В первых шестнадцати циклах вход состоит из 32 – битного слова данного блока. Для оставшихся 64 циклов вход состоит из сложения по модулю 2 нескольких слов.

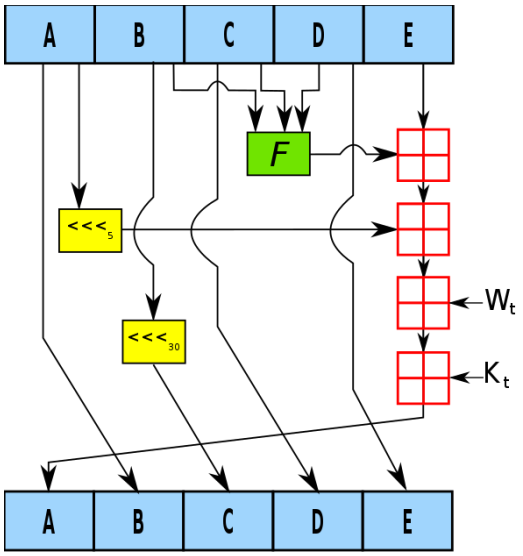

Рис. 3– Одна итерация цикла.

Длина дайджеста	160 бит
Размер блока обработки	512 бит
Число итераций	80
Число элементарных функций	3
Число дополнительных констант	4

Алгоритм прост в описании и реализации, однако достаточно устойчив к атакам грубой силы.

Список использованных источников:
1. D. Eastlake, P. Jones, RFC 3174 US Secure Hash Algorithm 1 (SHA1)// Cisco Systems – 2001.
2. Ярчук С. М. Конспект лекций по информационной безопасности/ С. Ярчук – Уральский Федеральный университет им. Б.Н. Ельцина «УПИ», 2011. - 75 с.

ОЦЕНКА РЕЧЕВОЙ МАСКИ ДЛЯ ИДЕНТИФИКАЦИИ ДИКТОРА

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Кручок Д.Н.

Петровский А.А. – д.т.н., профессор

В данной работе рассмотрены методы оценивания маски для речевого сигнала. Речевая маска позволяет использовать реконструированный сигнал для извлечения характеристического вектора и дальнейшей идентификации диктора в условиях акустических шумов.

Речевой сигнал, даже если он зашумлен, обладает значительной степенью избыточности, и информация о речевых характеристиках присутствует и в случае сильного зашумления. Методы, которые используют это наблюдение (англ. missingdataapproaches) используют речевые маски для маркировки сигнала для каждой точки время-частота по наличию речи или шума в ней [1]. В дальнейшем, маску используют для реконструирования сигнала, получения из него характеристического вектора для идентификации диктора. В случае, когда шум известен заранее, маску можно построить по критерию отношения сигнал-шум (англ. signaltonoise, SNR) для каждого временно-частотного компонента. Если такую маску использовать для идентификации, то распознавание диктора будет иметь высокую шумоустойчивость даже в присутствии сильных шумов и искажений [2]. На практике, отсутствие априорного знания о шуме толкает на осуществление оценки речевой маски.

Подходы оценивания речевой маски делятся на [1]:

1) Методы, основанные на оценке SNR;
2) Методы, основанные на слуховом восприятии;
3) Методы, основанные на классификации параметров оценивания.

При оценке SNR решение надежности того или иного временно-частотного компонента принимается

прямым измерением спектральной плотности мощности зашумлённого сигнала. Для этого применяется метод спектрального вычитания [3], где оценка значения средней спектральной мощности шума получается из нескольких первых фреймов выражения (при условии тишины, когда в начале выражения отсутствует энергия речи). Для зашумленного сигнала, предполагаемый спектр чистой речи можно получить по формуле [3]:

$$\hat{S}(t,f)=\begin{cases} X(t,f)-\hat{N}(t,f), если \quad X(t,f)-\hat{N}(t,f)>\gamma X(t,f), \\ \gamma X(t,f), \quad в \quad противном\ случае, \end{cases} \tag{1}$$

где,

$\hat{N}(t,f)=\frac{1}{T_{cpд}}\sum_{t=1}^{T_{cpд}}X(t,f)$ – спектр мощности для первых T кадров, а γ – малый масштабирующий коэффициент ($\gamma \approx 0$), необходим, чтобы значения спектра мощности были неотрицательными. Критерий для принятия решений [1]:

$$m_{tf}^{ss(\theta)}=\begin{cases} 1, \quad если \quad 10\log_{10}\left(\frac{\hat{S}(t,f)}{\hat{N}(t,f)}\right)>\theta, \\ 0, \quad в \quad противном\ случае \end{cases} \tag{2}$$

θ – порог в дБ.В работе [4] был предложен критерий отрицательной энергии, где временно-частотный компонент помечался как надёжный, если значение спектра зашумленного сигнала превышает значение спектра предполагаемого «шумового» сигнала. На рисунке 1 изображены спектрограмма зашумлённого речевого сигнала и полученная речевая маска с порогом надёжности θ= 0 дБ. Речевой сигнал содержит фразу: «штурман просил продолжать разворот», частота дискретизации равна 16 кГц, тип используемого шума – *babble*, уровень зашумления SNR составляет минус 5 дБ.

