4. Соколов В. М. Стандарты в управлении качеством образования / В. М. Соколов. – Н. Новгород: ННГУ, 1993. – 75 с.

5. Поташник М. М. Управление качеством образования / М. М. Поташник. – М.: Педагогическое общество России, 2006. – 448 с.

6. Система управління якістю. Основні положення та словник термінів (ISO 9000:2005, IDT) : ДСТУ ISO 9000:2007. – [Чинний від 2007-09-03]. – К.: Держспоживстандарт України, 2008. – 35 с. – (Національний стандарт України).

7. Система управління якістю. Вимоги (ISO 9001:2008, IDT) : ДСТУ ISO 9000:2009. – [Чинний від 2009-06-22]. – К.: Держспоживстандарт України, 2009. – 34 с. – (Національний стандарт України).

8. Варжина Н. В. Формирование системы управления качеством образовательных услуг: автореф. дис. на соискание уч. степени канд. экон. наук. спец. 08.00.05 – «Экономика и управление народным хозяйством (экономика, организация и управление предприятиями, отраслями, комплексами сферы услуг)» / Н. В. Варжина. – Екатеринбург, 2004. – 19 с.

9. Система управління якістю. Настанови щодо застосування ISO 9001:2000 у сфері освіти (IWA 2:2003, IDT) : ДСТУ-П IWA 2:2007. – [Чинний від 2007-09-03]. – К.: Держспоживстандарт України, 2008. – 70 с. – (Національний стандарт України).

10. Казимир В. В. Модельно-ориентированное управление интеллектуальными производствами и системами: диссертация на соискание ученой степени доктора технических наук: 05.13.06 / Казимир Владимир Викторович. – К., 2005. – 328 с.

УДК 681.324

В.А. Бичко, канд. фіз.-мат. наук
Д.О. Абламський, магістрант
Чернігівський державний технологічний університет, м.Чернігів, Україна

Dr. Ramit Azad, assistant professor
American International University-Bangladesh, Dhaka, Bangladesh

ПЕРЕТВОРЕННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ АКУСТИЧНИХ СИГНАЛІВ У ЗАДАЧУ РОЗПІЗНАВАННЯ ЗОРОВИХ ОБРАЗІВ

Розроблена комп'ютерна система аналізу акустичних сигналів, що дозволяє будувати зоровий образ акустичного потоку інформації. Проведений аналіз параметрів, які впливають на якість первинної обробки акустичного потоку. Реалізована можливість перетворення задачі розпізнавання акустичної інформації на задачу розпізнавання зорових та просторових образів.

Вступ

Останнім часом особливу актуальність набуває розвиток систем розпізнавання акустичних інформаційних потоків та їх синтезу. За допомогою таких систем вирішується широкий спектр завдань, зокрема, голосове управління різноманітними процесами, та наповнення баз даних і баз знань у системах штучного інтелекту.

При створенні таких систем актуальним залишається питання автоматизації процесу інтерпретації мови та пошуку таких форм її відображення, які забезпечували б просте і надійне виділення інформативних ознак сигналу. На ефективність акустичного аналізу впливає як вибір форми відображення акустичного сигналу, з яким працює система автоматичного розпізнавання мови, характер ознак, що використовуються для подальшої фонетичної обробки, а також ступінь надійності методу, який використовується для інтерпретації акустичного потоку інформації.

Існує безліч підходів і методів вирішення задачі розпізнавання мовлення [1; 2; 3]. Процес розпізнавання можна умовно розділити на три етапи: отримання та первинна обробка акустичного сигналу, розпізнавання фонем, слів і розуміння мови. Досить ву-

зьким містком цього ланцюга можна вважати другий етап, оскільки він пов'язаний з наявністю широкого різноманіття особистісних фізичних характеристик мови людини. З іншого боку, існує багато методів розпізнавання зорових образів з розвинутою методологією класифікації [4; 5]. На нашу думку, використання такої методології може виявитися досить ефективним при вирішенні задач розпізнавання мовлення.

