Для автоматизации процесса создания стенограмм заседаний предлагается система распределённого компьютерного документирования.

Поскольку стенографирование заседаний — задача, которая может быть достаточно просто распределена между многими исполнителями, система распределённого стенографирования должна поддерживать однопользовательский и многопользовательский режимы работы. На сегодняшний день в мире существует несколько систем распределённого документирования. В качестве примеров можно привести модуль стенографирования, входящий в состав системы поддержки принятия решений «Рада-3» [1], которая

используется для обеспечения законодательной деятельности Верховной Рады Украины и некоторых местных законодательных органов; системы стенографирования фирмы SRS (г. Киев) [2], которые широко используются для обеспечения протоколирования судебных заседаний; систему документирования «Нестор» фирмы «Центр речевых технологий» (г. Санкт-Петербург) [3] и некоторые другие.

Однако практически всем существующим в настоящее время системам свойственен ряд недостатков. Основной недостаток существующих систем стенографирования заседаний заключается в том, что они предъявляют высокие требования к аппаратному обеспечению, на котором работает серверная часть системы, и требует существенных затрат на внедрение системы в целом. Как правило, такие системы документирования устанавливаются стационарно в залах, где проходят заседания, и имеют жёсткую привязку к звукозаписывающему оборудованию зала, а также требуют установки отдельных серверов для обеспечения распределённого документирования, сервера звукозаписи и т.п. Также, как правило, в составе подобных систем документирования присутствует ещё и выделенный сервер баз данных. Очевидно, что системы, реализованные в соответствии с подобной архитектурой, требуют значительных усилий с точки зрения системного администрирования. Такие ограничения делают внедрение подобных систем стенографирования практически невозможным в небольших организациях, организациях с ограниченным ИТ бюджетом, а также для индивидуальных пользователей. Наличие серверной части в таких системах также делает практически невозможной мобильную работу с системой и требует постоянного подключения к локальной сети.

Ещё одним общим недостатком существующих систем является недостаток внимания, уделяемого разработчиками средствам повышения продуктивности работы операторов-стенографистов. Сегментация входящего звукового сигнала, как правило, осуществляется на фрагменты фиксированной длины. Фрагмент может обрываться на середине слова или фразы, возможности по предварительной цифровой обработке сигнала (шумоочистке, изменению скорости воспроизведения сигнала, и т.д.) также зачастую ограничены. В общем, основные недостатки существующих систем распределённого документирования могут быть сформулированы следующим образом:

- высокая стоимость внедрения и сопровождения, высокие требования к аппаратным ресурсам;
- недостаточная интеллектуальность при сегментации и распределении фрагментов входящего речевого сигнала;
- недостаточные возможности по предварительной цифровой обработке сигнала;
- отсутствие либо недостаточная развитость средств повышения продуктивности работы операторов-стенографистов.

Ниже предлагается к рассмотрению система распределённого компьютерного документирования речевых фонограмм, в основе программной архитектуры которой лежит концепция, не требующая реализации основных функций системы на выделенных серверах, где среди рабочих станций операторов, стенографистов одна может быть выделена как главная. На главной рабочей станции будет проводиться предварительная обработка звука, его сегментация и распределение между операторами. Таким образом, предложенная система сможет быть развёрнута на нескольких персональных компьютерах. В предложенной системе также поддерживается однопользовательский режим работы, при котором все функции системы будут реализованы на одном компьютере.

## Требования к системе распределённого документирования

Рассмотрим основные требования, которые выдвигаются к системе распределённого компьютерного документирования:

