

# КОМП'ЮТЕРНА ТЕХНОЛОГІЯ МОДЕЛЮВАННЯ ТА КЕРУВАННЯ ВІЗУАЛЬНИМИ ОБРАЗАМИ ЛЮДСЬКОГО ОБЛИЧЧЯ ПРИ СИНТЕЗІ МОВЛЕННЯ ЕМОЦІЙНО ЗАБАРВЛЕНОГО ТЕКСТУ<sup>1</sup>

Юрій Крак, Олександр Бармак

Київський національний університет імені Тараса Шевченка

64 вул. Володимирська, Київ 03022

Електронна пошта: [krak@unicyb.kiev.ua](mailto:krak@unicyb.kiev.ua), [barmak@svitonline.com](mailto:barmak@svitonline.com)

Yuriy Krak, Olexander Barmak, **Computer technologies modeling and control of human face visual images for speech synthesis with emotions.** The new technologies for modeling of processes animation of arbitrary text are proposed. Mathematical model of human head with possibilities of facial expression during conversation and synchronization with speech are created.

## Вступ

Світові тенденції розвитку сучасних технологічних рішень для моделювання і обробки аудіо та відео інформації спрямовані на створення систем візуалізації процесу промовляння за допомогою комп'ютерних засобів. Дані технології мають значне прикладне застосування в інтелектуалізації роботи з комп'ютером, кінематографії, телебаченні, телефонії, передачі інформації тощо [1-5].

Для реалізація такої технології необхідно створити програмне забезпечення, яке буде об'ємну модель голови людини та алгоритми, що реалізують динамічну генерацію візуальних образів людського обличчя при синтезі мовлення. Тобто потрібно вміти моделювати процес зміни образів обличчя людини, який синхронний звукам (фонемам), генерованим мовним синтезатором. Припускається, що є мовний синтезатор, який вхідний орфографічний текст перетворює у розмічений фонетичний текст (транскрипцію), тобто набір фонем з тривалістю кожної фонемі та генерує звук. Також припускається, що розмічений фонетичний текст містить маркери емоційного стану людини (сміх, горе тощо), який треба відтворити в процесі візуалізації.

## 1. NURBS-моделювання голови людини

В даній роботі розглянута альтернатива полігональному моделюванню голови людини, а саме – моделювання за допомогою нерівномірного раціонального базисного сплайну (NURBS) [6-9]:

$$p(u, v) = \frac{\sum_{i=0}^n \sum_{j=0}^m B_{i,d}(u) B_{j,d}(v) w_{i,j} p_{i,j}}{\sum_{i=0}^n \sum_{j=0}^m B_{i,d}(u) B_{j,d}(v) w_{i,j}}, \quad (1)$$

де  $p_{i,j}$  – масив опорних (керуючих) точок (*Control Points*);  $w_{i,j}$  – коефіцієнти для збільшення або зменшення “ваги” конкретної опорної точки;  $B_{i,d}(\cdot)$  – базисні функції, які обчислюються за методом Кокса-де Бура [7]:

$$B_{k,0} = \begin{cases} 1, & \text{якщо } u_k \leq u \leq u_{k+1} \\ 0 & \text{у іншому випадку} \end{cases} \quad (2)$$

$$B_{k,d} = \frac{u - u_k}{u_{k+d} - u_k} B_{k,d-1}(u) + \frac{u_{k+d+1} - u}{u_{k+d+1} - u_{k+1}} B_{k+1,d-1}(u)$$

При моделюванні обличчя за допомогою NURBS-поверхонь кількість точок, які потрібно зберігати зменшується на порядки, а деформації поверхні (для анімації) виходять більш плавні – подібні до деформацій реальних облич.

Для зображення NURBS-поверхні використовувалась рівномірна мозаїчна модель [8], для якої брались однакові кроки між мінімальними та максимальними значеннями для параметрів, визначених на поверхні. Отримані точки використовувались як вершини (вертекси) трикутників, які віддаються на вхід 3-х мірного API, типу OpenGL або Direct3D.

Для швидшого обчислення поверхні  $p(u, v)$ , один раз визначались  $B_{i,d}(u)$  та  $B_{j,d}(v)$  у точках поділу та зберігались у масиві, а потім використовувались, а не обчислювались кожного разу у внутрішньому циклі.

