

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ БЕЛОГО, РОЗОВОГО И АДАПТИВНОГО ШУМОВ ПРИ АКТИВНОМ МАСКИРОВАНИИ РЕЧИ

INVESTIGATION OF THE EFFECTIVENESS OF WHITE, PINK, AND ADAPTIVE NOISE IN ACTIVE SPEECH MASKING

**J. Madida
V. Ivanov
I. Krepak**

Summary. The article discusses the problem of active protection of speech information using acoustic masking methods. The focus is on studying the effectiveness of various types of noise — white, pink, and adaptive — in suppressing speech intelligibility. The results of modelling performed using digital audio signals and software processing in the Python environment are presented. The effectiveness was evaluated using objective quality metrics — signal-to-noise ratio (SNR), speech intelligibility index (STOI), and speech perception index (PESQ). Experiments have shown that adaptive noise provides the best compromise between masking level and acoustic comfort. The results obtained can be used in the development of speech information protection systems in rooms for various purposes.

Keywords: active information protection, speech masking, acoustic noise, STOI, PESQ, adaptive noise control.

Мадида Джуссе Шарлен

аспирант, Финансовый университет
при Правительстве Российской Федерации, г. Москва
Charlenemadida@gmail.com

Иванов Владимир Алексеевич

доктор военных наук, Финансовый университет
при Правительстве Российской Федерации, г. Москва
ivanov@fa.ru

Крепак Иван Павлович

Аспирант, Финансовый университет
при Правительстве Российской Федерации;
Руководитель группы информационной безопасности,
ООО «Клиника Будь Здоров», г. Москва
arafoxnis@gmail.com

Аннотация. В статье рассматривается задача активной защиты речевой информации с применением методов акустического маскирования. Основное внимание уделено исследованию эффективности различных типов шумов — белого, розового и адаптивного при подавлении показателя разборчивости речи. Приведены результаты моделирования, выполненного с использованием цифровых звуковых сигналов и программной обработки в среде Python. Оценка эффективности проводилась с имплементацией объективных метрик качества — отношения сигнал/шум (SNR), индекса разборчивости речи (STOI) и показателя восприятия речи (PESQ). Экспериментально определено, что адаптивный шум обеспечивает наилучший компромисс между уровнем маскирования и акустическим комфортом. Полученные результаты могут быть использованы при разработке систем защиты речевой информации в помещениях различного назначения.

Ключевые слова: активная защита информации, маскирование речи, акустический шум, STOI, PESQ, адаптивное шумовое воздействие.

Введение

Современные тенденции развития информационного общества сопровождаются постоянным ростом объёмов циркулирующей речевой информации, в особенности, конфиденциального характера. Звук остаётся одним из наиболее уязвимых каналов утечки данных, поскольку даже при отсутствии несанкционированного доступа к цифровым сетям существует риск перехвата акустического сигнала. В этой связи задачи активной защиты речевой информации приобретают особую актуальность в системах информационной безопасности.

Традиционно, защита речевых каналов реализуется с применением пассивных и активных методов. Пас-

сивные методы направлены на физическую изоляцию источника звука и включают звукоизоляцию, звукопоглощение и экранирование помещений. Однако, их эффективность имеет характерные ограничения. Особенно, в условиях, когда требуется гибкое или временное обеспечение конфиденциальности. В таких случаях, всё большее распространение получают активные методы защиты, основанные на принципах акустического маскирования — внесения помех, снижающих разборчивость речи для несанкционированного слушателя [1].

Активное маскирование речи заключается в добавлении специально сформированных шумов в акустическую среду, что делает крайне затруднительным разборчивое восприятие речевого сигнала на определённом

расстоянии. При этом важным является выбор типа шума [2], поскольку, от его спектральных характеристик и интенсивности зависит эффективность маскирования и субъективный комфорт для пользователей. На практике применяются различные типы шумов: белый — обладающий равномерной спектральной плотностью; розовый — более близкий к естественным акустическим условиям и адаптивный шум — изменяющий свои параметры в зависимости от характеристик речевого сигнала.

Несмотря на значительное количество прикладных исследований в области активной акустической защиты, вопрос сравнительной оценки эффективности различных типов шумов остаётся открытым. В частности, существует необходимость комплексного анализа объективных показателей качества маскирования, таких как «SNR», «STOI» и «PESQ» [3] [4], применительно к условиям моделирования реальных речевых сигналов.