- Система должна иметь возможность работать с входящим речевым сигналом из различных источников звуковой и видеоинформации. В упрощённом варианте система должна поддерживать различные форматы звуковых и видео файлов в качестве источников речевого сигнала (wav, mp3, wma, avi, mpeg и другие). Речевой сигнал также может сопровождаться видеосигналом, и при воспроизведении звуковой и видеосигнал должны быть синхронизированы.
- Система должна иметь возможность сегментации входящего звукового сигнала на равноценные сегменты, при этом они не должны начинаться и заканчиваться на середине слов. Длина сегментов не должна быть меньше определённого значения, но при этом сегменты должны быть достаточно короткими для удобного запоминания, чтобы в процессе стенографирования сводилась к минимуму необходимость повторного воспроизведения сегмента. Алгоритм сегментации должен надёжно работать в условиях нестационарного шума, который может присутствовать во входящем сигнале.
- При сегментации сигнала должна учитываться информация об изменении диктора, момент изменения диктора в сигнале должен совпадать с началом нового сегмента.
- В системе должны быть реализованы следующие функции предварительной цифровой обработки сигнала: уменьшение уровня шума, усиление уровня сигнала и изменение скорости воспроизведения сигнала, причём последнее не должно вызывать изменение тембра голоса говорящего.
- Интерфейс рабочего места оператора-стенографиста должен быть простым и интуитивно понятным и не требовать высокого уровня компьютерной грамотности от операторов системы. Это позволит сократить сроки обучения пользованию системой и получить более высокую продуктивность работы.
- Система должна иметь возможность работы как в однопользовательском, так и в многопользовательском режимах, при этом она не должна требовать комплексного внедрения и администрирования. Все компоненты системы должны иметь возможность быть установленными на распространённом компьютерном оборудовании. При этом скорость обработки сигнала и выполнения основных операций системы должны быть достаточно высокими для обеспечения комфортной работы с системой.
- В многопользовательском режиме фрагменты входящего речевого сигнала должны автоматически распределяться между операторами-стенографистами. Система должна иметь возможность работы с любым количеством операторов, которые могут подключаться к системе через локальную сеть или по сети Интернет.
- Исходя из особенностей восприятия человеком информации и кратковременной памяти человека [4], длина фрагментов, на которые разбивается входящий звуковой сигнал, должна составлять в среднем 5–9 слов. Аналогично количество основных элементов управления программы рабочего места оператора не должно превышать 9.
- Для повышения продуктивности работы операторов в системе должны быть реализованы функции автоматической проверки орфографии, автозамены и автоподстановки, автоматические всплывающие подсказки для наиболее часто встречающихся слов.

Ниже будут рассмотрены особенности реализации системы распределённого компьютерного документирования, удовлетворяющей перечисленным выше требованиям, которая была создана в Институте кибернетики НАН Украины им. В.М. Глушкова.

## Сегментация входящего речевого сигнала

Одной из важнейших задач системы распределённого компьютерного документирования является сегментация сигнала: разделение входящего речевого сигнала на равноценные сегменты. Для сегментации сигнала в рассматриваемой системе используется инфор-

мация о паузах, присутствующих в сигнале, а также о позициях в звуковом сигнале, где происходит смена диктора.

Для качественного определения пауз в сигнале предлагается алгоритм определения пауз, устойчивый к нестационарному уровню шума в сигнале. Поиск фрагментов сигнала, где присутствуют паузы, проводится путём сравнения энергии в анализируемом фрейме с пороговым значением. Для этого по сигналу проходим прямоугольным окном продолжительностью 50 мс таким образом, чтобы начало каждого последующего окна приходилось на середину предыдущего. Такая относительно большая продолжительность окна обусловлена тем, что для поиска пауз в речи нецелесообразно использовать паузы короче 50 мс, а большая продолжительность окна позволяет сократить общее количество итераций алгоритма. Таким образом, на каждом шаге алгоритма сигнал задаётся как

$$s_k[i] = \begin{cases} s[i], & 0.05 * D * k \leq i < 0.15 * D * k \\ 0, & else \end{cases} \quad (1)$$

Здесь  $D$  — частота дискретизации сигнала, а  $k$  — номер шага алгоритма.