Таким чином, задача цієї роботи полягає у розробці комп'ютерної системи аналізу звуку (КСАЗ), здатної виконувати первинну обробку акустичного сигналу та перетворювати його на зоровий образ.

Найпоширеніший підхід обробки акустичного сигналу базується на його спектральному представленні [6]. При використанні спектрального представлення застосовується дискретне перетворення Фур'є. При цьому акустичний сигнал ділять на вікна, кожне з яких відповідає певній ознаці фонеми. Але такий підхід дозволяє отримати і проаналізувати ознаки лише на окремій ділянці звукового сигналу. Проте, якщо весь фрагмент інтерпретованої фонеми представити графічно – у вигляді часового розгорнення (спектрограми), то це дозволить перейти від завдання розпізнавання акустичних сигналів (ЗРАС) до задачі розпізнавання зорових образів (ЗРЗО). При цьому з'явиться можливість застосовувати для інтерпретації існуючі алгоритми розпізнавання графічних образів [7; 8]. З нашої точки зору, такий підхід виглядає досить перспективним.

Методика отримання тривимірної моделі

В основу реалізації методу перетворення задачі розпізнавання звукової інформації на задачу розпізнавання зорових образів покладена можливість отримання послідовності елементів масиву, кожен з яких являє собою набір значень частотного спектру фонеми (рисунок 1). Упорядкований набір таких елементів утворює поверхню, яка є унікальним ідентифікатором фонеми. Фізично така поверхня залежить від амплітуди звукового сигналу двох параметрів: частоти та часу.

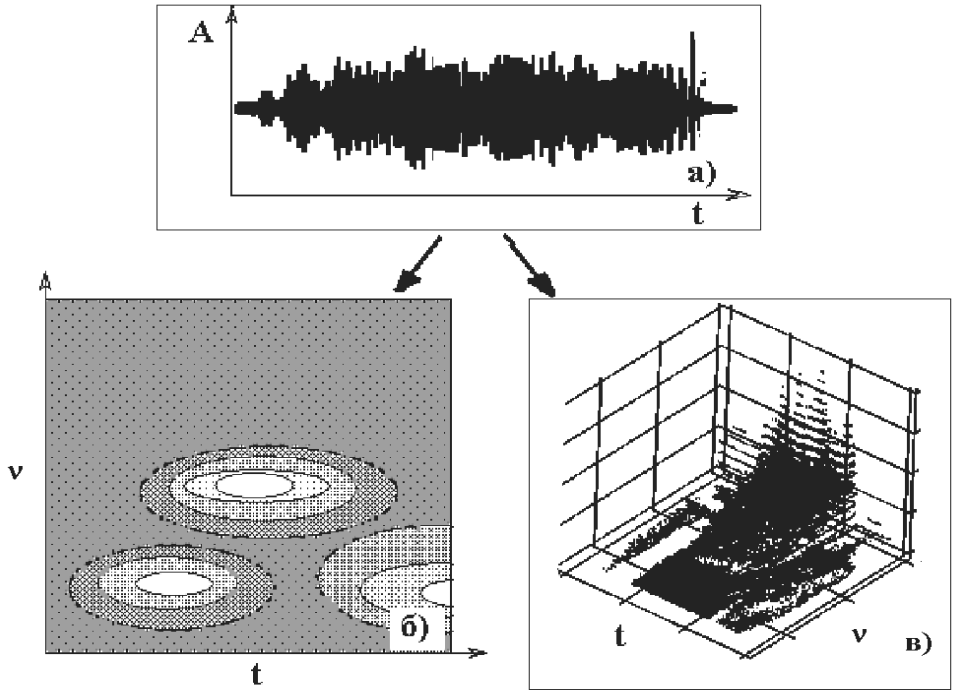

Рис. 1. Схема перетворення даних системою а – вхідний відбиток акустичного потоку; б – двовимірний образ фонеми; в – тривимірний образ фонеми

Залежно від подальшої необхідності, отриману поверхню можна використовувати або як двовимірний (рис. 1, *б*), або як тривимірний (рис. 1, *в*) унікальний образ послідовності фонем акустичного потоку інформації (рис. 1). У разі застосування її як зорового образу,

поверхню можна представити як плоский малюнок, на якому інтенсивність кольору на певній ділянці буде пропорційна відповідному значенню амплітуди акустичного сигналу (рис. 1, *б*).

