

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Резников А. Н. Тепловые процессы в технологических системах / А. Н. Резников, Л. А. Резников. – М. : Машиностроение, 1990. – 288 с.: ил.
- 2 Теория резания. Физические и тепловые процессы в технологических системах: учеб. для вузов. / Ящерицын П. И. [и др.]. – Минск : Высшая школа, 1990. – 512 с.
3. Проектирование технологий машиностроения на ЭВМ: учебник для вузов / О. В. Таратынов [и др.]; под ред. О. В. Таратынова. – М. : МГИУ, 2006. – 519 с.
4. Справочник технолога-машиностроителя. / Под ред. А. М. Дальского. М. : Машиностроение –1, 2001. – Том 2 – 944 с.

УДК 658.5.017.7

ПРОБЛЕМЫ РАСПОЗНАВАНИЯ ЗВУКОВОЙ ИНФОРМАЦИИ В НЕЙРОСЕТЕВОЙ СИСТЕМЕ ОПОВЕЩЕНИЯ ДЛЯ ЛЮДЕЙ С ОГРАНИЧЕННЫМИ СЛУХОВЫМИ ВОЗМОЖНОСТЯМИ

Рязанцева Н. В., Голдобина Т. А., Минин В. Е.

*Белорусский государственный университет транспорта,
г. Гомель, Республика Беларусь*

Одним из наиболее активно развивающихся направлений автоматизации в современных условиях являются технологии удаленного управления, такие как умный дом (англ. smart home). Домашняя автоматизация, способствующая повышению уровня комфорта и удобства жизни человека, – весьма гибкая система, которая предоставляет пользователю возможность самостоятельно конструировать, настраивать и дистанционно контролировать своё жизненное пространство в зависимости от насущных потребностей.

Однако системы оповещения, входящие в состав большинства современных систем домашней автоматизации общего пользования, не могут быть полнофункционально применены людьми с ограниченными возможностями. Несмотря на существующие мировые тенденции повышения комфортности для инвалидов, в настоящее время на рынке аппаратного и программного обеспечения представлено очень мало продуктов для слабослышащих или слабовидящих.

Авторами ведется разработка нейросетевой системы, которая в реальном времени анализирует окружающую обстановку, определяет источники и характер звуков, а также классифицирует данные звуки по категориям с дальнейшим уведомлением пользователя. Так как основными пользователями системы являются люди с ограниченными возможностями по слуху, категории звуков включают в себя основные бытовые события с возможностью дополнительного обучения (плач ребенка, звонок в дверь, звуковые уведомления бытовой техники и т. д.) на усмотрение пользователя.

Распознавание звуков является нетривиальной задачей, поскольку сложность сопоставления звуков в различных окружениях, а также иные источники шума способны существенно исказить звуковую волну, в связи с чем значи-

тельно увеличатся затраты на ее идентификацию. Существующие методы шумоподавления способны компенсировать шумы от сторонних объектов, однако их эффективность сильно варьируется от области применения.

Разрабатываемая система включает в себя аппаратную часть, программный модуль обработки звуков, программный модуль для формирования и обмена уведомлениями, а также приложение для мобильного телефона для обработки данных уведомления. Для решения поставленной задачи система должна обладать возможностью идентификации звука и сопоставления записанной звуковой волны с готовыми паттернами. Отсюда можно выделить несколько подзадач, которые требуется рассмотреть: получение оцифрованной звуковой волны, математическая обработка данных для получения цифрового отпечатка, обучение нейронной сети готовым набором цифровых последовательностей, верификация обученной нейронной сети. Качество обработки данных имеет ключевое значение и непосредственно влияет на результат распознавания, т. к. малое количество отсчетов при оцифровке сигналов или низкая разрешающая способность АЦП негативно скажутся на качестве выборки, что ухудшит качество распознавания нейронной сети в целом. Структурная схема системы представлена на рисунке 1.

