



Автоматичне розпізнавання, розуміння та синтез сигналів мовлення

Automatic Recognition, Understanding and Synthesis of Speech Signals

Використання ефекту згладжування у слуховій системі
для дискретного подання слова

Михайло Бондаренко, Валентина Чикіна

Харківський інститут радіоелектроніки

Україна, 310726 Харків
просп. Леніна, 14
Тел.: (057) 240-94-46

Досліджується зв'язок між неперервним поданням слова у вигляді фонограми звукового тиску та дискретним поданням слова (цифровим кодом). Відомо, що орган слуху має скінченну чутливість та скінченну смугу пропускання частот, тому він доставляє у відповідний відділ мозку лише скінченну інформацію. Це відкриває принципову можливість побудови скінченного дискретного образу звуку, який не відрізняється на слух від неперервного образу. Вибір вдалого засобу дискретизації звуку дозволяє у науковому плані просунутись у дослідженні проблеми аналізу і синтезу мовлення, а в технічному — удосконалювати акустичну апаратуру.

Первинна дискретизація фонограми заснована на роботах по ефекту згладжування у слуховому апараті, які проводились під керівництвом Ю.П. Шабанова-Кушнарєнка. Ефект згладжування формулюється таким чином. Нехай маємо звук α тривалістю T_0 , що має неперервну фонограму $A(t)$ на інтервалі $[0, T_0]$, і сім'ю звуків β_ϵ з фонограмами $B_\epsilon(t)$ на тому ж інтервалі $[0, T_0]$. Кожна фонограма $B_\epsilon(t)$ представляє собою константу, на якій розташовані вузькі, однакової форми і величини, прямокутні імпульси, площа під кожним з них дорівнює ϵ . Ефект згладжування полягає в тому, що якщо для будь-яких t_1 і t_2 ($t_2 > t_1$, $t_1, t_2 \in [0, T_0]$) виконується умова:

$$\lim_{\epsilon \rightarrow 0} \int_{t_1}^{t_2} B_\epsilon(t) dt = \int_{t_1}^{t_2} A(t) dt,$$

то існує таке значення ϵ_0 , що для будь-якого $\epsilon < \epsilon_0$ звук α не відрізняється на слух від звуку β_ϵ .

Перехід від звуку α з фонограмою $A(t)$ до звуку β_ϵ з фонограмою $B_\epsilon(t)$ (що еквівалентно $B_\omega(t)$, де ω — частота імпульсів) є важливим кроком у дискретизації неперервного акустичного коливання, і його подання B_ω є частотно-імпульсним кодом звуку. Однак цей же код ще не подається числовим кодом, оскільки моменти часу Θ_i , коли з'являються імпульси — довільні дійсні числа. Щоб завершити дискретизацію звуку, треба синхронізувати імпульси, тобто ввести дискретний час і таким чином затримати кожен імпульс, щоб його фронт попав на найближчий момент дискретного часу.

Сформулюємо загальне твердження. Нехай $B(t)$ — діаграма довільного звуку та $\{C_\psi(t)\}_\psi \in (0, \infty)$ — однопараметрична сім'я діаграм довільних звуків. Про функції $B(t)$ і $C_\psi(t)$ припускається лише те, що вони локально сумуються на інтервалі визначення $[0, T_0]$.

Якщо звук $B(t)$ та більшість звуків $\{C_\psi(t)\}_\psi \in (0, \infty)$ задовільняють умові

$$\lim_{\psi \rightarrow \infty} \int_0^{T_0} |C_\psi(t) - B(t)| dt = 0,$$

то існує таке число $\psi_0 > 0$, що для будь-якого $\psi > \psi_0$ відчуття звуку $C_\psi(t)$ збігається з відчуттям звуку $B(t)$. Назвемо дане твердження законом існування порогу. Зміст його полягає у тому, що будь-які два звуки, фонограми яких досить мало відрізняються одна від одної по формі, будуть невідмінні на слух.

Для проведення досліджень з експериментального вивчення критичних параметрів ω і ψ була сконструйована та виготовлена спеціальна експериментальна установка, названа квантувачем звукового сигналу.

