ОЦЕНКА ФУНДАМЕНТАЛЬНОЙ ПАРАЛЛЕЛЬНОСТИ В СИСТЕМАХ ОБРАБОТКИ ГОЛОСОВЫХ СИГНАЛОВ

Рождественский Ю.В., Дьяченко Ю.Г

Необходимость в проведении исследований скрытой параллельности в системах обработки голосовых сигналов возникла в процессе разработки архитектуры динамического многопоточного процессора цифровой обработки сигналов (ЦОС), ориентированного на решение, в том числе, и задач речевой технологии. К таким задачам относятся:

- сжатие и декомпрессия речевого сигнала,
- распознавание изолированных слов (команд),
- распознавание личности говорящего.

Очевидно, что максимальная эффективность архитектурного решения будет достигаться в том случае, если алгоритмическое распараллеливание вычислительных процессов будет соответствовать реальной (фундаментальной) параллельности модели речевого сигнала.

Поскольку функция сигнального процессора заключается в непосредственном формировании и восприятии звуковой составляющей речевого сообщения, объектами для изучения параллельности следует считать голосовой тракт и слуховой аппарат человека.

Сравнительный анализ этих двух объектов речевого обмена показывает, что ведущим является слуховой аппарат. Человеческий голосовой тракт способен формировать различные звуки, но только определенные их компоненты составляют человеческую речь. Именно слуховой аппарат производит первичный звуковой анализ и преобразует речевой сигнал в последовательность нейроимпульсов, формирующую речевой образ в мозгу человека.

Поэтому поиск скрытой, естественной параллельности был адресован к психоакустическим моделям слухового аппарата.

Исследования в этой области [1] указывают на то, что основным органом, осуществляющим преобразование звуковых колебаний в нервные импульсы, является расположенная в улитке внутреннего уха базилярная

мембрана с размещенным на ее внутренней стороне органом Корти. Базилярная мембрана состоит из нескольких тысяч поперечных волокон и имеет длину 32 мм. Она расположена в камере, заполненной жидкостью, и имеет увеличивающуюся ширину и сечение по мере продвижения от внутренней части к концу ушной улитки. Под воздействием звуковых колебаний, передаваемых через механизмы внешнего и среднего уха, в жидкости возникает бегущая волна, образующая зону пучности на базилярной мембране (рис.1). Расположение зоны пучности (максимального изгиба) определяется высотой звукового тона, а ее величина – интенсивностью звукового сигнала. Таким образом, осуществляется "спектральный анализ" входного звукового сигнала – размещением волн по длине мембраны.

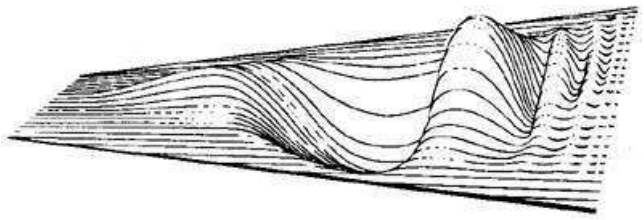

Рис.1. Бегущая волна

В органе Корти, представляющем собой гребенку нервных волокон, окруженных волосяными клетками, происходит непосредственное возбуждение нервных волокон слухового нерва волосяными клетками в области деформации базилярной мембраны. Интенсивность возбуждения пропорциональна степени деформации базилярной мембраны с ограничением примерно 1000 импульсов в сек.

Описанная структура позволяет характеризовать слуховой аппарат человека как анализатор спектра звуковых колебаний, преобразующий высоту тона звука в возбуждение определенных нервных волокон, позиционированных на базилярной мембране, а интенсивность – в частоту следования импульсов возбуждения.

