ОЦЕНКА ЭФФЕКТИВНОСТИ ЦИФРОВЫХ УСТРОЙСТВ ПОДАВЛЕНИЯ ШУМА МЕТОДОМ СПЕКТРАЛЬНОГО ВЫЧИТАНИЯ

Кастельянос Г., Коклеев С.К., Уртадо Х.

МТУСИ

Аннотация. Представлен метод подавления фонового шума для улучшения акустического анализа речевых сигналов. Данный метод компенсации помех осуществляется посредством спектрального вычитания. Исследование метода проводилось для двух его вариантов реализации: одноканальной и двухканальной схем. В статье, качество оценки параметров речи рассматривалось с точки зрения улучшения процедуры автоматического распознавания речи. Как результат статистического анализа оценок получено, что качество расчёта акустических параметров, которое зависит от конкретного вида помех искажающих речь по-разному, влияет на природу измеряемый акустического параметра. Так, например, параметр питч соответствующего основного гармонического компонента речи проявляет большую нестабильность в присутствии квазигармонических помех. В результате, выигрыш от использования метода спектрального вычитания неодинаков для всех параметров. В случае питч, его величина составляет порядка 10 дБ, а для шумовых параметров (джиттер и шиммер) выигрыш незначительный.

Введение

Акустический анализ речи состоит в определении колебательных параметров или акустических параметров (АП), характеризующих её гармоническую природу. В зависимости от желаемых для измерения акустические свойств АП могут быть разделены на две категории [1]: а) Квазигармонические АП – которые проявляют всевозможные виды периодичности имеющиеся в речевом сигнале. К этим параметрам относят питч, форманты и ширина их полосы. б) Шумовые АП – которые измеряют относительные характеристики шумового фона присущего в речевом сигнале. Примерами этой категории являются джиттер, шиммер и гармонический компонент шума.

Выбор АП подразумевает характеристики, легко измеряемые и слабо зависящие от помеховой обстановки, в частности от фонового шума. На практике электронная обработка речи деградирует из-за электронно-акустических устройств преобразования сигнала (микрофонов, АЦП, динамик и т.п.), и т.д. В данной работе рассмотрена компенсация помеховых составляющих, возникающих во время электронной записи речи при наличии стационарного или квазистационарного фонового шума.

Фоновый шум непосредственно приводит к ошибке в оценке АП, точность которой необходима для правильной классификации и автоматического распознавания речи (АРР). С другой стороны, эффективность оценки АП ухудшается, если не принимать меры по устранению помех искажающих речь. Так, например, если характер помех во время тренировки обучающей системы речевого классификатора отличается от таковых при оценке АП в момент распознавания, то работа АРР заметно ухудшается [2]. По этой причине необходимо применять методы улучшения входных речевых сигналов, которые могли быть менее чувствительные к помеховой обстановке. В данной статье, анализ оценок производится в смысле улучшения работы АРР. Это методика позволяет сохранить необходимое качество сигналов для их последующего распознавания.

Спектральное Вычитание Фонового Шума

Для описания фонового шума часто используется модель аддитивного стационарного гауссовского процесса $\eta(t)$, некоррелированного с речевым сигналом $x(t)$. Входная запись смеси речевого сигнала и фонового шума будет

$$y(t) = x(t) + \eta(t). \qquad (1)$$

Процедура оценки полезного сигнала $x(t)$, присутствующего во входной записи $y(t)$, включает в себя компенсацию шума, выполняемую на основе известной статистической модели такого фонового шума. В общем случае, методы оценки АП должны обладать следующими свойствами:

Улучшить качество восприятия записанной зашумленной речи.

Уменьшить влияния шума на оценку АП.

Улучшить работу устройства автоматического распознавания речи в присутствии шума.

