

Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

А. Петровский,

доктор технических наук, профессор

А. Борович,

доктор

М. Парфенюк,

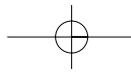
доктор

В контексте перцептуальной обработки речевых сигналов в статье рассматривается ДПФ с неравномерным частотным разрешением (англ. Warped Discrete Fourier Transform (WDFT)). Разработан метод минимизации ошибки реконструкции речевого сигнала на основе WDFT. Предложен алгоритм построения данного преобразования с избыточным базисом, обеспечивающим решение проблемы синтеза. Распределение мощности сигнала по шкале барков, вычисленное на основе WDFT, имеет точность приблизительно одинаковую в низкочастотных и высокочастотных диапазонах спектра. Это обуславливает возможность построения психоакустической модели в области WDFT-коэффициентов.

Abstract

Warped Discrete Fourier Transform (WDFT) allows non-uniform sampling of the z-transform of finite length sequence by using all-pass function. We focus on perceptual warping which allocates frequency samples in good accordance with the Bark scale. Namely, the problem of signal distortions generated by WDFT synthesis block is of interest. Spectral features of the reconstructed signal are analyzed and discussed in context of perceptual processing. Our proposition is to construct an overcomplete WDFT sinusoidal basis in order to minimize reconstruction error.





Введение

ДПФ является мощным средством для частотного анализа с постоянной разрешающей способностью. Однако в контексте перцептуальной обработки сигналов, основанной на психоакустической модели восприятия акустической информации человеком, возникает необходимость в неравнополосной частотной декомпозиции сигнала в соответствии со шкалой критических частотных полос (барков, ERB или MEL) [1, 2]. Примером могут служить системы перцептуального кодирования звука и речи, а также подавления шума окружающей среды в речевом сигнале [3, 4].

Идея построения преобразования с переменным частотным разрешением не нова. Впервые она была предложена почти 40 лет назад [5]. В настоящее время развивается ряд новых нестандартных подходов [6–8]. Одним из них является дискретное преобразование Фурье с неравномерным частотным разрешением (англ. Warped Discrete Fourier Transform (WDFT)) [8], позволяющее получить z -преобразование конечной последовательности входных отсчётов сигнала с неравномерным разложением коэффициентов преобразования на единичной окружности z -плоскости строго регулярно посредством использования фазового звена.

Определение WDFT

ДПФ с переменным частотным разрешением (англ. Nonuniform DFT (NDFT)) является наиболее обобщающим вариантом ДПФ [6]. Кроме единственности решения, NDFT не ограничивает расположение коэффициентов преобразования в z -плоскости никоим образом. Преобразование WDFT есть специальный случай NDFT, для которого коэффициенты преобразования располагаются неравномерно, но строго регулярно на единичной окружности z -плоскости.

WDFT последовательности $x[n]$ из N точек определяется по следующей формуле:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x[n] \hat{z}_k^{-n}, \quad k = 0, \dots, N-1, \quad (1)$$

где \hat{z}_k — изображения преобразованных фазовым звеном $A(z)$ равноотстоящих точек на единичной окружности в z -плоскости:

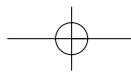
$$z_k^{-1} = e^{-j\frac{2\pi k}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k), \quad k = 0, \dots, N-1 \quad (2)$$

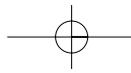
$A(z)$ — устойчивое фазовое звено произвольного порядка.

Простейший вариант WDFT основан на фазовом звене первого порядка с действительным коэффициентом a :

$$A(z) = \frac{z^{-1} - a}{1 - az^{-1}}. \quad (3)$$

Условием стабильности фильтра является $|a| < 1$. В зависимости от знака низкочастотный ($a > 0$) или высокочастотный диапазон спектра растягивается путём неравномерного распределения коэффициентов преобразования





Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

на единичной окружности в z -плоскости. Формально это может быть выражено следующим образом:

$$\hat{\omega} = \omega + 2 \arctan\left(\frac{a \sin \omega}{1 - a \cos \omega}\right) \text{ для } \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases}, \quad (4)$$

то есть осуществляется билинейное преобразование z -плоскости в новую искривлённую \hat{z} -плоскость (рис. 1).

Как обобщение ДПФ, WDFT также имеет свойства линейности, симметрии и сдвига [7, 8]. Сопряжённая симметрия для действительных данных имеет силу и для WDFT:

$$\hat{X}(z_{N-1-k}) = \hat{X}^*(z_k), \quad (5)$$

однако ряд важных свойств ДПФ теряется.

