УДК 004.934 DOI: 10.20535/2523-4455.2017.22.6.101929

Суб'єктивне оцінювання якості та розбірливості мовних сигналів, спотворених синтезованими шумами

Продеус А. М., д.т.н., проф. ORCID 0000-0001-7640-0850
e-mail am.prodeus@aae.kpi.ua
Вітик А. В., ORCID 0000-0002-6652-4152
e-mail andrvit8@gmail.com
Діденко Д. Ю., ORCID 0000-0001-7992-3235
e-mail dyu.didenko@aae.kpi.ua
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського" kpi.ua
Київ, Україна

***Анотація*—В даній роботі наведено результати оцінювання впливу стаціонарних та нестаціонарних синтезованих шумів на якість та розбірливість мовних сигналів. Для випадку стаціонарних шумів показано, що при малих відношеннях сигнал-шум білий шум поступається за маскувальною здатністю рожевому й коричневому шумам. Досліджено два простих, з обчислювальної точки зору, алгоритми формування нестаціонарних шумів, що забезпечують краще, у порівнянні з білим шумом, маскування мовних сигналів, а також менше забруднюють навколишнє середовище під час мовних пауз.**

Бібл. 12, рис. 8.

Ключові слова — розбірливість мови; якість мови; окрашений шум; мовоподібний шум; відношення сигнал-шум.

I. Вступ

Як відомо, спотворення мовних сигналів шумовими завадами негативно відображається на сприйнятті мовної інформації слухачами. Наприклад, в роботах [1], [2], [3] та [4], де наведено результати оцінювання ступеню впливу шуму на розбірливість мови в аудиторіях, показано, що шумова завада у вигляді розмов учнів, що сидять поруч, є набагато небезпечнішою за реверберацію. Причиною є висока інтенсивність завади, зумовлена близькістю її джерела. Ситуація посилюється тим, що шум розмови взагалі має високі маскувальні властивості внаслідок подібності спектрально-часових властивостей завади й сигналу.

Разом із тим, явище маскування мовних сигналів шумом може бути корисним. Наприклад, для нормальної роботи в офісних та бібліотечних приміщеннях вважається доцільним підтримування дещо підвищеного рівня шуму, на тлі якого розмови відвідувачів та працівників виявляються менш розбірливими для випадкових слухачів [5]. Більш того, в багатьох випадках розбірливість мови взагалі повинна бути мінімальною. Так, при конфіденційних переговорах потрібен ефективний, пасивний або активний, захист приміщень від витоку мовної інформації. Тому на сьогодні існує велика кількість різних систем активного віброакустичного маскування мовної інформації, які генерують акустичні завади у вигляді стаціонарних та нестаціонарних шумів [6].

Оцінку якості систем акустичного маскування, що генерують стаціонарні шуми, можна здійснити формантним методом [6] та [7]. При цьому розбірливість мовних сигналів доречно використовувати як міру якості маскування.

Згідно із формантним методом, формантну розбірливість A обчислюють за формулою:

$$A = \sum_{k=1}^{K} p_k \cdot P(E'_k),$$

де p_k - імовірність перебування формант в k-тій смузі частот ($k = 1,...,K$):

$$p_k = F_1(f_{вк}) - F_1(f_{нк}),$$

$F_1(f)$ – функція розподілу ймовірностей формант за частотою; $P(E'_k)$ – коефіцієнт сприйняття мовлення; $E'_k = B_{pk} - \Delta B_k - B_{шk}$ – ефективний рівень відчуття формант в k-тій смузі частот; B_{pk} і $B_{шk}$ – рівні спектрів потужності мовленнєвого сигналу й шуму, відповідно, в k-тій смузі частот; $\Delta B_k = B_{pk} - B'_{pk}$, де B'_{pk} – спектр формант. Словесну розбірливість W оцінюють, використовуючи відому функціональну залежність $W(A)$ [6] та [7].