На рисунку 2 зображений процес послідовності обробки акустичного потоку інформації аналітичною системою.

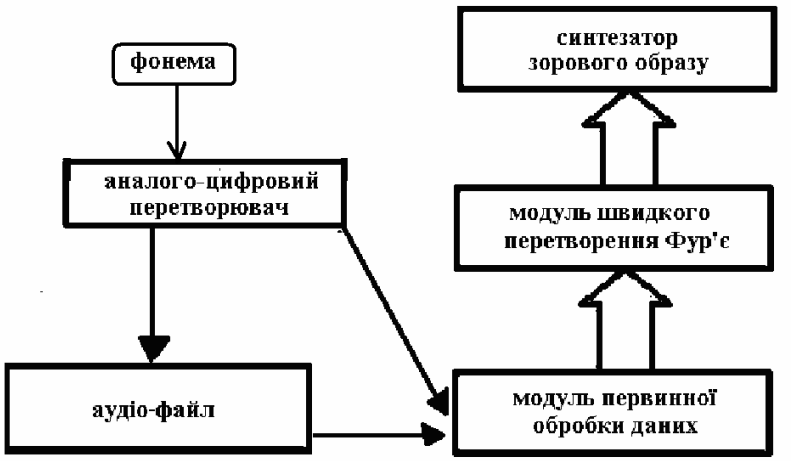

Рис. 2. Схема отримання та перетворення даних системою

Виходячи з вищесказаного, аналітична система повинна забезпечити можливість виконання наступних дій:

1) отримання звукової інформації;
2) попередню обробку звуку;
3) перетворення первинних даних у двовимірний масив даних спектрограми;
4) відтворення графічного зображення спектрограм;
5) зберігання кінцевих результатів у файл.

У розробленій КСАЗ за допомогою мікрофону виконується перетворення звукових коливань в електричні коливання, які надалі можуть бути посилені, відфільтровані від перешкод, а потім переведені у цифровий формат для подальшої обробки.

Для зберігання первинних даних необхідно було вибрати придатний формат звукового файлу. Відповідні формати файлів, що існують на цей час, можна поділити на дві групи, які відрізняються методом стиснення даних, а саме: з втратами та без втрат даних.

MP3 є одним з найпоширеніших форматів. У форматі MP3 використовується алгоритм стиснення з втратами, розроблений для істотного зменшення розміру даних. Принцип стиснення полягає в зниженні точності частин звукового потоку, що практично не впливає на якість сприйняття більшості людей [9]. Цей метод називають кодуванням сприйняття.

WAV – формат файлу-контейнера для зберігання цифрового аудіопотоку. Цей формат найчастіше використовується як оболонка для нестисненого звуку (PCM), коли для кожного значення амплітуди сигналу виділяється певне число біт [9]. Проте в контейнер WAV можна помістити звук, стиснений майже будь-яким кодеком.

У КСАЗ використовується WAV-формат, оскільки він більш простий у використанні і дозволяє зберігати дані без втрат.

При виборі несучої частоти аналого-цифрового перетворення, згідно з теоремою Котельникова [6], необхідно врахувати, що така частота повинна бути вдвічі вище за максимальну частоту спектра перетворюваного сигналу. У різних людей межі діапазону чутних частот можуть відрізнятися. Проте вважається, що частота звукових коливань, які здатна відчувати людина, не перевищує 20000 Гц [9]. Отже, для аналого-цифрового перетворення без втрати якості звукового сигналу потрібно вибрати частоту перетворення, не меншу ніж 40000 Гц.

