

И.Ш. НЕВЛЮДОВ, д-р техн. наук, А.М. ЦЫМБАЛ, канд. техн. наук,
С.С. МИЛЮТИНА, канд. техн. наук, В.Ю. ШАРКОВСКИЙ

ПРИМЕНЕНИЕ АКУСТИЧЕСКОЙ МОДЕЛИ ДЛЯ УПРАВЛЕНИЯ МОБИЛЬНЫМ РОБОТОМ

Современной тенденцией развития роботизированных систем является широкое применение аудио- и визуальных средств ввода и обработки информации. Например, в рамках автоматизированной транспортно-складской системы мобильный погрузочно-разгрузочный робот работает в тесном сотрудничестве с людьми – сотрудниками склада, что вызывает необходимость разработки эффективного человеко-машинного интерфейса. Применение голосового интерфейса, при котором команды оператор подаёт при помощи голоса и также получает подтверждение от робота, позволит уменьшить затраты времени на постановку задачи и управлять роботом с минимальным дополнительным оборудованием.

Система распознавания речи (рис. 1) для голосового управления состоит, как правило, из трех основных компонентов: акустической модели, языковой модели и декодера [1].



Рис. 1

Акустическая модель позволяет оценить распознавание речевого сегмента с точки зрения схожести на звуковом уровне. Известно несколько методов построения дикторонезависимой акустической модели [2]. Одни основаны на обработке сигнала от приемника звука с целью выделения признаков, характерных для определенных слов и независимых от диктора. Другие акустические модели для того чтобы обеспечить инвариантность к произнесению звуков людьми разного пола, возраста, с разным тембром и акцентом, «тренируются» на специально подобранных и отсегментированных речевых базах большого объема, включающих речь сотен различных людей. В результате, несколько тысяч моделей фонем в разных фонетических контекстах являются основой дикторонезависимого пофонемного распознавания речи на определенном языке.

Использование чисто акустической информации недостаточно для осуществления качественного распознавания речи. Например, в реальных условиях (при наличии посторонних шумов и искажений речевого сигнала) ни одни, даже самые точные, акустические модели не смогут отличить слово «крюк» от слова «трюк». В такой ситуации важна информация о контексте (теме разговора) и, что еще более важно, о тех словах, которые уже были распознаны ранее. Например, если ранее было распознано слово «железный», то в этой ситуации гораздо вероятнее ожидать произнесения слова «крюк», чем «трюк». Подобная оценка и осуществляется языковой моделью. Модели языка бывают двух основных видов: на основании грамматик и статистические [3].

Декодер – программный компонент системы распознавания, который совмещает данные, получаемые в ходе распознавания от акустических и языковых моделей, и на основании их объединения определяет наиболее вероятную последовательность слов, которая и является конечным результатом распознавания

Акустическое колебание, формируемое в речевом тракте человека, является непрерывно изменяющимся процессом. С математической точки зрения его можно описать функцией непрерывного времени t . Аналоговые (непрерывные во времени) сигналы будут обозначаться через $x_a(t)$. Речевой сигнал можно представить и последовательностью чисел, например $x(n)$. Если последовательность чисел представляет собой последовательность мгновенных значений, аналогового сигнала, взятых периодически с интервалом T , то эта операция дискретизации обозначается через $x_a(nT)$. На рис. 2 показан пример речевого сигнала в аналоговой форме и в виде последовательности отсчетов, взятых с частотой дискретизации 8 кГц .

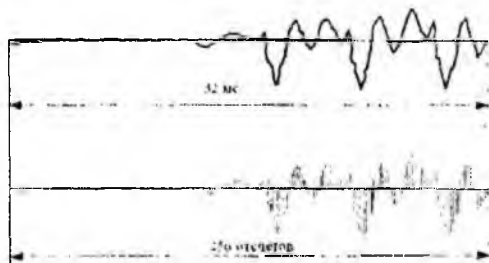


Рис. 2

Обработка сигналов включает преобразование их в форму, удобную для дальнейшего использования. Таким образом, интерес представляют дискретные системы или, что то же самое, преобразования входной последовательности в выходную. Подобные преобразования далее изображаются на структурных схемах. Многие системы анализа речевых сигналов разработаны для оценивания переменных во времени параметров по последовательности мгновенных значений речевого колебания. Подобные системы имеют многомерный выход, т. е. одномерная последовательность на входе, представляющая собой речевой сигнал, преобразуется в векторную последовательность на выходе.

