Адаптивный алгоритм принятия решения «ТОН–НЕ ТОН», синхронный с основным тоном

И.А. Архипов,
кандидат технических наук

В.Б. Гитлин,
доктор технических наук

Д.А. Лузин

Признак «ТОН–НЕ ТОН» (Т/НТ) указывает на наличие или отсутствие вокализации в речевом сигнале. Он определяет способ образования звука [1] и служит одним из признаков параметрического описания речи. Его точная оценка необходима в системах анализа и синтеза речи [1], [2], [3].

Основными признаками, на основе знания которых принимается решение Т/НТ, служат следующие признаки [4].

1. *Энергия звука в различных областях спектра*: для вокализованных звуков она сосредоточена в низкочастотном диапазоне, для невокализованных — в высокочастотном. Энергия вокализованных звуков сконцентрирована в формантных областях, энергия невокализованных — распределена по спектру более равномерно [1], [2], [3].

2. *Энергия вокализованных звуков пульсирует с частотой основного тона* (ОТ), невокализованных — более равномерна, кроме взрывных /п/, /т/, /к/ и аффрикат /ц/, /ч/ [5], [6].

3. *Распределение вероятностей* мгновенных значений сигнала невокализованных звуков близко к гауссовскому закону, распределение для вокализованных звуков отлично от гауссовского. Отсчёты вокализированного сигнала существенно коррелированы между собой, корреляция отсчётов невокализированного сигнала слабее [4], [7].

4. *Частота пересечений нуля* сигналом вокализованных звуков ниже частоты пересечений нуля сигналом невокализованных звуков [1]. В общем слу-

чае частота пересечений нуля не служит надёжным признаком для принятия решения Т/НТ [4]. Это вызвано низкой помехоустойчивостью этого признака, широкой изменчивостью параметров фонового шума, большой зоной перекрытия распределений частоты переходов через нуль двух рассматриваемых классов («ТОН», «НЕ ТОН») [2].

Энергия вокализованных звуков выше энергии невокализованных звуков и пауз [1]. Алгоритмы принятия решения Т/НТ по энергии сигнала с фиксированным порогом имеют относительно низкую надёжность, поскольку принятие решения в существенной мере зависит от уровня сигнала и уровня шума [4]. Уровни сигнала и шума не остаются постоянными даже во время произнесения достаточно короткого текста [8]. Динамический диапазон акустического сигнала речи может достигать 80 дБ [1]. Для компенсации изменений сигнала по амплитуде используют адаптивный порог или нормализацию речевого сигнала [1].

Принятие решения по энергии в некоторой полосе частот, составляющей часть от полного спектра сигнала, позволяет учесть способ образования звука и тем самым повысить надёжность принятия решения [4], [9]. Однако ряд фрикативных и аспирированных шумных звуков, например, /ф/, /х/, имеют довольно мощные составляющие в низкочастотной части спектра, что может вызвать сбои систем принятия решения по энергии в полосе частот [4].

В работах [10], [11] Атал и Рабинер исследовали следующие признаки: нормированный коэффициент корреляции с единичной задержкой $R(1)$, первый коэффициент модели линейного предсказания a_1 при числе полюсов M=12 в ковариационном методе линейного предсказания и нормализованную ошибку линейного предсказания E_p.

Для вокализованной речи [11] $R(1)$ близко к единице, для невокализованной речи и шума $R(1)$ близко к нулю. Первый коэффициент линейного предсказания a_1 связан с $R(1)$ и зависит от порядка модели M, т.е. от формантной структуры звука. Нормализованная ошибка линейного предсказания E_p отражает степень близости спектра сигнала к спектру белого шума: чем спектр равномернее, тем ошибка больше. Для вокализованной речи E_p меньше, для невокализованной — больше.

Атал и Рабинер в работе [11] делают следующие выводы.

1. При принятии решения Т/НТ дополнительно появляются ошибки на интервалах паузы из-за изменчивости фонового шума, который различен для обучающей и контрольной выборок.
2. Большинство ошибок появляются на границе между классами. Ошибки возникают в случае, когда внутрь одной рамки анализа попадают два разных класса звуков.

Периодичность сигнала, связанную с основным тоном, можно оценить по виду спектра. Спектр вокализованных звуков неравномерен и концентрируется на гармониках. Спектр невокализованных звуков более равномерен [4]. Недостаток оценки периодичности сигнала по виду спектра — низкая помехоустойчивость, поскольку искажения и фоновые шумы могут существенно исказить истинный спектр [10].

Можно принимать решение Т/НТ по оценке периодичности сигнала путём перехода к анализу колебательности временной функции. Однако по данным работы [4] оценка степени колебательности временной функции речи не обеспечивает надёжного принятия решения Т/НТ.

В процессе выделения основного тона довольно часто вычисляют функции, которые могут служить мерой оценки периодичности, связанной с ОТ. Такими функциями могут быть:

значение максимума автокорреляционной функции, значение минимума разностной функции, величина пика кепстра и ряд других [12] [13], [14]. Недостатком данного способа принятия решения Т/НТ является зависимость указанных параметров от формантной структуры сигнала, от длины кадра анализа, от величины фонового шума и от ряда других факторов [7].

Повысить надёжность принятия решения Т/НТ можно путём увеличения количества признаков, по которым принимают решение. Повышение надёжности возможно в том случае, когда признаки независимы или, по крайней мере, слабо коррелированы относительно ошибок принятия решения Т/НТ [16]. Если решение Т/НТ принимают в многомерном пространстве признаков, то процедура принятия решения существенно усложняется, отсутствует наглядность представления распределений признаков, необходимо увеличение обучающей выборки. Для упрощения этой процедуры можно использовать методы теории распознавания образов [16]. Выбранная система признаков должна в совокупности обеспечить необходимую надёжность принятия решения при минимальной стоимости принятия решения.