Наступним етапом після обробки сигналу є попередній аналіз даних. Результатом аналізу сигналу є послідовність мовних кадрів. Звичайно, кожен мовний кадр – це результат аналізу сигналу на невеликому відрізку часу (порядку 10 мс). Вибір оптимального значення розміру вікна було здійснено емпіричним шляхом.

При такій обробці щодо послідовності потоку даних переміщується вікно відбору, розмір якого дорівнює N елементів цифрової послідовності даних (рисунок 3). Значення N відповідає розміру вікна дискретного перетворення Фур'є (ДПФ). Під час аналізу вікно переміщується на ΔN позицій відносно попереднього положення. У кожному новому положенні воно заповнюється новим набором даних звукового сигналу.

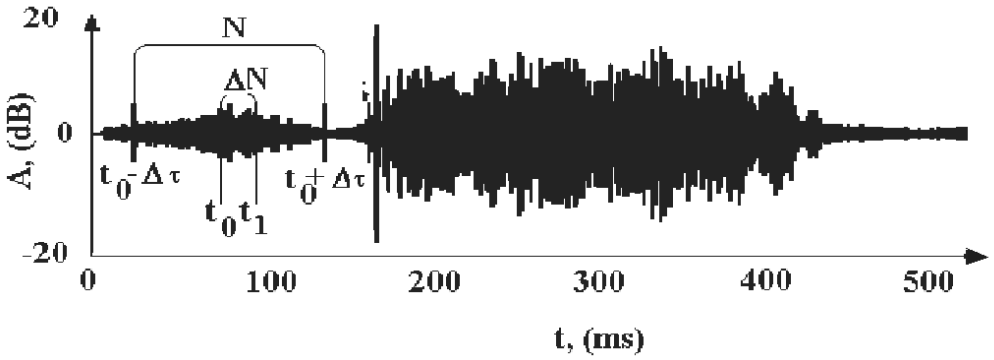

Рис. 3. Отримання сегмента даних

На практиці ДПФ обчислюється за формулою 1:

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn}, \qquad (1)$$

де X_k – амплітуди частот коливань.

При підрахунках інформація про фазу акустичного сигналу відкидається, оскільки при вирішенні цієї проблеми вона є надлишковою. З метою пришвидшення обчислень у КСАЗ було використане перетворення за алгоритмом швидкого перетворення Фур'є.

У результаті масштабування первинних даних та ДПФ відтворюється крива спектра, яка відображає залежність амплітуди звукової хвилі від частоти у момент часу t_0 (рис. 4).

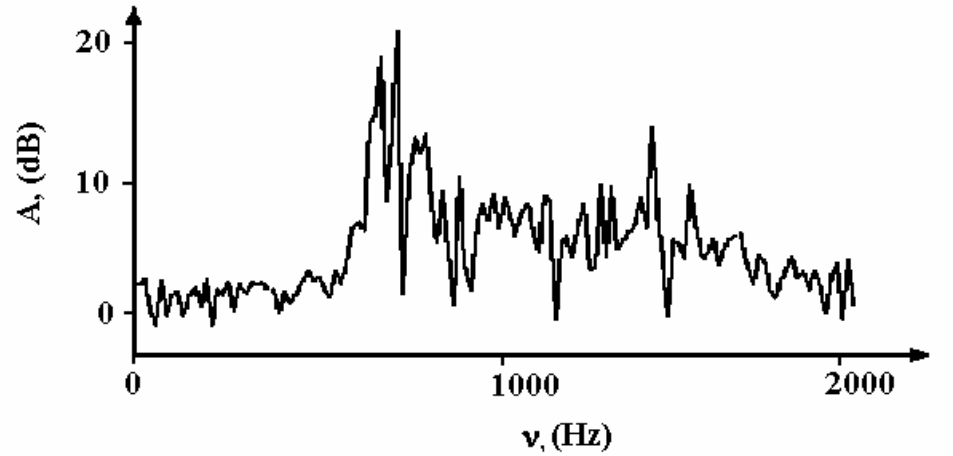

Рис. 4. Результат ДПФ звукового сигналу для певного моменту часу

Після виконання перетворень над послідовністю сегментів звукового сигналу з набору спектральних кривих утворюється часова розгортка – спектрограма, що являє собою у загальному вигляді нерегулярну поверхню (рис. 5).