В матричной записи (с $\hat{X}[k]$, обозначающим $\hat{X}(z_k)$), WDFT может быть представлено следующим выражением:

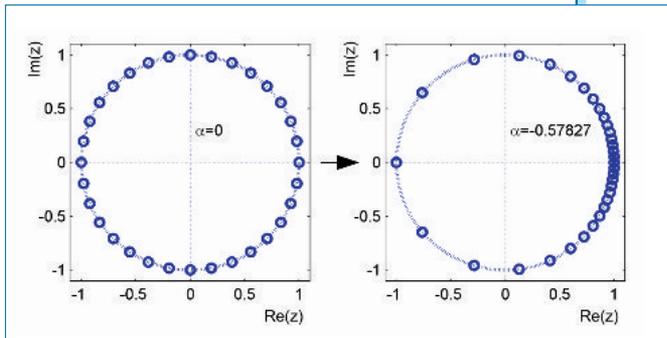


Рис.1. Расположение коэффициентов преобразования на единичной окружности в z -плоскости: ДПФ (слева) и WDFT (справа)

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & A(z_0) & \cdots & A(z_0)^{N-1} \\ 1 & A(z_1) & \cdots & A(z_1)^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & A(z_{N-1}) & \cdots & A(z_{N-1})^{N-1} \end{bmatrix}}_{\mathbf{D}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}. \quad (6)$$

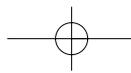
Обратное преобразование WDFT

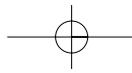
В случае неравномерного частотного разрешения матрицы Вандермонда, к классу которых относится и матрица \mathbf{D} для WDFT, обладают свойствами, делающими невозможным получение точной обратной матрицы. А именно, матрица WDFT может рассматриваться как сингулярная, поскольку между некоторыми её строками существуют почти линейные зависимости. Этот факт представляется как очень малое значение её детерминанта $\det \mathbf{D} = \prod_{i < j} (\hat{z}_i^{-1} - \hat{z}_j^{-1})$ для определённого \hat{z}_k . Другими словами, WDFT-матрица явля-

ется плохо обусловленной. Это означает, что любой численный алгоритм обращения матрицы, применяемый к подобной матрице, является очень чувствительным к малым изменениям данных.

Чтобы оценить усиление ошибки, можно использовать число обусловленности матрицы полного ранга (собственное значение матрицы):

$$\text{cond}(\mathbf{D}) = \|\mathbf{D}\| \|\mathbf{D}^{-1}\| = \sigma_{\max} / \sigma_{\min}, \quad (7)$$





Петровский А., Борович А., Парфенов М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

где $\|\cdot\|$ означает произвольную норму матрицы (в общем случае евклидову), а σ_{\max} и σ_{\min} — соответственно наибольшее и наименьшее сингулярные числа.

Как показано на рисунке 2, число обусловленности преобразуемой матрицы зависит от её размерности и величины коэффициента деформации α и имеет очень большие значения даже при малоразмерных преобразованиях, слегка отличающихся от ДПФ. Плохая обусловленность является неотъемлемым свойством матриц Вандермонда, связанных с реальными проблемами. Единственным исключением является случай отсутствия деформирования частотной оси, когда WDFT становится обычным ДПФ.

Существует ряд методов для аппроксимации обращений плохо обусловленных матриц. Данные методы используют разложение по сингулярным числам матрицы (англ. Singular Value Decomposition — SVD) [9], которое определяется так:

$$\mathbf{D} = \mathbf{U}\mathbf{Y}\mathbf{V}^H = \sum_{i=1}^N \mathbf{u}_i \sigma_i \mathbf{v}_i^H, \quad (8)$$

где

$$\begin{aligned} \mathbf{U} &= [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N] \\ \mathbf{V} &= [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N] \end{aligned} \quad (9)$$

— это матрицы с ортогональными столбцами

$$\mathbf{U}^H \mathbf{U} = \mathbf{V} \mathbf{V}^H = \mathbf{I}_N, \text{ а}$$

$$\mathbf{Y} = \text{diag}(\sigma_1, \dots, \sigma_N) \quad (10)$$

является диагональной матрицей, состоящей из сингулярных чисел, отсортированных по убыванию: $\sigma_1 \geq \dots \geq \sigma_N \geq 0$. Левый и правый столбцы матриц \mathbf{U} и \mathbf{V} — сингулярные векторы, а N обозначает эрмитово транспонирование матрицы. Степень неполноты матрицы \mathbf{D} (наличие некоторых линейных зависимостей между её столбцами) проявляется в существовании почти нулевых сингулярных чисел σ_i . Распределения сингулярных значений для нескольких различных матриц WDFT проиллюстрированы на рисунке 3.

Как показывает рисунок 3, значения достаточно большого количества сингулярных чисел близки к нулю. Распределение компонент SVD может дать много полезных пояснений плохой обусловленности матриц, а также использоваться для формирования псевдообратной матрицы:

$$\mathbf{D}^\dagger = \sum_{i=1}^N f_i \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^H. \quad (11)$$

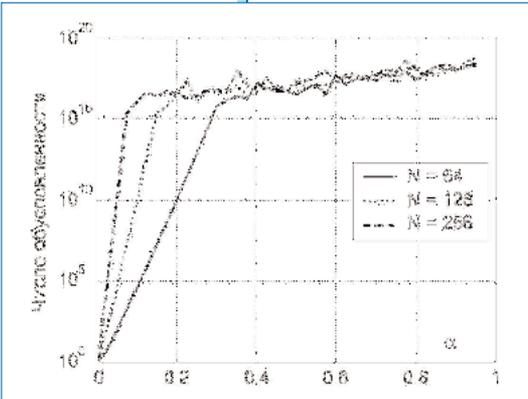


Рис. 2. Зависимость числа обусловленности матрицы от коэффициента деформации α для различных размеров WDFT