Сегментацию речи на тональные интервалы выполняют синхронно [20], [21] и асинхронно с ОТ [1]…[3]. Асинхронная с ОТ обработка предполагает фиксированный размер кадра анализа. Согласно [17] оптимальная длительность интервала усреднения для энергии равна 10 мс. Текущий кадр анализа располагается случайным образом, и возможно попадание участков с разным типом возбуждения речевого тракта в один кадр. Решение о принадлежности данного кадра к какому-либо способу возбуждения во многом зависит от соотношения длительностей участков с разным способом возбуждения, попавших в данный кадр. На рис.1 показаны обобщённые схемы формирования признака Т/НТ синхронно и асинхронно с ОТ. На рис. 1*а* исходный сигнал сегментируют на тональные интервалы, а затем только тональные интервалы подвергают выделению ОТ. При сегментации речи асинхронно с ОТ кадры анализа имеют длительность, превышающую длительность периода ОТ, и следуют с перекрытием.

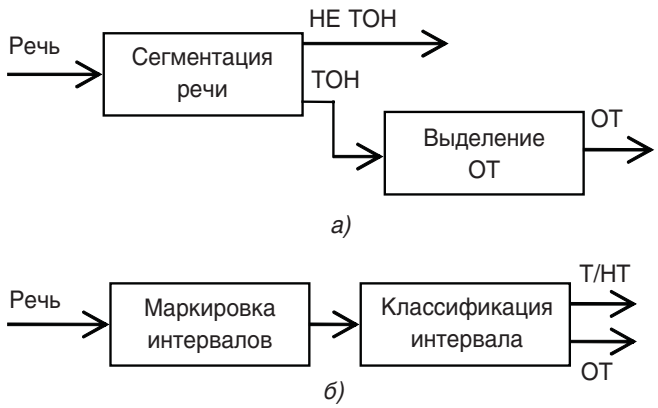

Рис.1. Способы классификации речи по признаку Т/НТ: а) асинхронно с ОТ; б) синхронно с ОТ

В обработке синхронной с ОТ кадры анализа привязаны к периодам ОТ. Привязка интервалов анализа к периодам ОТ позволяет избежать указанную выше неопределённость в расположении кадра анализа. Под кадром анализа здесь следует понимать участок сигнала между соседними марками, соответствующими началам новых периодов ОТ. Длительность каждого тонального интервала можно принимать за оценку периода ОТ. Кадры анализа следуют без перекрытия, за счёт чего существенно повышается скорость обработки.

Для простановки марок в началах периодов ОТ без предварительной сегментации на вокализированные и невокализированные интервалы необходимо использовать

локальный алгоритм выделения ОТ, в качестве которого выбран алгоритм, работающий по методу GS [18]. Синхронный с ОТ анализ ограничивает набор признаков, которые могут быть использованы для принятия решения Т/НТ, только такими, интервал вычисления которых может быть равен периоду ОТ (от 2 мс до 20 мс [1]). По этой причине из набора признаков, указанных выше, взяты только три признака [19]: нормированный коэффициент корреляции с единичной задержкой $R(1)$, логарифм частоты пересечения нулевого уровня и логарифм энергия сигнала в полосе частот 20…1500 Гц.

Нормированный коэффициент корреляции с единичной задержкой определяли следующим образом:

$$(\mathbf{R}\ 1) = \mathbf{K}_{\mathrm{r}} \cdot \left(1 + \sum_{\mathrm{i}=0}^{\mathrm{N}-2} \mathbf{S}_{\mathrm{i}} \cdot \mathbf{S}_{\mathrm{i}+1} \Big/ \sum_{\mathrm{i}=0}^{\mathrm{N}-1} \mathbf{S}_{\mathrm{i}}^{\ 2}\right), \qquad (1)$$

где $K_{r\_}$ нормирующий множитель, S_i — отсчёт входного речевого сигнала, не прошедшего этап предварительной обработки, N — число отсчётов на анализируемом периоде ОТ. Эксперименты показывают, что паузы в речи обычно заполнены слабыми, относительно случайными колебаниями, спектр которых зависит от спектра фонового шума. Поведение функции $R(1)$ в данном случае непредсказуемо.

На рис. 2*а*, 2*б* представлены осциллограмма слова «четыре» и функция $R(1)$ данного произнесения. На рис. 2*б* тональный и шумовой участки можно надёжно разделить по значениям функции $R(1)$. Поведение функции $R(1)$ на паузе (между марками 3–4) нестабильно и не позволяет классифицировать этот сегмент как невокализированный. Для лучшего разделения паузы и вокализованного сигнала по $R(1)$ необходимо приблизить спектр паузы к спектру невокализованных звуков. Для этой цели в работах [7], [11] предложено смешивать сигнал с шумом определённого уровня с подъёмом в сторону высоких частот.

Проведены эксперименты по оценке надёжности принятия решении Т/НТ по $R(1)$.