Для знаходження поверхневих нормалей (щоб 3-х мірне API змогло зробити освітлення) потрібно створити дві дотичні (одну в  $u$  напрямі та одну в  $v$  напрямі) та взяти векторний добуток для отримання поверхневої нормалі. Якщо записати систему рівнянь для обчислення коефіцієнтів у вигляді поліноміального рівняння, еквівалентного  $B_{i,d}(u)$ , а потім брати похідну від  $B_{i,d}(u)$ , то це буде просте множення на коефіцієнт та зменшення степені на одиницю:

$$\frac{d(Ax^n + Bx^m)}{dx} = nAx^{(n-1)} + mBx^{(m-1)}. \quad (3)$$

<sup>1</sup> Робота виконана в рамках ДНТІ України “Образний комп'ютер”

Отже, потрібно обчислити коефіцієнти базисних функцій (3), коли вони подані у вигляді полінома:

$$B_{i,d}(u) = C_{i,d,d}(u)u^d + C_{i,d,d-1}(u)u^{d-1} + \dots + C_{i,d,1}(u)u + C_{i,d,0}(u). \quad (4)$$

Обчислення коефіцієнтів  $C_{i,d,k}(u)$ , маючи на увазі (2), виглядає наступним чином:

$$\begin{aligned} C_{i,0,0}(u) &= B_{i,0}(u) \\ C_{i,d,0}(u) &= \frac{u_{i+d+1}C_{i+1,d-1,0}(u) - u_i C_{i,d-1,0}(u)}{u_{i+d+1} - u_i} \\ C_{i,d,d}(u) &= \frac{C_{i,d-1,d-1}(u) - C_{i+1,d-1,d-1}(u)}{u_{i+d} - u_i} \\ C_{i,d,k}(u) &= \frac{C_{i,d-1,k-1}(u) - u_i C_{i,d-1,k}(u)}{u_{i+d} - u_i} - \frac{C_{i+1,d-1,k-1}(u) - u_{i+d+1} C_{i+1,d-1,k}(u)}{u_{i+d+1} - u_{i+1}}, \end{aligned} \quad (5)$$

для  $0 < k < d$

Тепер похідна  $B_{i,d}(u)$  обчислюється за формулою:

$$\frac{dB_{i,d}(u)}{du} = dC_{i,d,d}(u)u^{d-1} + (d-1)C_{i,d,d-1}(u)u^{d-2} + \dots + C_{i,d,1}(u) \quad (6)$$

При реалізації цього підходу, якщо розглядати поліноміальну форму завдання кривої на стандартному інтервалі:

$$p(u) = \sum_{i=0}^n C_i u^i, \quad 0 \leq u \leq 1 \quad (7)$$

то, замість того, щоб обраховувати незалежно кожний член, який містить  $u^k$ , їх можна згрупувати наступним чином (схема Горнера):

$$p(u) = c_0 + u(c_1 + u(c_2 + u(\dots + c_n u))) \quad (9)$$

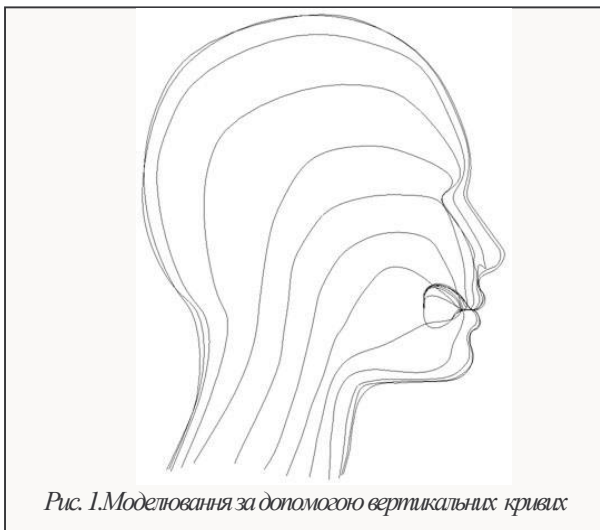


Рис. 1. Моделювання за допомогою вертикальних кривих

Моделювати голову людини запропоновано з використанням вертикальних кривих (Рис. 1), які беруть початок на внутрішній поверхні роту [17]. Вони виходять з роту на поверхню, повторюючи риси обличчя, та закінчуються біля основи шиї. Цей підхід ефективно спрацьовує при створенні окремої каркасної моделі голови, яка визначає більшу частину рис обличчя. Оскільки, зазвичай, напрямком

кривих співпадає з напрямком м'язів обличчя, це робить легкою анімацію міміки.