Анализ сигналов и расчет систем значительно облегчаются при их описании в частотной области. В этой связи полезно кратко остановиться на представлении сигналов и систем в дискретном времени с использованием преобразования Фурье и z -преобразования.

Прямое и обратное z -преобразование последовательности определяется двумя известными уравнениями:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}, \quad (1)$$

$$x(n) = \frac{1}{2\pi i} \oint_C X(z)(z)^{n-1} dz. \quad (2)$$

Прямое z -преобразование $x(n)$ определяется уравнением. В общем случае $X(z)$ – бесконечный ряд по степеням z^{-1} ; последовательность $x(n)$ играет роль коэффициентов ряда. В общем случае подобные степенные ряды сходятся к конечному пределу только для некоторых значений z . Достаточное условие сходимости имеет вид

$$\sum_{n=-\infty}^{\infty} |x(n)| |z^{-n}| < \infty. \quad (3)$$

Описание сигнала в дискретном времени с помощью преобразования Фурье задаётся в виде

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}, \quad (4)$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(\omega) e^{j\omega t} d\omega. \quad (5)$$

Преобразование Фурье получается вычисления z-преобразования на единичной окружности, т. е. подстановкой $z=e^{j\omega}$. Частота ω может быть интерпретирована как угол на z-плоскости. Достаточное условие существования преобразования Фурье можно получить, подставляя $|z|=1$

$$\sum_{n=-\infty}^{\infty} |x(n)| < \infty. \quad (6)$$

Важная особенность преобразования Фурье последовательности состоит в том, что оно является периодической функцией ω с периодом 2π . С другой стороны, поскольку $X(e^{j\omega})$ представляет собой значение $X(z)$ на единичной окружности, оно должно повторяться после каждого полного обхода этой окружности, т. е. когда ω изменится на 2π рад.

Для дискретных сигналов, как и в случае аналоговых, если последовательность периодическая с периодом N , т. е.

$$x(n) = x(n + N), \quad -\infty < n < \infty. \quad (7)$$

то $x(n)$ можно представить в виде суммы синусоид, а не в виде интеграла. Преобразование Фурье для периодической последовательности имеет вид[4]

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \quad (8)$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{-j2\pi kn/N}. \quad (9)$$

Это точное представление периодической последовательности. Однако основное преимущество описания заключается в возможности несколько иной интерпретации уравнений. Рассмотрим последовательность конечной длины $x(n)$, равную нулю вне интервала $0 \leq n \leq N - 1$. В этом случае z-преобразование имеет вид

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n}. \quad (10)$$

Если записать $X(z)$ в N равноотстоящих точках единичной окружности, т. е. $z_k=e^{j2\pi k/N}$, $k=0, 1, \dots, N-1$, то получим

$$X(e^{j2\pi k/N}) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \quad k=0, 1, \dots, N-1. \quad (11)$$

Дискретное преобразование Фурье (ДПФ) со всеми его особенностями является важным способом описания сигналов по следующим причинам:

- ДПФ можно рассматривать как дискретизированный вариант z-преобразования (или преобразования Фурье) последовательности конечной длительности;
- ДПФ очень сходно по своим свойствам (с учетом периодичности) с преобразованием Фурье и z-преобразованием;
- N значений $X(k)$ можно вычислить с использованием эффективного (время вычисления пропорционально $N \log N$) семейства алгоритмов, известных под названием быстрых преобразований Фурье (БПФ).

Дискретное преобразование Фурье широко используется при вычислении корреляционных функций, спектров и при реализации цифровых фильтров, а также часто используется и при обработке речевых сигналов[4].