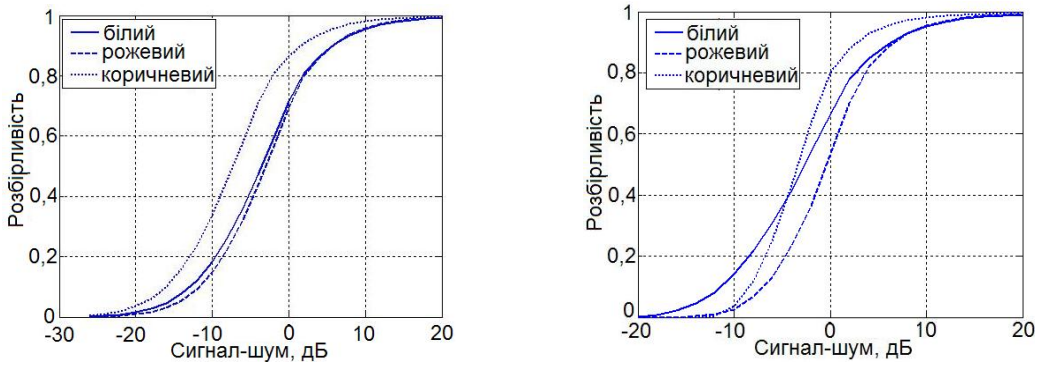


Рис. 1 Розрахункові залежності $W(SNR)$ для коефіцієнтів сприйняття Н. Б. Покровського (а) та М. А. Сапожкова (б) для смуги частот 180…5600 Гц [6] та [8]

Проте, як вказано в [8], факт існування кількох версій формантного методу (версії Н. Б. Покровського [7], М. А. Сапожкова та Ю. С. Бикова) призводить до можливої неоднозначності результатів оцінювання розбірливості мови. Розбіжність версій полягає, по-перше, в різній трактовці поняття «формантний спектр». По-друге, відрізняється форма коефіцієнтів сприйняття $P(E_k')$. Нарешті, не співпадають думки авторів версій й стосовно форми функції розподілу ймовірностей формант за частотою $F_1(f)$. Співставлення зазначених версій формантного методу дозволило дійти висновку, що найбільш принциповим є коректне визначення коефіцієнтів сприйняття [8].

Дійсно, аналізуючи наведену в [6] залежність W від інтегрального відношення сигнал-шум SNR для коефіцієнтів $P(E_k')$, визначених за Н. Б. Покровським [7], можна було б дійти висновку, що білий шум за маскувальною здатністю майже не поступається рожевому шуму (рис. 1а).

Проте в [8] показано, що цей висновок є некоректним, а графіки рис. 1а є невірними внаслідок невірного визначення Н. Б. Покровським коефіцієнтів сприйняття $P(E_k')$.

Якщо обчислення проводити із використанням більш коректно обчислених коефіцієнтів $P(E_k')$, визначених М. А. Сапожковим (рис. 1б), доходимо висновку, що маскувальні властивості білого шуму є найгіршими при $SNR < -5$ дБ.

Разом із тим, графіки рис. 1б також не слід вважати остаточними, оскільки для спрощення обчислень М. А. Сапожков припустив незалежність $P(E_k')$ від смуги частот. Цей недолік усунено в роботах [8] та [9], де враховано залежність $P(E_k')$ від смуги частот (рис. 2).

Зокрема, в [8] наведено графіки залежностей $W(SNR)$ для смуги частот 180…5600 Гц (рис. 2а), а в [9] одержано залежності $W(SNR)$ для смуги частот 90…10500 Гц (рис. 2б).

Зазначені уточнення не змінюють загального висновку щодо поганої маскувальної здатності білого шуму при малих ($SNR < -5$ дБ) відношеннях сигнал-шум.

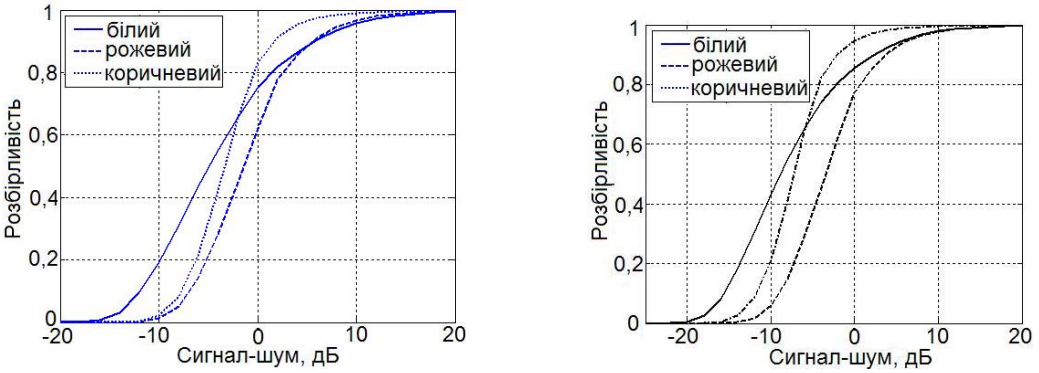


Рис. 2 Відкориговані розрахункові залежності $W(SNR)$ для смуг частот 180…5600 Гц (а) та 90...10500 Гц (б) [8] та [9]

