# Моделювання мовного апарату людини в задачах комп'ютерного синтезу

*І.О. Стеля*

Київський національний університет імені Тараса Шевченка, Україна
igor.stelia@gmail.com

## Резюме

The method of synthesis speech information problem solving got development in work, that allowed to made a decision the task of artificial formation of sounds of man voice, on the basis of the general use of physical models of vocal source and voice tract. On the base of the developed algorithms a complex of programmes is created for research of influencing of parameters of models in the system «a vocal source is a voice tract» on speech production.

## 1. Вступ

Розвиток комп'ютерних мовних синтезаторів призвів до розробки декількох підходів до їх побудови, а саме:

– синтез на основі конкатенації відрізків записаних мовленнєвих сигналів, що зберігаються у мовленнєвих базах даних;

– формантний синтез, що не використовує зразки людської мови для синтезу. Замість цього, синтезована мова створюється з використанням акустичних моделей. Параметри, такі як набір значень частот основних формант та рівні шуму, змінюються в часі та створюють сигнал штучної мови;

– артикуляторний синтез, що створює штучну мову шляхом безпосереднього моделювання людського мовного апарату. Цей метод являє собою синтез мови на базі обчислювальних методів, заснованих на моделях людського мовного апарату і артикуляційних процесах, що відбуваються у ньому. Моделі мовного апарату, як правило, включають модель мовного тракту і модель голосових зв'язок.

Для побудови та дослідження математичних моделей голосового джерела, що моделюють коливні процеси та моделей мовного тракту, що моделюють розповсюдження акустичних хвиль, використовуються чисельні методи. Для задачі відновлення параметрів мовного тракту за вимірюваним сигналом на виході використовується математичний апарат розв'язання обернених задач. Проблемі розв'язання мовних обернених задач присвячені, зокрема, роботи [1, 2].

## 2. Модель голосових зв'язок

У моделі Ішізаки і Фланагана [Error: Reference source not found] кожна з голосових зв'язок описується двома масами, що зв'язані пружинами як із стінками, так і між собою. Більш того, припускається, що зв'язки є двосторонньо симетричними. Маси, що моделюють зв'язки, здійснюють коливання у поперечному до руху повітря напрямку. Система рівнянь для двох мас, що коливаються, записується у вигляді:

$$m_1 \frac{d^2 x_1(t)}{dt^2} + r_1 \frac{dx_1(t)}{dt} + k_1(x_1(t) - x_{01}) + k_c(x_1(t) - x_2(t)) = l_g d_1 p_{m1}(t),$$

$$m_2 \frac{d^2 x_2(t)}{dt^2} + r_2 \frac{dx_2(t)}{dt} + k_2(x_2(t) - x_{02}) - k_c(x_1(t) - x_2(t)) = l_g d_2 p_{m2}(t),$$

(1)

де $m_1$ і $m_2$ – маси; $x_1(t)$, $x_2(t)$ – зміщення мас $m_1$ і $m_2$; $x_{01}, x_{02}$ – початкове положення мас $m_1$ і $m_2$; $t$ – час; $r_1$ і $r_2$ – коефіцієнти демпфування; $k_1$, $k_2$ – пружність пружин для мас $m_1$ і $m_2$; $k_c$ – пружність пружини, що з'єднує маси $m_1$ і $m_2$; $d_1, d_2$ – товщина мас $m_1$ і $m_2$; $l_g$ – діюча довжина голосових зв'язок; $l_g d_1$, $l_g d_2$ – поверхні мас $m_1$ і $m_2$, на які діють тиски $p_{m1}(t)$ і $p_{m2}(t)$, відповідно.