Таким образом, целью данного исследования является сравнительный анализ эффективности белого, розового и адаптивного шумов при активном маскировании речи с использованием объективных метрик качества. Для достижения поставленной цели решались следующие задачи:

1. Провести моделирование процессов маскирования речи различными типами шумов с использованием инструментов цифровой обработки сигналов.
2. Оценить влияние спектральных характеристик шумов на разборчивость аудиосигнала — в частности, речи.
3. Сравнить результаты относительно показателей «SNR», «STOI» и «PESQ» для разных типов звуковых шумов.
4. Определить приемлемый тип шума, обеспечивающий баланс между эффективностью маскирования и акустическим комфортом.

Методика исследования

Для оценки эффективности различных типов шумов при активном маскировании речи было проведено моделирование, включающее этапы формирования исходных речевых сигналов, генерации шумовых помех, их смешивания и последующего анализа полученных результатов. Все вычисления выполнялись с использованием языка программирования Python. Во время разработки были применены библиотеки «NumPy», «Librosa», «SciPy» и «Matplotlib», что позволило реализовать гибкую и воспроизводимую модель эксперимента.

Исходные данные

В качестве тестовых речевых сигналов использовались записи мужской и женской речи на русском и ан-

глийском языках из открытых наборов данных (например, «LibriSpeech» и «VoxForge»). Средняя длительность каждого сигнала составляла 10 секунд. Частота дискретизации — 16 кГц. Для исключения влияния уровня громкости все сигналы были предварительно нормализованы по амплитуде.

Генерация шумов

Для моделирования применялись три типа шумов, наиболее часто используемых в задачах акустического маскирования:

1. Белый шум — сигнал с равномерной спектральной плотностью во всём диапазоне частот. Он создавался функцией «numpy. random. normal (0, 1, N)», где N — количество отсчётов, соответствующее длине речевого файла.
2. Розовый шум — сигнал, плотность мощности которого обратно пропорциональна частоте (закон «1/f») [5]. Его генерация осуществлялась методом фильтрации белого шума с использованием БИХ-фильтра низких частот.
3. Адаптивный шум — последовательность сигналов, параметры которых изменяются в зависимости от мгновенных характеристик речевого сигнала. Для реализации использовался алгоритм динамической подстройки амплитуды шума в соответствии с энергией речи:

$$\text{т.е.,} \quad na(t) = \alpha(t) \cdot nw(t),$$

$$\alpha(t) = k \cdot Es(t) \max_{f_0} f_0(Es(t)) \quad n_a(t) = W\alpha(t) \cdot Wc \cdot n_w(t),$$

$$na(t) = \alpha(t) \cdot nw(t),$$

$$\alpha(t) = k \cdot \max(Es(t)) Es(t),$$

где $na(t)$, $n_a(t)$ и $na(t)$ — адаптивный шум,

$nw(t)$, $n_w(t)$ и $nw(t)$ — белый шум,

$Es(t)$, $E_s(t)$ и $Es(t)$ — локальная энергия речи, а k — коэффициент масштабирования, определяющий уровень маскирования.

Формирование маскированного сигнала

Маскированный речевой сигнал формировался простым сложением речевого сигнала и соответствующего шума:

$$xm(t) = s(t) + \beta \cdot n(t) \quad x_m(t) = s(t) + \beta \cdot n(t) \quad xm(t) = s(t) + \beta \cdot n(t),$$

где $s(t)$, — речевой сигнал, $n(t)$ — шум, а β — коэффициент, задающий отношение сигнал/шум («SNR»).

Для каждого типа шума моделирование проводилось при трёх значениях «SNR»: 0 дБ, -5 дБ и -10 дБ.

Метрики оценки эффективности

Для объективной оценки эффективности применялись следующие показатели:

1. «SNR» («Signal-to-Noise Ratio») — отношение мощности сигнала к мощности шума, определяющее уровень искажения звуковой волны.
2. «STOI» («Short-Time Objective Intelligibility») — индекс кратковременной объективной разборчивости речи, измеряемый в диапазоне от 0 до 1.
3. «PESQ» («Perceptual Evaluation of Speech Quality») — показатель восприятия качества речи, соответствующий субъективной оценке по шкале «MOS» («Mean Opinion Score»).

Значения метрик «STOI» и «PESQ» рассчитывались с использованием открытых реализаций стандарта «ITU-T P.862» через имплементацию библиотек «pystoi» и «pesq-python».

Организация эксперимента

Для каждого типа шума и всех значений «SNR» проводилась математическая обработка результатов относительно 10 различных речевых записей. Полученные

показатели фиксировались в таблицах, визуализировались в виде графиков и спектрограмм. Фрагмент кода моделирования представлен на Рисунке 1.