Основоючись на описаному вище підході, використовуючи парадигму об'єктно-орієнтованого проектування, створено еталонну модель та абстрактний клас *TNURBSSurface*, який реалізує алгоритм побудови NURBS поверхні цієї моделі у реальному часі (Рис.2). В рамках цієї моделі, розглянуті наступні питання:

- основні результати моделювання процесу анімації еталонною моделлю;
- нанесення текстури на NURBS-поверхню голови.

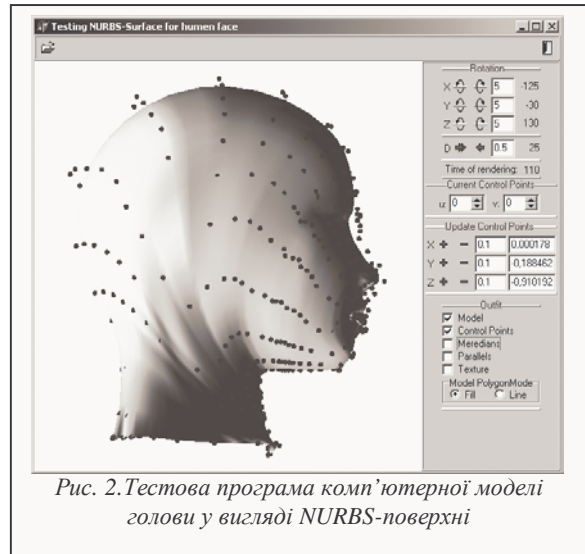


Рис. 2. Тестова програма комп'ютерної моделі голови у вигляді NURBS-поверхні

## 2. Моделювання процесу анімації

Існує властивість NURBS-поверхонь [6], яка робить їх більш привабливими у використанні (з точки зору анімації), ніж полігональні поверхні. Мова йде про те, що при зміні координат керуючих точок – поверхня залишається гладкою, тобто зміна координат керуючих точок приводить до “людиноподібних”, плавних змін на поверхні. Виходячи з цієї властивості, шляхом модифікації керуючих точок, створюються множини різних емоційних станів, як варіації базової моделі. Створено множину виразів обличчя при промовлянні фонем та множину типових емоційних станів. Комбінації моделей, утворені з цих множин, використовуються для анімації промовляння тексту.

Для моделювання процесу анімації використано алгоритми морфінгу та блендингу (Рис. 3). Введені означення цих операцій для NURBS-поверхонь.

**Означення 1.** Під *NURBS-морфінгом* розуміється плавний перехід (перетворення) NURBS-поверхні з одного стану у інший. Тобто, генерація (інтерполяція) між цими станами потрібної кількості проміжних станів. NURBS-поверхню певної топології однозначно означає множина керуючих точок та вектор вузлів. Під час переходу від одного стану до іншого приймається, що кількість керуючих точок не міняється, а змінюються лише координати

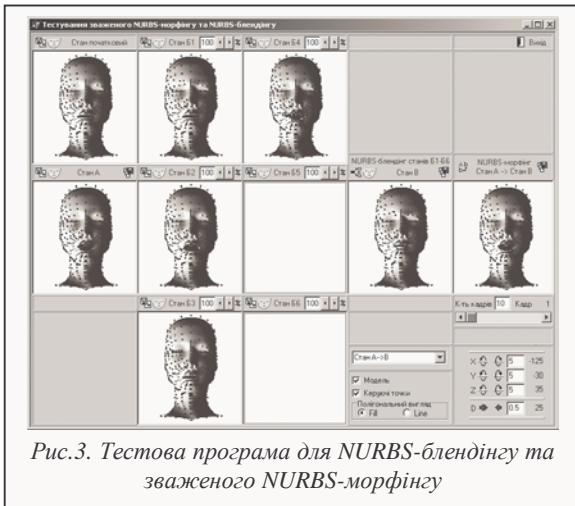


Рис.3. Тестова програма для NURBS-блендінгу та зваженого NURBS-морфінгу

цих точок (при фіксованому векторі вузлів). Використовується та властивість сплайнових поверхонь, що довільна зміна координат керуючих точок приводить до плавної зміни на поверхні. Виходячи з цього, визначається *NURBS-морфінг*, як лінійна інтерполяція між парними керуючими точками. Тобто, ставиться задача: є множина із  $N$  трьохмірних керуючих точок  $p_i = [x_i, y_i, z_i]^T$ , яка означає початковий стан (стан **A**) поверхні, та множина керуючих точок  $p_i^* = [x_i^*, y_i^*, z_i^*]^T$ , яка означає кінцевий стан (стан **B**), і потрібно перейти зі стану **A** у стан **B** за певну ( $k$ ) кількість кроків. Для розв'язку знаходяться направляючі вектори для кожної пари початкових та кінцевих керуючих точок:

$$\Delta p_i = \frac{p_i^* - p_i}{k}, i = 1 \dots N. \quad (9)$$

Ці вектори направлені від точок, що означають стан **A**, до точок, що означають стан **B**, і мають таку довжину, щоб за  $k$  кроків подолати відстань від стану **A** до стану **B**. Тепер проміжні стани визначаються як:

$$p_j^i = p_{j-1}^i + \Delta p_i, i = 1 \dots N, j = 1 \dots k, \text{ де } p_0^i = p_i. \quad (10)$$