Розподіл тиску в голосовій щілині апроксимується послідовними дискретними кроками $p_{ij}$ на кожному $j$-у кінці кожної $i$-ї маси. Перше падіння тиску $p_s - p_{11}(t)$ виводиться з рівняння Бернуллі для ідеальної рідини у стаціонарному режимі. Вздовж кожної з мас падіння тиску $p_{11}(t) - p_{12}(t)$ та $p_{21}(t) - p_{22}(t)$ визначається в'язкими втратами і пропорційні зсувовій в'язкості повітря. Падіння тиску на стику мас $p_{12}(t) - p_{21}(t)$ еквівалентне зміні кінетичної енергії в одиниці об'єму газу. І остаточно різке розширення на верхньому кінці голосової щілини призводить до відновлення тиску до атмосферного тиску $p$. Таким чином, система рівнянь для змін тиску записується у вигляді [Error: Reference source not found]:

$$p_s - p_{11}(t) = 0.69\rho \frac{u_g^2(t)}{A_{g1}^2(t)} + \int_0^{l_c} \frac{\rho}{A_c(x)} dx \cdot \frac{du_g}{dt},$$

$$p_{11}(t) - p_{12}(t) = 12\nu d_1 \frac{l_g^2 u_g(t)}{A_{g1}^3(t)} + \frac{\rho d_1}{A_{g1}} \cdot \frac{du_g}{dt},$$

$$p_{12}(t)-p_{21}(t)=\frac{1}{2}\rho u_g^2(t)\left(\frac{1}{A_{g2}^2(t)}-\frac{1}{A_{g1}^2(t)}\right),$$

(2)

$$p_{21}(t)-p_{22}(t)=12\nu d_2\frac{l_g^2 u_g(t)}{A_{g2}^3(t)}+\frac{\rho d_2}{A_{g2}}\cdot\frac{du_g}{dt},$$

$$p_{22}(t)-p=\frac{1}{2}\rho\frac{u_g^2(t)}{A_{g2}^2(t)}\left[2\frac{A_{g2}(t)}{A_1}\left(1-\frac{A_{g2}(t)}{A_1}\right)\right],$$

де $\rho$ – густина повітря; $\nu$ – зсувова в'язкість повітря; $A_1$ – площа голосового тракту на вході; $A_{gi}$ – площа голосової щілини під $i$-ю масою; $u_g(t)$ – об'ємна швидкість потоку повітря;

$$A_{gi}(t)=(A_{g0i}+2l_g x_i(t)),\qquad i=1,2,$$

де $x_1(t)\geq x_{01}$, $x_2(t)\geq x_{02}$; $A_{g01}$, $A_{g02}$ – залишкові площі в момент змикання голосових зв'язок.

Для визначення значень тиску $p_{m1}(t)$, $p_{m2}(t)$ використовуються співвідношення:

$$p_{m1}(t)=\frac{1}{2}(p_{11}(t)+p_{12}(t)),$$

$$p_{m2}(t)=\frac{1}{2}(p_{21}(t)+p_{22}(t)). \quad (3)$$

Шуканим розв'язком системи (1)–(3) є функція $u_g(t)$, яка визначає потік повітря на виході з голосової щілини.

Для реалізації двомасової моделі голосових зв'язок Ішізаки – Фланагана розроблений удосконалений чисельний алгоритм, в основу якого покладена комбінація методу розв'язання системи рівнянь коливання двох мас і методу розв'язання нелінійної системи для змін тиску. Отриманий за допомогою розробленого чисельного алгоритму розв'язок та його похідна можуть бути використані як голосове джерело для моделі мовного тракту без додаткової обробки. Чисельний алгоритм дозволяє відтворювати різні режими коливання голосових зв'язок, що відповідають різним типам фонації. Значення вихідних параметрів моделі та розраховані характеристики голосового джерела знаходяться у фізіологічно допустимих межах. Приклади розрахунків наведені на Рис. 1 та 2.