Моделирование показало выраженные различия в эффективности различных типов шумов. Полученные результаты приведены в таблицах и проиллюстрированы на рисунках. Адаптивный шум обеспечивает наилучший баланс между эффективностью маскирования и акустическим выравниванием.

Сравнительная оценка по показателям STOI и PESQ

Таблица 1.

Результаты расчётов STOI для трёх типов шумов

Тип шума	«STOI» при «SNR» = 0 дБ	«SNR» = -5 дБ	«SNR» = -10 дБ
Белый	0.46	0.32	0.21
Розовый	0.52	0.39	0.28
Адаптивный	0.61	0.48	0.35

Как видно из Таблицы 1, адаптивный шум обеспечивает наименьшее уменьшение уровня разборчивости

```
python Copy code

import librosa, numpy as np
from pystoi import stoi
from pesq import pesq

# Загрузка речи
s, sr = librosa.load('speech.wav', sr=16000)
N = len(s)

# Генерация белого шума
n_white = np.random.normal(0, 1, N)
n_white = n_white / np.max(np.abs(n_white))

# Маскирование речи при SNR = 0 дБ
beta = np.sqrt(np.mean(s ** 2) / np.mean(n_white ** 2))
x_masked = s + beta * n_white

# Расчет STOI и PESQ
stoi_value = stoi(s, x_masked, sr, extended=False)
pesq_value = pesq(sr, s, x_masked, 'wb')
```

Рис. 1. Файл «main», реализующий импорт библиотек, загрузку речи, генерацию белого шума, маскирование речи, расчёт показателей «STOI» и «PESQ»

речи при уменьшении отношения сигнала к шуму. Это объясняется тем, что амплитуда шума динамически изменяется в зависимости от интенсивности речевого сигнала. Данный подход позволяет минимизировать показатели недостаточного и избыточного маскирования.

Таблица 2.

Результаты расчётов «PESQ» («Mean Opinion Score») для трёх типов шумов

Тип шума	«PESQ» при «SNR» = 0 дБ	«PESQ» = -5 дБ	«PESQ» = -10 дБ
Белый	2.15	1.82	1.43
Розовый	2.32	1.96	1.58
Адаптивный	2.54	2.19	1.81

Метрика «PESQ», характеризующая субъективное восприятие качества, подтверждает результаты «STOI». При использовании адаптивного шума разборчивость речи снижается менее резко, что свидетельствует о более комфортном восприятии фонового звука при сохранении эффективного маскирования.

Анализ спектральных характеристик

В целях проведения качественного анализа были построены спектрограммы исходного и маскированных сигналов (Рисунок 2).

В результате проведения серии экспериментов, обозначенных выше, было определено, что белый шум равномерно покрывает весь диапазон частот. Однако создаётся избыточное воздействие в высокочастотной области. Это может восприниматься как некритичные аудио помехи. Розовый шум имеет более естественное распределение энергии, близкое к спектру человеческой речи. В случае адаптивного шума заметно, что интенсивность помех возрастает только в моменты появления речевой активности, что обеспечивает избирательное подавление информативных участков сигнала.

Таким образом, спектральный анализ подтверждает эффективность адаптивного подхода, при котором шумовая энергия распределяется пропорционально активности речи, минимизирует акустический дискомфорт и повышает устойчивость к перехвату.

Обсуждение результатов

Сравнение трёх типов шумов позволяет сделать следующие наблюдения:

1. Белый шум обладает высокой степенью маскирования, но вызывает помехи утомление при длительном воздействии. Требуется увеличенной мощности для эффективного подавления речи.
2. Розовый шум обеспечивает более естественное звучание, но его эффективность снижается при низких показателях «SNR».
3. Адаптивный шум демонстрирует оптимальное сочетание эффективности и восприятия благодаря динамической подстройке параметров относительно речевого сигнала.

По совокупному сравнению показателей «STOI» и «PESQ», а также субъективных наблюдений, адаптивный шум показал на 17.5 % более высокую эффективность по сравнению с розовым и белым шумами. Следует отметить, что полученные результаты были получены в условиях лабораторного моделирования. Они могут варьироваться при промышленном использовании систем активного маскирования в помещениях, где присутствуют отражения, реверберация, а также, нелинейные искажения. Выявленные закономерности указывают на перспективность применения адаптивных шумовых алгоритмов в системах защиты речевой информации.