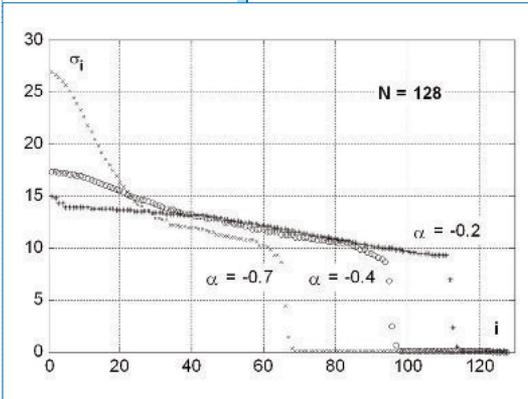
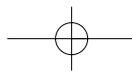
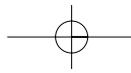


Рис. 3. Распределение сингулярных чисел как функция коэффициента деформации α частотной шкалы для фиксированного размера преобразования $N=128$





Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

Символы f_i обозначают так называемые коэффициенты ослабления фильтра и должны быть все равны единице для получения точного обращения матрицы. Теория регуляризации матриц [9] рекомендует исключать влияние малых сингулярных чисел, ослабляя их вклад в (11). Это реализуется путём установки соответствующих значений f_i . Так как сингулярные векторы с наибольшими индексами имеют ярко выраженный колебательный характер, это действие изменяет спектральный состав данных.

Основное различие между известными методами регуляризации заключается в выборе коэффициентов ослабления фильтра. В простейшем подходе усечённого SVD (англ. Truncated SVD — TSVD) [10] сумма (11) ограничивается обычным отбрасыванием термов, имеющих сингулярные числа меньше определённого порога:

$$f_i = \begin{cases} 1 & \sigma_i > \lambda, \\ 0 & \text{в противном случае.} \end{cases} \quad (12)$$

Менее радикальным является демпфированное SVD (англ. Damped SVD — DSVD), в котором коэффициенты ослабления фильтра постепенно изменяются в сторону нуля, обеспечивая более сглаженное отсечение:

$$f_i = \frac{\sigma_i}{\sigma_i + \lambda}. \quad (13)$$

В обоих случаях параметр регуляризации λ должен быть выбран с осторожностью. Решение должно быть стабильным при спектральных изменениях, ограниченных определённым минимумом.

Оценка ошибки реконструкции сигнала

В зависимости от приложения, где применяется WDFТ, ошибка реконструкции речевого сигнала (вычисление обратного WDFТ) может иметь решающее влияние на качество синтезированного сигнала (число артефактов), например, в системах редактирования шумов окружающей среды в речевом сигнале на основе психоакустически мотивированного правила взвешивания спектра зашумленной речи [11].

Используя матричную форму записи, вектор сигнала ошибки \mathbf{d} можно определить как разность между оригинальным сигналом \mathbf{x} и реконструированным сигналом $\hat{\mathbf{x}}$:

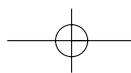
$$\mathbf{d} = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{D}^\dagger \mathbf{D}) \mathbf{x} \quad (14)$$

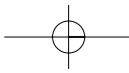
Соответствующая мера в частотной области может быть выражена как спектральная плотность мощности (СПМ) сигнала ошибки:

$$S_{dd}(\omega) = \frac{1}{N} E \left\{ \left| \mathbf{e}(\omega)^H \mathbf{d} \right|^2 \right\} = \frac{1}{N} \mathbf{e}(\omega)^H \mathbf{R}_{dd} \mathbf{e}(\omega)^H, \quad (15)$$

где

$$\mathbf{e}(\omega) = \left[1 \quad e^{-j\omega} \quad e^{-j2\omega} \quad \dots \quad e^{-j\omega(N-1)} \right]^T \quad (16)$$





Петровский А., Борович А., Парфенов М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

является вектором-столбцом синусоидального базиса ДПФ, а \mathbf{R}_{dd} обозначает ковариационную матрицу сигнала ошибки. Положим, что входной речевой сигнал \mathbf{x} является случайным вектором с нулевым математическим ожиданием и известной ковариационной матрицей $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\}$. Если $\mathbf{Q} = \mathbf{I} - \mathbf{D}^\dagger \mathbf{D}$, матрица \mathbf{R}_{dd} может быть записана следующим образом:

$$\mathbf{R}_{dd} = \mathbf{Q}\mathbf{R}_{xx}\mathbf{Q}^H. \quad (17)$$

Очевидно, что спектральное искажение (15) зависит от характеристик входного речевого сигнала и качества аппроксимации обратного WDFТ. Теоретически точное обращение матрицы возможно, но данное решение будет очень нестабильным вследствие большого числа обусловленности матрицы и не найдёт практического применения в системах обработки речи. Однако, если удаётся получить стабильное решение обратной задачи, мера величины ошибки (15) может быть использована для регулирования величины артефактов в синтезированном речевом сигнале. Так, если псевдообратная матрица \mathbf{D}^\dagger определена с помощью техники SVD, представленной выше, то матрица \mathbf{Q} может быть вычислена напрямую из выражения

$$\mathbf{Q} = \mathbf{I} - \mathbf{D}^\dagger \mathbf{D} = \mathbf{I} - \sum_{i=1}^N f_i \mathbf{v}_i \mathbf{v}_i^H. \quad (18)$$