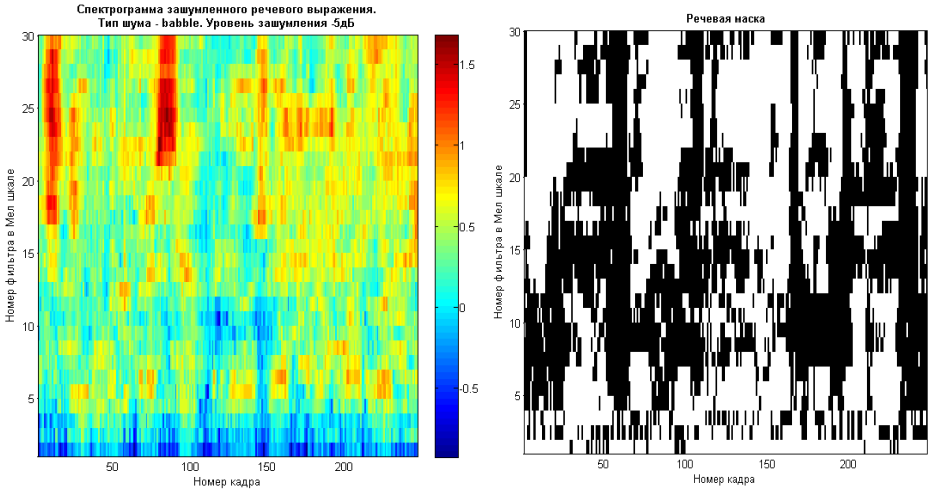

Рисунок 1. Спектрограммы речевого сигнала зашумленного шумом *babble* с уровнем SNR= = -5дБ (слева) и полученная речевая маска с порогом надёжности θ = = 0 дБ (справа)

Недостатком методов, основанных на оценке SNR – это отсутствие шумоустойчивости к нестационарным и переходным шумам.

Методы оценки, основанные на слуховой системе человека, используют свойства разделения целевых сигналов от источников шумов. В работе [5] был предложен метод группировки гармоник, основанный на том, что энергия внутри голосового сигнала находится возле гармоник частоты основного тона. В работе [6] используется метрика – голосовая дистанция – евклидово расстояние между формами спектра кратковременного сигнала и анализируемого окна. Это расстояние потом сравнивается с порогом, и сходно с формулой 2, выносится решение по каждому временно-частотному компоненту. Также методы могут использовать пространственное местонахождение источников сигнала для определения временно-частотных регионов в слуховой сцене с помощью бинауральной слуховой модели [7, 8].

Методы, основанные на классификации параметров оценивания, используют модель обучения для оценивания речевой маски.В байесовском подходе эта классификация достигается путем моделирования распределений набора признаков надежности. Если признаки сконструированы таким образом, что они используют характеристики самого речевого сигнала (а не характеристики искажающего шума), то вероятностная оценка значений признаков будет показывать, надежна ли связанная с ними временно-частотная точка.

Рассмотрим спектральный компонент x_{tf} и набор соответствующих признаков надёжности ss_{tf}. В соответствии с байесовским подходом решение о бинарной надежности для x_{tf} можно получить [1]:

$$m_{tf}{}^{Bayes}=\begin{cases} 1 \quad если \quad p(\lambda_{m=1}^{Bayes}(f))p(\vec{ss}_{tf}\mid\lambda_{m=1}^{Bayes}(f))>p(\lambda_{m=0}^{Bayes}(f))p(\vec{ss}_{tf}\mid\lambda_{m=0}^{Bayes}(f)), \\ 0, \quad в \quad противном\ случае \end{cases} \tag{3}$$

$\lambda_{m=1}^{Bayes}(f)$ и $\lambda_{m=0}^{Bayes}(f)$ являются обученными распределениями для надежного и ненадежного решения,

соответственно, в частотном канале *f*. А $p(\vec{ss}_{tf} \mid \lambda_{m=1}^{Bayes}(f))$ и $p(\vec{ss}_{tf} \mid \lambda_{m=0}^{Bayes}(f))$ соответствующие плотности вероятности надежности и ненадёжности для канала *f*, основанные на признаках $ss_{tf.}$

В качестве признаков используют различные параметры: в работе [9] – спектральная энергия полосы и её производные, в работе [8] были использованы: *коэффициент гребенчатого фильтра* для сравнения энергии в вокализованных областях с энергией в негармонических областях; *коэффициент автокорреляции*, для измерения периодичности сигнала; *отношение энергии поддиапазона к энергии полной полосы*, представляющее спектральную форму; *оценка энергии шума*; *коэффициент эксцесса*, используемый для измерения «остроты» пика сигнала; *коэффициент тональности*, для измерения SNR.