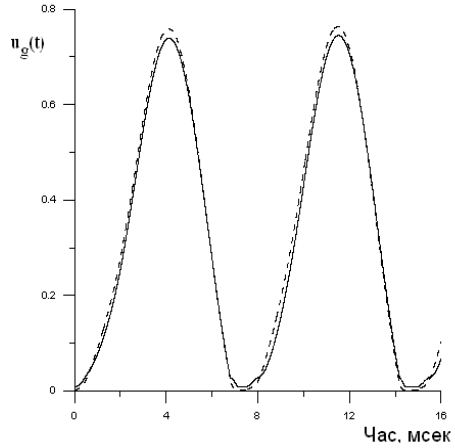


Рисунок 1: Потік повітря $u_g$, обчислений для нейтрального (пунктирна лінія) та придихального (суцільна лінія) типу фонації

З метою дослідження здійснене порівняння потоку повітря, обчисленого за двомасовою моделлю, та потоку повітря, отриманого за аналітичною R-моделлю [4]:

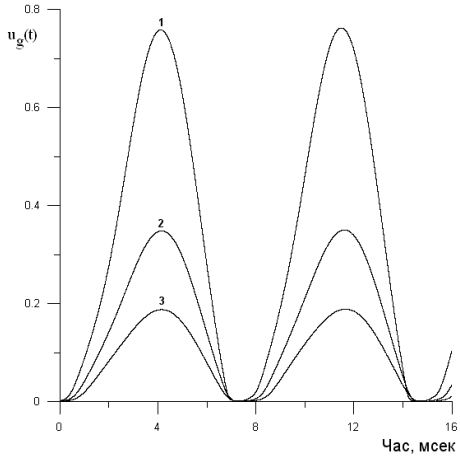


Рисунок 2: Результати розрахунків потоку повітря $u_g(t)$ для трьох значень тиску на вході у голосову щілину $p_{s_1}$ = 0.02, $p_{s_2}$ = 0.0012, $p_{s_3}$ = 0.008

$$u_g(t)=\begin{cases}\alpha\left(3\left(\frac{t}{t_p}\right)^2-2\left(\frac{t}{t_p}\right)^3\right), & 0\leq t\leq t_p,\\ \alpha\left(1-\left(\frac{t-t_p}{t_p}\right)^2\right), & t_p\leq t\leq t_p+t_n\leq t_0,\end{cases}$$

де $\alpha$ – амплітуда, $t_p$ – час відкриття, $t_n$ – час закриття, $t_0$ – період. Результати порівняння моделей наведені на Рис. 3.

Для вивчення голосового джерела, окрім функції потоку повітря, використовується її похідна. Звичайний набір параметрів для кількісного опису коливань виводиться із залежності похідної потоку від часу і включає: амплітуду від'ємного піка, тривалість відкриття, тривалість закритої фази, проміжок часу між додатним і від'ємним піками, час повернення в початкове положення.

Проведений порівняльний аналіз отриманої чисельної похідної потоку повітря за двомасовою моделлю та похідної потоку повітря, отриманою за аналітичною LF моделлю [5]:

$$\frac{dy(t)}{dt} = \begin{cases} E_0 e^{\alpha t} \sin(\omega_g t), & 0 \le t < t_e, \\ -\dfrac{E_e}{\varepsilon t_a}\left[e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)}\right], & t_e \le t < t_c, \\ 0, & t_c \le t < t_0, \end{cases}$$

де $t_0$ – період; $t_e$ – відмітка мінімального значення похідної; $t_a$ – визначається як точка, в якій тангенс перетинає вісь координат; $t_c$ – момент, в який похідна під час фази закриття майже досягає нуля; $E_e$ – абсолютне значення мінімуму похідної; $\alpha$, $\varepsilon$, $\omega_g$ – деякі параметри. Результати порівняння моделей наведені на Рис. 4.

