

# ЗАЛЕЖНІСТЬ ПОКАЗНИКІВ ЯКОСТІ СИСТЕМ ОСЛАБЛЕННЯ РЕВЕРБАЦІЙНОЇ ЗАВАДИ ВІД СТУПЕНЯ СПОТВОРЕННЯ МОВЛЕННЕВОГО СИГНАЛУ

**А. Продеус**, доктор технічних наук, професор,  
**О. Ладощко**, аспірант, кафедра акустики та акустoeлектроніки,  
Національний технічний університет України «КПІ», м. Київ

**Зависимость показателей качества систем подавления реверберационной помехи от степени искажения речевого сигнала**

А. Продеус, доктор технических наук, профессор,  
О. Ладощко, аспирант, кафедра акустики и акустoeлектроники,  
Национальный технический университет Украины «КПИ», г. Киев

**Dependence of Dereverberation Systems Quality Indicators on Speech Signal Distortion Degree**

A. Prodeus, Doctor of Technical Sciences, Professor,  
O. Ladoshko, Postgraduate Student, Acoustics and Acoustoelectronics Department,  
National Technical University of Ukraine «KPI», Kiev

*Із застосуванням експериментальних методів виконано порівняльний аналіз наскрізних та проміжних показників якості систем ослаблення ревербераційної завади, що їх використовують як препроцесори в системах автоматичного розпізнавання мовлення.*



А. Продеус



О. Ладощко

## ВСТУП

Із поширенням мобільних засобів зв'язку, таких як стільникові телефони та комп'ютерні планшети, увагу розробників та споживачів було повернуто до систем автоматичного розпізнавання мовлення (САРМ) як важливого елементу систем комунікації та цифрового оброблення мовленнєвої інформації. Проте широкому застосуванню САРМ

суттєво заважає недостатньо висока якість їх роботи, спричинена, зокрема, їх сприйнятливостю до дії досить широкого кола завад, таких як шум навколишнього середовища, реверберація, спотворення сигналу під час його кодування та під час проходження через низку перетворювачів (мікрофони, підсилювачі тощо). Тому протягом багатьох років дослідники приділяють значну увагу питанням підвищення

несприйнятливості (робастності) САРМ до дії завад різної природи [1, 2]. Зрозуміло, що ці питання тісно пов'язані із завданнями оцінення якості роботи та сертифікації як окремих САРМ, так і САРМ у складі ліній зв'язку.

В останні роки спостерігається хвиля інтересу дослідників до розв'язання завдання ослаблення дії ревербераційної завади, що є спотворенням акустичного сигналу, обумовленого багатократним його відбиттям від стін, підлоги, стелі та навколишніх предметів (рис. 1). Дія ревербераційної завади є малопомітною, якщо відстань між джерелом звуку та мікрофоном становить кілька сантиметрів, як це буває за звичайної телефонної розмови. Проте зі збільшенням відстані до десятків та сотень сантиметрів, що є звичним, наприклад, для телеконференцій, дією ревербераційної завади на мовленнєвий сигнал вже не можна знехтувати, оскільки якість мовленнєвого сигналу помітно погіршується. Відповідно, суттєво погіршується й якість розв'язання, за допомогою САРМ таких важливих завдань, як автоматизація стенографування, виявлення ключових слів, ідентифікація диктора тощо.



Рис. 1. Розповсюдження сигналу у замкненому приміщенні

На жаль, таке спотворення завжди буває під час застосування систем коригування мовленнєвих сигналів. Сам факт спотворення дуже легко пояснити, враховуючи, що системи коригування мовленнєвих сигналів фактично є фільтрами, сконструйованими для ослаблення того чи іншого виду завад. Проте, ослаблюючи заваду, такі фільтри неминуче спотворюють й мовленнєвий сигнал. І хоча фільтри розраховують з огляду на мінімізацію цих спотворень, повністю позбутися їх не вдається. Наслідком цих залишкових спотворень є погіршення якості роботи САРМ.

Метою статті є порівняння показників якості систем ослаблення ревербераційної завади, залежних від ступеня спотворення мовленнєвого сигналу.