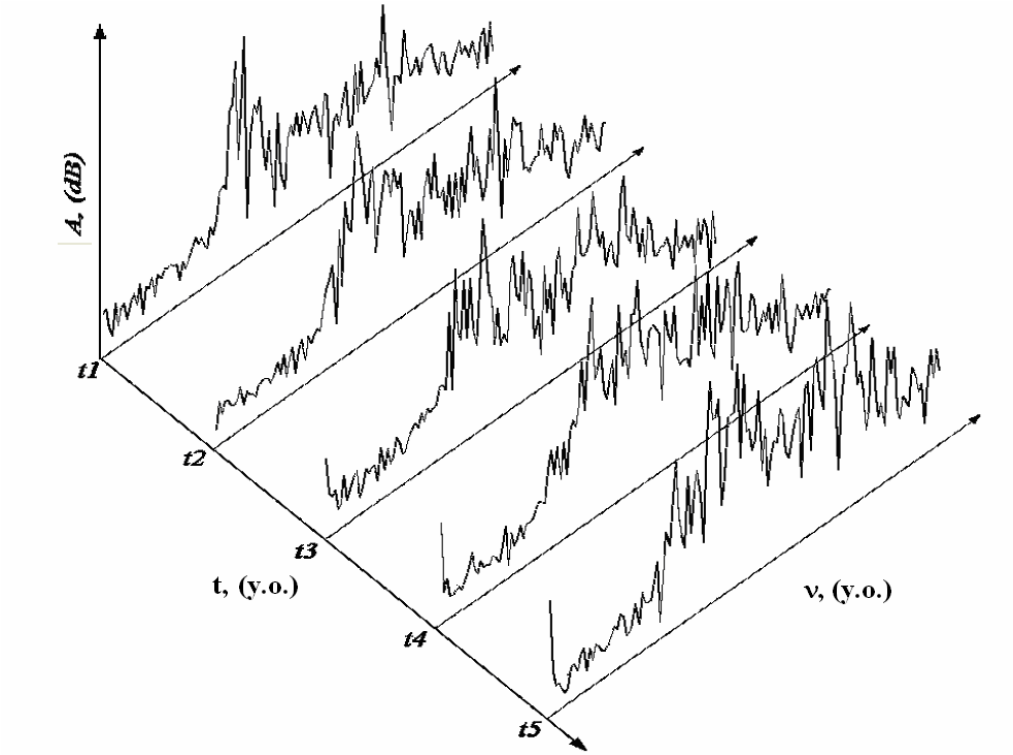

Рис. 5. Об'єднання сегментів даних БПФ у спектральну поверхню зорового образу

Результати та обговорення

У ході роботи були отримані різні варіанти спектрограми фонемних послідовностей з метою визначення оптимальних параметрів, необхідних для представлення та зберігання таких спектрограм. Приклади отриманих двовимірних та трьохвимірних зображень спектрограм представлені на рисунках 6 та 7 відповідно. При двомірному відтворенні, з міркувань доцільності, була реалізовано залежність, в якій значення амплітуди сигналу пропорційне інтенсивності світлого кольору на відповідній ділянці поверхні спектрограми (рис. 6). Вочевидь, що темним кольором позначені місця з низьким рівнем амплітуди.