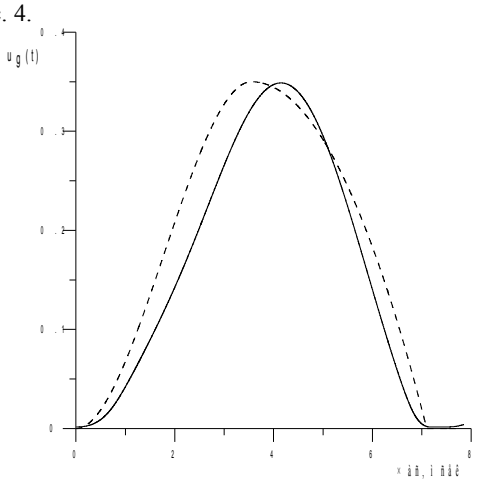


Рисунок 3: Потік повітря $u_g$ на виході з голосових зв'язок, обчислений за двомасовою моделлю (суцільна лінія), та R-моделлю (пунктирна лінія)

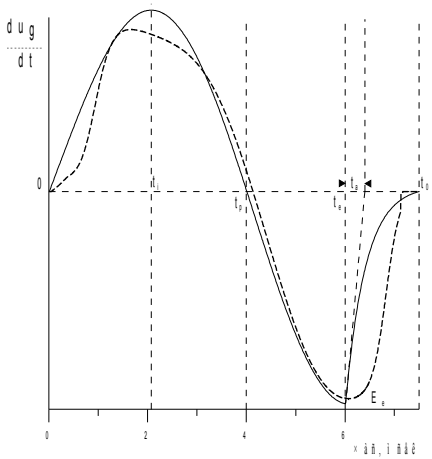


Рисунок 4: Чисельна похідна від потоку повітря, обчисленого за двомасовою моделлю (пунктирна лінія) та похідна, обчислена за LF моделлю (суцільна лінія)

## 3. Моделювання мовного тракту людини

Для моделювання розповсюдження акустичних хвиль у мовному тракті як у неоднорідній акустичній трубі, що починається між голосовими зв'язками та закінчується губами, використовується система рівнянь акустики в частинних похідних, яка записується у вигляд [6]:

$$-\frac{\partial p}{\partial x} = \frac{\rho}{S(x)}\frac{\partial u}{\partial t}, \qquad -\frac{\partial u}{\partial x} = \frac{S(x)}{\rho c^2}\frac{\partial p}{\partial t}, \tag{4}$$

де $0 \le x \le L$, $t > 0$, $L$ – довжина мовного тракту; $p(x,t)$ – тиск у тракті в момент часу $t$; $u(x,t)$ – об'ємна швидкість потоку; $\rho$ – густина повітря в тракті; $c$ – швидкість звуку; $S(x)$ – функція площі поперечного перерізу.

Оскільки тракт має неоднорідний поперечний переріз, він розбивається на циліндричні секції однакової довжини з постійною площею перерізу (Рис. 5).

В якості крайової умови на вході в тракт вибирається потік $u_g(t)$, знайдений з (1)–(3). Виходячи з цього, маємо крайову умову: $u(0,t) = u_g(t)$. На протилежному кінці тракту задаємо умову $p(L,t) = 0$.

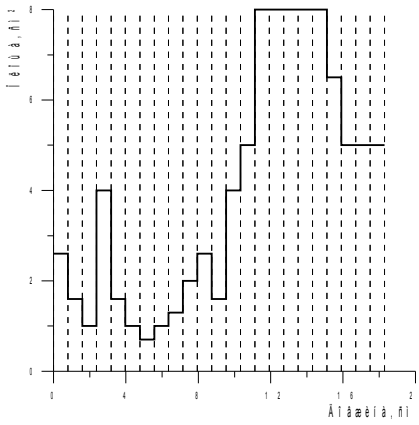

Рисунок 5: Наближення мовного тракту циліндричними секціями однакової довжини

Різницева задача для апроксимації системи рівнянь (4) будується на рознесеній сітці. Для розв'язання використовується явний метод «чехарда». Цей алгоритм має високу обчислювальну ефективність і може використовуватися для моделювання у реальному часі [7].