**1. Показники якості**

Характеризуючи якість роботи САРМ, зазвичай використовують такий показник, як «точність правильного розпізнавання слів» [3]:

$$Acc\% = \frac{N - D - S - I}{N} \cdot 100\%$$

де  $N$  — загальна кількість слів, що їх розпізнають;  $D$  — кількість помилкових видалень слів;  $S$  — кількість заміненних слів;  $I$  — кількість помилково вставлених слів.

Якщо системі автоматичного розпізнавання мовлення передує препроцесор у вигляді системи ослаблення ревербераційної завади (деревербератор), або якщо САРМ є складовою частиною цифрової лінії зв'язку, що містить деревербератор, тоді показник  $Acc\%$  доцільно використовувати як наскрізний (інтегральний) показник якості сукупності систем оброблення мовленнєвої інформації (рис. 2).

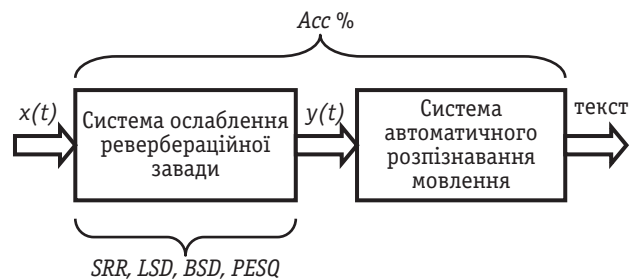


Рис. 2. Інтегральний показник якості сукупності систем оброблення мовленнєвої інформації

Між тим, зрозумілим є бажання використати простіші, з точки зору моделювання ситуації, «проміжні» показники якості. Наприклад, для оцінювання якості деревербераторів часто використовують такі показники як відношення сигнал-ревербація (Signal-to-Reverberation Ratio — SRR), логарифмічно-спектральне спотворення (Logarithmic Spectral Distortion — LSD), барк-спектральне спотворення (Bark Spectral Distortion — BSD), перцептуальна оцінка якості мовлення (Perceptual Evaluation of Speech Quality — PESQ) тощо [2].

Дійсно, якщо для оцінювання інтегрального показника  $Acc\%$  потрібно моделювати не лише деревербератор, але й САРМ, то для оцінювання показників SRR, LSD, BSD та PESQ достатньо було б моделювати лише деревербератор. При цьому співвідношення для обчислення показників SRR, LSD, та BSD мають такий вигляд:

$$SRR = \frac{1}{L} \sum_{l=1}^L 10 \lg \left[ \frac{\sum_{n=Rl}^{Rl+N-1} x^2(l, n)}{\sum_{n=Rl}^{Rl+N-1} [x(l, n) - y(l, n)]^2} \right],$$

$$LSD = \frac{2}{KL} \sum_l \sum_{k=0}^{\frac{K}{2}-1} |G\{X(l, k)\} - G\{Y(l, k)\}|,$$

$$G\{X(l, k)\} = \max\{20 \lg(|X(l, k)|), \delta\},$$

$$\delta = \max_{l,k} \{20 \lg(|X(l,k)|)\} - 50,$$

$$BSD = \frac{\sum_{l=1}^L \sum_{k=0}^{\frac{K}{2}-1} [B\{X(l,k)\} - B\{Y(l,k)\}]^2}{\sum_{l=1}^L \sum_{k=0}^{\frac{K}{2}-1} [B\{X(l,k)\}]^2},$$

де  $x(l, n)$  та  $y(l, n)$  —  $n$ -на вибірка  $l$ -го фрейму вхідного та вихідного сигналів  $x(n)$  та  $y(n)$ , відповідно, системи ослаблення реверберації;  $X(l, k)$  та  $Y(l, k)$  — амплітудні спектри  $l$ -го фрейму сигналів  $x(n)$  та  $y(n)$ , відповідно;  $B\{X(l, k)\}$  та  $B\{Y(l, k)\}$  — барк-спектри  $l$ -го фрейму сигналів  $x(n)$  та  $y(n)$ , відповідно.