В случае перцептуальной деформации частотной шкалы выбор параметров регуляризации ограничивает влияние на уровень спектральных искажений. Другими словами, если даже нестабильная аппроксимация обратной матрицы WDFТ приводит к относительно высокой ошибке синтеза, то дальнейшая стабилизация решения только увеличит спектральные искажения. Например, положим, что коэффициент ослабления фильтра f_i вычисляется по методу DSVD (13) с параметром регуляризации $\lambda = 0,001$. СПМ ошибки WDFТ-синтеза, вычисленная по формуле (15), показана на рисунке 4. В качестве входного сигнала в эксперименте использовался стационарный окрашенный гауссовый шум, который моделировался с заданной априори теплицевой ковариационной матрицей $(\mathbf{R}_{xx})_{i,j} = p^{|i-j|}$ для $p = 0,9$. Можно заметить, что уровень искажений на заданной частоте зависит от расстояния между соседними WDFТ-коэффициентами. Сигнал полностью восстанавливается только в точках преобразования (частотах, определённых выбором коэффициента α фазового звена (3), и спектральные искажения особенно заметны в растянутых частотных диапазонах, в то время как в сжатых частотных диапазонах ошибка синтеза имеет приемлемый уровень.

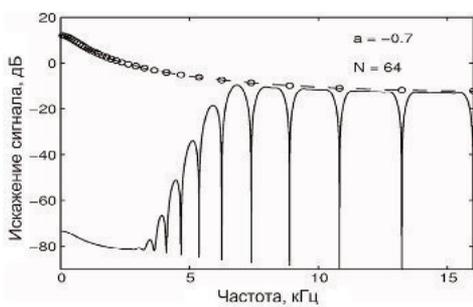
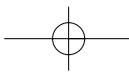


Рис. 4. СПМ оригинального сигнала (кружочки) и соответствующей ошибки синтеза (сплошная линия)

Этот эффект наблюдается также и на спектрограмме реконструированного речевого сигнала в виде узких спектральных «дыр», локализованных между точками преобразования (рисунок 56). Таким образом, частотная характеристика синтезированного речевого сигнала в некоторых деталях не восстанавливается. Заметим, что проявление данного эффекта в высокочастотной части частотного диапазона ($\alpha < 0$) обусловлено перцептуальной деформацией частотной шкалы, но ситуация меняется, и эффекты ошибки WDFТ-синтеза проявляются в низкочастотной части спектра ($\alpha > 0$). Однако данный случай неинтересен для перцептуальных систем обработки речевых сигналов.



Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

Известно, что высокочастотный диапазон спектра речевого сигнала влияет на качество речи, в то время как низкочастотный — на разборчивость речи. Таким образом, при выборе параметров настройки WDFT необходимо искать некое компромиссное решение между требуемым частотным разрешением и ошибкой синтеза. Можно попытаться проектировать фазовое звено с коэффициентом α , зависящим от времени, но данное решение имеет очень высокую вычислительную сложность. Более того, для широкополосного сигнала невозможно одновременно обеспечить его обработку с хорошей разрешающей способностью в низкочастотной части и перфективную реконструкцию его высокочастотных компонент, даже если удастся построить управляемое фазовое звено. Единственный путь минимизации спектрального искажения — это модификация матрицы преобразования \mathbf{D} таким образом, чтобы количество почти нулевых сингулярных чисел было уменьшено.

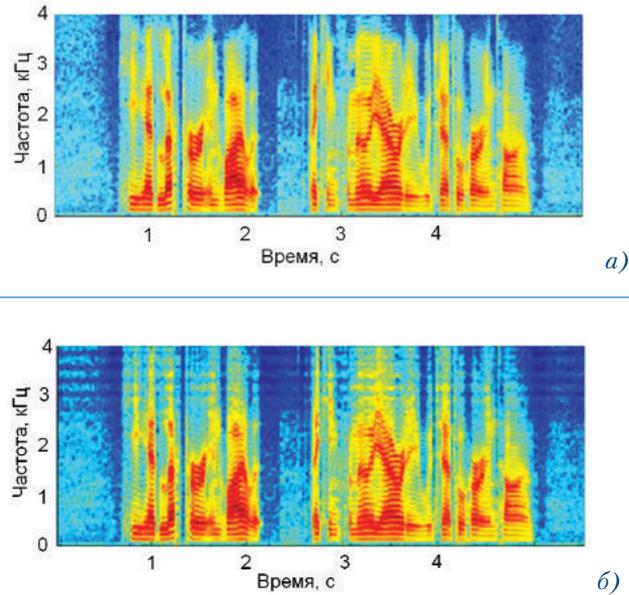


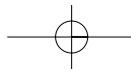
Рис. 5. Спектрограммы оригинального (а) и реконструированного (б) речевого сигнала

Избыточный синусоидальный базис

В работе [11] было показано, что формирование избыточного неортогонального базиса может внести значительный вклад в коррекцию ошибки реконструкции, а именно, при подходящем выборе векторов синусоидального базиса можно модифицировать распределение сингулярных чисел для соответствующей матрицы преобразования. Новая WDFT-матрица \mathbf{D} не является квадратной: количество строк $M > N$ увеличивается. Матричное представление WDFT с избыточным базисом может быть записано следующим образом:

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[M-1] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \hat{z}_0^{-1} & \dots & \hat{z}_0^{-N+1} \\ 1 & \hat{z}_1^{-1} & \dots & \hat{z}_1^{-N+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{z}_{M-1}^{-1} & \dots & \hat{z}_{M-1}^{-N+1} \end{bmatrix}}_{\mathbf{D}_{M \times N}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}. \quad (19)$$

Задача поиска обратной матрицы для данной прямоугольной матрицы $\mathbf{D}_{M \times N}$ [12, 13] может быть так же решена с использованием SVD-процедуры, как и для квадратной матрицы \mathbf{D} . Следует отметить, что стабильность новой матрицы выше, чем у матрицы стандартного WDFT. Избыточный базис с добавленными новыми векторами уменьшает эксцентricность SVD-эллипсоида, который является отображением единичной сферы в N -мерном пространстве.



Петровский А., Борович А., Парфенов М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

Соотношение между форматом WDFT с избыточным базисом и собственными числами для различных значений коэффициента деформации частотной шкалы α показано на рисунке 6. Видно, что если $M \rightarrow \infty$, то число обусловленности матрицы $\mathbf{D}_{M \times N}$ близко к единице.

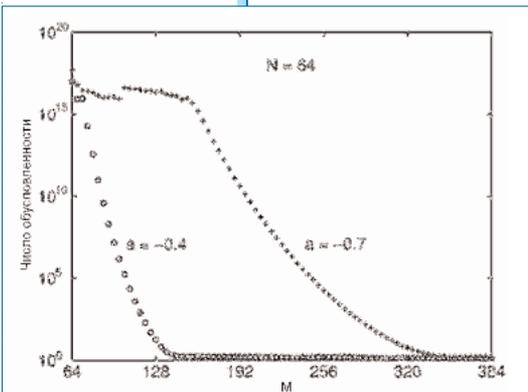


Рис. 6. Соотношение между форматом WDFT с избыточным базисом и собственными числами матриц $\mathbf{D}_{M \times N}$

С практической точки зрения необходимо выполнить два условия при формировании избыточного базиса WDFT. Во-первых, должна сохраниться регулярность расположения коэффициентов преобразования в соответствии с деформацией частотной шкалы (например, шкалой барков). Во-вторых, размер нового синусоидального базиса должен быть как можно меньшим, чтобы вычислительная сложность WDFT оставалась приемлемой.

Так как любое направление в пространстве комплексных векторов, определяемое базисным вектором, соответствует конкретному частотному диапазону, WDFT можно представить как максимально децимированный банк фильтров. При этом k -ая строка WDFT-матрицы может рассматриваться как КИХ-фильтр с передаточной функцией, определяемой по следующему выражению:

$$H_k(z) = \sum_{n=-\infty}^{\infty} A(z_k) z^{-n}, \quad k = 0 \dots N-1, \quad (20)$$

где $A(z)$ — фазовое звено первого порядка; $H_k(z)$ — полосовой фильтр с центральной частотой

$$\hat{\omega}_k = 2 \tan^{-1} \left(\frac{1+a}{1-a} \tan \left(\frac{\omega_k}{2} \right) \right), \quad \omega_k = \text{angle}(z_k) \quad (21)$$

и полосой $2\pi/N$.

АЧХ банка фильтров $H_k(z)$ для $k = 0, \dots, N-1$ представлены на рисунке 7.

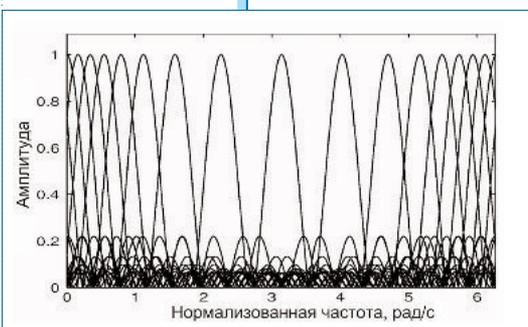
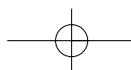


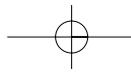
Рис. 7. АЧХ критически децимированного банка фильтров WDFT

Новый избыточный базис WDFT должен конструироваться из M векторов соответствующих импульсных характеристик КИХ-фильтров с центральными частотами

$$z_k = \exp(j \frac{2\pi k}{M}), \quad k = 0, \dots, M-1, \quad M > N, \quad (22)$$

регулярность которых на единичной окружности гарантирована, потому что вектор новых коэффициентов преобразования в точности совпадает со стандартным WDFT, но для формата преобразования $M-1$. Другими словами, размер нового избыточного базиса должен быть равен числу перекрывающихся полосовых фильтров (20), переопределённых для $k = 0, \dots, M-1$. Если коэффициент фазового звена в (3) отрицателен ($a < 0$), то максимальное расстояние между центральными частотами КИХ-фильтров составит:





Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

$$\Delta\omega_{\max} = \pi - 2 \tan^{-1} \left(\frac{1+a}{1-a} \tan \left(\frac{\pi - 2\pi/M}{2} \right) \right). \quad (23)$$

