

КОНТИНУАЛЬНО-ДИСКРЕТНАЯ МОДЕЛЬ РАСПРЕДЕЛЕНИЯ ЗНАЧЕНИЙ ОЖИДАЕМОЙ ДЛИННЫ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ТОНАЛЬНЫХ ЗВУКОВ СЛИТНОЙ РЕЧИ

Калашников Д. М., Иванов А.И., Бойков И.В. (Пенза)

Распознавание слитной речи предполагает то, что нейросетевой автомат будет способен выделять отдельные звуки. В простейшем случае для этой цели может быть использован классификатор тон-шум, построенный на линейном предсказателе [1].

На данный момент эффективного фрагментатора слитной речи на звуки не существует. Одно из причин, затрудняющей решение этой задачи, является то, что не создано эффективных прогнозистов ожидаемой длины очередного звука слитной речи. Для того, чтобы создать прогнозиста времени звуков, необходимо знать статистику их распределения. Простейшим способом получения этой статистики является использование классического линейного предсказателя тон-шум [1].

Был проделан следующий эксперимент, состоящий в прочтении 15 строк, русскоязычного текста одним диктором. Далее звуковой файл был разбит на тональные звуковые фрагменты. Гистограмма распределений интервалов длительности тональных фрагментов звука приведена на рисунке 1.

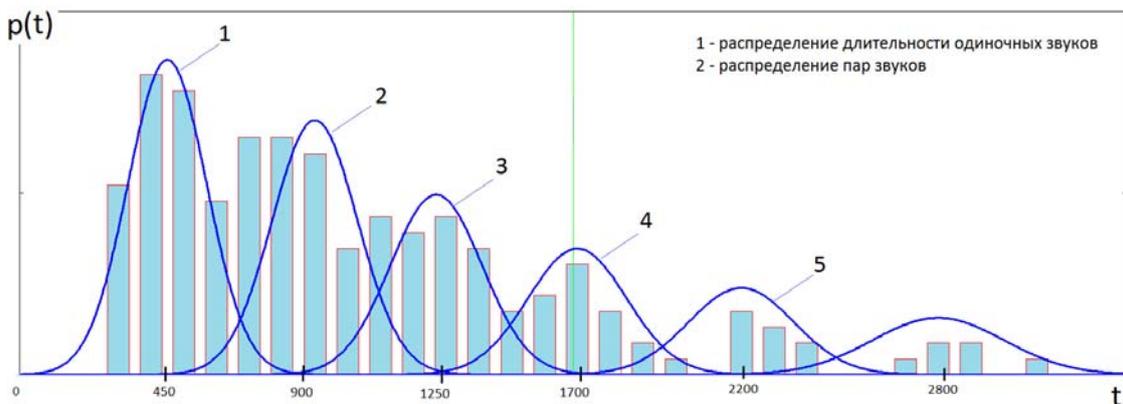


Рис.1 Смесь нормальных законов распределения значений длительности звуков тональной слитной речи одного из дикторов

На рисунке 1 дана аппроксимация гистограммы смесью из 6-и нормальных законов распределения значений разных длин интервалов. Аппроксимация осуществлена, опираясь на зрительное восприятие. Мы видим, что наиболее вероятное значение длин одного, двух, трех, ..., звуков кратно друг другу. Это условие может быть записано следующей системой уравнений (1):

$$\begin{cases} E(T_1) = C_{\text{Бuo}} \\ E(T_2) = 2C_{\text{Бuo}} \\ \dots\dots\dots \\ E(T_n) = nC_{\text{Бuo}} \end{cases} \quad (1),$$

где $E(T_n)$ - математическое ожидание нормальных распределений.

Статистическая модель предсказателя длины звуков существенно упрощается, если принять гипотезу одинаковых значениях среднеквадратических отклонений распределений длин звуков:

$$\sigma_1 = \sigma_2 = \dots = \sigma_n = \sigma_{\text{Бuo}}. \quad (2)$$

Еще одним параметром статистической модели является отношение значений максимумов нормальных распределений. Из рисунка 1 видно, что наибольшим является 1-ое распределение, а последующее уменьшается. Значения P_1, P_2, \dots, P_n монотонно убывают, однако:

$$P_1 + P_2 + \dots + P_n \approx 1 \quad (3).$$

Последнее обусловлено тем что каждое из нормальных распределений появляются со своей вероятностью P_i . Сумма всех вероятностей событий должна быть единичной. Оценит вероятность появления каждого из n -и нормальных законов можно путём выделения равных интервалов вокруг математических ожиданий и подсчета, попавших в эти интервалы значений:

$$P_i = \frac{1}{\sqrt{2\pi}\sigma} \int_{E(T_i)-\sigma}^{E(T_i)+\sigma} e^{\frac{-(E(T_i)-U)^2}{2\sigma^2}} dU \quad (4)$$

В формуле 4 принят интервал размером в 2 среднеквадратических отклонений от i -го математического ожидания.

Вычисления по (4) могут быть проведены однократно, т.к. отношения рассматриваемых вероятностей, не зависит от диктора, являясь параметрами языка, на котором говорит диктор: $\frac{P_i}{P_1} = \text{const}_i \quad (5).$

Таким образом, опираясь на описанную выше статистическую модель, можно создать достаточно эффективные предсказатели длин звуков слитной речи конкретного человека. Эта модель должна постоянно подстраиваться по своим параметрам при анализе речи. По нашим данным человек способен надежно предсказывать интервалы звуков, опираясь на 2-3 слова, произнесенных прослушиваемым диктором. Искусственный автомат, предсказывающий длину звука должен работать, опираясь на статистику порядка 20-30 слов, произнесенных диктором.

Литература

1. Маркл Грей Линейное предсказание речи. М. Изд-во: Связь, 1980 г. 154 с.

Материал поступил 12.04.2013, опубликован по положительной рецензии д.т.н., проф. Малыгина А.Ю.