С другой стороны, при компенсации шума очень важно соблюдать принцип минимального искажения сигнала, при котором все параметры алгоритма фильтрации должны наименьшим образом реагировать на подавление самого шума. Одним из более эффективных алгоритмов подавления фонового шума в речевых сигналах является метод спектрального вычитания (СВ),

который основан на том, что амплитудно-частотная характеристика, рассчитываемая посредством процедуры кратковременного Фурье преобразования, несёт большую информацию по сравнению с фазо-частотной характеристикой. При этом статистические характеристики спектра предполагаются либо известными, либо доступными для оценки по той же кратковременной обработке в реальном времени. Таким образом, для заданной модели (1), в предположении нормального распределения плотности мощности смеси, оценка плотности мощности полезного сигнала (ПМС) $\hat{S}_x(\omega)$ получается как результат вычитания из плотности мощности смеси $\hat{S}_y(\omega)$ плотности мощности фонового шума $\hat{S}_\eta(\omega)$, по следующему правилу [3]:

$$\hat{S}_x(\omega) = \left[\left| \hat{S}_y(\omega) \right|^\beta - \alpha \left| \hat{S}_n(\omega) \right|^\beta \right]^{1/\beta} \exp j\varphi_y(\omega), \qquad (2)$$

где α - масштабный фактор, взвешивающий оценку шума, β - фактор который настраивается с целью получения оптимального решения компенсатора шума (часто используется β=2). Для получения полной оценки ПМС комплексной записи $\hat{S}_x(\omega)$ полезного сигнала амплитуда представлена в (2) дополняется со значением фазы $\exp j\varphi_y(\omega)$ входного сигнала, записанного в (1). Наконец, путём обратного преобразования Фурье образуется оценка полезного сигнала $\hat{x}(t)$ во времени.

При использовании выражения (2) могут быть получены отрицательные значения для амплитудно-частотной характеристики. По этому необходимо использовать дополнительное правило. Например:

$$\hat{S}_x(\omega) = \max\{S_y(\omega) - \alpha S_\eta(\omega), \lambda\}, \qquad (3)$$

где $\lambda \geq 0$ настраиваемый порог. Поскольку большинство систем обработки речи используют спектральные представления, то метод СВ не нуждается в дополнительном комплексном преобразовании, что является одним из его достоинств. Однако, самое жесткое требование к использованию СВ состоит в наличии представительных выборок для оценки с необходимой точностью статистических параметров шума. Одним из решений этой проблемы является оценка в тех интервалах времени, при которых речь отсутствует и, следовательно, $y(t)=\eta(t)$. В общем случае, метод СВ может быть осуществлен как по одноканальной схеме так и по двухканальной схеме. В первом случае, смесь сигнала и шума поступают по одному тракту, а статистика шума оценивается из этого же канала. Во втором случае, имеется дополнительный канал, с помощью которого снимается отдельная некоррелированная оценка шума. В данной работе рассматривались работы обеих схем.

Эффективность оценки АП после СВ

Критерии качества оценки АП были построены с использованием, предлагаемой в [4], метрики «спектральных искажений фильтрации»:

$$SD = \left[\frac{1}{\Delta\omega} \int_0^{\Delta\omega} \{S_x(\omega) - S_y(\omega)\}^2 d\omega \right]^{1/2}$$

где $\Delta\omega$ полоса полезного сигнала; и метрики «функции когерентности»:

$$\hat{\gamma}^2(f) = \frac{\left| \sum_{n=1}^{N} X_n(f) Y_n^*(f) \right|^2}{\sum_{n=1}^{N} \left| X_n(f) \right|^2 \sum_{n=1}^{N} \left| Y_n(f) \right|^2}$$

где $X_n(f)$ и $Y_n(f)$ – спектры входного и выходного отфильтрованного сигналов. Оценка минимального допустимого отношения сигнал/шум, при котором ещё можно считать оценки удовлетворительными, проводилась экспериментально. Расчёт точности оценки каждого АП ζ_i i=1,... N_ζ соответствовал ожидаемому значению относительной ошибки,

$$\bar{\delta}_i(l) = E\left[\frac{\left| \zeta_i(l) - \tilde{\zeta}_i \right|}{\zeta_i(l)} \right], \forall\, l = 1,.., N_\zeta.$$

Относительная эффективность качества оценки АП рассчитывалась по формуле

$$\rho_i = \frac{\mathrm{var}(\zeta_i)}{\mathrm{var}(\tilde{\zeta}_i)},\ i=1,\ldots, N_\zeta,$$

где var(ζ_i) - дисперсия зашумленной оценки ζ_i-ого АП, а var($\tilde{\zeta}_i$) – дисперсия оценки того же ζ_i-ого АП после процедуры СВ.