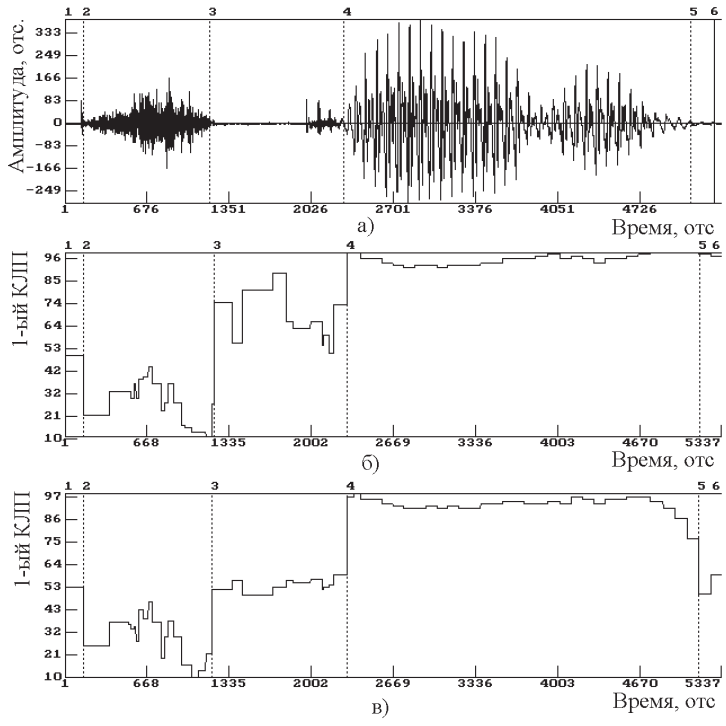

Рис. 2. Нормированный коэффициент корреляции с единичной задержкой:
а) осциллограмма слова «четыре»; б) функция нормированного коэффициента корреляции с единичной задержкой; в) функция нормированного коэффициента корреляции с единичной задержкой, вычисленного при добавлении шума с размахом 20 отсчётов

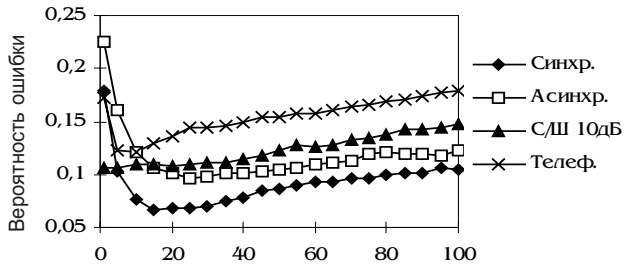

Рис. 3. Зависимость вероятности ошибки классификации Т/НТ по функции R(1) от уровня добавляемого шума

В качестве речевого материала использовали по одному произнесению фраз «Не видали мы такого невода», «Саша кусал сало», «На ухабе» и «Жирные сазаны ушли под палубу». В эксперименте принимали участие 12 дикторов (6 мужчин и 6 женщин).

Испытания проводили для чистого сигнала, для сигнала с аддитивным шумом при отношении С/Ш=10дБ и для сигнала, ограниченного полосой телефонного канала 300...3400Гц. Первоначально все фразы были вручную сегментированы на вокализированные и невокализированные сегменты.

К каждому произнесению был добавлен шум интенсивностью от 0 до 100 отсчётов уровней квантования с шагом через 5 отсчётов при максимуме сигнала в 32768 отсчётов. Нулевая интенсивность соответствует отсутствию шума. Результаты эксперимента показаны на рис. 3. Ошибку принятия решения Т/НТ определяли для синхронного с ОТ и асинхронного с ОТ методов принятия решения Т/НТ. Для телефонного сигнала и сигнала с аддитивным шумом при С/Ш=10 дБ анализ проводили синхронно с ОТ.

Кривая для синхронного способа вычисления признака $R(1)$ при всех уровнях добавляемого шума проходила ниже асинхронной кривой. Область минимума ошибки принятия решения Т/НТ была близка для всех типов исследованных сигналов, кроме сигнала с аддитивным шумом при С/Ш=10 дБ. В среднем, при добавлении оптимального значения шума синхронный с ОТ анализ по сравнению с асинхронным позволяет снизить вероятность суммарных ошибок классификации на 11%.

Энергия вокализованных звуков, как правило, выше энергии невокализованных звуков и пауз. Значение энергии определяли по формуле:

$$E = K_e \cdot \lg\left(\mathrm{e} + \sum_{i=1}^{N} x_i^{\,2}\right), \qquad (2)$$

где x_i — отсчёт речевого сигнала на выходе фильтра низких частот (ФНЧ) с частотой среза f_c, а K_E — нормирующий множитель.

На рис. 4. представлены осциллограммы произнесения слова «четыре», функция энергии исходного произнесения и функция энергии исходного произнесения, прошедшего через ФНЧ с частотой среза 1000 Гц. Энергию вычисляли синхронно с ОТ. Участок сигнала между марками 2–3 соответствует шумовому звуку «ч». Из рис. 4*б* видно, что звук «ч» имеет энергию, сравнимую с энергией вокализованных звуков. На рис. 4*в* энергия звука «ч» в значительной степени подавлена фильтром нижних частот. В данном случае можно легко отделить шипящий звук «ч» от вокализованных звуков. Эксперименты показывают, что с ростом частоты среза ФНЧ для значений, превышающих 1000 Гц, вероятность ошибки класси-