Результати дослідів показали, що критична частота практично не залежить від виду звуку $A(t)$, на неї мало впливає вибір аудитора. Вона навіть точно збігається з максимальною частотою синусоїдального сигналу, який сприймає вухо, та складає в дослідях величину 16–20 кГц. Критична частота у різних дослідях суворо корелюється з верхнім порогом частоти, що сприймається. Як тільки частота надходження імпульсів у сигналі $B(t)$ знижується настільки, що попадає у чужий діапазон, вона починає сприйматися вухом. Підтвердженням може служити той факт, що у випадку, коли мінімальна частота надходження імпульсів у сигналі $B_\omega(t)$ хоча б на невеликих інтервалах часу знижується до величини $\omega_{кр}$, і звуки $A(t)$ і $B_\omega(t)$ стають відмінними на слух, то ця відмінність з'являється лише у тому, що у відчутті звуку $B_\omega(t)$ на фоні звука $A(t)$ з'являється високочастотний писк.

Були проведені досліді по встановленню змін критичного тактового інтервалу $T_{кр}$ у залежності від параметрів T_{max} і T_{min} максимального та мінімального інтервалу між сусідніми імпульсами у синусоїдальному сигналі.

Різниця між відчуттям звуків збільшується при зменшенні величини $T_{кр}$, а при її зростанні — зменшується. Різниця у звучанні виявляється наявністю у звукові $C_\psi(t)$ шумів того чи іншого виду: виск, шурхіт, потріскування, постукування. Висота, сила шуму, а також його тембр змінюються при зміні значень параметрів T_{max} , T_{min} , $T_{кр}$. При наближенні до критичної точки шум у звукові $C_\psi(t)$ поступово слабшає, а при переході через неї зникає зовсім. При нерівності відчуттів звук $C_\psi(t)$ сприймається по силі, висоті та тембру точно таким же, як і звук $B(t)$. Різниця відчуттів полягає в тому, що звук $B(t)$ відчувається як мелодичний, а звук $C_\psi(t)$, будучи по суті тим же самим звуком, забрудненим шумом. При великому рівні шуму звук $B(t)$, як компонент відчуття звуку $C_\psi(t)$, заглушується.

Досліді показали, що при збільшенні амплітуди коливань дискретизованого синусоїдального сигналу чистота його звучання погіршується. При підвищенні середньої частоти надходження імпульсів у сигналі $B(t)$ якість звуку погіршується. Таким чином, чим рідше надходять імпульси у коді звуку, тим кращою буде якість його звучання. За збільшення амплітуди коливань звукового сигналу на вході порогового елемента доводиться розплачуватися дорогою ціною: у стільки ж разів треба підвищити частоту синхронізації, якщо необхідно отримати незашумлене звучання сигналу $A(t)$.

Щоб переконатися у загальності отриманих закономірностей, були проведені вибіркові дослідження, в яких у ролі звуку $A(t)$ вживалися синусоїдальні сигнали різної частоти, дикторська мова та музичні твори. Результати дослідів були однакові: при не дуже високій частоті синхронізації сигнал $C_{\psi}(t)$ сприймається як сигнал $A(t)$, який звучить на фоні шуму. Збільшення частоти синхронізації призводить до зникнення шуму, після чого звук $C_{\psi}(t)$ не відрізняється на слух від звуку $A(t)$. Критична частота синхронізації цілком визначається мінімальним інтервалом між імпульсами у сигналі $B(t)$. У процесі проведення дослідів варіювалась також ширина імпульсів, що складала звуки $B(t)$ і $C_{\psi}(t)$, у межах від 0,9 до 6 мкс. Було встановлено, що помітного впливу на величину критичної частоти синхронізації зміна ширини імпульсів не має.