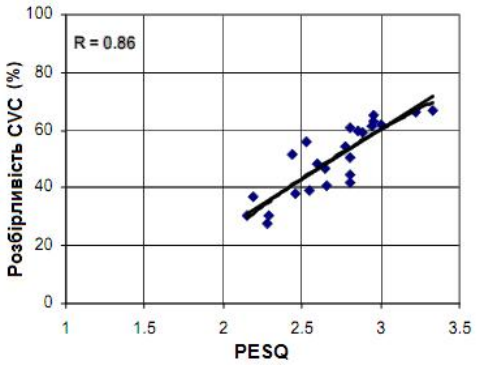


Рис. 3 Зв'язок між значеннями оцінок розбірливості CVC та PESQ [10]

Практична цінність такого висновку навряд чи викликає сумнів. На жаль, його справедливість й досі не було перевірено суб'єктивним методом. Тому одним із завдань даної роботи було усунення зазначеного недоліку.

Що стосується «мовоподібних» завад, які відносяться до класу нестаціонарних випадкових процесів, в [6] відзначено, що їх маскувальна здатність є близькою до такої для рожевого шуму. Проте, на відміну від випадку стаціонарного шуму, оцінку ефективності "мовоподібних" завад доводиться здійснювати виключно із застосуванням артикуляційного методу [6], який є трудомістким та дорогим. Тому іншою метою даної роботи був пошук можливостей усунення цього недоліку.

II. Організація оцінювання маскувальних властивостей шумів суб'єктивним методом

Оцінювання розбірливості мови артикуляційним (суб'єктивним) методом є процесом дуже трудомістким. Наприклад, згідно вимогам стандартів ГОСТ 16600-72 та ГОСТ 7153-85 до участі в експериментах треба залучити не менше трьох дикторів та трьох слухачів. При цьому кожний диктор повинен зачитати не менше чотирьох таблиць звукових сполучень, кількість яких сягає 50. Якщо ми хочемо одержати залежності, подібні до представлених на рис. 1 та рис. 2, кількість експериментів різко зростає, оскільки бажано варіювати як значеннями SNR, так і видом шумової завади. Зрозуміло, що виконання такого завдання є вкрай утрудненим.

Тому в даній роботі замість суб'єктивного оцінювання розбірливості мови суб'єктивно оцінюється якість мовних сигналів. Обґрунтуванням такої заміни є близькість таких характеристик як розбірливість мови та якість мовних сигналів, хоча, як відомо, ці характеристики не є тотожними [10], [11] та [12].

Зокрема, в [11] знаходимо: «Розбірливість є лише одним із специфічних атрибутів інтегральної якості мовних сигналів». В той же час, в [10] знаходимо: «… кореляція між оцінками PESQ MOS та CVC є вельми високою, $R = 0,86\ldots$». Тут PESQ є об'єктивною (інструментальною) мірою якості мовних сигналів [10], а CVC визначається як процент слів, в яких слухачами вірно розпізнано голосні та приголосні звуки. Зв'язок між оцінками PESQ та CVC показано на рис. 3, де спостерігаємо монотонну та майже лінійну залежність цих мір для низьких та середніх рівнів розбірливості мови.

Разом із тим, відомим є високий ступінь кореляції між PESQ та результатами суб'єктивного оцінювання якості мовних сигналів ($R \approx 0,92$) [10].

Пов'язуючи між собою ці факти та враховуючи монотонність та практично-лінійну залежність між інтегральними розбірливістю та якістю мовних сигналів, вважаємо достатньо обґрунтованою можливість використання методу суб'єктивного оцінювання якості мовних сигналів для суб'єктивного оцінювання їх розбірливості.

Оскільки мовні сигнали, що піддаються тестуванню, можуть бути спотвореними дуже слабо, було застосовано спеціальний варіант тесту з прослуховування. Згідно із цим тестом, для кожного окремого випробування слухачі мали порівнювати зразковий неспотворений мовний сигнал із спотвореним сигналом. При цьому слухачів просили оцінити за 5-бальною шкалою погіршення якості спотворених сигналів. Якість сигналів оцінювалася за шкалою Degradation Mean Opinion Score (DMOS): 5 (спотворення нечутно), 4 (чутно, але не дратує), 3 (трохи дратує), 2 (дратує) і 1 (дуже дратує) [11].

При тестуванні використано 8 записів мовних сигналів (16 біт, 22050 Гц), серед яких було 4 записи чоловічих та 4 записи жіночих голосів. Тестування відбувалося 10 слухачами віком 21-23 років за допомогою спеціальної комп'ютерної програми, при цьому слухачам у випадковому порядку пред'являлися записи спотворених сигналів. Прослуховування сигналів виконувалося за допомогою навушників, що в певній мірі сприяло підвищенню достовірності оцінювання. Довжина сигналів складала 20-30 с, проте слухачі могли власноруч її регулювати.