Алгоритм обчислення показника PESQ є громіздкішим, тому за бажанням його практичного використання доречно звернутися до праць [4, 5]. Проте, як у випадку показників SRR, LSD, та BSD, вихідними даними для алгоритму обчислення PESQ є вибірки сигналів  $x(n)$  та  $y(n)$ .

## 2. Порівняння показників якості

Порівняємо зазначені вище показники якості деревербераторів, залежні від ступеня спотворення мовленнєвого сигналу.

Зазначимо, що в разі створення деревербераторів, перш за все, йдеться про ослаблення дії так званої пізньої реверберації. На рис. 3 показано важливу акустичну характеристику кожного приміщення — імпульсну характеристику  $h(t)$  приміщення, що відображає форму сигналу на виході мікрофона як відгук на випромінювання джерелом звуку короткого імпульсу (звук плеску в долоні, проколу гумової кульки, пострілу із стартового пістолета тощо). Початкова частина імпульсної характеристики описує так звані ранні відбиття та має вигляд дискретної послідовності коротких імпульсів. Кінцева частина імпульсної характеристики називається пізньою реверберацією та має неперервний стохастичний характер.

Оскільки акустичний сигнал  $x(t)$  довільної форми (зокрема мовленнєвий сигнал) можна зобразити суперпозицією щільно притиснених один до одного коротких імпульсів із різними ваговими коефіцієнтами, сигнал на виході мікрофона  $y(t)$  також можна представити суперпозицією щільно притиснутих та зважених такими самими ваговими коефіцієнтами імпульсних характеристик:

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau.$$

Якби функція  $h(t)$  була дуже короткою (кілька десятків мілісекунд) та складалася лише із ранніх відбиттів, тоді форма сигналу  $y(t)$  мало б відрізнялася від форми сигналу  $x(t)$ . Але оскільки імпульсні характе-

ристики приміщень зазвичай довші за декілька сотень мілісекунд внаслідок існування пізньої реверберації, сигнал  $y(t)$  за формою буде суттєво відрізнятися від сигналу  $x(t)$ .

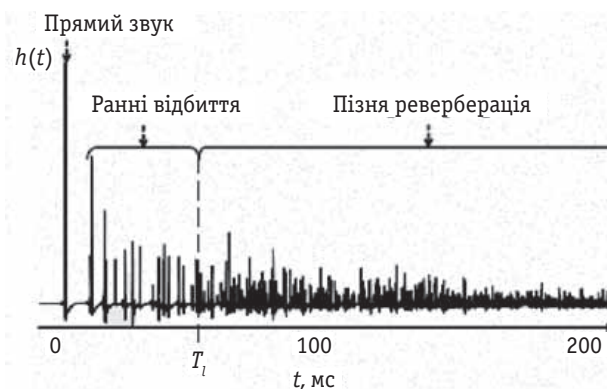


Рис. 3. Імпульсна характеристика приміщення

Таким чином, межа  $T_l$  між вказаними двома частинами імпульсної характеристики є важливим параметром, що разом із параметром  $T_{60}$  (часом реверберації — часом згасання імпульсної характеристики до рівня мінус 60 дБ відносно свого максимуму) характеризує завадову частину реверберованого сигналу  $y(t)$  на виході мікрофона.

Довільна система коригування спотвореного сигналу, що перетворює сигнал  $y(t)$  в скоригований сигнал  $\tilde{x}(t)$ , потребує апріорних даних про параметри завади. Зокрема, під час побудови деревербераторів потрібно знати параметри  $T_{60}$  та  $T_l$ , щоб оцінити спектр потужності шумової завади, спричиненої наявністю пізньої реверберації [2]. Параметр  $T_{60}$  оцінюють, заздалегідь проаналізувавши імпульсну характеристику приміщення, або «наосліп», спеціальним обробленням реверберованого сигналу  $y(t)$ . Що стосується параметру  $T_l$ , навіть з рис. 3 зрозуміло, що зазначити його точне значення практично неможливо. В науковій літературі наводять можливі значення  $T_l \approx 30 - 100$  мс [1, 2]. Проте слід враховувати, що ці значення було свого часу одержано для визначення здатності приміщень якісно передавати на відстань мовленнєві або музичні сигнали. Що стосується завдання ослаблення впливу пізньої реверберації, оптимальні значення  $T_l$  до останнього часу не були одержані. Більш того, до останнього часу навіть не було відомо, чи такі значення взагалі можна назвати.