Дальнейшее изучение этого механизма восприятия звука привело к обнаружению эффекта "маскирования". Суть этого эффекта состоит в том, что из двух звуков большей и меньшей интенсивности, близко расположенных в частотной области, слабый звук не воспринимается человеком (маскируется

сильным звуком). Исследования, проведенные по маскированию тонального звука "белым" шумом [2], позволили установить ширину "критической" полосы частот, в которой наблюдается этот эффект. Оказалось, что "критической" полосе частот соответствует линейный участок базилярной мембраны длиной 1 – 2 мм, а ширина "критической" полосы зависит от значения маскируемой тональной частоты (рис.2). В пределах этого участка человек воспринимает только высоту тона наиболее мощной компоненты сложного сигнала, а остальные его составляющие либо не слышны, либо влияют только на тембр тона наиболее мощной компоненты.

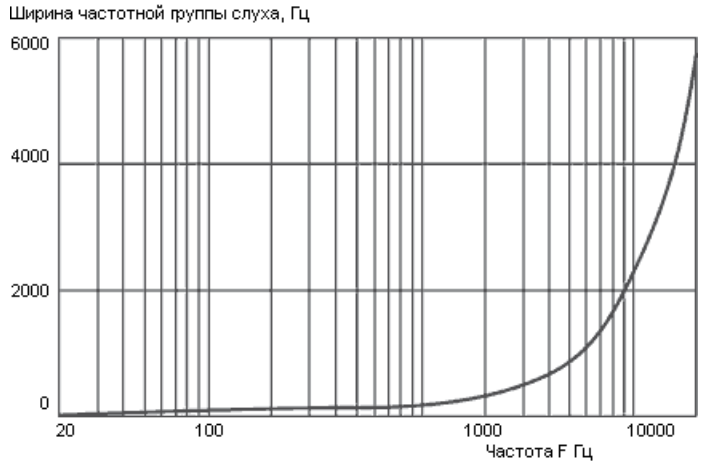

Рис.2. Зависимость ширины "критической" полосы от значения маскируемой тональной частоты

Процессы "маскирования" происходят в высших разделах головного мозга и механизм этого явления до сих пор до конца не изучен, однако он позволил создать модель слухового аппарата человека в виде набора полосовых фильтров с определенными характеристиками. Количество таких фильтров определяется физиологией слухового аппарата человека. На базилярной мембране умещаются 24 зоны, соответствующие "критическим" полосам частот. Внутри каждой такой зоны интенсивности всех компонент сложного сигнала суммируются и приписываются тону, центрированному в соответствующей "критической" полосе частот.

Вне этой полосы звуки воспринимаются как самостоятельные. В результате сложный речевой сигнал воспринимается как состоящий из нескольких звуков. Следовательно, модель слухового аппарата человека хорошо описывается линейкой из 24 полосовых фильтров (рис.3). Здесь $W_{нсу}$ – линейная

передаточная функция, моделирующая органов внешнего и среднего уха; W_i , i = 1,…,24 – передаточные функции критических зон внутреннего уха, последовательно расположенных на базилярной мембране; $\int_i$ – блоки интегрирования энергии сигнала в критической полосе. В общем случае фильтры W_i нелинейные и аппроксимируются параметрическими линейными фильтрами [3].