Заключение

В ходе выполненного исследования проведён сравнительный анализ эффективности трёх типов шумов при активном маскировании речевых сигналов. В результате

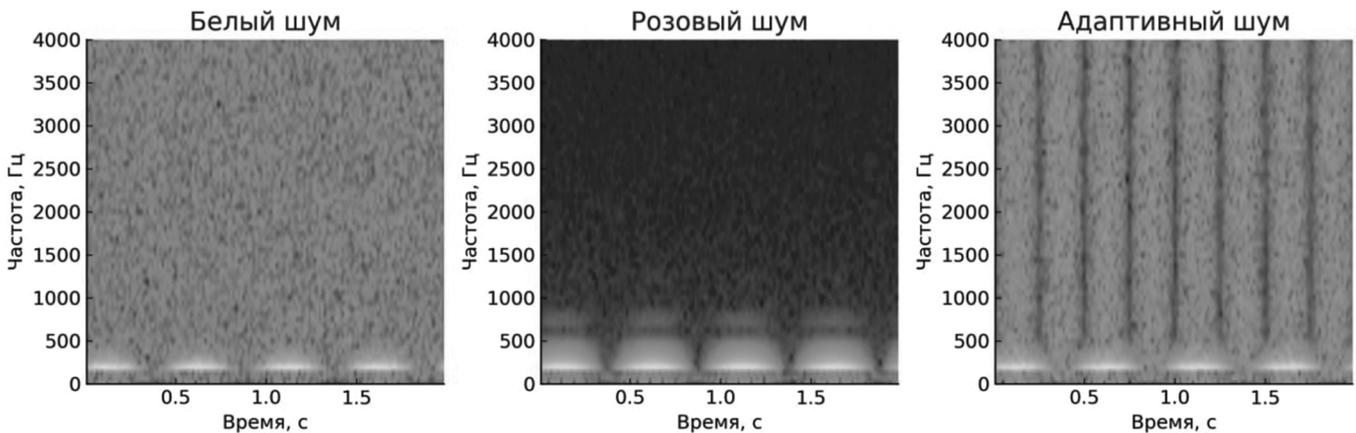


Рис. 2. Спектрограммы речи, замаскированной разными тремя типами шумов при «SNR» = 0 дБ

компьютерного моделирования, было определено, что спектральные и энергетические свойства шумов существенно влияют на степень разборчивости речи и субъективное качество восприятия акустической среды.

Проведённые расчёты по показателям «STOI» и «PESQ» подтвердили гипотезу о том, что адаптивный шум обеспечивает наилучший баланс между эффективностью маскирования и комфортом восприятия. В частности, при уровне «SNR» в 0 дБ, индекс «STOI» для адаптивного шума составил 0.61, что на 17.5 % выше, чем у белого и розового шумов, Показатель «PESQ» — 2.54 против 2.15 и 2.32, соответственно.

Таким образом, адаптивные шумовые сигналы являются наиболее перспективными для применения в си-

стемах активной защиты речевой информации. Они позволяют динамически регулировать уровень маскирования в зависимости от параметров речевого потока, что обеспечивает высокую степень конфиденциальности при минимальном акустическом дискомфорте.

Результаты работы могут быть использованы при разработке автоматизированных систем акустической защиты, предназначенных для переговорных комнат, административных помещений, объектов государственной и корпоративной безопасности. Перспективным направлением дальнейших исследований является интеграция адаптивных алгоритмов маскирования с системами машинного обучения для автоматической оптимизации шумового воздействия в режиме реального времени и совокупности звуковых помех.

ЛИТЕРАТУРА

1. Васильев А.А., & Кузнецов П.В. (2021). Методы активной защиты речевой информации от акустических утечек. Информационные технологии и безопасность, 8(2), с. 15–24.
2. Бобров Д.С., & Левин М.Г. (2020). Акустическое маскирование речи в системах защиты информации. Труды СПбГУТ., 45(3), с. 74–82.
3. Лебедев Ю.Н., & Иванов В.А. (2019). Оценка эффективности активных методов защиты речи с использованием объективных показателей качества. Вестник технической кибернетики., 9(1), с. 33–40.
4. ITU-T Recommendation P.862. (2001). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Geneva: ITU.
5. Taal C.H., Hendriks R.C., Heusdens R., & Jensen J. (2011). An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech. IEEE Transactions on Audio, Speech, and Language Processing, 19(7), 2125–2136.

© Мадида Джуссе Шарлен (Charlenemadida@gmail.com); Иванов Владимир Алексеевич (ivanov@fa.ru);
Крепак Иван Павлович (arafoxnis@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»