III. Результати експериментів

В якості стаціонарних шумів використовувалися дискретний білий, рожевий та коричневий шуми, які формувалися програмно, в середовищі MATLAB®. Забарвлені (рожевий та коричневий) шуми формувалися шляхом пропускання білого шуму через гребінку із 7 смугових октавних фільтрів з центральними частотами 125, 250, …, 8000 кГц. Спектри синтезованих шумів мали ступінчасту форму, потрібний рівень ступенів спектру встановлювався коефіцієнтами підсилення в кожному із частотних каналів. Зауважимо, що саме такий спосіб формування шумів використовувався в [8] та [9], де були одержані графіки рис. 2.

Для одержання бажаного відношення сигнал-шум SNR_0, мовний сигнал $x(t)$ підсумовувався із синтезованим шумом $n(t)$:

$$y(t) = x(t) + k_1 \cdot n(t), \qquad (1)$$

де $k_1 = 10^{0.05(SNR_1 - SNR_0)}$ – коригувальний коефіцієнт, SNR_1 – відношення сигнал-шум для ненормованого шуму.

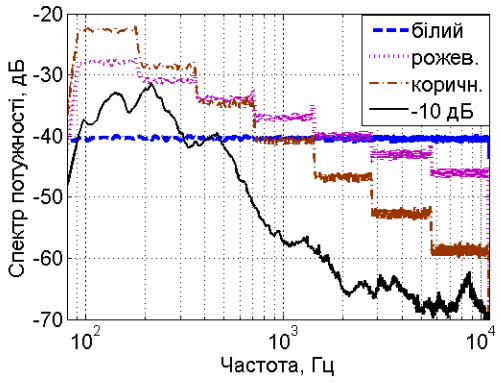


Рис. 4 Спектри потужності сигналу та шумів для ситуації $SNR_0 = -10$ дБ

Форму оцінок білого, рожевого та коричневого шумів однакової потужності наведено на рис. 4, де також чорною неперервною лінією показано усереднену за 8 дикторами оцінку довготривалого спектру мовного сигналу, потужність якого на 10 дБ менша за таку для шумів (оцінки спектрів формувалися за методом Бартлета за однохвилинними реалізаціями сигналів та шумів і довжині сегментів 0,185 с).

Наведені на рис. 4 графіки добре узгоджуються із попереднім висновком про низьку маскувальну здатність білого шуму при низьких ($SNR < -5$ дБ) відношеннях сигнал-шум [8] та [9]. Поведінка графіків рис. 5а дозволяє вважати експериментально підтвердженими ці висновки, а також прогнозні висновки, що маскувальна здатність рожевого шуму має бути найкращою при всіх відношеннях сигнал-шум, а коричневий шум займає проміжне місце між білим та рожевим шумами при $SNR < -5$ дБ, хоча поступається білому шуму при $SNR > 0$ дБ. В середньому стандартне відхилення оцінок DMOS близьке до 0,7 (рис. 5б).

Зазначимо також, що одержані експериментальні результати узгоджуються із висновками праць [10] та [12] щодо принципової можливості використання мір якості мовних сигналів для оцінювання розбірливості мови.

Що стосується дослідження маскувальних властивостей «мовоподібних» шумів, такі нестаціонарні завади можуть бути сформовані як із застосуванням сигналу, що піддається маскуванню, так і з довільних фрагментів мовлення. В даній роботі було розглянуто лише перший підхід, тобто завади формувалися із застосуванням сигналу, що піддавався маскуванню.

В [6] вказано на наступні 3 найпоширеніші способи формування мовоподібних завад:

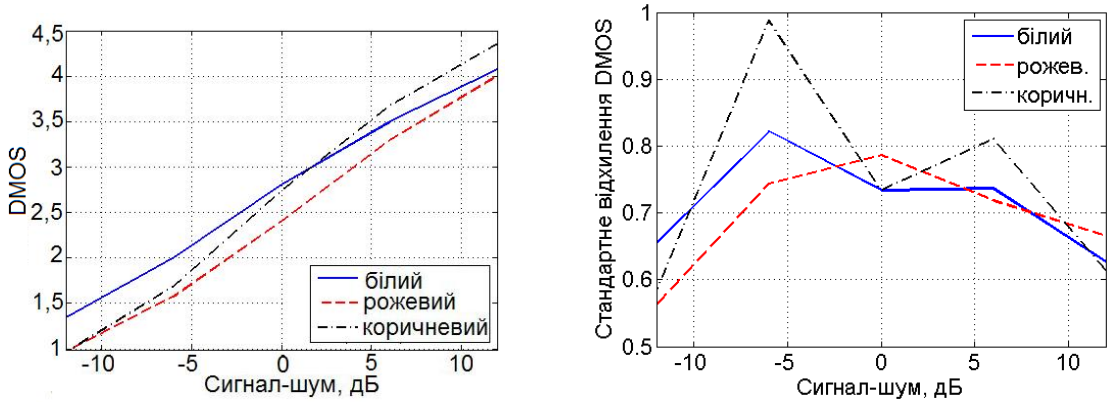
- імітація реверберації шляхом багаторазового накладення мовного сигналу з різними рівнями;
- складна інверсія спектру мовного сигналу;
- комбінований спосіб, за яким спільно використано як інверсію спектру, так і імітацію реверберації.