**Означення 2.** Зважений *NURBS-морфінг* є механізм, при якому множина керуючих точок, яка визначає кінцевий стан, має ваговий коефіцієнт  $w$  (від 0 до 100%). Тобто, зі стану **A** у стан **B** треба перейти не повністю, а на  $w$  %. Розв'язок задачі зваженого *NURBS-морфінгу* такий самий, як і у випадку *NURBS-морфінгу* але з дещо модифікованими формулами (9):

$$\Delta p_i = \frac{p_i^* - p_i}{k} * \frac{w}{100}, i = 1 \dots N. \quad (11)$$

**Означення 3.** Під *NURBS-блендінгом* розуміється утворення поверхні, як змішування декількох поверхонь. Тобто, з декількох ( $m$ ) поверхонь, означених множинами керуючих точок  $p_i^l = [x_i^l, y_i^l, z_i^l]^T, i = 1, \dots, N, l = 1, \dots, m$  (причому одна з поверхонь (нехай  $i=1$ ) є домінуючою, тобто, вклад

точок цієї поверхні у результуючу поверхню - домінуючий) створюється одна, результуюча поверхня, яка визначається множиною керуючих точок  $p_i^{blend} = [x_i^{blend}, y_i^{blend}, z_i^{blend}]^T$ . Для цього пропонується наступний алгоритм:

1. Якщо відповідні точки поверхонь ( $i$ -тий індекс – фіксований) мають однакові координати, то результуюча поверхня має  $i$ -ту точку з такими ж координатами.
2. В іншому випадку (тобто коли на основі множини точок з різними координатами, потрібно розрахувати точку результуючої поверхні), пропонується наступне:

- з'єднати відрізками точку домінуючої поверхні (домінуюча точка 1-го кроку) з відповідними точками інших поверхонь (якщо їх координати не рівні);
- на наступному кроці, замість цих точок взяти множину нових точок, котрі визначено, як середини відрізків, означених у попередньому пункті (розмірність цієї множини буде на одиницю менша);
- продовжувати цей процес (за домінуючу точку вибирати ту точку, відстань від якої до домінуючої точки 1-го кроку найменша) до тих пір, поки не залишиться одна точка – вона й буде відповідною точкою результуючої поверхні.

На Рис. 4. показано таблицю послідовностей кадрів візуалізації промовляння фонему **A**, при умові, що перед нею була пауза. Фонема промовляється з емоційним станом усміненого обличчя.



Рис. 4. Послідовність кадрів візуалізації промовляння фонему **A** (з емоційним забарвленням "усміненого обличчя")

### 3. Нанесення текстур на NURBS-поверхню голови людини

Взагалі кажучи, нанесення текстур на NURBS-поверхню, у загальному випадку, є досить складною задачею [13]. Проблема полягає у тому, що текстура накладається на вершини полігонів, у яких прораховані нормалі для правильного освітлення. А для NURBS-поверхні таких вершин немає – вони прораховуються при рендерінгу.

У даній роботі пропонується для рендерінгу NURBS-поверхні використовувати, так звану, рівномірну мозаїку (uniform tessellation). У цьому випадку, у порції поверхні, утвореної між кожними чотирма керуючими точками NURBS-поверхні, є однакова кількість вершин полігонів для рендерінгу.

Звідси, алгоритм нанесення текстур на таку поверхню буде наступним:

- 1) у пакеті трьохмірного моделювання з базової NURBS-поверхні голови (з віднормованими

координатами) утворюється полігональна поверхня з “uniform tessellation”;

2) наноситься на неї текстура з віднормованої фотографії;

3) запам'ятовуються координати вершин поверхні та віднормовані текстурні координати.

Зауважимо, що при анімації, нанесена таким чином текстура буде також “людиноподібною”, тобто буде синхронно мінятися, тому що, як зазначалося, при “uniform tessellation” кількість відповідних точок вершин полігонів між порціями поверхні, утвореними кожними чотирма керуючими точками, залишиться тією ж. На Рис. 5. наведені кадри анімації NURBS-моделі з нанесеною на неї текстурою.