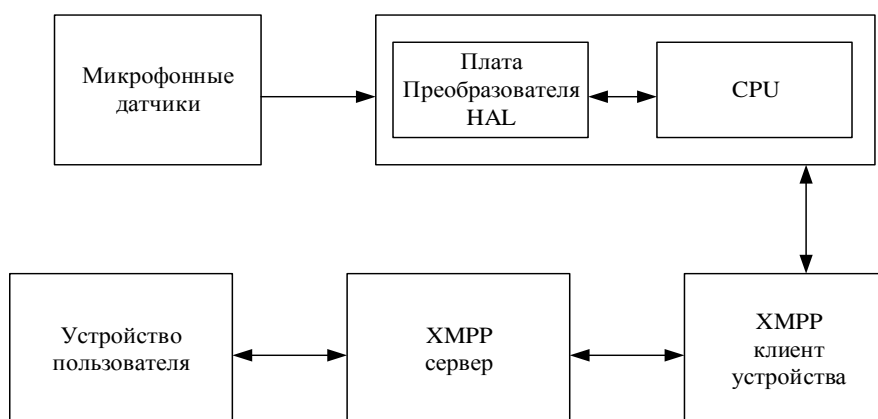


Рисунок 1 – Структурная схема системы идентификации звуков

Аппаратная часть разрабатываемого комплекса представлена микрофонным модулем, подключенным к микропроцессорной системе под управлением ОС Linux. В основе модуля лежит микросхема WM8960, которая представляет собой аудиокодек с низким энергопотреблением. Чтобы обеспечить полный программный контроль над всеми функциями, WM8960 использует двухпроводной интерфейс управления. Он полностью совместим и является идеальным партнером для широкого спектра стандартных микропроцессоров, контроллеров и DSP. Неиспользуемые схемы можно отключить с помощью программного обеспечения для экономии энергии. В качестве микропроцессорного модуля был выбран Raspberry Pi 3 Model B – полноценный бесшумный компьютер размером с банковскую карту с 64-битным четырёхядерным процессором ARM Cortex-A53 на однокристальном чипе Broadcom BCM2837 [3].

Задачи, которые выполняет ПО, – это аппаратный захват аудиосигнала, предобработка, сопоставление с образцами, принятие решения, формирование уведомления, работа с пользователем. Разработка программного обеспечения вы-

полнена на языке Python. Программное обеспечение включает в себя управляющий модуль, написанный на языке Python, отвечающий за непрерывный захват и предобработку аудиопотока, полученного непосредственно с микрофонных датчиков. Также в состав ПО входит модуль цифровой обработки и распознавания, в его зону ответственности входит формирование цифровых отпечатков, передача их в нейронную сеть. За отправку уведомлений отвечает модуль обмена, на котором реализован клиент XMPP IoT.

Разрабатываемая система непрерывно фиксирует акустическое окружение посредством HAL-модуля, при этом ключевыми параметрами являются быстродействие и малая ресурсоемкость подсистемы захвата. Обработка через аудиострим непосредственно с HAL не должна вносить задержку в обработку, превышающую 100 мс, применение быстрого преобразования Фурье и анализа превышения порога звука может исключить подсистему предобработки в некоторых пороговых случаях и сформировать уведомление непосредственно на этапе захвата. Предобработка включает в себя шумоподавление, спектральный анализ, формирование отпечатка аудиосигнала, который в дальнейшем будет сопоставлен с готовыми образцами, полученными на этапе обучения системы. Если аудиосигнал по отпечатку удалось идентифицировать, система формирует уведомление для пользователя, которое включает в себя временные отметки, место и характер сигнала. Сформированное уведомление высылается пользователю через IOT-XMPP клиент, сконфигурированный заранее.

При разработке мобильного приложения использовался готовый стандарт для передачи сообщений. В качестве транспорта для передачи данных был использован протокол XMPP. Данное решение обладает рядом преимуществ, в число которых входят: открытость программного кода для написания клиентов и серверных модулей, высокая масштабируемость, наличие плотной сети преднастроенных серверов. Клиентские XMPP приложения для Android и IOS также широко распространены, что позволяет значительно снизить стоимость разработки прототипа. В основе механизма обмена лежит XMPP (Extensible Messaging and Presence Protocol) – расширяемый протокол обмена сообщениями и информацией о присутствии, основанный на XML, свободный для использования протокол для мгновенного обмена сообщениями и информацией о присутствии в режиме, близком к режиму реального времени. Изначально спроектированный легко расширяемым протокол помимо передачи текстовых сообщений поддерживает передачу голоса, видео и файлов по сети. Когда серверное и клиентское приложения запущены, пользователь имеет возможность конфигурации сервера через клиентскую часть простыми командами чата, это позволяет вручную конфигурировать интервалы уведомлений, приходящих со стороны программно-аппаратного комплекса, а также конфигурировать типы событий, которые могут активировать те или иные уведомления. Данный механизм основан на классе XMPP нотификатор и может быть промасштабирован при переходе от прототипа к разработке реального устройства.