В представленій статті «мовоподібні» завади формувалися за третім, комбінованим, способом, згідно із співвідношенням:

$$n_1(t) = [x(t) \cdot m(t)] \otimes h(t), \qquad (2)$$

де $x(t)$ – мовний сигнал; $m(t)$ – періодичне коливання заданої форми; $h(t)$ – імпульсна характеристика гребінчастого фільтру; $\otimes$ – символ згортки.

Добуток $x(t) \cdot m(t)$ описує процедуру балансної модуляції, котра забезпечує складну інверсію спектру мовного сигналу. Вибір періоду несучого коливання $m(t)$ дозволяє керувати зсувом по осі частот спектру $X(f)$ сигналу $x(t)$. А вибір форми несучого коливання $m(t)$ дозволяє керувати ваговими коефіцієнтами при багатократному накладенні зсунутих спектрів $X(f)$.



а) б)

Рис. 5 Оцінки $DMOS(SNR)$ для стаціонарних шумів (а) та стандартні відхилення цих оцінок (б)

Важливою особливістю алгоритмів, що описуються співвідношенням (2), є те, що відношення сигнал-шум майже не змінюється в часі. Як наслідок, під час мовних пауз шум практично не чутний, що дозволяє зменшити звукове забруднення навколишнього середовища.

Для експериментальних досліджень використано наступний конкретний алгоритм формування мовоподібної завади:

$$n_1(t) = y(t-t_0) + y(t-t_1),$$
$$y(t) = x(t) \cdot \mathrm{mdr}(2\pi f_0 t). \quad (3)$$

де $\mathrm{mdr}(2\pi f_0 t)$ – меандр частотою $f_0 = 1000$ Гц; t_0 та t_1 – величини затримки сигналу $y(t)$, що дорівнюють 10 та 125 мс, відповідно. Адитивна суміш мовного сигналу та мовоподібної завади формувалася згідно виразу (1). Вигляд відповідних спектрів потужності для $SNR_0 = -12$ дБ показано на рис. 6.

На рис. 6 добре видно, що балансна модуляція, застосована при формуванні завади, спричинила зсув спектру мовного сигналу на величину f_0. Завдяки використанню меандру в якості несучого коливання, спостерігаємо також додатковий зсув на величину $3f_0$, що посилює маскувальну властивість синтезованої завади.

Крім того, в даній роботі досліджено маскувальну властивість іншого нестаціонарного процесу, алгоритм формування котрого полягав в амплітудній модуляції шумового сигналу $w(t)$ обвідною $o(t)$ мовного сигналу $x(t)$:

$$n_2(t) = o(t) \cdot w(t).$$

Обчислення значень обвідної $o(t)$ в дискретні моменти часу $t_r = r/F_s$, $r = 1,2,3,\ldots$, F_s – частота дискретизації, виконувалося шляхом ковзного усереднення модуля $|x(t)|$ мовного сигналу:

$$o(n) = (1-\beta) \cdot o(n-1) + \beta \cdot |x(n)|,$$

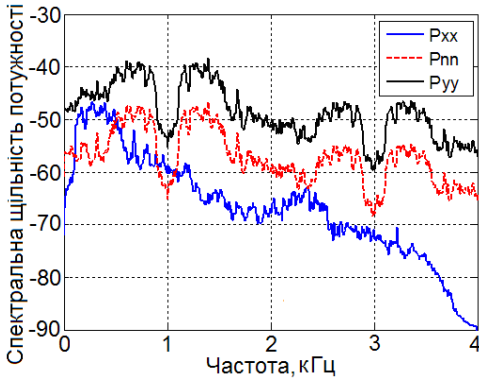
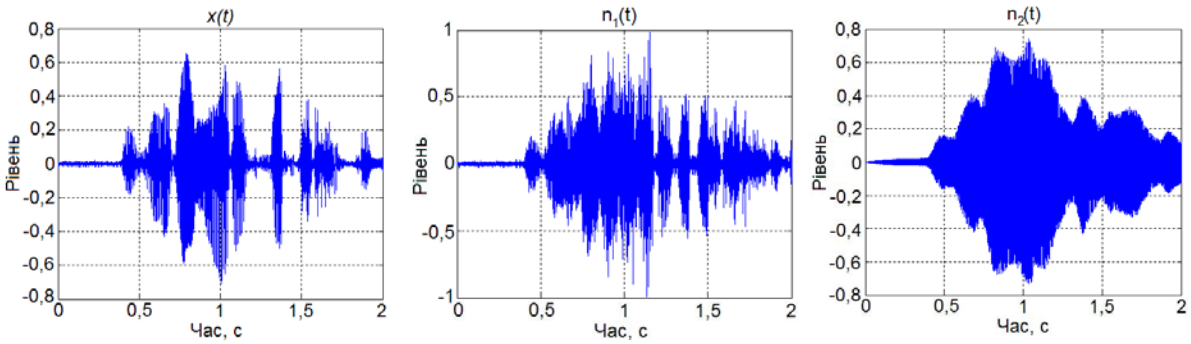


Рис. 6 Спектри потужності сигналу $x(t)$, шуму $n_1(t)$ та суміші $y(t)$ для $SNR_0 = -12$ дБ

де вибір значення $\beta = 3.6e-4$ забезпечував інтервал усереднення 0,125 с, близький до середньої довжини звуків мови [7].

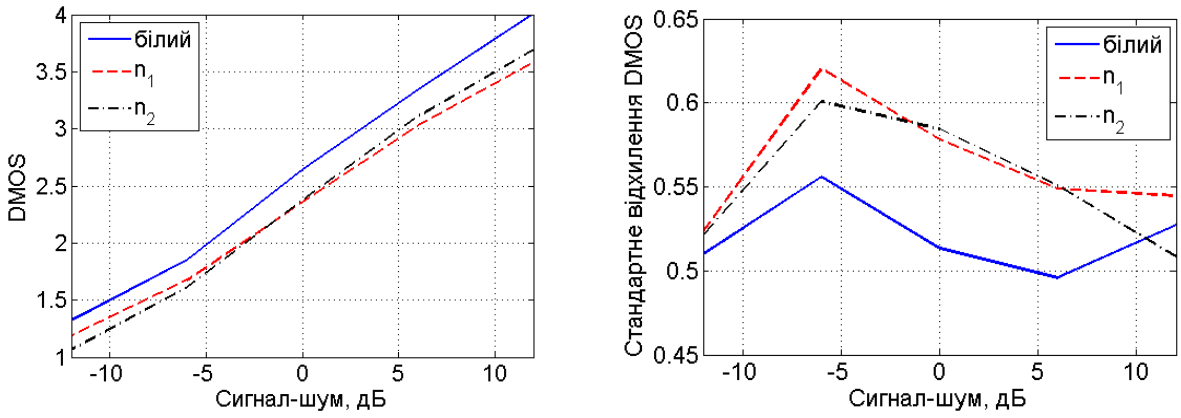
Для експериментальних досліджень в даній роботі в якості шуму $w(t)$ використано дискретний білий шум, як найбільш простий для практичної реалізації. Хоча, виходячи з наведених вище результатів дослідження маскувальних властивостей стаціонарних шумів, логічно очікувати, що застосування рожевого або коричневого шуму є більш перспективним. Форму мовного сигналу $x(t)$ та синтезованих шумів $n_1(t)$ та $n_2(t)$ показано на рис. 7.

Аналізуючи рис. 7, неважко бачити, що і у випадку використання завади $n_2(t)$ відношення сигнал-шум майже не змінюється у часі. В результаті таку заваду також не буде чутно під час мовленнєвих пауз. Результати оцінювання якості мовних сигналів, що маскуються завадами $n_1(t)$ і $n_2(t)$, у порівнянні із таким для білого шуму, показано на рис. 8а. У даному випадку в середньому стандартне відхилення оцінок DMOS близьке до 0,55 (рис. 8б).



а) б) в)

Рис. 7 Фрагменти мовного сигналу (а), шуму $n_1(t)$ (б) та шуму $n_2(t)$ (в)



а) б)

Рис. 8 Оцінки $DMOS(SNR)$ для завад $w(t)$, $n_1(t)$ і $n_2(t)$ (а) та стандартні відхилення цих оцінок (б)

Як бачимо, запропоновані в даній роботі завади $n_1(t)$ та $n_2(t)$, незважаючи на відносну простоту алгоритмів їх формування, мають кращі маскувальні властивості порівняно із білим шумом в діапазоні значень $SNR = -16\ldots+16$ дБ. Цей результат можна пояснити тим, що для розглянутих завад відношення сигнал-шум є практично постійним у часі, оскільки потужність шуму зростає майже одночасно із підвищенням потужності сигналу.