Полагаем, что на участке сигнала, длиной 10 с должна быть по крайней мере одна пауза. Следовательно, для адаптации алгоритма к текущему соотношению «сигнал/шум» будем использовать информацию о предыдущих 10 с звучания сигнала. Также, будем считать, что уровень энергии сигнала в участках, соответствующих паузам, ниже, чем в участках, где присутствует голосовая активность. Энергию сигнала определим как дисперсию амплитуды сигнала в заданном окне:

$$E_{s_k} = \log_{10} \left( \frac{1}{N} \sum_{i=1}^N s_k[i]^2 - \left( \frac{1}{N} \sum_{i=1}^N s_k[i] \right)^2 \right) \quad (2)$$

На рис. 1 и 2 приведены графики амплитудно-временного представления зашумленного сигнала и соответствующий ему график уровня энергии в сигнале. Уровень энергии в участке, который соответствует паузе, существенно ниже.



Рис. 1. АЧП зашумленного сигнала с паузой

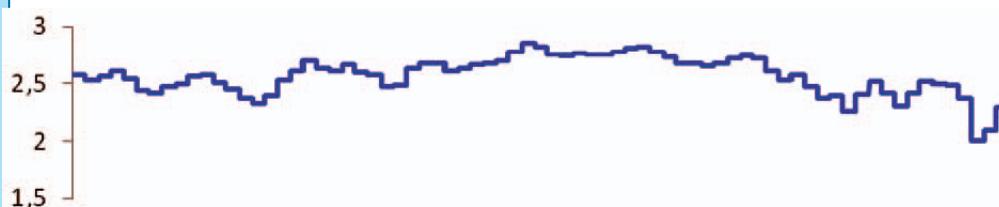


Рис. 2. Энергия сигнала, рассчитанная как дисперсия амплитуды

Для устранения негативного влияния случайных возмущений на измерения к полученным уровням энергии применяется метод медианного сглаживания 5-го порядка в соответствии с формулой:

$$s[i] = \text{med}\{s[i-2], s[i-1], s[i+1], s[i+1]\}. \quad (3)$$

Для принятия решения о том, соответствует ли анализируемый фрейм паузе, значение энергии в этом фрейме сравнивается с порогом. Поскольку окружающие условия и уровень шума в сигнале могут меняться со временем, существует необходимость динамического расчёта порога в процессе обработки сигнала. Предложен следующий алгоритм адаптивного вычисления порога энергии для пауз.

На участке сигнала протяжённостью 10 с, предшествующем анализируемому фрейму, находятся минимальный и максимальный уровни энергии для данного участка:  $E_{\min}$  и  $E_{\max}$ . Далее, уровень энергии в текущем фрейме сравнивается с полученными минимальным и максимальными значениями. Решение о том, что текущий фрейм принадлежит паузе, принимается, если выполняется следующее условие:

$$E < E_{\min} \vee \frac{E - E_{\min}}{E_{\max} - E_{\min}} < 0.2 \quad (4)$$

Минимальное и максимальное значения уточняются на каждом шаге алгоритма с учётом предыдущих 10с звучания. Фреймы с низким уровнем энергии, расположенные последовательно один за другим, объединяются в одну паузу. Паузы, длина которых меньше некоторой заданной длины, исключаются из рассмотрения, так как они наиболее вероятно соответствуют участкам с низкой энергией в середине слова (например, шипящим согласным).

Графически работа алгоритма адаптивного вычисления порога проиллюстрирована на рис. 3.