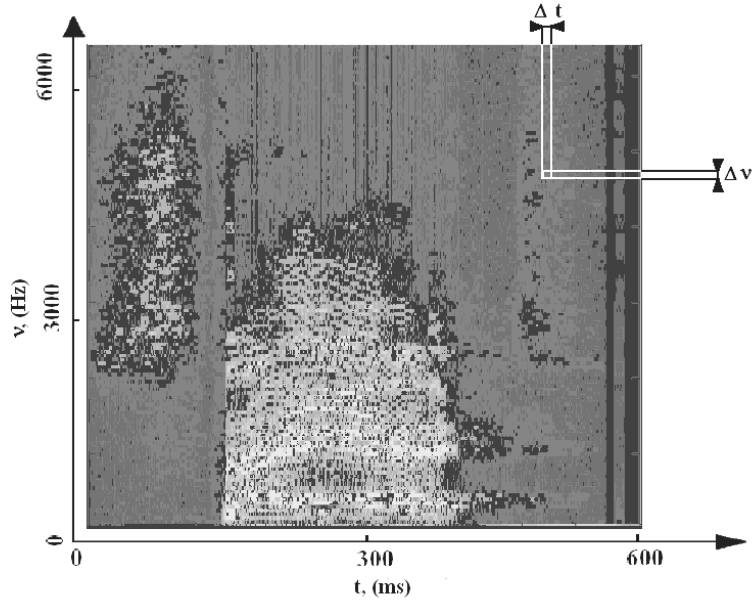

Рис. 6. Двовимірне відображення спектрограми звукового сигналу, що відповідає фонемі слова „стій"

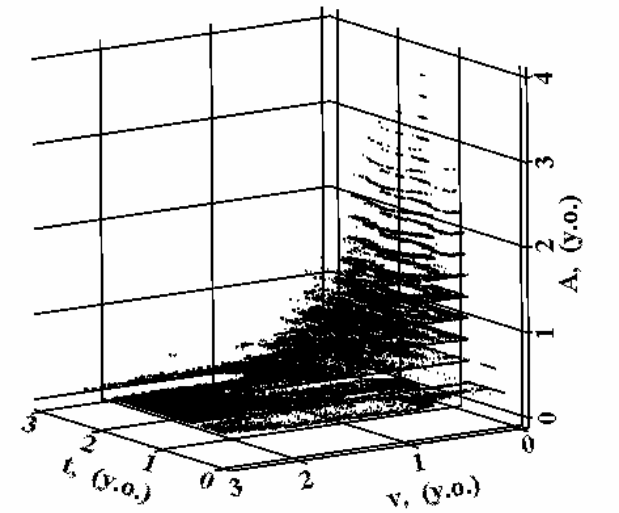

Рис. 7. Тривимірне відображення спектрограми звукового сигналу, що відповідає фонемі слова „стій”

Для визначення оптимального формату представлення спектрограм було проаналізовано наступні параметри: крок дискретизації частоти Δν спектрограми, розмір кроку зміщення Δt вхідного вікна для ДПФ (рис. 6). Слід нагадати, що значення цих двох параметрів визначають розмір елементарної ділянки спектрограми, та, відповідно, обумовлюють роздільну здатність спектрограми.

Вочевидь, більша роздільна здатність точніше відображає інформаційне поле. Проте надмірно часті вимірювання можуть привести до невиправданого зростання кількості даних і даремної витрати апаратних ресурсів при їх обробці.

Розрахунок оптимального кроку дискретизації частоти Δν обумовлений наступними міркуваннями. Дані про частотні характеристики акустичного сигналу, який здатна відчувати людина, утворені різними джерелами, відрізняються між собою. Так, самим широким можна вважати діапазон 16-20000 Гц. Але акустична частота людської мови лежить у діапазоні частот 150-4000 Гц [2].

Крім того, дослідження мовних фонем людини показали, що для них інформативність різних частин спектра неоднакова. А саме у низькочастотній області спектра міститься більше інформації, ніж у високочастотній області. Тому для економнішого використовування обчислювальних ресурсів і збільшення продуктивності роботи системи можливо частково зменшити кількість даних з високочастотної області спектра. Для цього у даній системі застосований найпоширеніший і найпростіший метод — логарифмічне стиснення, або mel-стиснення.

Логарифмічне стиснення спектра виконується за формулою:

$$m = a \times \log(f \times b + 1), \tag{2}$$

де f – частота в спектрі; m – частота в новому стислому частотному просторі a,b – параметри стиснення a і b.