Для получения того же частотного разрешения в высокочастотном диапазоне, что и у обычного ДПФ, расстояние между центральными частотами не должно быть больше $2\pi/N$. Подставляя $\Delta\omega_{\max} = 2\pi/N$ в (21) и решая относительно M , получаем:

$$M = M_{opt} = 2\pi \left[\pi - 2 \tan^{-1} \left(\frac{1-a}{1+a} \tan \left(\frac{\pi - 2\pi/N}{2} \right) \right) \right]^{-1}. \quad (24)$$

Отметим, что для $\alpha=0$ (нет деформации частотной шкалы) $M=N$, а следовательно, $M \geq N$. Соотношение между размером избыточного базиса M и коэффициентом деформации α для заданного числа столбцов матрицы $\mathbf{D}^\dagger \mathbf{D}$ WDFT с избыточным базисом иллюстрируется на рисунке 8.

Спектральные искажения (15) могут быть вычислены для прямоугольной матрицы так же, как и для квадратной, потому что продукт $\mathbf{D}^\dagger \mathbf{D}$ всегда является $N \times N$ матрицей. На рисунке 9 показаны СПМ входного сигнала и соответствующей ошибки синтеза, вычисленные для прямоугольной матрицы $\mathbf{D}_{M \times N}$ WDFT. Как видно, ошибка синтеза уменьшается при увеличении M и может не учитываться при $M \geq M_{opt}$ (меньше -75 дБ). На практике число строк матрицы $\mathbf{D}_{M \times N}$ можно слегка уменьшить, так как перекрытие частотных характеристик полосовых фильтров (20) выбиралось случайно и без всякого психоакустического критерия. Однако прослушивание речевых тестов показало, что спектральные искажения не слышны в реконструированном сигнале речи для $M \approx M_{opt}$.

Вычислительные аспекты WDFT

Алгоритм, по эффективности сравнимый с БПФ, не может быть построен для WDFT из-за асимметричности WDFT-матрицы. Тем не менее, прямой алгоритм работы с комплексной матрицей может быть значительно оптимизирован. Наиболее усовершенствованный алгоритм, несмотря на сложность $O(N^2)$, использует факторизацию WDFT-матрицы в произведение трёх матриц: действительной, ДПФ (вычисляемой через БПФ) и комплексной диагональной матрицы [8]. Данный метод хорошо подходит для обработки изображений, где все данные поступают одновременно. В обработке речевых сигналов, где отсчёты следуют один за другим, может быть использована даже прямая реализация преобразования WDFT. Учитывая выражение (5), формула (6) может быть записана в следующем виде:

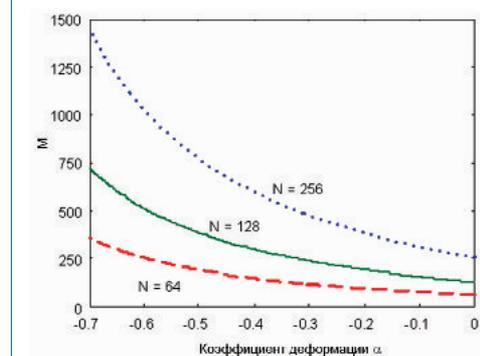


Рис. 8. Соотношение между размером избыточного базиса и коэффициентом деформации α для заданного числа столбцов матрицы $\mathbf{D}_{M \times N}$

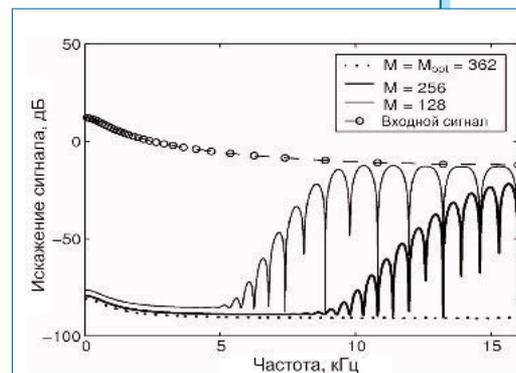
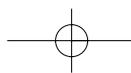
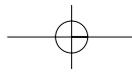


Рис. 9. СПМ входного сигнала (кружочки) и соответствующих спектральных искажений (сплошная линия) для разных форматов WDFT с избыточным базисом ($N=64$)





Петровский А., Борович А., Парфенов М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[N-1] \end{bmatrix} = \sum_{n=0}^{N-1} \begin{bmatrix} A(z_0)^n \\ A(z_1)^n \\ \vdots \\ A(z_{N-1})^n \end{bmatrix} x[n]. \quad (25)$$

Каждый элемент в данной сумме относится только к одному входному отсчёту. Он может быть рассчитан при поступлении входного отсчёта, а результат аккумулируется. Коэффициенты преобразования успешно будут вычислены после N шагов. При этом вычислительная нагрузка, приведённая к входному отсчёту, равна $O(N)$.