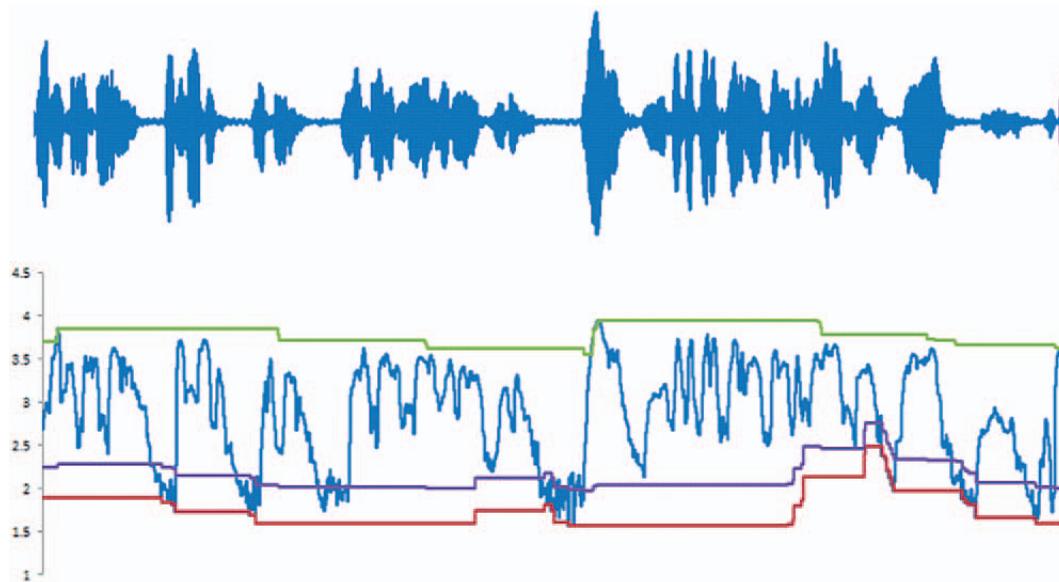


Рис. 3. Адаптивный порог при поиске пауз в сигнале

Для повышения интеллектуальности сегментации сигнала целесообразно учитывать позиции в сигнале, где происходит изменение диктора. Для нахождения таких позиций предлагается подход, рассмотренный ниже.

Полагаем, что изменение диктора в сигнале происходит в районе паузы. Иными словами, после того, как заканчивает говорить первый диктор и начинает говорить второй, есть

пауза. На практике это не всегда так, и дикторы могут перебивать друг друга, однако, такие ситуации сложно учитывать при сегментации сигнала, и в данной работе они не рассматриваются.

Положим, что  $X = \{x_1, x_2, \dots, x_n\}$  — множество характеристических векторов, которые соответствуют участку сигнала до паузы, а  $Y = \{y_1, y_2, \dots, y_n\}$  — множество характеристических векторов, которые соответствуют участку сигнала между текущей паузой и последующей.  $N_x$  и  $N_y$  — соответственно количество точек в первом и втором множествах. Характеристические векторы в данном случае представляют собой 13 мел-кепстральных коэффициентов, рассчитанных на участке сигнала продолжительностью 30 мс, и частота основного тона в качестве 14-го элемента вектора.

Пусть  $Z = X \cup Y$  — объединение множеств характеристических векторов с количеством точек  $N = N_x + N_y$ . Множества  $X$  и  $Y$  сравниваются при помощи некоторой меры различия, и если они отличаются значительно, принимается положение о том, что в анализируемом участке сигнала присутствует изменение диктора.

Задачу определения наличия изменения диктора можно сформулировать в виде задачи проверки гипотезы. Пусть  $H_0$  — гипотеза о том, что изменение диктора отсутствует, а  $H_1$  — гипотеза о том, что происходит изменение диктора. Положим также, что векторы, из которых состоят множества  $X$  и  $Y$  являются независимыми и имеющими одинаковое распределение случайными величинами. Пусть  $\Theta_Z$  — параметры распределения для множества  $Z$ , рассчитанные при помощи метода максимального правдоподобия. В таком случае логарифмическое соотношение правдоподобия для множества наблюдений  $Z$  при условии выполнения гипотезы  $H_0$  запишется как:

$$L_0 = \sum_{i=1}^{N_x} \log p(x_i | \Theta_Z) + \sum_{i=1}^{N_y} \log p(y_i | \Theta_Z) \quad (5)$$

Здесь  $p(x | \Theta)$  — вероятность того, что  $x$  выполняется при условии  $\Theta$ . Функция плотности распределения находится при помощи метода Gaussian Mixture Models (GMM).