Выводы

Для анализа эффективности оценки АП после процедуры СВ был использован ансамбль из 30 выборок выбранных в результате медицинского экспертиза, которые принадлежали к одному и тому же классу сигналов речи (взрослому населению, мужского пола и без особых речевых патологий). В качестве фонового шума рассматривались шумы от мощного двигателя, гармоническая помеха (с фиксированной частотой и со скользящей частотой тона). Как результат статистического анализа оценок получено, что качество расчёта акустических параметров, которое зависит от конкретного вида помех искажающих речь, по-разному влияет на природу измеряемого АП. Так, например, параметр питч соответствующего основного гармонического компонента речи проявляет наибольшую нестабильность в присутствии квазигармонических помех. В результате, выигрыш от использования метода спектрального вычитания неодинаков для разных АП. В случае питч, его величина составляет порядка 10 дБ, а для шумовых параметров (джиттер и шиммер) выигрыш незначителен.

Литература

[1] Castellanos, G. Vargas F., Análisis Acústico en la Clasificación de señales de voz empleando RNA, Congreso Internacional de Inteligencia Computacional., pp 107-11. 2001

[2] PROAKIS, John y G. MANOLAKIS, Dimistris G. Tratamiento Digital de Señales. Tercera Edición. Prentice-Hall. Madrid, 1998

[3] DELLER, John R, et al. Discrete Time Processing Of Speech Signals. Macmillan Publishing Company. NJ, 1993.

[4] FURUI, Sadaoki and SONDHI, Mohan. Advanced In Speech Signal Processing. Marcel Dekker. New York, 1992

ESTIMATION OF PARAMETER EFFICIENCY AFTER NOISE REDUCTION BY SPECTRAL SUBTRACTION

Castellanos G., Kokleev S., Hurtado J.

ABSTRACT

A method of background noise reduction is developed for quality improving of Acoustic voice Feature (AF) extraction. The background noise directly causes errors in the AF estimates, whose accurate calculation is needed for classification and recognition purposes [3,5]. This paper is limited to the study of the signal degradation caused by either quasi-stationary or non-stationary background noise, such as that produced by alarms, motors or whispering. These techniques are purposed to produce a high quality signal allowing an accurate analysis in the recognition phase [4]. This paper is focused on the Spectral Subtraction (SS) algorithm, which has proved to be the most efficient in a survey performed in [5]. This algorithm is based on the fact that the spectral amplitude, as computed by the Short Time Fourier Transform (STFT), is more important than the phase form an information point of view. The statistics of the spectrum are assumed to be known or that they can be estimated on-line. Since the majority of noise reduction systems use any kind of signal spectral representation, the subtraction does not need an additional transformation. However, the most stringent requirement for applying this method lies in the need of having good statistical measures of the noise. A solution to this problem could be to estimate them in intervals in which the voice is absent. Generally speaking, the SS technique can be implemented by one- or two-channel approaches. Both of them were analyzed in the present research. The latter was selected because it offers a better performance of the ASR.

The analysis of the behavior of the SS algorithm was performed for a population of 30 samples of the same gender category, which were regarded as normal by an expert in Phonoaudiology. The influence of the following kinds of disturbances was analyzed: a fuel motor, a synthetic white noise and a fixed harmonic disturbance for the following fixed values of SNR: 10, 15 and 20 dB. The following criteria were used for measuring the error in the estimation Spectral Distortion (SD) and the sample coherence function (CF).

The statistical analysis of the estimation error showed that the influence of the perturbation is dependent on its type as expected. For instance, the pitch, which arose as the most important parameter from the feature extraction analysis, showed to be highly sensitive to the harmonic disturbance. This is somewhat obvious because the pitch is an estimation of the frequency having the highest energy. However, its perturbation amplitude and frequency measures (jitter and shimmer) showed to be severely affected by any type of disturbance even for SNR of 20 dB. In the case of formants and their respective bandwidth the error is so high that the their recovering is almost impossible, regardless the intensity of the perturbation. Generally speaking, the employment of the SS filtering for the pitch and energy estimation can produce a process gain up to 10 dB. In the jitter and shimmer cases other, more robust algorithms should be considered, while for the formants and their bandwidths LPC algorithms with noise compensation should be used.