У роботі [6] вперше експериментально досліджено поведінку функціональних залежностей  $Acc \% (T_l)$  та  $PESQ (T_l)$  й показано, що залежність  $Acc \% (T_l)$  має виражений екстремум, тоді як із залежністю  $PESQ (T_l)$  питання є проблематичним: в одній із трьох розглянутих ситуацій спостерігався слабо виражений екстремум, а в інших двох ситуаціях екстремум не спостерігався зовсім.

У рамках цієї роботи було проведено подальші детальніші дослідження залежностей  $Acc \% (T_l)$  та  $PESQ (T_l)$ . Так, зокрема, було зменшено крок змінної  $T_l$  з 48 до 16 мс. Крім того, розглянуто залежність функцій  $Acc \% (T_l)$  та  $PESQ (T_l)$  від параметру  $\eta_z^d$  (на графіках позначено як  $etzd$ ), що його застосовують під час оцінювання спектру ревербераційного сигналу [2, 6]. Усереднені (за параметром  $T_{60}$ , що набув значень 0,74 с, 0,89 с та 1,1 с) результати цих досліджень, наведені на рис. 4 та 5, дають змогу одержати рекомендації з оптимізації деревербератора не лише за параметром  $T_l$ , але й за параметром  $\eta_z^d$ . Так, аналіз

рис. 4 дає змогу зробити висновок, що за умови  $\eta_z^d \approx 0,66 - 0,75$  максимум  $Acc \%$  досягається для  $T_l \approx 100$  мс.

Складніше, як видно із рис. 5, є поведінка показника  $PESQ (T_l)$ . Максимальну якість мовлення  $PESQ$  одержуємо за умови  $\eta_z^d \approx 0,5 - 0,66$ , проте залежність  $PESQ (T_l)$  при цьому має монотонно спадний характер, тобто не містить екстремуму. Екстремуми залежності  $PESQ (T_l)$  є лише за  $\eta_z^d \approx 0 - 0,2$ , та відповідають значенням  $T_l \approx 50 - 80$  мс.

На нашу думку, наявність та виразність екстремумів показників  $Acc \% (T_l)$  та  $PESQ (T_l)$  доцільно тракту-