Для проверки гипотезы  $H_1$  рассчитываются параметры индивидуальных распределений для наборов наблюдений  $X$  и  $Y$ , которые соответственно обозначаются как  $\Theta_X$  и  $\Theta_Y$ . При этом логарифмическое отношение правдоподобия для гипотезы запишется как:

$$L_1 = \sum_{i=1}^{N_x} \log p(x_i | \Theta_X) + \sum_{i=1}^{N_y} \log p(y_i | \Theta_Y) \quad (6)$$

Меру различия для множеств  $X$  и  $Y$ , в таком случае, можно задать как байесовский информационный критерий:

$$d_1 = L_1 - L_0 - \frac{\lambda}{2} \Delta K \log N \quad (7)$$

Здесь  $\Delta Z = N_x - N_y$ , а  $\lambda$  — параметр, который подбирается экспериментально. Решение о наличии изменения диктора в анализируемом участке сигнала принимается, если заданная таким образом мера различия превышает некоторый порог, который задаётся экспериментально.

Далее в процессе сегментации сигнала учитывается информация о найденных паузах и позициях изменения диктора. Процесс сегментации показан на рис. 4.



Рис. 4. Процесс сегментации сигнала

### Предварительная цифровая обработка сигнала

В рассматриваемой системе в качестве метода предварительной цифровой обработки сигнала реализованы методы уменьшения уровня шума и изменения скорости воспроизведения звукового сигнала без изменения тембра голоса говорящего.

Шумы, которые, как правило, присутствуют в сигналах, подаваемых на вход системы распределённого документирования, могут считаться аддитивными в спектральной области. Следовательно, для фильтрации таких шумов могут быть применены распространённые методы спектрального вычитания или виннеровской фильтрации. В случае виннеровской фильтрации фильтр задаётся следующим образом [5]:

$$H(\omega) = \frac{S_X(\omega)}{S_X(\omega) + S_N(\omega)} \quad (8)$$

Здесь  $S_X(\omega)$  — спектр сигнала, а  $S_N(\omega)$  — спектр шума.

Функция фильтрации для фрейма  $m$  может быть задана как:

$$f(Y_m) = \frac{S_X(m)}{S_X(m) + S_N(m)} Y_m \quad (9)$$

Если спектр шума известен, то спектр очищенного от шума сигнала можно рассчитать как

$$S_X(m) = \begin{cases} |Y(m)|^2 - S_N(m), & |Y(m)|^2 > S_N(m) \\ 0 & \text{else} \end{cases} \quad (10)$$

Отсюда функция фильтрации шума может быть записана следующим образом:

$$f(Y(m)) = \begin{cases} \frac{|Y(m)|^2 - S_N(m)}{|Y(m)|^2} Y(m), & |Y(m)|^2 > S_N(m) \\ 0, & \text{else} \end{cases} \quad (11)$$

Для работы описанного метода фильтрации необходимо знать спектр шума. Предположим, что участки сигнала, которые соответствуют паузам, содержат только шум. Следовательно, для аппроксимации шума можно использовать паузы, которые были найдены при помощи алгоритма, описанного в «Сегментации речевого сигнала».

Для изменения скорости воспроизведения сигнала с сохранением тембра голоса диктора необходимо убедиться, что продолжительность сигнала изменяется, но частота основного тона говорящего сохраняется. Обеспечить это возможно при помощи использования ал-

горитмов типа PSOLA (pitch-synchronous overlap and add), которые широко применяются в системах искусственного синтеза речи. Для реализации таких алгоритмов сперва решается задача обнаружения периодов псевдопериодичности в звуковом сигнале (питч-периодов). Для этого исходный звуковой сигнал пропускается через низкочастотный и высокочастотный фильтры с конечными импульсными характеристиками. На рис. 5 и 6 приведен пример слога «ма» до и после фильтрации соответственно.