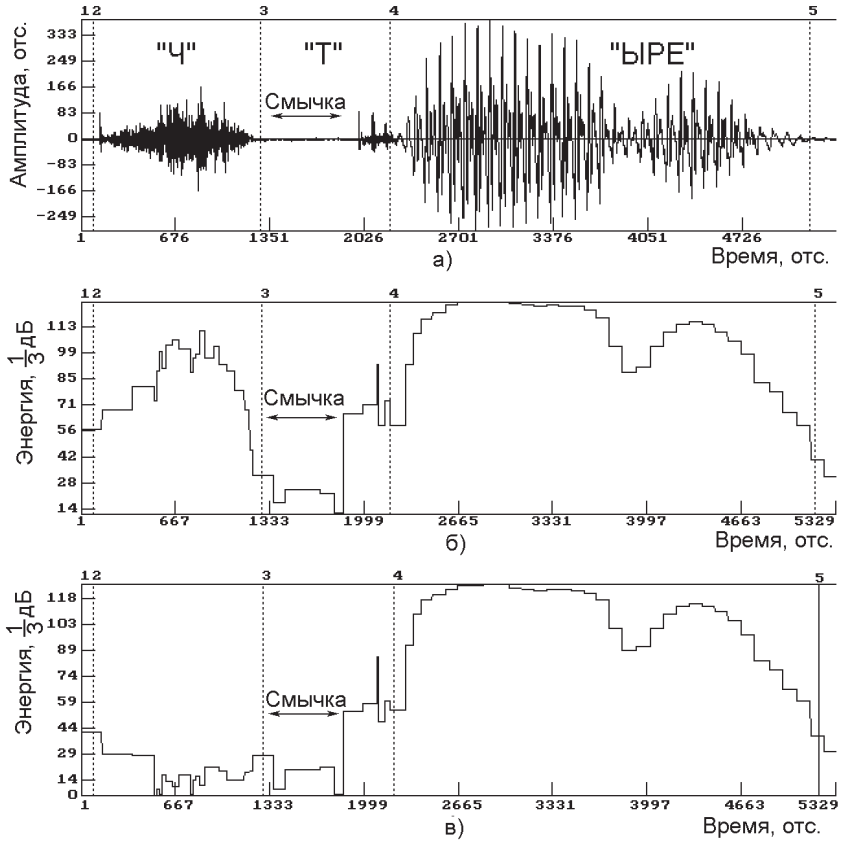

Рис. 4. Слово «четыре» (диктор — мужчина):
а) осциллограмма исходного произнесения; б) функция энергии исходного произнесения; в) функция энергии исходного произнесения, прошедшего ФНЧ с частотой среза 1000 Гц

фикации Т/НТ медленно монотонно возрастает. В последующих экспериментах мы ограничились частотой среза ФНЧ f_c=1500 Гц, определяемой требованиями алгоритма GS [22].

Вычисление энергии синхронно с ОТ приводит к снижению вероятности суммарной ошибки классификации по сравнению с асинхронным способом вычисления. Суммарная вероятность ошибки снижается на величину от 1,5% до 3,3% при минимальной ошибке классификации по энергии около 10% в зависимости от ширины полосы частот, в которой вычисляют энергию.

Частота пересечений нулевого уровня сигналом (ЧПН) имеет большой динамический разброс значений [1], [2], [23], вследствие чего предпочтительно в качестве признака классификации Т/НТ использовать логарифм частоты пересечения через ноль (ЛЧПН):

$$Z_{cr} = K_z \lg(M/T_0), \qquad (3)$$

где K_z — нормирующий коэффициент; M — количество пересечений нулевого уровня на периоде основного тона.

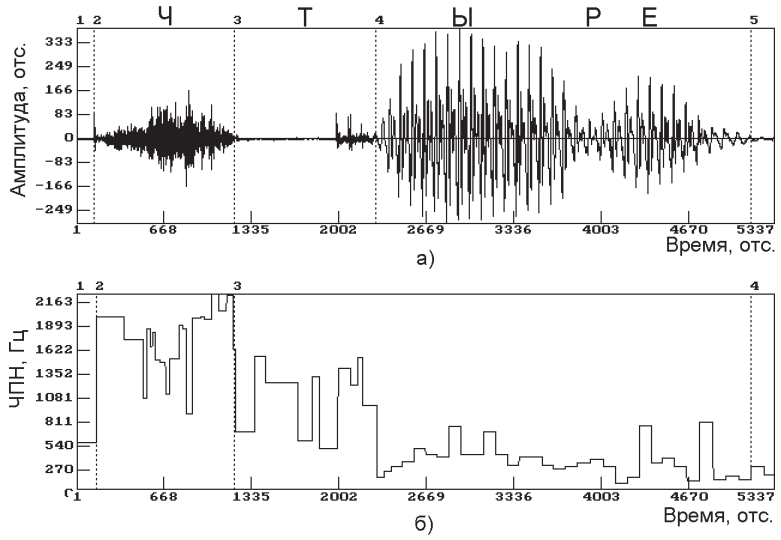

Рис. 5. Частота пересечения нулевого уровня речевого сигнала: а) осциллограмма слова «четыре» (диктор — мужчина); б) ЧПН сигнала

На рис. 5 изображены осциллограмма изолированного слова «четыре» и соответствующий ей график ЧПН. Марки 3, 4, 5 и 6 установлены на границах интервалов вокализации. Частота пересечений нуля вокализованных звуков ниже частоты пересечений нуля невокализованных звуков.

Из рис. 5*б* видно, что график признака ЧПН значительно изрезан, как на вокализированном, так и на невокализированном участках. Изрезанность графика ЧПН говорит о том, что короткие интервалы анализа при синхронном с ОТ способе вычисления ЧПН недостаточно сглаживают значения ЧПН, что приводит к указанному выше расширению динамического диапазона значений ЧПН. Распределения ЧПН вокализированных и невокализированном интервалов перекрываются даже на стационарных интервалах.

На рис. 6 представлены гистограммы распределений ЧПН и ЛЧПН вокализированных и невокализированных интервалов без добавления шума. По гистограммам видно, что диапазон возможных значений функции ЛЧПН значительно уже значений функции ЧПН. Гистограммы вокализированных и невокализированных интервалов в значительной степени перекрываются, причём область перекрытия для ЛЧПН меньше, чем для ЧПН.