Дослідження з розпізнавання ключових слів у потоці зв'язного мовлення

Тарас Вінцюк, Святослав Біднюк, Анатолій Куляс, Валерій Пилипенко

Інститут кібернетики АН України

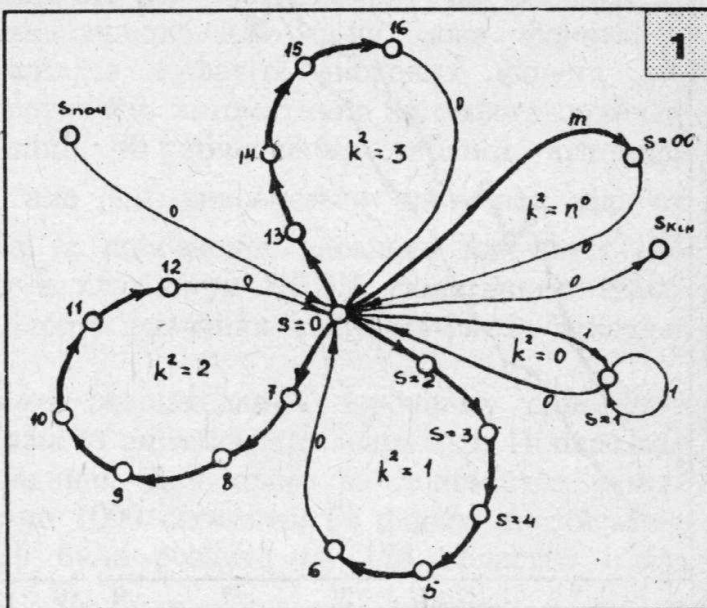
Україна, 252207 Київ
просп. Академіка Глушкова, 40
Тел.: (044) 266-43-56

E-mail address: vintsiuk@c.golos.kiev.ua.@relay.ussr.eu.net

Задача розпізнавання окремих (ключових) слів у потоці зв'язного мовлення все більше привертає увагу дослідників [1, 2], і це обумовлено насамперед тим, що в реальних умовах взаємодії оператора та ЕОМ обмеження словника вживаних слів завдає оператору великих незручностей. В той же час наявність автоматичного розпізнавання тільки окремих слів та словосполучень, що визначають зміст висловлювань, розковує оператора та надає йому можливість вільно будувати фрази, користуючись усім словниковим запасом мови. Окрім того, актуальною є проблема створення спеціальних систем автоматичного реагування на певні слова та словосполучення, наприклад систем автоматизованого управління при аваріях літаків та морських суден.