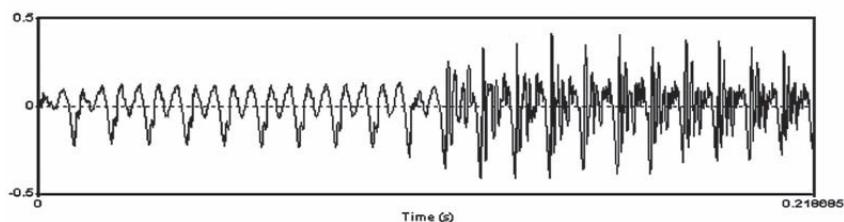


Рис. 5. Слог «ма» до фильтрации

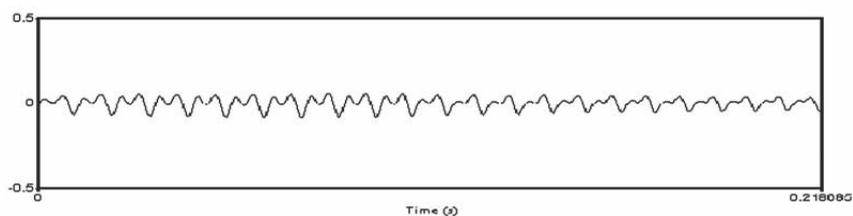


Рис. 6. Слог «ма» после фильтрации

Далее для сглаживания сигнала каждый элемент вектора исходного сигнала заменяется на взвешенное среднее четырёх окружающих его элементов по формуле:

$$d[i] = \frac{3x[i-2] + x[i-1] - x[i+1] - 3x[i+2]}{10}. \quad (12)$$

К полученному сигналу применяется медианное сглаживание порядка  $y = 199$  (каждый элемент вектора заменяется на медиану вектора, состоящего из  $n$  элементов, окружающих текущий элемент). Вид сигнала после сглаживания представлен на рис. 7.

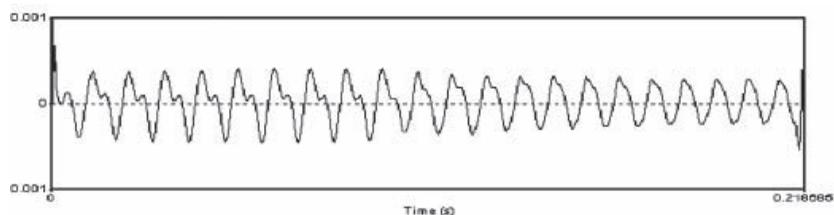


Рис. 7. Слог «ма» после фильтрации и медианного сглаживания

После этого в полученном сигнале обнаруживаются точки, где последовательность, состоящая из элементов вектора сигнала, изменяет знак с «-» на «+» и такие точки обозначаются как границы питч-периодов. Среди определённых таким образом границ обнаруживаются и исключаются точки, расположенные слишком близко друг к другу, а для участков сигнала, где нет явной псевдопериодичности назначаются условные границы с некоторым постоянным интервалом.

После определения границ периодов псевдопериодичности можно изменять акустические характеристики сигнала. Исходный сигнал можно представить в виде функции периодов основного тона  $x_i[n]$  :

$$x[n] = \sum_{i=-\infty}^{\infty} x_i[n - t_a[i]], \quad (13)$$

где  $t_a[i]$  — границы периодов псевдопериодичности сигнала, т.е. разница между двумя соседними границами  $P_a[i] = t_a[i] - t_a[i - 1]$  равняется периоду основного тона в момент времени  $t_a[i]$ . Питч-период определим через исходный сигнал помноженный на оконную функцию:

$$x_i[n] = w_i[n]x[n], \quad (14)$$

где окна  $w_i$  удовлетворяют условию:

$$\sum_{i=-\infty}^{\infty} w_i[n - t_a[i]] = 1, \quad (15)$$

что достигается использованием оконных функций типа Хэннинга или трапецевидным окном длиной в два периода основного тона.