Вероятность ошибки классификации для логарифмического масштаба частот пересечения нуля оказалась на 10–15% меньше, чем для линейного. Вероятность ошибки классификации Т/НТ по ЛЧПН для разных типов сигнала и различных дикторов изменялась в пределах 11%...21%. Добавление шума, подобно добавлению шума к признаку $R(1)$, несколько снижало ошибку классификации Т/НТ для чистого сигнала. Для других типов сигнала добавление шума практически не влияло на надёжность принятия решения Т/НТ. Выбирая уровень добавляемого шума при вычислении ЛЧПН, следует придерживаться тех

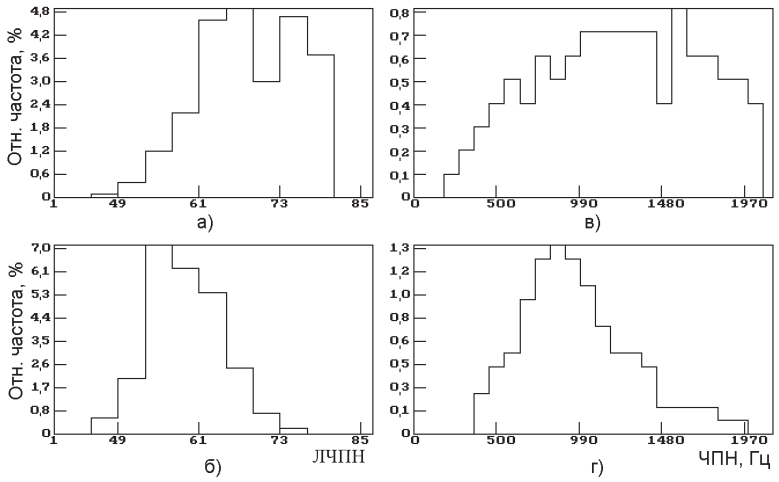

Рис. 6. Гистограммы распределений ЧПН и ЛЧПН:
а) невокализованные интервалы (ЛЧПН); б) вокализованные интервалы(ЛЧПН);
в) невокализированные интервалы (ЧПН); г) вокализированные интервалы (ЧПН)

же рекомендаций, что и при вычислении признака *R*(1). В обоих случаях можно использовать единый генератор шума (для речи без искажений z=30 отс.; для телефонной речи z=15 отс.; для зашумлённой речи добавление шума нецелесообразно). Различия в поведении вероятности ошибки классификации были незначительны для синхронного с ОТ и асинхронного с ОТ способов вычисления признака ЛЧПН. С этой точки зрения, не имеет значения, каким способом вычислять ЛЧПН — синхронно или асинхронно с ОТ.

Принятие решения Т/НТ по совокупности признаков в многомерном пространстве признаков лишено наглядности представления распределений и требует больших вычислительных затрат, существенно большей обучающей выборки, а также процесса переобучения при изменении условий произнесения [11]. Для упрощения процедуры классификации решено объединить три указанных выше признака в один, исходя из следующих соображений. Коэффициент *R*(1) и энергия в полосе частот имеют максимальные значения на тональных интервалах. ЛЧПН на тональных интервалах минимальна. Тогда обобщённый признак, по которому выполняют классификацию Т/НТ, может быть записан следующим образом:

$$G = \frac{R(1) \cdot E}{Z_{cr}}. \qquad (4)$$

На рис. 7 изображены осциллограмма фразы «Саша кусал сало» и соответствующий ей график обобщённого признака Т/НТ. Марки 2-11 установлены на границах вокализации. Обобщённый признак Т/НТ вокализованных звуков имеет большие значения по сравнению с признаком на невокализованных звуках.

В таблице 1 приведены значения вероятности ошибки классификации для обобщённого признака Т/НТ, а также для отдельных признаков классификации Т/НТ. Результатом объединения трёх признаков стало повышение точности классификации. Тем не менее, вероятность появления ошибки классификации остаётся достаточно высокой (см. табл. 1). Повышения точности распознавания можно достичь путём привлечения дополнитель-

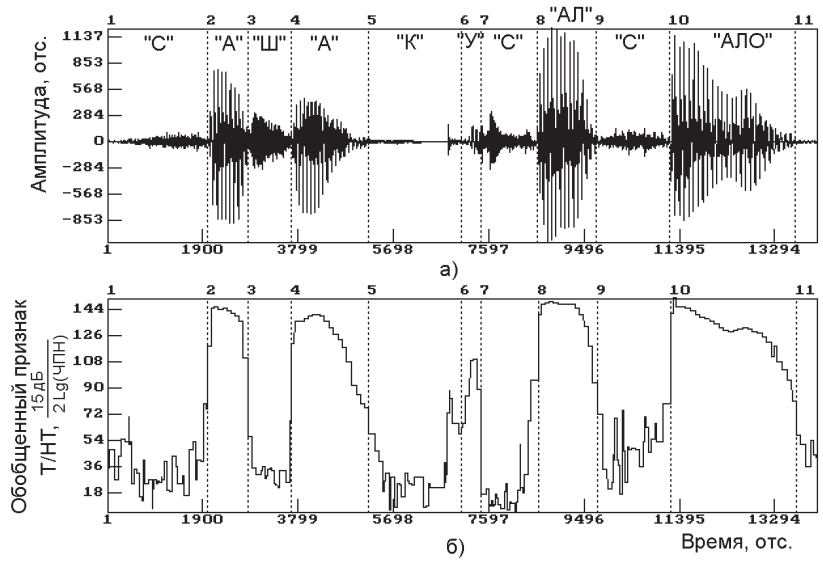

Рис. 7. Обобщённый признак Т/НТ речевого сигнала:
а) осциллограмма фразы «Саша кусал сало» (диктор — мужчина);
б) обобщённый признак Т/НТ сигнала