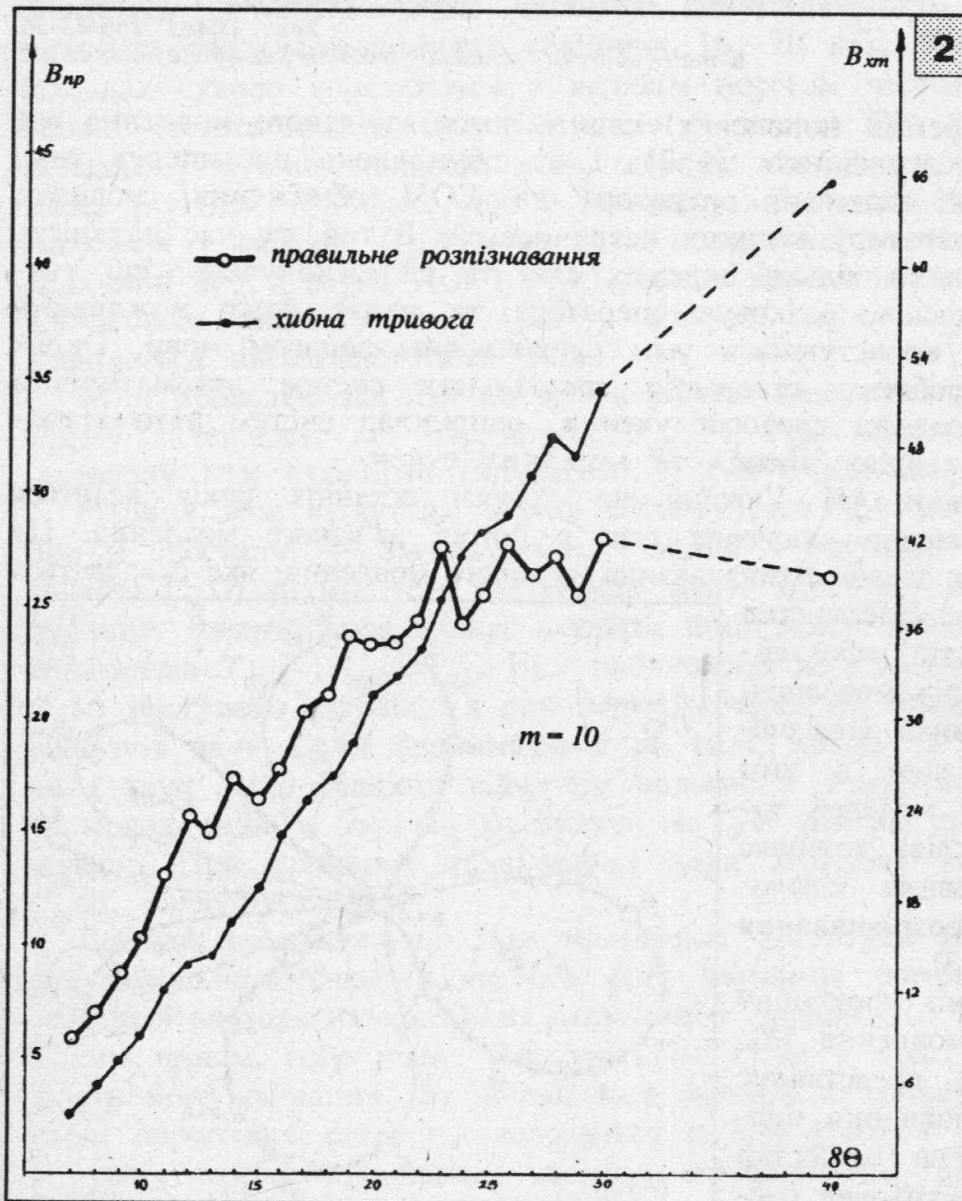
В Інституті кібернетики АН України на протязі останніх років ведуться дослідження по розпізнаванню ключових слів у потоці зв'язного мовлення. Ця задача більш складна, ніж задача розпізнавання зв'язного мовлення, яке складається зі слів певного словника. Насамперед це обумовлено відсутністю ефективних засобів опису тих слів мовлення, що не входять до словника визначених ключових слів. Разом з тим введення досить простих моделей таких слів (нейтральних слів) дозволяє звести задачу розпізнавання ключових слів до задачі розпізнавання зв'язного мовлення [2, 3].

На рис. 1 зображено фрагмент квітки-графу зв'язного мовлення [3], в якій кожна пелюстка представляє одне слово. Суттєве доповнення полягає в тому, що введена пелюстка $k^2 = n^0$, яка моделює нейтральне сло-



во. Усього на рис. 1. зображено п'ять слів: три ключових $k^2 = 1, 2, 3$, нейтральне слово $k^2 = n^0$ та слово-пауза $k^2 = 0$. Кожна пелюстка містить у собі декілька станів s , кожному з яких приписана відповідна фонема $k_s^1 \in K^1$ з множини K^1 усіх фонем. Послідовність фонем у слові $k^2 \in K^2$ (K^2 — множина слів) визначається фонетичною транскрипцією $k^2 = (k_1^1, k_2^1, \dots, k_s^1, \dots, k_{q_k}^1)$. Кожна фонема k_s^1 реалізується послідовністю елементів $(j_{\mu+1}, j_{\mu+2}, \dots, j_\nu)$ — сегментом $J_{\mu\nu}$ в мовному сигналі $J_{0l} = (j_1, j_2, \dots, j_{\mu+1}, j_{\mu+2}, \dots, j_\nu, \dots, j_l)$.

Елементи j_i , послідовність яких описує мовний сигнал, є номерами областей, на які розбито весь багатовимірний простір мовних сигналів [3, 4]. Поява на i -му місці в послідовності J_{0l} , що спостерігається, символу j_i означає, що відповідний багатовимірний вектор ознак x_i , що обчислюється в момент часу i , належить до області простору сигналів, яка має номер j_i . «Відстань» між символом j та фонемою $k^1 \in K^1$ визначається як $g_{jk^1} = -\ln p(j/k^1)$, де $p(j/k^1)$ — ймовірність появи символу j при умові реалізації фонемою k^1 . Розбиття простору мовних сигналів на певну кількість областей відбувається шляхом розв'язування задачі кластеризації великої вибірки мовних сигналів.