Для моделювання розповсюдження акустичних хвиль також може використовуватися рівняння Вебстера [Error: Reference source not found]:

$$S(x)\frac{\partial^2 P}{\partial t^2} = c^2 \frac{\partial}{\partial x}\left(S(x)\frac{\partial P}{\partial x}\right) \qquad (5)$$

де $x$ – просторова координата уздовж середньої лінії тракту в середньо-сагітальній площині; $t$ – момент часу; $p(x,t)$ – шуканий тиск у тракті; $S(x)$ – профіль площ поперечного перерізу вздовж тракту; $c$ – швидкість звуку в тракті.

В якості крайової умови на вході в тракт вибирається похідна від потоку повітря $P(0,t) = -\frac{\rho}{S(0)}\frac{du_g(t)}{dt}$.

Для розв'язання задачі (5) використовується скінченнорізницевий метод. Для розв'язання системи різницевих рівнянь використовується ітераційний метод послідовної верхньої релаксації.

Розглянемо задачу відновлення форми мовного тракту за виміряними акустичними параметрами сигналу на базі акустичного рівняння Клейна – Гордона [8]. Для цього вводиться нова змінна $\varphi(x,t)$, яка визначається виразом:

$$\varphi(x,t) = P(x,t)S(x)^{1/2}. \qquad (6)$$

Акустичне рівняння у формі Клейна – Гордона записується:

$$\frac{\partial^2 \varphi(x,t)}{\partial t^2} = c^2 \frac{\partial^2 \varphi(x,t)}{\partial x^2} - c^2 U(x)\varphi(x,t), \qquad (7)$$

$$0 < x < L \quad 0 < t \le T.$$

Рівняння (7) має форму хвильового, де функція $U(x)$ визначена в термінах площі поперечного перерізу мовного тракту як

$$U(x) = \frac{d^2 S(x)^{1/2} / dx^2}{S(x)^{1/2}}. \qquad (8)$$

## 4. Мовна обернена задача

Мовна обернена задача визначається як задача знаходження функції *S(x)* за виміряними параметрами мовного сигналу на виході з тракту. Математично ця задача розв'язується як задача пошуку мінімуму деякого функціоналу за різного роду обмежень.

Нехай на виході з тракту вимірюється тиск $P(L,t)$, зв'язаний з розв'язком рівняння Клейна – Гордона співвідношенням (6). Позначимо $\Phi(t)$ функцію, яка вимірюється на виході з тракту.

Задача зводиться до мінімізації функціоналу

$$J(U) = \int_0^T \left(\Phi(t) - \varphi_U(L,t)\right)^2 dt, \qquad (9)$$

де $\varphi_U(L,t)$ – розв'язок задачі (9) при заданій функції $U(x)$.

Для мінімізації функціоналу (9) використовується градієнтний метод [9]. Приріст функціоналу записується у вигляді:

$$\Delta J(U) = J(U+h) - J(U) =$$

$$= \int_0^T 2(\Phi(t) - \varphi_U(L,t))\Delta\varphi\, dt + \int_0^T (\Delta\varphi)^2 dt,$$

де $\Delta\varphi = \varphi_{U+h}(x,t) - \varphi_U(x,t)$. Для визначення градієнта функціоналу будується спряжена задача, яка записується у вигляді:

$$\frac{\partial^2 \Psi(x,t)}{\partial t^2} = c^2 \frac{\partial^2 \Psi(x,t)}{\partial x^2} - c^2 U(x)\Psi(x,t),$$

$$0 < x < L \quad 0 < t \le T.$$

Градієнт функціоналу визначається через розв'язок спряженої задачі за формулою

$$J' = -\varphi\Psi.$$

Після визначення $U(x)$ функція $S(x)$ визначається з (8).

У тестових розрахунках була використана наступна стратегія перевірки працездатності побудованого алгоритму та створеного програмного забезпечення. Розв'язувалась пряма задача та визначався сигнал на виході. Далі для розв'язування оптимізаційної задачі цей сигнал використовувався як виміряний. Для початкового наближення функції $U(x)$ використовувалась довільна функція, представлена графіком на Рис. 6.