Рис. 5. Кадри анімації NURBS-моделі з нанесеною на неї текстурою.

## Висновки

Програмна реалізація запропонованої NURBS-технології підтвердила її життєздатність на звичайних мультимедійних комп'ютерах. Результати тестування показали можливість імплементації запропонованої технології у програмне забезпечення операційних систем (чи надбудов над операційними системами) для організації нових типів інтерфейсів користувача для спілкування з комп'ютером (тобто інтерфейсів у яких спілкування користувача з комп'ютером та комп'ютера з користувачем відбувається з допомогою звичайної людської мови). Важливою перевагою запропонованої технології є те, що не потрібно зберігати великі бази даних зі станами моделей для візуалізації промовляння, тому що запропонований механізм вирішення цієї проблеми дозволяє динамічно створювати модель з потрібною мімікою на основі невеликого вектора параметрів NURBS-поверхні. Для тестування використовувалась модель, яка генерувалась з допомогою 750-ти керуючих точок. Для моделювання артикуляції та емоційного забарвлення використовувались варіації лише двох десятків (із

750) точок. Процес анімації між генерованими станами проходить за допомогою лінійної інтерполяції між керуючими точками, що також є перевагою цієї технології.

## Література

1. Крак Ю.В., Бармак О.В. Анімація віртуальних образів людського обличчя при синтезі мовлення // Искусственный интеллект. – 2002. – № 4. – с.458-462.
2. Крак Ю.В., Бармак О.В. Анімація віртуальних образів людського обличчя при синтезі мовлення // Искусственный интеллект–2002 //Материалы Международной научно-технической конференции. Таганрог: Изд-во ТРТУ, 2002. Том 2. с.138-142.
3. Крак Ю.В., Кириченко М.Ф., Бармак О.В. Моделювання візуальних образів людського обличчя в процесі мовлення // Вісник Київського університету. Серія: фіз.-мат. науки. – 2002. – 2. – с. 228-232.
4. Крак Ю.В., Вінцюк Т.К., Кириченко М.Ф., Гаращенко Ф.Г., Бармак О.В. Розробка комп'ютерних технологій моделювання та керування візуальними образами людського обличчя при синтезі мовлення // Шоста всеукраїнська міжнародна конференція “Оброблення сигналів і зображень та розпізнавання образів” УКРОБРАЗ’2002, 8-12 жовтня 2002 р., Київ, Україна. Праці.: Видання УАОІРО, Київ 2002. ст. 23-26
5. Бармак А.В. Информационная компьютерная технология для моделирования и управления манипуляционными и виртуальными системами //Проблемы управления и информатики. – 2003. №5. – с.143-154
6. Piegl, Les and Tiller, Wayne. The NURBS Book, 2nd Edition, – Berlin, Germany: Springer-Verlag, 1996.
7. C. deBoor, "A Practical Guide to Splines", 1978, New York, Springer-Verlag
8. James D. Foley et al., "Introduction to Computer Graphics", 1994, Addison-Wesley
9. Les Piegl "On NURBS: A Survey", Jan 01, 1991, IEEE Computer Graphics and Applications, Vol. 11, No. 1, pp. 55 - 71
10. David F. Rogers, Rae A. Earnshaw (editors), "State of the Art in Computer Graphics - Visualization and Modeling", 1991, New York, Springer-Verlag, pp. 225 - 269
11. Alan Watt, Mark Watt, "Advanced Animation and Rendering Techniques", 1992, New York, AMC press, Addison-Wesley
12. Эйнджел Э. Интерактивная компьютерная графика. – М.: Вильямс, 2001
13. Macri D. Using NURBS Surfaces in Real-time Applications. <http://www.gamasutra.com/features>, 2001
14. Foley J.D., van Dam A., Feiner S.K., Hughes J.F. Computer Graphics, Second Edition. – Addison-Wesley, Reading, MA, 1990 (C Version 1996)
15. Фоли Дж., ван Дэм А. Основы интерактивной машинной графики: в 2-х книгах. – М.: Мир, 1985
16. Краснов М.В. OpenGL. Графика в проектах Delphi. – СПб.: БШВ-Петербург, 2001
17. Ratner P. 3-D Human Modeling and Animation. – New York: John Wiley & Sons, 1999.
18. Флеминг Б., Доббс Д. Методы анимации лица. Мимика и артикуляция. Пер. с англ. – М.: ДКМ Пресс, 2002.