В результате работы алгоритма необходимо получить сигнал  $y[n]$ , который имеет одинаковые с  $x[n]$  спектральные характеристики, но отличается от него основным тоном и/или продолжительностью. Чтобы достичь этого, заменяем аналитические границы питч-периодов  $t_a[i]$  границами  $t_b[i]$ , а аналитические периоды основного тона  $x_i[n]$  — периодами  $y_i[n]$  согласно

$$y[n] = \sum_{j=-\infty}^{\infty} y_j[n - t_b[j]]. \quad (16)$$

Таким образом, достаточно лишь задать границы  $t_b[i]$ , соответствующие продолжительности и основному тону, которые необходимо получить. Результирующий период основного тона получаем подстановкой ближайшего соответствующего аналитического периода  $x_i[n]$ . Графически работа алгоритма представлена на рис. 8.

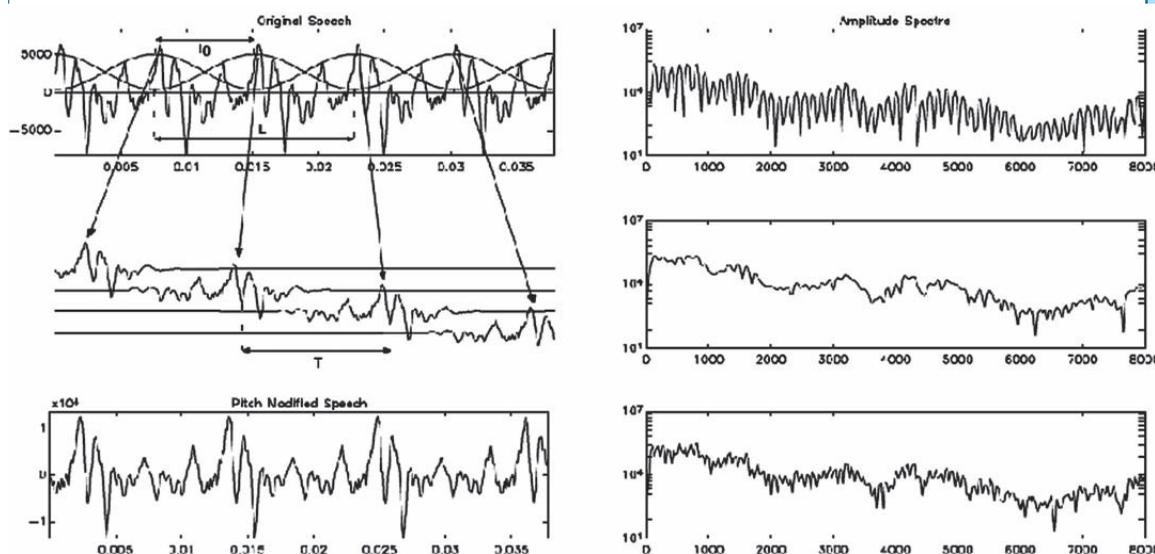


Рис. 8. Схема работы алгоритма для модификации продолжительности и основного тона

## Программная реализация системы

Программная архитектура рассматриваемой системы распределённого компьютерного документирования построена без использования централизованного сервера. Такой подход обеспечивает высокую мобильность системы и невысокие требования к аппаратному обеспечению. Система состоит из двух типов модулей: главного модуля и модулей, которые являются рабочими местами операторов-стенографистов. На главном модуле осуществляется получение звукового сигнала, его предварительная обработка и сегментация. Также главный модуль управляет распределением сегментов между операторами. На нём осуществляется компоновка полученных от операторов результатов обработки сегментов в единый результирующий документ-стенограмму. В процессе обработки фонограммы оператор главного модуля может также отслеживать текущие статусы сегментов. Кроме всех перечисленных функций, в главном модуле системы доступны все функции, которые доступны прочим операторам-стенографистам: прослушивание звукового сигнала, редактирование текстового представления сегмента, изменение скорости воспроизведения и т.д.

На рабочем месте оператора-стенографиста доступны функции: получения по сети следующего доступного сегмента от главного модуля, прослушивание сегмента, редактирование текстового представления сегмента, отправка результата на сервер. Для повышения продуктивности работы операторам предоставляется возможность автоматической проверки орфографии, автозамены и автоподстановки, а также автоматических всплывающих подсказок для длинных слов.

Как главный модуль, так и рабочие места операторов могут быть установлены на обычных компьютерах под управлением ОС Microsoft Windows XP, Vista или Windows 7.