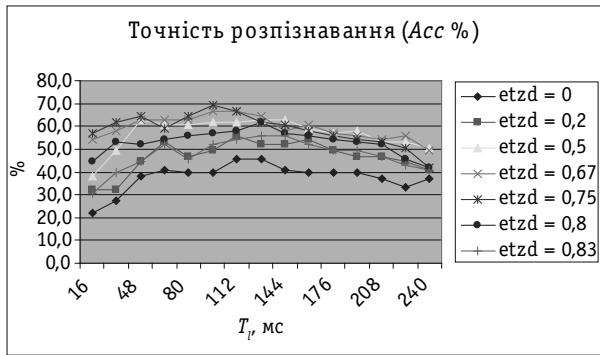


Рис. 4. Графік залежності точності розпізнавання  $Acc \%$  від параметру  $T_l$

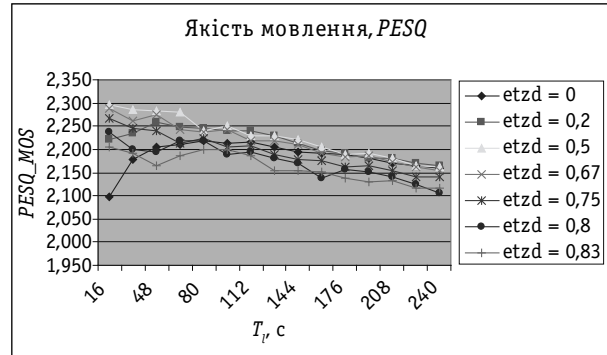


Рис. 5. Графік залежності показника  $PESQ$  від параметру  $T_l$

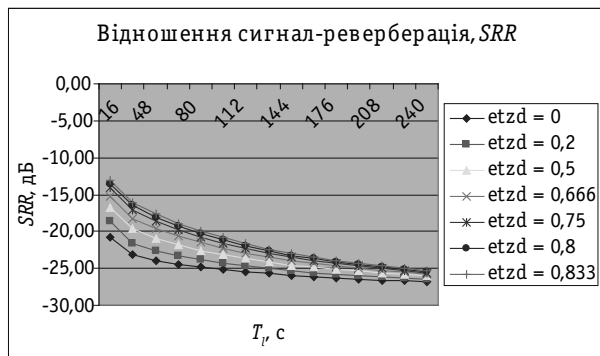


Рис. 6. Графік залежності показника  $SRR$  від параметру  $T_l$

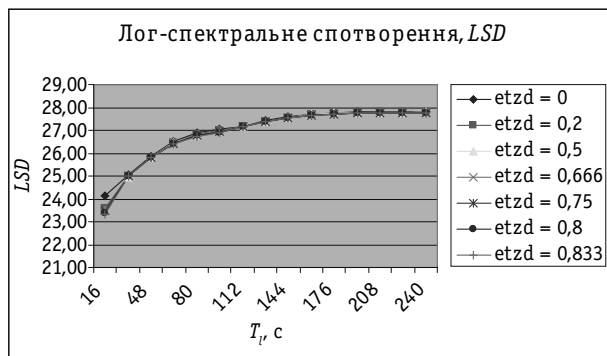


Рис. 7. Графік залежності показника  $LSD$  від параметру  $T_l$

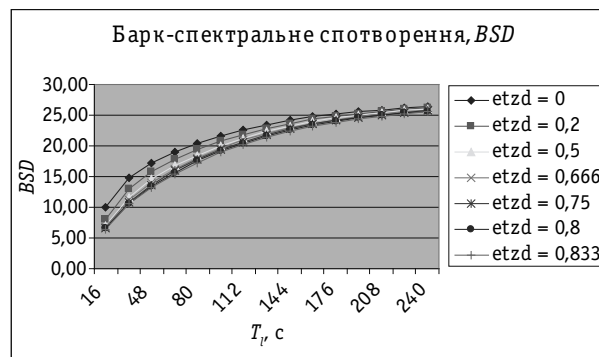


Рис. 8. Графік залежності показника  $BSD$  від параметру  $T_l$

вати як відображення ними ступеня спотворення мовленнєвого сигналу, спричиненого дією системи ослаблення пізньої реверберації. Цей висновок базується на тому факті, що використання малих значень  $T_i$  в алгоритмі ослаблення пізньої реверберації призводить до посиленого ослаблення ревербераційної завади, що є бажаним. Однак при цьому й сильніше спотворюється мовленнєвий сигнал — і це є небажаним. Із збільшенням  $T_i$  бажане ослаблення ревербераційної завади слабшає, проте й зменшується небажане спотворення мовленнєвого сигналу. Таким чином, максимум функції  $Acc\ % (T_i)$  не лише сам по собі вказує на оптимальне значення параметру  $T_i$ , але й є ознакою того, що досягнуто рівновагу між позитивним та негативним наслідками процедури дереверберації.

Особливий характер поведінки залежності  $PESQ (T_i)$  можна пояснити тим, що хоча алгоритм оцінювання параметру  $PESQ$  сконструйовано так, щоб він відображав діяльність слухової системи людей-експертів, проте особливості сприйняття слуховою системою реверберованого сигналу було враховано не повною мірою [4, 5].

Таким чином, порівнюючи залежності  $Acc\ % (T_i)$  та  $PESQ (T_i)$ , доходимо висновку, що інтегральний показник  $Acc\ %$  значно краще відображає спотворення сигналу системою ослаблення пізньої реверберації.

Результати експериментального оцінювання усереднених за параметром  $T_{60}$  залежностей  $SRR(T_i)$ ,  $LSD(T_i)$  та  $BSD(T_i)$  наведено на рис. 6—8, відповідно.