Спектральный анализ является методом обработки сигналов, который позволяет выявить частотный состав сигнала. Поскольку анализируемые сигналы во многих случаях имеют случайный характер, то важную роль в спектральном анализе играют методы математической статистики. Частотный состав сигналов определяют путем вычисления оценок спектральной плотности мощности (СПМ). Задачами вычисления СПМ являются обнаружение гармонических составляющих в анализируемом сигнале и оценивание их параметров. Для решения указанных задач требуется соответственно высокая разрешающая способность по частоте и высокая статистическая точность оценивания параметров. Эти требования противоречивы. Аргументы в пользу выбора высокого разрешения или высокой точности оценки СПМ зависят от того, что интересует исследователя: устойчивые оценки в пределах всего диапазона частот или высокая степень обнаруживаемости периодических составляющих.

Все методы цифрового спектрального анализа можно разделить на две группы [4]: классические методы, базирующиеся на использовании преобразований Фурье, и методы параметрического моделирования, в которых выбирается некоторая линейная модель формирующего фильтра и оцениваются его параметры. К первой группе относят корреляционный и периодограммные методы. Ко второй группе относят методы оценивания СПМ на основе авторегрессии скользящего среднего и др.

Периодограммный метод обеспечивает вычисление оценки СПМ непосредственно по числовой последовательности $x[nT_0]$, формируемой путем дискретизации стационарного эргодического случайного процесса $x(t)$. Периодограммная оценка СПМ

$$\bar{S}_x(\omega) = \frac{T_0}{N} \left| \sum_{n=0}^{N-1} x[nT_0] e^{-j\omega n T_0} \right|^2. \quad (12)$$

Выражение соответствует возможности вычисления СПМ с помощью преобразования Фурье непосредственно по реализации исходного сигнала.

Вычисленная оценка СПМ несостоятельна, т.е. с увеличением N она не улучшается. Для получения состоятельной оценки ее необходимо сглаживать. Кроме этого, при выполнении преобразования Фурье последовательности $x[nT_0]$ конечной длины N происходит «размывание» спектра, которое также оказывает влияние на состоятельность оценки СПМ.

Ограничение последовательности $x[nT_0]$ конечным числом значений равносильно умножению исходной бесконечной последовательности $x_0[nT_0]$ на другую последовательность

$$w[nT_0] = \begin{cases} 1, & 0 \leq n \leq N-1, \\ 0, & n < 0, n > N-1 \end{cases} \quad (13)$$

которую называют прямоугольным окном. Тогда можно записать:

$$x[nT_0] = x_0[nT_0] w[nT_0]. \quad (14)$$

На основании указанного метода обработки звуковой информации и приведенных формул создано программное обеспечение, реализующее распознавание голосовой информации. Для реализации функции распознавания голосовых команд голосового управления промышленным роботом создана функция *OpenData*. После того, как данные о записанной информации занесены в массив, функция вызывает функции, отвечающие за преобразование этого массива. Прежде всего, вызывается функция *Noise*, которая на основании разделения вокализованного звука и невокализованного переписывает массив с данными о звуке в новый массив, содержащий лишь участки вокализованного звука.

График, отражающий модули значений массива до преобразования, приведен на рис. 3.

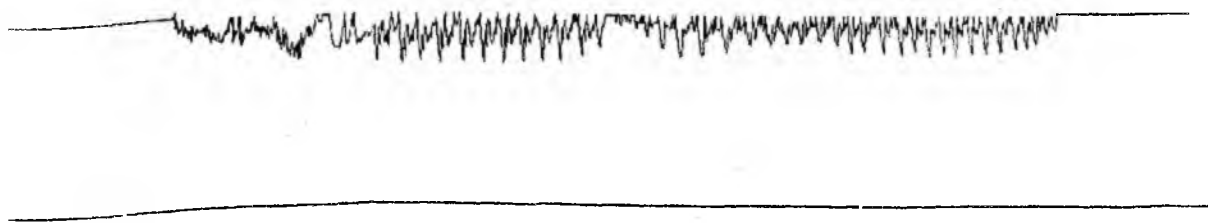


Рис. 3

График, отражающий модули значений массива после обработки с помощью функции Noise, представлен на рис. 4.