Распределение сегментов между операторами-стенографистами осуществляется следующим образом: при подключении к главному модулю системы оператор получает следующий необработанный сегмент из очереди. При этом на главном модуле системы такой сегмент помечается как «находящийся в обработке», и он становится заблокированным и недоступным для других пользователей системы. После завершения обработки фрагмент может принять один из статусов «завершён успешно» или «при обработке возникла проблема». Сегмент также может стать доступным автоматически для других операторов системы в случае отключения оператора, который его заблокировал.

## Выводы

После реализации прототипа системы был проведён эксперимент, целью которого было сравнение эффективности работы одного стенографиста и группы стенографистов при использовании предложенной системы и без такового, т.е. при использовании лишь традиционных средств (таких как Windows Media Player для воспроизведения звукового сигнала и Microsoft Office Word для набора текста стенограммы). В качестве входящего речевого сигнала для эксперимента была выбрана запись защиты докторской диссертации продолжительностью около 2 часов. В результате эксперимента получены следующие результаты:

- При работе одного оператора-стенографиста на стенографирование записи с использованием предложенной системы было потрачено около 4 часов. Для стенографирования этой же записи при использовании стандартных средств оператор тратит в среднем 12–16 часов.

- Группе стенографистов из 5 человек для обработки записи понадобилось около 40 минут, после чего полученный текст стенограммы был направлен на обработку корректору. Вместе с коррекцией в общей сложности расшифровка стенограммы заняла около 1 часа.

Проведённый эксперимент демонстрирует эффективность предложенной системы в сравнении с традиционными средствами. Вместе с остальными преимуществами, среди которых отсутствие необходимости администрирования и внедрения, простота в использовании и качественная предварительная обработка сигнала, предложенная система является достаточно эффективным средством для автоматизации процесса создания и расшифровки стенограмм заседаний для небольших и крупных организаций, а также для индивидуальных пользователей.

## Литература

1. Морозов А.О. «Рада-3» — система підтримки прийняття рішень для законотворчого процесу Верховної Ради України та рад інших рівнів. / А.О. Морозов, Л.Б. Баран, В.В. Копейчиков, В.Л. Косолапов // Математичні машини і системи. 2008. № 1. С. 3–22.
2. Система стенографирования SRS Report. Сайт компанії SRS. Електронний ресурс. Режим доступа: [http://srs.kiev.ua/index.php?option=com\\_content&view=article&id=43%3Asrs-report&catid=3%3A2009-07-15-06-10-55&Itemid=13&lang=ru](http://srs.kiev.ua/index.php?option=com_content&view=article&id=43%3Asrs-report&catid=3%3A2009-07-15-06-10-55&Itemid=13&lang=ru)
3. Система распределённого компьютерного документирования устной речи Нестор. Общее описание системы. Электронный ресурс. Режим доступа: <http://www.speechpro.ru/sites/default/files/product/docs/description.pdf>
4. Miller G. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. // The Psychological Review. 1956. Vol. 63. P. 81–97.
5. Rabiner L. A comparative performance study of several pitch detection algorithms. / L.R. Rabiner, M.J. Cheng, A.E. Reseonberg, C.A. McGonegal // IEEE Tran. on. Acoustics, Speech and Signal Processing. № 24(5), 1976. P. 399–418.

## Сведения об авторе

### **Юрий Васильевич Крак —**

доктор физико-математических наук, профессор Киевского национального университета имени Тараса Шевченко, старший научный сотрудник института кибернетики им.В.М.Глушкова НАН Украины. Специалист в области искусственного интеллекта, анализа и синтеза голосовой и жестовой коммуникационной информации. Автор около 300 научных работ.

### **Александр Сергеевич (Семёнович) Загваздин —**

закончил факультет кибернетики в 2007 году. Научный сотрудник института кибернетики им. В.М.Глушкова НАН Украины. Круг научных интересов включает цифровую обработку речевых сигналов, синтез и распознавание речи.