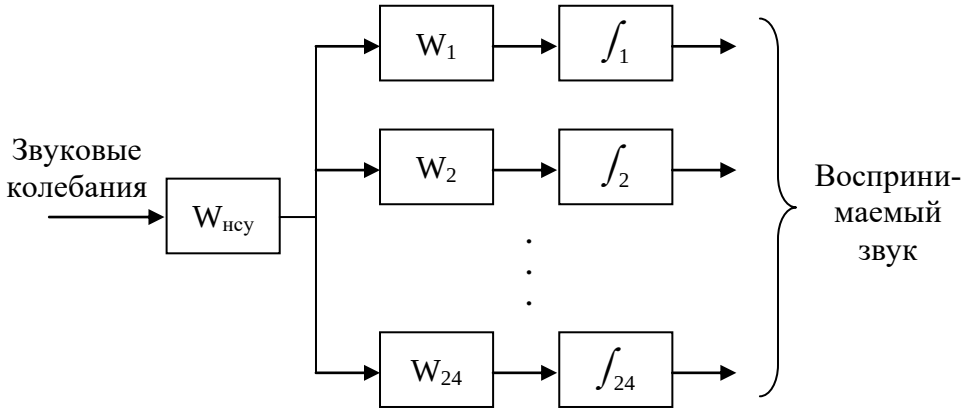

Рис.3. Модель слухового восприятия человеком уровня звукового сигнала

Структуры программ распознавания речи, личности говорящего и оценки качества компрессии/декомпрессии речи имеют в своем составе общий блок (рис.4). Этот блок, выполняющий функции слухового аппарата человека, преобразует звуковые колебания, порожденные голосовым трактом человека в конкретные цифровые параметры и характеристики звуков, дискретизированные во времени. (Частота дискретизации звуков обычно связана с минимальной частотой колебания голосовых связок человека и выбирается в диапазоне 10 – 35 мсек.)

Большинство систем распознавания в настоящее время имеют коммерческое применение и работают совместно с телефонными каналами передачи голосовых сообщений. Это системы распознавания банковских клиентов, идентификационных пин-кодов, информационные системы о наличии товаров, расписаний рейсов и т.д. Для них характерен ограниченный частотный диапазон звукового сигнала (300 Гц – 3.6 кГц), что соответствует 15 – 16 критическим полосам в аппарате звукового восприятия. Объемы моделей в этом случае также невелики и до 80% вычислительных ресурсов приходится на блок распознавания звука. Поэтому процессор ЦОС с параллельной архитектурой

для эффективного решения задачи распознавания слов или идентификации личности, работающий с телефонными средствами передачи голосовых сообщений, должен иметь архитектурное решение, поддерживающее параллельную обработку не менее чем 16 процессов.

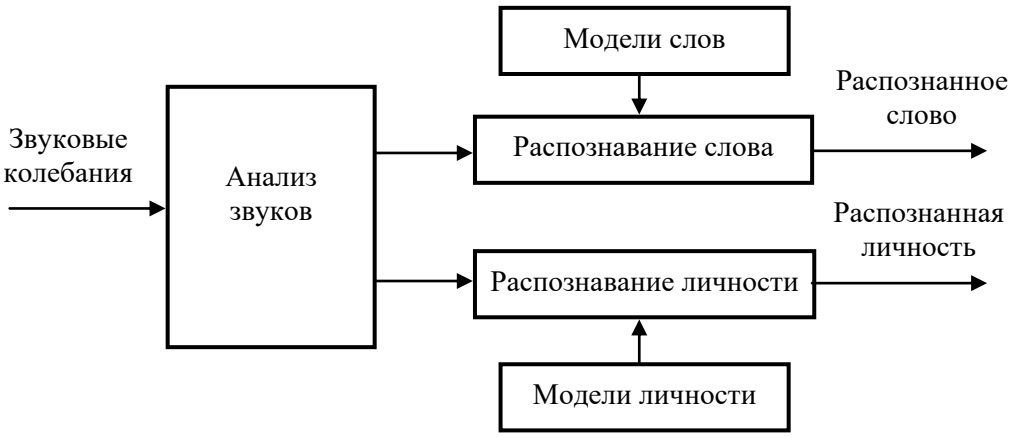

Рис.4. Блок-схема алгоритмов распознавания речи и личности говорящего

Выводы.

В структуре человеческой речи на психофизиологическом уровне присутствует параллельность звукового восприятия, равная 24.

Архитектуры процессоров ЦОС для обработки речевой информации в телефонных телекоммуникационных каналах должны предполагать эффективную обработку 16 параллельных вычислительных процессов.

Литература:

1. Алдошина И., "Основы психоакустики. Часть 1" / "Звукорежиссер", 1999, N6.
2. Алдошина И., "Основы психоакустики. Часть 12" / "Звукорежиссер", 2000, N9.
3. B.C.J. Moore, B.R. Glasberg and T. Baer, "A model for the prediction of thresholds, loudness and partial loudness" / J. Audio Eng. Soc. 45, 1997, pp. 224-240.