Відсутність екстремумів в цих залежностях дає змогу вважати їх несприйнятливими до спотворення сигналу, зумовленого дією систем ослаблення ревербераційної завади. Це дає змогу дійти висновку, що

використання показників  $SRR$ ,  $LSD$  та  $BSD$  для оцінювання якості деревербератора як препроцесора системи автоматичного розпізнавання мовлення є недоцільним.

## ВИСНОВКИ

За допомогою комп'ютерного моделювання експериментально одержано залежності показників якості  $Acc\ %$ ,  $PESQ$ ,  $SRR$ ,  $LSD$  та  $BSD$  від параметрів системи ослаблення пізньої реверберації, що використовується як препроцесор системи автоматичного розпізнавання мовлення. Одержані результати є корисними в науковому сенсі, оскільки демонструють обмеженість можливостей багатьох досить популярних показників якості. Зокрема, показано, що показник  $PESQ$  лише обмежено дає змогу контролювати ступінь спотворення сигналу, зумовленого дією деревербератора. А показники  $SRR$ ,  $LSD$  та  $BSD$  взагалі виявилися неспроможними до відображення спотворень сигналу, що робить недоцільним їх використання в разі сертифікації ліній зв'язку, що містять деревербератори як препроцесори систем автоматичного розпізнавання мовлення.

Практична користь одержаних результатів полягає в знаходженні оптимальних, в сенсі максимуму точності автоматичного розпізнавання мовлення, значень параметрів системи ослаблення пізньої реверберації.

Одним із можливих перспективних завдань є дослідження ширшого набору імпульсних характеристик приміщень із різними значеннями часу реверберації, з метою виявлення особливостей, притаманних приміщенням з малим та великим часом реверберації.

## ЛІТЕРАТУРА

1. Speech Processing in Modern Communication / Edited by Israel Cohen, Jacob Benesty and Sharon Gannot // Springer-Verlag Berlin Heidelberg 2010, 342 p. (Обробка мовлення в сучасних системах зв'язку / Під редакцією Ізраеля Коена, Джекоба Бенесті та Шарона Геннота // Шпрінгер-Верлаг Берлін Гейдельберг 2010б. — 342 с.)
2. Naylor P., Gaubitch N. Speech Dereverberation. — Springer, 2010. — 399 p. (Нейлор П., Гаубіч Н. Деревевербація мовлення. — Шпрінгер, 2010. — 399 с.)
3. Young S. et al, The HTK Book // Cambridge University Engineering Department, 2005, 354 p. — Online: <http://www.aprodeus.narod.ru/Information/Rech/htkbook.zip> (Янг С. та ін. Посібник з НТК // Інженерний факультет Кембріджського університету, 2005. — 354 с. — Режим доступу: <http://www.aprodeus.narod.ru/Information/Rech/htkbook.zip>)
4. Loizou P. Speech enhancement: Theory and Practice. — Boca Raton: CRC Press, 2007. — 632 p. (Лоїзоу П. Ко-
- рекція мовлення: теорія та практика. — Бока Рейтон: CRC Press, 2007. — 632 с.)
5. Measurement of speech intelligibility based on the PESQ approach // Beerends J., Larsen E., Lyer N., and J. van Vugt / In Proc. Int. Conf. Meas. Speech Audio Quality Netw., 2004, 4 p. (Вимірювання розбірливості мовлення, що базується на використанні PESQ // Бірендс Дж., Ларсен Е., Ліер Н., Дж. ван Вугт / Праці конференції з якості мовлення в мережах, 2004. — 4 с.)
6. Prodeus A., Ladoshko O. On existance of optimal boundary value between early reflections and late reverberation // Proc. of IEEE 34th International Scientific Conference «Electronics and Nanotechnology», Kyiv, Ukraine, April 15—18, 2014. — P. 442—446. (Продеус А., Ладощко О. Про існування оптимальної межі між ранніми відбиттями та пізньою реверберацією // Праці IEEE 34-ї міжнародної конференції «Електроніка та нанотехнології». — Київ, Україна. — Квітень 15—18, 2014. — С. 442—446). ■