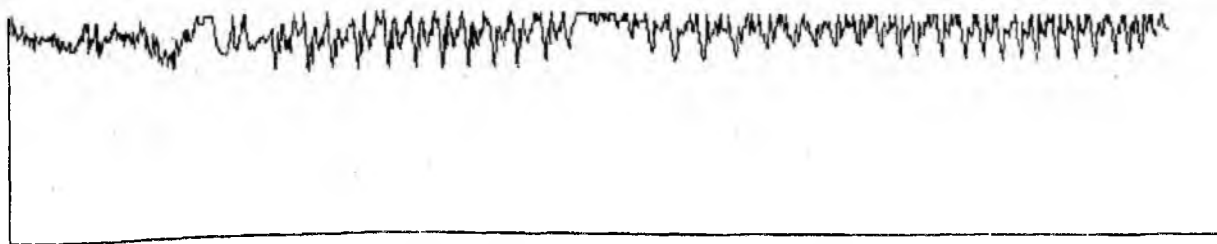


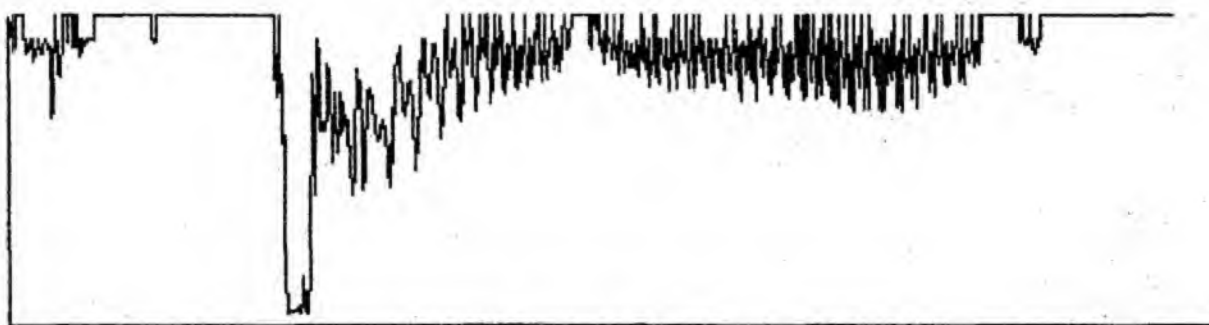
Рис. 4

После этого проводится обработка в окне Хэмминга. График, полученный в результате этого преобразования, приведен на рис. 5.



Рис. 5

Далее выполняется преобразование Фурье для коэффициентов, взвешенных окном Хэмминга. После этого производится фильтрация. Затем проводится обратное преобразование Фурье. Далее следует операция клиппирования, на основании графика которой делается вывод о том, какое слово было произнесено. На рис. 6 приведены графики сигнала для слова «вперёд»: рис. 6, а – график вокализованной части сигнала; рис. 6, б – график сигнала после обработки в окне Хэмминга; рис. 6, в – график сигнала после клиппирования.



а



б



в

Рис. 6

Таким образом, созданное программное обеспечение позволяет распознавать управляющие команды произнесенные голосом для дальнейшего формирования задания системе управления мобильным роботом. Применение данной программы для управления роботом позволит в значительной степени сократить время необходимое для формулировки задания системе управления, а также обеспечивает возможность управления роботом без необходимости в специализированных технических средствах и длительном обучении оператора. Кроме того, к достоинствам данного подхода следует отнести независимость от оператора-диктора. В то же время, недостатком предложенного метода распознавания команды является отсутствие возможности обучения программы, отнесение всех произнесенных слов лишь к набору заложенному в программе, а также высокие требования к вычислительной мощности системы управления.

Список литературы: 1. *Рассел, Стюарт.* Искусственный интеллект: Современный подход : пер. с англ./ Стюарт Рассел, Питер Норвинг. – 2-е изд. – М.: Изд. дом «Вильямс» 2006. – 1408 с. 2. *Jurafsky, D.* Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition / D. Jurafsky, J.H. Martin. – New Jersey:Prentice-Hall, 2000. – 934 с. 3. *Искусственный интеллект: Применение в интегрированных производственных системах / под ред. Э. Кьюсиака:* пер. с англ. – А.П.Фомина / под ред. А.И. Дашенко, Е.В. Левнера. – М.: Машиностроение, 1991. – 544с. 4. *Рабинер, Л.* Теория и применение цифровой обработки сигналов / Л. Рабинер, Б. Голд. – М.: Мир, 1978. – 848 с.

Харьковский национальный
университет радиоэлектроники

Поступила в редколлегию 12.01.2011