Зазначимо, що, з огляду на обмеженість обсягу експериментальних даних, одержані результати слід вважати попередніми. Крім того, в майбутньому доцільно використати, окрім шкали якості мовлення DMOS, ще й шкалу якісної оцінки ступеню захищеності мовної інформації від витоку [6].

При впровадженні запропонованих алгоритмів синтезу мовоподібних та нестаціонарних шумових завад слід враховувати, що наслідком їх обчислювальної простоти може бути досить легке декодування замаскованих сигналів.

Висновки

Одержані експериментальним шляхом суб'єктивні оцінки якості мовних сигналів, що маскуються білим, рожевим та коричневим шумами, добре узгоджуються із прогнозними оцінками розбірливості мовних сигналів, одержаними шляхом комп'ютерного моделювання. Зокрема, підтверджено прогноз щодо низької маскувальної здатності білого шуму, порівняно із рожевим та коричневим шумами, при низьких значеннях відношення сигнал-шум. Даний факт також свідчить про принципову можливість використання відносно простих способів суб'єктивного оцінювання якості мовних сигналів замість трудомісткого та дорогого артикуляційного методу оцінювання розбірливості мови.

Експериментальні дослідження маскувальних властивостей двох простих для обчислень алгоритмів формування нестаціонарних шумів показали, що завдяки синхронізації миттєвих потужностей сигналу та шуму вдається досягти подвійного ефекту. По-перше, маскувальна властивість нестаціонарних шумів є помітно вищою за таку для білого шуму. По-друге, під час мовних пауз шум відсутній, що дозволяє не забруднювати шумом навколишнє середовище. Разом із тим, при впровадженні запропонованих алгоритмів слід враховувати, що наслідком їх обчислювальної простоти може бути досить легке декодування замаскованих сигналів.

Одержані в даній роботі результати є попередніми, з огляду на обмеженість експериментальних даних. Крім того, в майбутньому доцільно використати, окрім шкали якості мовлення DMOS, ще й шкалу якісної оцінки ступеню захищеності мовленнєвої інформації від витоку.

Перелік посилань

[1] J. S. Bradley, R. D. Reich and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1820-1828, August 1999. **DOI:** 10.1121/1.427932

[2] H. Sato and J. S. Bradley, "Evaluation of acoustical conditions for speech communication in working elementary school classrooms," *The Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. 2064-2077, April 2008. **DOI:** 10.1121/1.2839283

[3] J. S. Bradley and H. Sato, "Speech Intelligibility Test Results for Grades 1, 3 and 6 Children," in *18th International Congress on Acoustics*, Kyoto, Japan, 2004. **URL:** https://www.icacommission.org/Proceedings/ICA2004Kyoto/pdf/Tu4.B1.2.pdf

[4] W. Yang and J. S. Bradley, "Effects of room acoustics on the intelligibility of speech in classrooms for young children," *The Journal of the Acoustical Society of America*, vol. 125, no. 2, pp. 922-933, February 2009. **DOI:** 10.1121/1.3058900

[5] Golden Harvest LLC, "Sound masking systems," [Online]. Available: http://www.ghaa-g.com/sms.html.

[6] A. A. Horev and Y. K. Makarov, «Otsenka effektivnosti sistem vibroakusticheskoy zaschity [On assessment of the effectiveness of vibroacoustic protection],» bnti.ru, [Online]. Available: http://www.bnti.ru/showart.asp?aid=874&lvl=04.02.03..

[7] N. B. Pokrovskiy, Raschet i izmerenie razborchivosti rechi [Prediction and measurement of speech intelligibility], Moskow: Svyazizdat, 1962, p. 390.

[8] A. N. Prodeus, V. S. Didkovskiy nd M. V. Didkovskaya, Akusticheskaya ekspertiza kanalov rechevoy kommunikatsii. Monografiya [Acoustic examination of the speech communication channels. Monograph], Kyiv: Imeks-Ltd, 2008, p. 420.

[9] A. N. Prodeus, L. B. Dronzhevskaya, V. A. Klimkov and D. A. Shagitova, «Modelirovanie algoritmov formantno-modulyatsionnogo metoda otsenivaniya razborchivosti rechi [Modeling of the algorithms of the formant-modulation technique for evaluating speech intelligibility],» *Electronics and Communications*, vol. 16, no. 2, pp. 79-85, 2011.