Враховуючи вищесказане, для оцінки значень частотного кроку дискретизації Δν було розраховано параметри стиснення a і b для двох крайніх діапазонів 10-44000 Гц [9] та 150-4000 Гц [2]. Параметри a,b визначаються з системи рівнянь. Приклад такого розрахунку для ширшого діапазону наведено нижче (формули 3,4):

$$\begin{cases} a \times \log(10\,b + 1) = 1 \\ a \times \log(44000\;b + 1) = 255 \end{cases} \tag{3}$$

Розв’язок цієї системи рівнянь дає значення відповідно: a = 135,385, b = 0,001715:

$$m = 135 \times \log(0{,}001715 \times f + 1) \tag{4}$$

Як видно з рисунку 8, використання вужчого частотного діапазону дає можливість суттєво зменшити розширення інтервалу чутливості при використанні логарифмічного стиснення. У цьому разі роздільна здатність на всьому діапазоні не буде перевищувати 20 Гц, що є співрозмірним з роздільною здатністю при ідентифікації акустичного сигналу людиною.

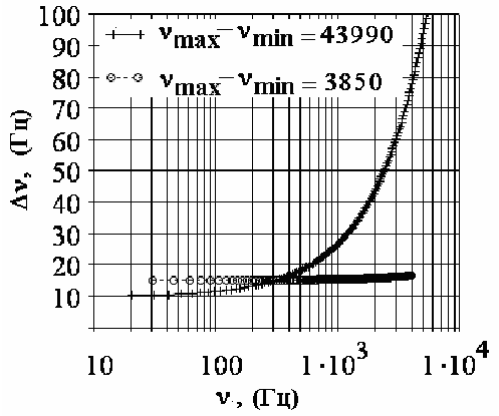

Рис. 8. Залежності роздільної здатності частоти спектрограми від інтервалу частотного діапазону спектрограми при фіксованому значенні N=256частотних ділянок спектрограми

Визначення оптимального часового кроку Δt спектрограми на міркуваннях, що максимально можлива роздільна частота сприйняття головного мозку людини зорових зображень не перевищує 30 Гц [2]. Тобто, щоб людина здатна була б ідентифікувати та запам'ятати акустичний сигнал як окремий, він повинен тривати не менш 25 мс з запасом для уникнення помилок [2].

На рисунку 9 наведена залежність розміру файлу даних спектрограми фіксованої тривалості (0,7 с) від М – кількості ділянок за частотою та тривалістю часового кроку Δt. З рисунку видно, що розмір файлу має мультиплікативну залежність від роздільності спектрограми за часом та частотою. Вочевидь, така залежність суттєво впливає на кінцевий розмір файлу даних спектрограми. Такі параметри відіграють суттєву роль при аналізі довготривалих потоків акустичної інформації. Саме тому визначенню оптимальних параметрів дискретизації спектрограми теж слід приділяти належну увагу. В попередніх розрахунках цієї роботи було використано значення М=256.

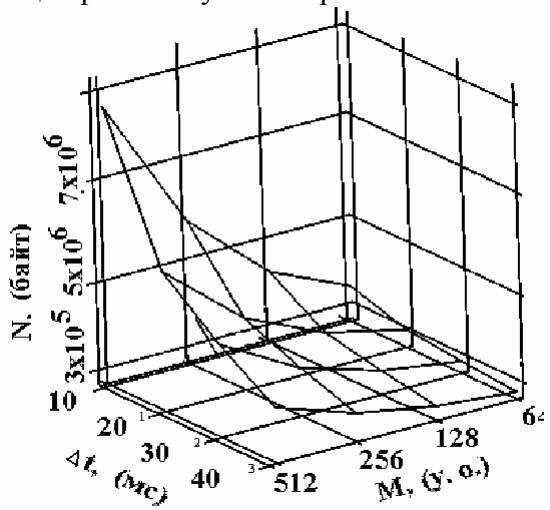

Рис. 9. Залежність розміру файлу даних (тривалість 0,7 с) від розміру часової та частотної дискретизації