Таблица 1

Параметры классификации речи по коэффициенту R(1), энергии в полосе частот, ЛЧПН и обобщённому признаку для разных способов их вычисления

Способ вычисления признака	Признак классификации	Вероятность ошибки классификации
Чистый сигнал синхронно с ОТ	Коэффициент R(1)	0,0695
	Энергия	0,0735
	ЛЧПН	0,1135
	Обобщённый	0,059
Чистый сигнал асинхронно с ОТ	Коэффициент R(1)	0,098
	Энергия	0,104
	ЛЧПН	0,1125
	Обобщённый	0,104
Телефонный сигнал синхронно с ОТ	Коэффициент R(1)	0,1445
	Энергия	0,1465
	ЛЧПН	0,204
	Обобщённый	0,121
С/Ш 10 дБ синхронно с ОТ	Коэффициент R(1)	0,111
	Энергия	0,129
	ЛЧПН	0,175
	Обобщённый	0,101

ных признаков, определяемых предысторией процесса и длительностью интервалов, классифицированных как вокализированные или невокализированные [14].

В работе [19] при принятии решения Т/НТ с помощью порогов g_0, g_1 и g_2 область значений обобщённого признака разбивали на четыре области: «уверенно НЕ ТОН», «неуверенно НЕ ТОН», «неуверенно ТОН», «уверенно ТОН». Пороги g_0 и g_2 устанавливали так, что вероятности попадания вокализированного звука в невокализированную область и невокализированного звука в вокализированную не превышала 2%. При неопределённом решении о вокализации дополнительную информацию извлекали из априорных данных и известных значений длительностей предполагаемых периодов ОТ. Области «неуверенно НЕ ТОН», «неуверенно ТОН» относили к вокализированным или к невокализированным в ходе последующей обработки. Порог g_1, разделяющий области «неуверенно НЕ ТОН», «неуверенно ТОН», устанавливали из условия минимума вероятности суммарной ошибки классификации с учётом последующей обработки.

В таблице 2 представлены значения порогов классификации g_1, g_0 и g_2 для разных условий вычисления обобщённого признака. Значения порогов зависели от типа сигнала, а также от диктора и отдельных произнесений сигнала. Такая зависимость требует подстройки значений порогов для конкретных произнесений. Подобный способ установки порогов не способен учесть все возможные изменения произнесений и окружающей диктора обстановки. По этим причинам принято решение выполнять классификацию Т/НТ за два прохода.

Таблица 2

Значения порогов классификации

	Порог g0	Порог g1	Порог g2
Чистый сигнал синхронно с ОТ	67	76	97
Чистый сигнал асинхронно с ОТ	71	87	128
Телефонный синхронно с ОТ	48	65	143
С/Ш 10 дБ синхронно с ОТ	74	86	134

В предлагаемой модификации алгоритма на первом проходе вычисляют значение обобщённого признака G по формуле (4) для каждого периода ОТ. Эту процедуру выполняют как на вокализированных, так и на невокализированных участках речевого сигнала. На невокализированных участках сигнала за интервал анализа принимают интервал между двумя марками, проставленными алгоритмом *GS* [18] случайным образом. После окончания первого прохода для всего произнесения в целом строят гистограмму значений признака G (рис. 8) и вычисляют среднее значение признака G_t для данного произнесения. Эксперименты показывают, что величину G_t можно принять за первоначальную оценку границы между значениями обобщённого признака, соответствующими вокализированным ($G>G_t$) и невокализированным ($G<G_t$) интервалам речевого сигнала.

Для интервала значений $G<G_t$ (предположительно невокализированные звуки) вычисляют среднее значение обобщённого признака G_{uv} и среднеквадратическое отклонение σ_{uv}. Аналогично, для предположительно вокализированных звуков ($G>G_t$) вычисляют среднее значение обобщённого признака G_v и среднеквадратическое отклонение σ_v.

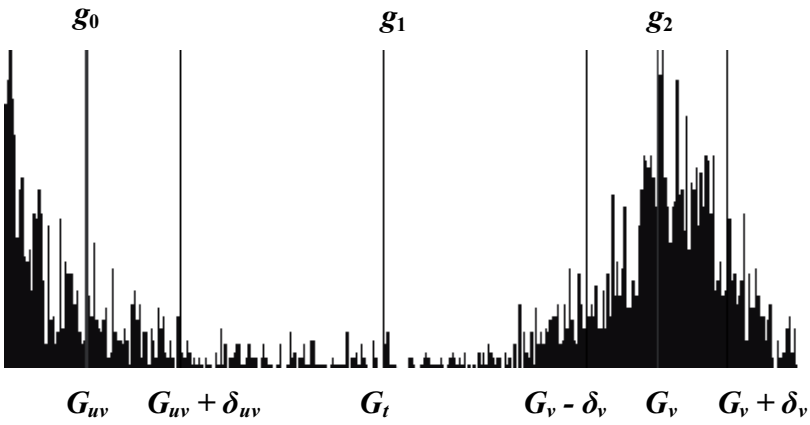

Рис. 8. Гистограмма обобщённого признака Т/НТ всех произнесений диктора AIO

Таблица 3

Значение обобщённой ошибки при различных значениях g0, g1, g2

g_0	g_1	g_2	Обобщённая ошибка
G_{uv}	G_t	G_v	2,76
G_{uv}	G_t	$G_v+\delta_v$	3,13
G_{uv}	G_t	$G_v - \delta_v$	3,08
$G_{uv}+\delta_{uv}$	G_t	G_v	4,50
$G_{uv}- \delta_{uv}$	G_t	G_v	3,54
$G_{uv} - \delta_{uv}$	G_t	$G_v-\delta_v$	3,50
$G_{uv}- \delta_{uv}$	G_t	$G_v + \delta_v$	3,57
$G_{uv} + \delta_{uv}$	G_t	$G_v - \delta_v$	4,47
$G_{uv}+\delta_{uv}$	G_t	$G_v+\delta_v$	4,57