В разрабатываемой системе для получения аудиоданных используется два микрофона, при захвате аудио формируется два независимых аудиоканала, которые в дальнейшем формируют двумерный numpy-массив, участвующий в эталонном распознавании.

В структуру модуля цифровой обработки и распознавания входят искусственная нейронная сеть (ANN), реализация поставленной задачи включает в себя формирование структуры ANN, обучение тестовыми последовательностями, а также последующую верификацию. Построение структуры ANN выполнено посредством библиотеки PyTorch, использование данной библиотеки обусловлено ее доступностью и большим объемом обучающего материала.

В основе нейронной сети лежит последовательная модель на 4 реляционных слоя с тремя выходными нейронами, в качестве эталонов использованы стереозаписи плача ребенка в различном окружении, звонка колокольчика и звонка в дверь.

При обучении модели были использованы различные комбинации входных наборов, сводная информация приведена в таблице 1.

Таблица 1 – Ключевые параметры результатов обучения

Входной набор	Эпохи	Точность	Границы срабатывания	Задержка преобразования
Спектрограмма	128	68 %	0.74 сек	0.2 сек
Спектрограмма+ активные фильтры	128	78 %	0.74 сек	1.2 сек
Спектрограмма + MFCC	128	74 %	1.4 сек	0.5 сек
Спектрограмма + хроматограмма	128	79 %	1.3 сек	0.4 сек
Спектрограмма + MFCC + хроматограмма	128	86 %	1.5 сек	1.4 сек
Спектрограмма + MFCC + активные фильтры	128	88 %	2.0 сек	1.9 сек

При настройке модели в зависимости от комбинации входных наборов для количества эпох равных 128 средняя точность составила 78.8 %, что является достаточно хорошим результатом. Однако, при увеличении объема входных данных, существенно увеличивается время обработки, например, время срабатывания при анализе спектрограммы отличается от времени срабатывания полного набора данных с применением активных фильтров в 5.2 раза. В данной ситуации также немаловажно учитывать производительность конечной системы. По итогу обучения максимальная эффективность для соотношения точность – производительность была получена для набора спектрограмма + хроматограмма (точность 79 %, задержка 1.7 секунды).

Исходя из полученных данных, можно предположить, что концепция динамически конфигурируемой нейронной сети как продукта для конечного пользователя не является жизнеспособной – разбежка результатов даже для изолируемого набора данных требует оптимизации и системного подхода к обучению сети. Оптимальным решением представляется создание системы с жестко запрограммированным программным обеспечением, включающего максимальное количество распознаваемых паттернов. Для расширения количества идентифицируемых объектов потребуются дополнительная разработка, поэтому для исходного ПО нужна система версионирования, поддержка обновлений, а также возможность восстановления исходной версии программного обеспечения. Перспективным направлением дальнейшей деятельности может быть применение, адаптация и комбинирование алгоритмов распознавания, входящих в со-

став библиотек системы компьютерной математики Matlab и языка программирования R.

Таким образом, подход к идентификации звуков посредством распознавания ресурсами нейронной сети выглядит реалистичным, однако при переходе от прототипа к реальному устройству потребуется оптимизация алгоритмов обработки ввиду значительной ресурсоемкости преобразования уже на этапе создания прототипа.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Bosch Smart Home: Security and comfort from an single source [Electronic resource]. – Mode of access: <https://www.bosch-smarthome.com/uk/en/>. – Date of access: 05.05.2022.

2. Ubisys Sicherheits-Starterpaket Smart Home [Electronic resource]. – Mode of access: <https://www.smarthome-store.de/de/starter-kit/sicheheitsstarerpaket.html>. – Date of access: 22.09.2022.

3. Raspberry Pi Documentation [Electronic resource]. – Mode of access: <https://www.raspberrypi.com/documentation/>. – Date of access: 25.09.2022.