У ході проведеного аналізу даних слід зазначити, що використаний метод трансформації задачі розпізнавання акустичного сигналу в задачу розпізнавання зорових образів є перспективним з наступних міркувань:

- по-перше: при такому перетворенні не відбувається невідворотна втрата інформації, яка має місце при використанні інших методів відображення акустичного сигналу [10]. Навпаки, використовуючи обернене перетворення Фур'є, можливо поновити початкову фонему. При цьому ступінь відповідності оригіналу є підконтрольним процесом;
- по-друге: подальше розв'язання проблеми ідентифікації морфем можливо розглядати у рамках напрацьованих розробок та розроблених алгоритмів у галузі ЗРЗО [4; 5; 7; 8].

Висновки

Спираючись на отримані результати, можна зробити висновок, що в ході виконання роботи вдалося перетворити задачу розпізнання звукової інформації на задачу розпізнання зорових образів. Наступним кроком розв'язання цієї проблеми буде створення класифікатора зорових образів, що відповідають мовним фонемам.

Подальші кроки розвитку запропонованого підходу передбачають розробку ефективних алгоритмів часового та частотного нормування спектрограми з метою створення бібліотеки еталонних морфем [11]. Необхідність таких перетворень пов'язана як з наявністю різного темпу мовлення, так і з індивідуальними особливостями тональності людини. Ефективним інструментом для реалізації таких алгоритмів є використання нейронних мереж.

Список використаних джерел

1. Lawrence R. Rabiner, Biing-Hwang Juang / Fundamentals of Speech Recognition. – Prentice Hall, 1993. – 496 p.
2. Вишнякова О. А. Автоматическая сегментация речевого сигнала на базе дискретного вейвлет-преобразования / О. А. Вишнякова, Д. Н. Лавров //Математические структуры и моделирование. – 2011. – Вып.23. – С.43-48.
3. Ермоленко Т. В. Разработка системы распознавания изолированных слов русского языка на основе вейвлет-анализа / Т. В. Ермоленко // Искусственный интеллект. – 2005. – № 4. – С. 595-601.
4. F.R. Bach, M.I. Jordan, Kernel Independent Component Analysis, Journal of Machine Learning Research, Vol. 3, 2002, pp. 1-48.
5. M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neurosicence, Vol. 3, No. 1, 1991, pp. 71-86.
6. Сергиенко А. Б. Цифровая обработка сигналов / А. Б. Сергиенко – СПб.: Питер, 2002. – 608 с.
7. K. Etemad, R. Chellappa, Discriminant Analysis for Recognition of Human Face Images, Journal of the Optical Society of America A, Vol. 14, No. 8, August 1997, pp. 1724-1733.
8. L. Wiskott, J.-M. Fellous, N. Krueuger, C. von der Malsburg, Face Recognition by Elastic Bunch Graph Matching, Chapter 11 in Intelligent Biometric Techniques in Fingerprint and Face Recognition, eds. L.C. Jain et al., CRC Press, 1999, pp. 355-396.
9. Фролов А. В. Мультимедиа для Windows. Библиотека системного программиста. Т. 15. / А. В. Фролов, Г. В. Фролов. – М.: Диалог-МИФИ, 1994.
10. Данченко О. И. Использование динамических портретов звука при распознавании речевого сигнала / О. И. Данченко, Д. В. Николаенко // Искусственный интеллект. – 2008. – № 1. – С. 139-144.
11. Ермоленко Т. В. Методика формирования эталонов фонем, базирующаяся на вейвлет-преобразовании Морле / Т. В. Ермоленко // Таврический вестник информатики и математики. – 2006. – № 1. – С. 127-132.
12. Грабовая В. А. О системе компьютерного распознавания русской речи с автоматическим построением эталонов / В. А. Грабовая, Е. Е. Федоров, В. Ю. Шелепов // Искусственный интеллект. – 2000. – № 1. – С. 76-81.
13. Круглов В. В. Искусственные нейронные сети. Теория и практика / В. В. Круглов, В. В. Борисов. – М.: Горячая линия-Телеком, 2002.