Исследовано несколько экспериментальных правил задания значений порогов g_0, g_1, g_2. Эти правила сведены в таблицу 3. В этой же таблице показаны значения обобщённой ошибки (ОШ), получаемые двухпроходным алгоритмом для каждого из выбранных правил задания порогов. Обобщённая ошибка учитывает значения ошибок «ТОН–НЕ ТОН», ошибок «НЕ ТОН» и больших ошибок, оцениваемых путём сравнения измеренного контура ОТ с эталонным по правилу, изложенному в работе [24]. Из таблицы 3 следует, что минимальная обобщённая ошибка (ОШ=2,76%) получена в том случае, когда значения порогов g_0, g_1, g_2 устанавливали из соотношений:

$$\begin{cases} g_0 = G_{uv} \\ g_1 = G_t \\ g_2 = G_v \end{cases}. \qquad (5)$$

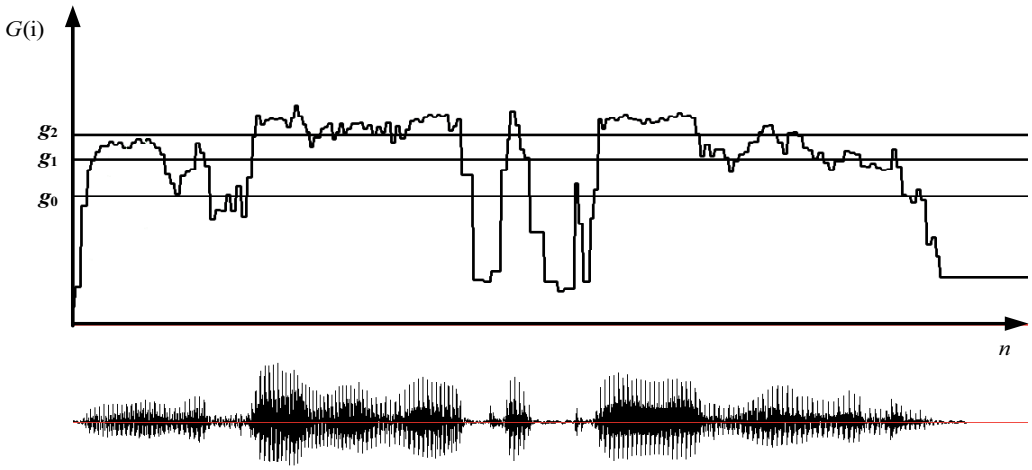

Рис. 9. Речевой сигнал (внизу) и обобщённый признак Т/НТ G(i) (вверху) с отображёнными порогами g0, g1, g2 для диктора AIO (фраза: «Не видали мы такого невода»)

На рис. 9 представлен пример осциллограммы предложения «Не видали мы такого невода» (диктор AIO); траектория обобщённого признака $G(i)$ (i — порядковый номер периода ОТ) для данного произнесения и значения порогов g_0, g_1, g_2, выбранные по правилу (5).

Окончательные решения Т/НТ получали путём коррекции предварительных решений «уверенно ТОН», «уверенно НЕ ТОН», «неуверенно ТОН», «неуверенно НЕ ТОН». При окончательном решении Т/НТ по предварительной оценке «неуверенно ТОН», «неуверенно НЕ ТОН», учитывали относительную нестабильность соседних периодов ОТ. Вокализированные участки длительностью меньше 20 мс относили к невокализированным.

В таблице 4 представлены результаты сопоставительных испытаний двухпроходного алгоритма классификации Т/НТ, совмещённого с выделителем ОТ, по методу *GS* с шестью выделителями ОТ и признака Т/НТ, реализованных в системе SIS [23]: с пиковым, фильтровым, автокорреляционным, кепстральным методами, с методом Голда-Рабинера и методом ЛЛК.

Таблица 4

Результаты испытаний алгоритмов выделения ОТ для общей группы голосов (15 дикторов, 38 произнесений)

Выделитель ОТ	Ошибка ТНТ, %	Ошибка НТТ, %	ТНТср %	Большие ошибки, %	Малые ошибки, %	Обоб. ошибка, %	Отношение ТНТср/ ОШ
Чистый сигнал							
GS2	1.97	2.37	2.17	1.70	6.16	2.76	0.79
Пиковый	0.62	7.27	3.94	1.23	10.06	4.13	0.95
Кепстральный	3.89	5.50	4.70	3.76	19.95	6.02	0.78
АКФ	1.67	21.44	11.55	2.98	10.00	11.93	0.97