Аппроксимация шкалы барков

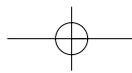
Для достижения высокого качества реконструированных сигналов в перцептуальных системах обработки речи требуется эффективная психоакустическая модель. В известной работе Джонстона [14] психоакустическая модель основана на ДПФ: расчёт ДПФ взвешенного временным окном сегмента сигнала, группировка коэффициентов преобразования в группы, соответствующие критическим частотным полосам, и расчёт энергии в данных частотных полосах. Достижение приемлемого спектрального разрешения в критических частотных полосах, расположенных в низкочастотной части частотного диапазона, требует использования ДПФ с достаточно длинным временным окном. Поэтому концептуальная простота и эффективность нивелируются недостаточным временным разрешением, неприемлемым для анализа такого более тонкого феномена как маскирование назад («pre-masking») [1].

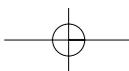
Первый шаг при использовании WDFT в психоакустической модели — проектирование соответствующего фазового преобразования. Частотные коэффициенты z -преобразования должны быть распределены равномерно на перцептуальной шкале. В работе [15] было показано, что фазовое звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу барков, при этом значение коэффициента фазового фильтра для заданной частоты дискретизации определяется по следующему выражению:

$$a_{Bark} = 0.1957 - 1.048 \cdot \left[\frac{2}{\pi} \arctan \left(0.07212 \frac{f_s}{1000} \right) \right]^{\frac{1}{2}}. \quad (26)$$

Для частоты дискретизации $f_s=16$ кГц коэффициент $a_{Bark} = -0.57827$.

Так как ширина критических частотных полос строго изменяется с их местоположением на частотной шкале, различное количество коэффициентов преобразования ассоциируется с конкретной критической частотной полосой. В части А таблицы количество коэффициентов в группах варьируется от 3 до 38 для ДПФ, в то время как в части Б для WDFT той же размерности не отдаётся предпочтение ни одной из полос, т.е. все коэффициенты преобразования распределены практически равномерно.



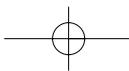


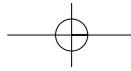
Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

Сравнение распределения коэффициентов ДПФ и WDFT в критических частотных полосах

Крит. полоса	Часть А (размер ДПФ = 512, $f_s = 16$ кГц)			Часть Б (размер WDFT = 512, $f_s = 16$ кГц)		
	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц
1	1–3	3	31–94	1–12	12	8–100
2	4–6	3	125–188	13–24	12	109–202
3	7–9	3	219–281	25–36	12	210–305
4	10–13	4	313–406	37–48	12	314–412
5	14–16	3	438–500	49–60	12	421–523
6	17–20	4	531–625	61–73	13	533–650
7	21–24	4	656–750	74–85	12	660–776
8	25–29	5	781–906	86–97	12	787–912
9	30–34	5	938–1063	98–110	13	923–1073
10	35–40	6	1094–1250	111–123	13	1086–1254
11	41–46	6	1281–1438	124–135	12	1269–1443
12	47–54	8	1469–1688	136–148	13	1460–1680
13	55–62	8	1719–1938	149–161	13	1700–1961
14	63–73	11	1969–2281	162–174	13	1985–2302
15	74–86	13	2313–2688	175–186	12	2331–2690
16	87–102	16	2719–3188	187–198	12	2726–3174
17	103–122	20	3219–3813	199–210	12	3220–3792
18	123–145	23	3844–4531	211–221	11	3851–4513
19	146–173	28	4563–5406	222–231	10	4588–5328
20	174–205	32	5438–6406	232–242	11	5419–6412
21	206–243	38	6438–7594	243–252	10	6520–7533
22	244–256	13	7625–8000	253–256	4	7650–8000

WDFT в его оригинальной форме не сохраняет энергию сигнала в соответствующих частотах единичной окружности до и после фазового звена. Вследствие этого каждый коэффициент WDFT должен масштабироваться в соответствии с коэффициентом $\sqrt{1-a^2}/(1-az)$ [16], чтобы можно было корректно вычислить уровни энергии в критических частотных полосах.





Петровский А., Борович А., Парфенов М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

На рисунке 10 приведён пример обработки широкополосного сигнала, который состоит из голосового сообщения и музыкального фрагмента (верхний рисунок). Распределение мощности в шкале барков для данного сигнала, вычисленное на основе спектрограммы WDFТ ($a_{\text{Bark}} = -0.57827$) (нижний рисунок), показывает, что оценка энергии в низкочастотном и высокочастотном диапазонах может быть получена приблизительно одинаковой, в отличие от измерения по спектрограмме, полученной на основе ДПФ (средний рисунок).

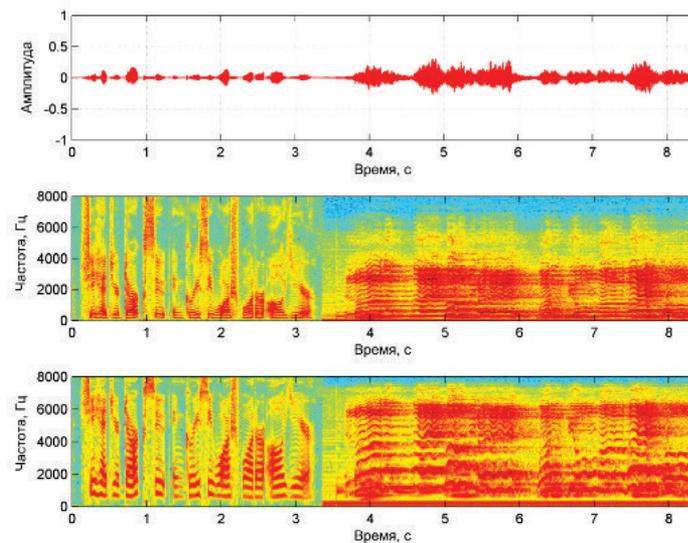


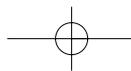
Рис. 10. Широкополосный сигнал во временной области (сверху) и его спектрограммы, полученные при помощи ДПФ (в центре) и WDFТ (снизу)