Используя предполагаемые маски, отдельные классификаторы обучаются для невокализованных и вокализованных типов внутри каждого канала.В оценках для шумоустойчивого распознавания оценки классификатора превосходят традиционные оценки спектрального вычитания во всех условиях шума, но особенно для нестационарных случаев [10].

Таким образом, методы, основанные на слуховой системе человека, так и основанные на классификации речевых параметров могут обеспечить лучшую эффективность распознавания в сравнении с методами на основе оценки только SNR. Преимуществом аудиторных подходов в оценке маски является их способность строить решения, основанные на свойствах спектра речевого сигнала. Это позволяет более точно идентифицировать доминирующие речевые спектральные области по сравнению с подходами на основе SNR, которые предполагают обобщение шумовых характеристик, наблюдаемых в небольшом числе свободных от речи кадров. Недостатком подхода, основанного на классификации речевых параметров, является его слабость к шуму, которые имеют сходные спектральные характеристики с характеристиками речевого сигнала. В этом случае речевые параметры не смогут отличить речевые и шумовые компоненты доминирующей частоты, что приводит к низкой точности маскировки.

Список использованных источников:

1. Togneri R., Pullella D. An Overview of Speaker Identification Accuracy and robustness Issues // IEEE Circuits and systems magazine. 2011. P. 23–58.

2. M. Cooke, P. Green, and M. Crawford, “Handling missing data in speech recognition,” in Proc. IEEE Int. Conf. Acoustics Speech Signal Processing (ICASSP), 1994, pp. 1555–1558.

3. S. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust. SpeechSignalProcess., vol. 27, no. 2, pp. 113–120, 1979.

4. M. El-Maliki and A. Drygajlo, “Missing features detection and handling for robust speaker verification,” in Proc. European Conf. Speech Communication Technology (Eurospeech), Budapest, Hungary, 1999, pp. 975–978.

5. J. Barker, M. Cooke, and P. Green, “Robust asr based on clean speech models: An evaluation of missing data,” in Proc. European Signal Process. Conf. (EUSIPCO), Aalborg, Denmark, 2001, pp. 213–216.

6. P. Janˇcoviˇc and M. Kцkьer, “Estimation of voicing-character of speech spectra based on spectral shape,” IEEE Signal Process. Lett., vol. 14, no. 1, pp. 66–69, 2007.

7. K. J. Palomdki, G. J. Brown, and D. Wang, “A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation,” S peech Commun., vol. 43, no. 4, pp. 361–378, 2004.

8.Кручок, Д. Н. Эффект бинауральной маскировки для идентификации диктора в акустических шумах / Д.Н. Кручок // Современные технологии в науке и образовании – СТНО-2017 [текст]: сб. тр. междунар. науч.-техн. и науч.-метод. конф.: в 8 т. Т.3./ под общ. ред. О.В. Миловзорова. – Рязань: Рязан. гос. радиотехн. ун-т, 2017; Рязань. – 292 с. – С. 165–168.

9.B. Raj, “Reconstruction of incomplete spectrograms for robust speech recognition,” Ph.D. dissertation, Pittsburgh, PA, Carnegie Mellon Univ., 2000.

10. M. L. Seltzer, B. Raj, and R. M. Stern, “A Bayesian classifi er for spectrographic mask estimation for missing feature speech recognition,” Speech Commun., vol. 43, no. 4, pp. 379–393, 2004.

СТИЛИЗАЦИЯ ГОЛОСА С ИСПОЛЬЗОВАНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Крылов Н.Д.

Одной из самых сложных задач для вычислительной техники является обработка естественного языка. В частности, одной из таких задач является имитация человеческого голоса. Алгоритмы машинного обучения являются наиболее эффективным инструментом в этой области.

Для обучения нейронной сети необходим правильно составленный набор данных. Будет ли правильным обучающим примером для данной задачи сказанная одна, двумя разными людьми, фраза? Фраза может быть произнесена с различной скоростью, громкостью и темпом, входная и выходная последовательность будет иметь различную длину. Обучение нейронной сети на таких данных затруднительно, в таком случае хорошим решением является использование двух нейронных сетей. Первая нейронная сеть используется в качестве кодировщика и переводит данные в иное представление, вторая в качестве декодировщика, позволяет получить требуемый результат.