Рабинер-Гоулд	1.93	8.52	5.23	2.87	9.59	5.96	0.88
Фильтровой	0.11	14.83	7.47	1.16	9.02	7.56	0.99
ЛЛК	0.64	6.29	3.47	1.08	6.63	3.63	0.95
Сигнал с аддитивным шумом С/Ш = 5 дБ							
GS2	2.01	18.18	10.10	15.58	20.72	18.56	0.54
Пиковый	1.47	36.22	18.84	7.03	24.26	20.11	0.94
Кепстральный	1.95	37.90	19.92	11.85	36.75	23.18	0.86
АКФ	2.63	36.30	19.46	3.35	13.56	19.75	0.99
Рабинер-Голда	2.10	32.57	17.33	4.00	17.23	17.79	0.97
Фильтровой	1.38	41.44	21.41	2.81	22.55	21.59	0.99
ЛЛК	1.01	45.93	23.47	2.13	20.61	23.57	1.00
Сигнал ограничен полосой телефонного канала							
GS2	0.61	14.74	7.67	6.19	6.38	9.86	0.78
Пиковый	0.90	19.50	10.20	12.08	9.89	15.81	0.65
Кепстральный	4.98	15.29	10.14	2.91	19.03	10.55	0.96
АКФ	1.06	46.60	23.83	5.36	11.06	24.42	0.98
Рабинер-Голда	2.37	19.00	10.69	28.76	5.22	30.68	0.35
Фильтровой	0.10	37.56	18.83	10.10	6.81	21.37	0.88
ЛЛК	0.10	37.56	18.83	10.10	6.81	21.37	0.88

Литература

1. Сапожков М.А. Речевой сигнал в кибернетике и связи. М.: Связьиздат, 1963. 472 с.

2. Гитлин В.Б. Основной тон речевого сигнала / Деп. В ВИНИТИ, 1998. №1206-В98. 739 с.

3. Сапожков М.А., Михайлов В.Г. Вокодерная связь М.: Радио и связь, 1983. 248 с.

4. Вокодерная телефония / Под ред. Пирогова А.А. М.: Связь, 1974. 536 с.

5. Miller N.J. Pitch detection by data reduction // IEEE Symp. speech recogn. Carnague-Mellon Univ., 1974. Contrubut Pap. P.122–130.

6. Friedman D.H. Multidimensional Pseudo-Maximum Likelihood pitch estimation // IEEE Trans. Acoust., Speech and Signal Process. 1978. Vol.26. N3. P.185–196.

7. Маркел Дж. Д., Грэй А.Х. Линейное предсказание речи. М.: Связь, 1980. 308 с.

8. De Souza P. A statistical approach to the design of an adaptive self-normalising silence detector / IEEE Trans. Acoust., Speech and Signal Process. 1983. 31. N3. P.678–684.

9. Foo S.W., and Turner L.F. Application of sub-band energy ratio to Voiced-Unvoiced-Silence classification of speech signals // Proc. MELECON'83 Mediterr. Electrotechn.Conf. Athens, 24-26, May, 1983, Vol. 2. S1. Sa. 1983. C3.05/1 — C3.05/2.

10. Atal B.S. Speech signal pitch detector using prediction error date. Pat. N 3740476 USA. G10L 1/04. 19.06.73.

11. Atal B.S., Rabiner L.R. A pattern recognition approach to voiced-unvoiced-silence classification with application to speech recognition // IEEE Trans. Acoust., Speech and Signal Process. 1976. 24. N3. P.201–202.

12. Hebid M.K., and Robinson D.M., Sincoscie W.D. Real Zeros in pitch detection // IEEE Int. Conf. Acoust., Speech and Signal Process. Record. Tulsa, Okla, 1978. New York, N.Y. 1978. P.31–34.

13. Кельманов А.В. Алгоритм классификации тон/шум, основанный на критерии адекватности модели авторегрессии // Вычислительные системы. Методы обработки информации. Новосибирск, 1978. Вып.74. С. 129–148.

14. Кельманов А.В. Алгоритм классификации тон/шум по частотным автокорреляциям // Вычислительные системы. Эмпирическое предсказание и распознавание образов. Новосибирск, 1980. Вып.83. С. 67–73.

15. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. 485 с.

16. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. 512 с.

17. Баронин С.П. Автокорреляционный метод выделения основного тона речи // Сб. тр. Гос. НИИ Министерства связи СССР. 1961. 3(24). С. 93–102.

18. Архипов И.О., Гитлин В.Б. Метод выделения основного тона на основе понятия о генерируемом солитоне // Распознавание образов и анализ изображений: новые информационные технологии. 4-я Всероссийская с международным участием конференция. РОАИ-98. 1998 г. Новосибирск, 1998. Часть 1. С. 23–27.

19. Архипов И.О., Гитлин В.Б. Формирование признака ТОН/НЕ_ТОН синхронно с основным тоном // Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. М.: ГЕОС, 1999. С. 43–46.

20. Архипов И.О., Гитлин В.Б. Добавление шума при сегментации речи на тональные участки // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск: изд. ИПМ УрО РАН, 1997. с. 63–69.

21. Архипов И.О., Гитлин В.Б. Сегментация речи по первому коэффициенту линейного предсказания синхронно с основным тоном // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск, изд. ИПМ УрО РАН, 1998. С. 17–19.

22. Архипов И.О., Гитлин В.Б. Оценка частоты среза ФНЧ, используемого для выделения основного тона // Труды научно-молодёжной школы «Информационно-измерительные системы на базе наукоёмких технологий». Ижевск: изд. ИПМ УрО РАН, 1998. С. 12–16.

23. Методические рекомендации по практическому использованию программы SIS при работе с речевыми сигналами / Центр речевых технологий. Санкт-Петербург, 1997. 394 с.

24. Архипов И.О., Гитлин В.Б. Оценка точности выделения основного тона методом GS // Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. М.: ГЕОС, 1999. С. 38–42.

Архипов Игорь Олегович,
кандидат технических наук, доцент кафедры «Программное обеспечение ЭВМ» Ижевского технического университета (426069, Ижевск, ул. Студенческая, 7).

Гитлин Валерий Борисович,
доктор технических наук, профессор кафедры «Вычислительная техника» Ижевского технического университета. E-mail: vbg_istu@mail.ru, vbg@mitm.ru.

Лузин Дмитрий Александрович,
аспирант кафедры «Вычислительная техника» Ижевского технического университета.