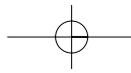
Выбор WDFТ малого формата с успехом может заменить ДПФ с большой длиной выборки благодаря тому, что коэффициенты WDFТ-преобразования равномерно распределены в критических частотных полосах, поэтому в психоакустической модели на базе WDFТ могут быть уравновешены частотное и временное разрешения.

Данная работа частично финансировалась Белостокским политехническим институтом, грант W/WI/8/08.

Литература

1. E. Zwicker, H. Fastl. Psychoacoustics: facts and models. Berlin: Springer-Verlag, 1990.
2. Ковалгин Ю.А., Вологдин Э.И. Цифровое кодирование звуковых сигналов. СПб.: КОРОНА-принт, 2004.
3. Speech Coding and Synthesis; edited by W.B. Kleijn, K.K. Paliwal. Elsevier, Netherlands, 1998.
4. A Noise Reduction Preprocessor for Mobile Voice Communication / R. Martin [et al.] // EURASIP Journal on Applied Signal Processing. August, 2004. P. 1046–1058.
5. A.V. Oppenheim, D.H. Johnson, K. Steiglitz. Computation of spectra with unequal resolution using the FFT // Proceedings IEEE. February 1971. Vol. 59. Pp. 299–301.





Петровский А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением

6. S. Bagchi, S.K. Mitra. The Nonuniform Discrete Fourier Transform and Its Applications in Filter Design // IEEE Trans. Circuits Systems II. 1996. Vol. 43, no. 6. Pp. 422–433.
7. Cho, Ik. Nam, S.K. Mitra. Warped discrete cosine transform and its application in image compression // IEEE Trans. Circ. Systems Video Technol. 2000. Vol.10, no.8. Pp. 1364–1373.
8. A. Makur, S.K. Mitra. A Warped Discrete-Fourier Transform: Theory and Applications // IEEE Trans. Circuits Systems I. 2001. Vol. 48, no. 9. P. 1086–1093.
9. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. М.: Мир, 1980.
10. Hansen, P.C., The truncated SVD as a method for regularization / P.C. Hansen // BIT. 1987. Vol. 27. Pp. 534–553.
11. A. Borowicz, M. Parfieniuk, A.A.Petrovsky. An application of the warped discrete Fourier transform in the perceptual speech enhancement. Speech Communication, vol. 48, ELSEVIER. 2006. Pp. 1024–1036.
12. R. Penrose. A generalized inverse for matrices // Proc. Camb. Philos. Soc. 1955. no.51. Pp. 406–413.
13. C.R. Rao, S.K. Mitra. Generalized inverse of matrices and its applications. New York.: Wiley, 1971.
14. J.D. Johnston. Transform coding of audio signals using perceptual noise criteria // IEEE Transactions on Selected Areas Communication. February, 1988. Vol. 6. Pp. 314–323.
15. J.O. Smith, J.S. Abel. Bark and ERB Bilinear Transforms // IEEE Transactions on Speech and Audio Processing. June 1999. Vol. 7. Pp. 697–708.
16. M. Parfieniuk, A. Petrovsky. Warped DFT as the basis for psychoacoustical model // The proc. of the IEEE International conference on Acoustic, Speech, Signal processing, ICASSP-2004, Montreal, 2004. Pp. IV–185–188.

Петровский Александр Александрович —

доктор технических наук, профессор. Работает в Учреждении образования «Белорусский государственный университет информатики и радиоэлектроники», кафедра «Электронные вычислительные средства». Закончил Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники» по специальности «Электронные вычислительные машины». Главные научные интересы лежат в области цифровой обработки сигналов речи и звука для целей компрессии, распознавания, редактирования шума, а также в области проектирования проблемно-ориентированных средств вычислительной техники реального времени для систем мультимедиа. Член НТО РЭС им. А.С.Попова, IEEE, EURASIP, AES.

Борович Адам —

доктор, ассистент. Работает в Белостокском политехническом институте, г. Белосток, Польша. Основное образование: Белостокский политехнический институт, факультет информатики. Специальность — информатик. Область интересов: цифровая обработка речевых сигналов для целей редактирования шума, проектирование систем мультимедиа.

Парфенюк Марек —

доктор, доцент. Работает в Белостокском политехническом институте, г. Белосток, Польша. Основное образование: Белостокский политехнический институт, факультет информатики. Специальность — информатик. Область интересов: цифровая обработка сигналов (многоскоростная обработка, анализ/синтез банков фильтров), проектирование систем мультимедиа. Член общества IEEE.