[10] J. G. Beerends, E. Larsen, N. Iyer and J. M. v. Vugt, "Measurement of speech intelligibility based on the PESQ approach," in *Measurement of Speech and Audio Quality in Networks*, Prague, Czech Republic, 2004.
URL: http://wireless.feld.cvut.cz/mesaqin2004/papers/3.PDF

[11] N. Côté, Integral and Diagnostic Intrusive Prediction of Speech Quality, Springer-Verlag Berlin Heidelberg, 2011, p. 250. **DOI:** 10.1007/978-3-642-18463-5

[12] J. Ma, Y. Hu and P. C. Loizou, «Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions,» *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387-3405, 2009. **DOI:** 10.1121/1.3097493

Надійшла до редакції 18 травня 2017 р.

УДК 004.934

Субъективное оценивание качества и разборчивости речевых сигналов, искаженных синтезированными шумами

Продеус А. Н., д.т.н., проф. ORCID 0000-0001-7640-0850
e-mail am.prodeus@aae.kpi.ua
Витык А. В., ORCID 0000-0002-6652-4152
e-mail andrvit8@gmail.com
Диденко Д. Ю., ORCID 0000-0001-7992-3235
e-mail dyu.didenko@aae.kpi.ua
Национальный технический университет Украины
"Киевский политехнический институт имени Игоря Сикорского" kpi.ua
Киев, Украина

***Аннотация*—В данной работе приведены результаты оценивания влияния стационарных и нестационарных синтезированных шумов на качество и разборчивость речевых сигналов. Для случая стационарных шумов показано, что при малых отношениях сигнал-шум белый шум уступает розовому и коричневому шумам по маскировочной способности. Исследовано два простых, с вычислительной точки зрения, алгоритма формирования нестационарных шумов, обеспечивающих лучшую, в сравнении с белым шумом, маскировку речевых сигналов, а также меньше загрязняющих окружающую среду во время речевых пауз.**
Библ. 12, рис. 8.

Ключевые слова—разборчивость речи; качество речи; окрашенный шум; речеподобный шум; отношение сигнал-шум.

UDC 004.934

Subjective evaluation of quality and intelligibility of speech distorted by synthesized noise

A. M. Prodeus, Dr.Sc.(Eng.), Prof. ORCID 0000-0001-7640-0850
e-mail am.prodeus@aae.kpi.ua
A. V. Vityk, ORCID 0000-0002-6652-4152
e-mail andrvit8@gmail.com
D. Yu. Didenko, ORCID 0000-0001-7992-3235
e-mail dyu.didenko@aae.kpi.ua
National technical university of Ukraine "Igor Sikorsky Kyiv polytechnic institute" kpi.ua
Kyiv, Ukraine

***Abstract*—The distortion of the speech signal by noise interferences negative impacts on the perception of speech information by listeners, and a noise disturbance in the form of people conversations has the best masking ability. This phenomenon is usually used when the intelligibility of speech should be minimal. Therefore, there are nowadays many different acoustical systems generating acoustic noise in the form of stationary or non-stationary noise for active masking of speech information. Assessment of acoustic masking quality for systems generating stationary noise can be made by means of formant technique and speech intelligibility can be used as a measure of masking quality for such systems. Previously, it was theoretically shown that masking property of white noise is worst at low signal-to-noise ratio. However, this result was not tested by subjective testing. Moreover, masking ability of nonstationary noise was not tested too. In this paper, this gap has been eliminated and the results of subjective estimation of the effect of stationary and nonstationary synthesized noise on the quality and intelligibility of speech signals are presented. Degradation Mean Opinion Score (DMOS) measure of speech quality was used for the estimation. It was used the fact of high correlation (about 0.9) between speech quality and intelligibility upon results interpretation. For the case of stationary noise, it was shown that white noise is inferior to pink and brown noise by masking ability for signal-to-noise ratios below minus 5 dB. This result is in a good agreement with previously theoretically predicted one. Two simple, from the computational point of view, non-stationary noise generation algorithms were studied also. The first algorithm uses both spectrum inversion and reverberation simulation. Second algorithm is based on formation of nonstationary process as result of noise carrier amplitude modulation by envelope of speech signal. It was found that these nonstationary processes provide a better, in comparison with white noise, masking of speech signals. These nonstationary processes have the significant advantage compare to stationary ones because they provide less environmental pollution during speech pauses.**
Ref. 12, fig. 8.

Key words — speech intelligibility; speech quality; colour noise; speech-like noise; signal-to-noise ratio.