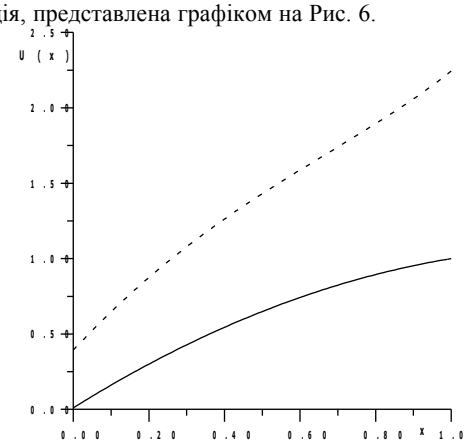


Рисунок 6: Початкове наближення для функції $U(x)$ (суцільна лінія), уточнена функція $U(x)$ (пунктирна лінія)

Точність розв'язку оцінюється за процедурою ресинтезу: синтезований за знайденим розв'язком сигнал має мало відрізнятися від вихідного сигналу, за параметрами якого вирішувалася обернена задача. Результати розрахунків наведені на Рис. 7.

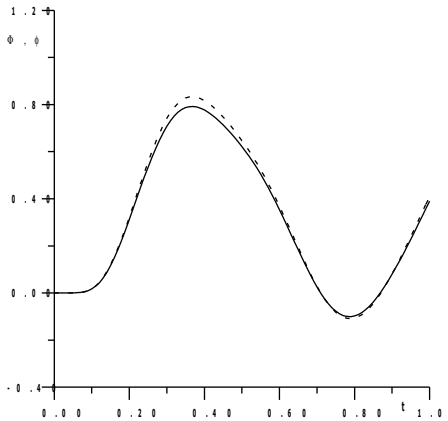


Рисунок 7: Виміряний сигнал (суцільна лінія) $x = L$ та його наближення, як розв'язок оберненої задачі (пунктирна лінія)

## 5. Висновки

У статті висвітлена єдина система, яка дозволяє комп'ютерне відтворення звуків, що утворює людина на основі спільного використання фізичних моделей голосового джерела та мовного тракту. Також сформульовано і розв'язано обернену задачу відновлення параметрів мовного тракту.

## 6. Література


[1] Леонов А. С. Обратная задача для управления артикуляцией /А. С. Леонов, В. Н. Сорокин// Доклады Академии Наук, 2000. – Т. 374, № 6. – с. 749–753.

[2] Артикуляторный ресинтез фрикативных / [Леонов А. С., Макаров И. С., Сорокин В. Н., Цыплихин А. И.] // Инф. процессы, –2004. –т.4, №2. –С. 141–159.

[3] Rudnicky, A. I., Polifroni, Thayer, E H., and Brennan, R. A. "Interactive problem solving with speech", *J. Acoust. Soc. Amer.*, *Vol. 84, 1988, p S213(A).*

[4] Rosenberg A. Effect of glottal pulse shape on the quality of natural vowels /A. Rosenberg// J. Acoust. Soc. Am., – 1971. –Vol. 49. –pp. 583–590.

[5] Fant G. The lf-model revisited. Transformations and frequency domain analysis / G. Fant // STL-QPSR, Tech. Rep., – 1995. – Vol.2-3. – pp. 119–156.

[6] Fundamentals of Acoustics / L. E. Kinsler, A. E. Frey, A.B. Coppens, J.V. Saunders. – San Diego: Academic Press, – 1982. – 496 pp. – ISBN: 9780471029335.

[7] Van den Doel K. Real-time numerical solution of Webster's equation on a non-uniform grid /K. van den Doel, U. Ascher// IEEE Trans. Audio, Speech and Language Processing, –2008. –Vol. 16. – pp. 1163–1172.

[8] Forbes B.J., Pike E.R, Sharp D.B. The accoustical Klein-Gordon equation: The wave-mechanical step and barrier potential functions, // J.Acooust.Soc.Am. – 2003. – 114(3). – pp.1291-1302.

[9] Васильев Ф.П. Методы решения экстремальных задач. – Москва: Наука, 1981.