При розпізнаванні ключових слів на графі зв'язного мовлення, у відповідності з [2, 3], знаходять таку послідовність фонем і знаходять таке розбиття сигналу J_{0l} на сегменти, що відповідають цій послідовності фонем, щоб забезпечити найменшу відстань між сигналом як послідовністю символів та послідовністю фонем. Та послідовність фонем, що забезпечує найменшу відстань, визначає послідовність з ключових та нейтральних слів у мовному сигналі.

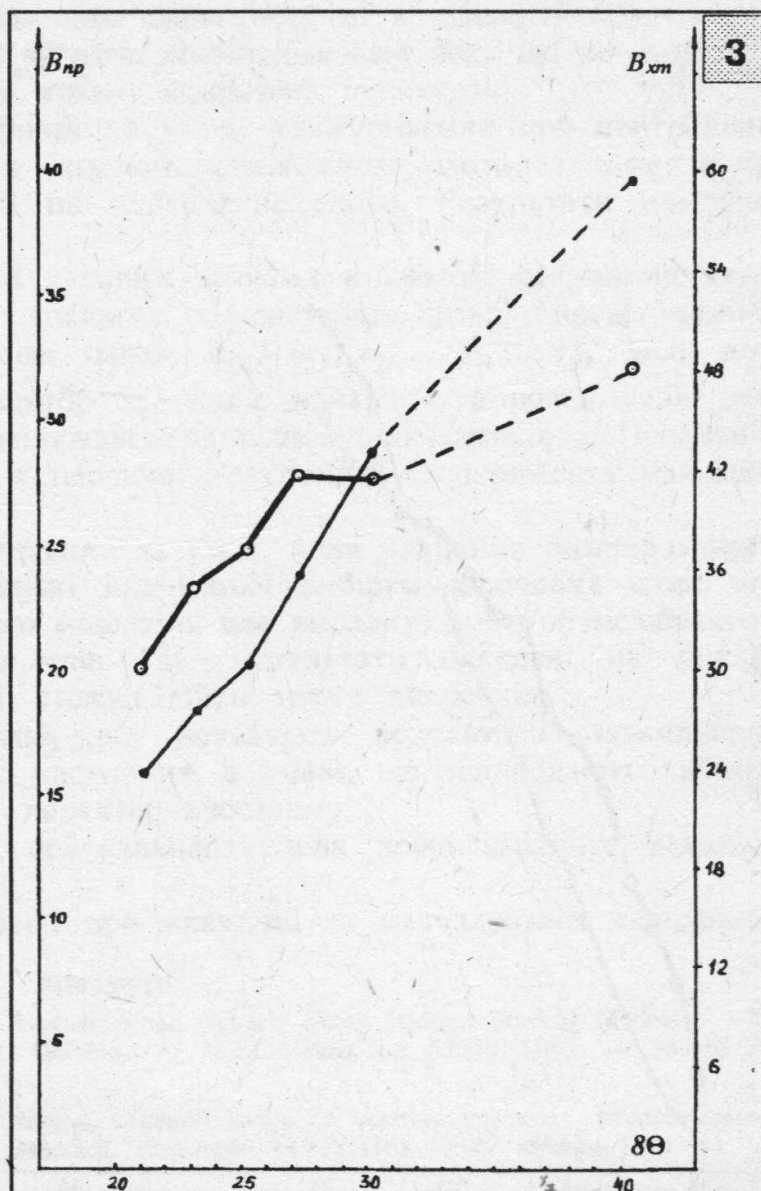
Окремо слід сказати про пелюстку, що відповідає нейтральним словам. Вона містить m станів, у відповідність яким поставлена деяка узагальнена фонема нейтрального слова k_0^1 . Вивчалися два способи визначення відстані

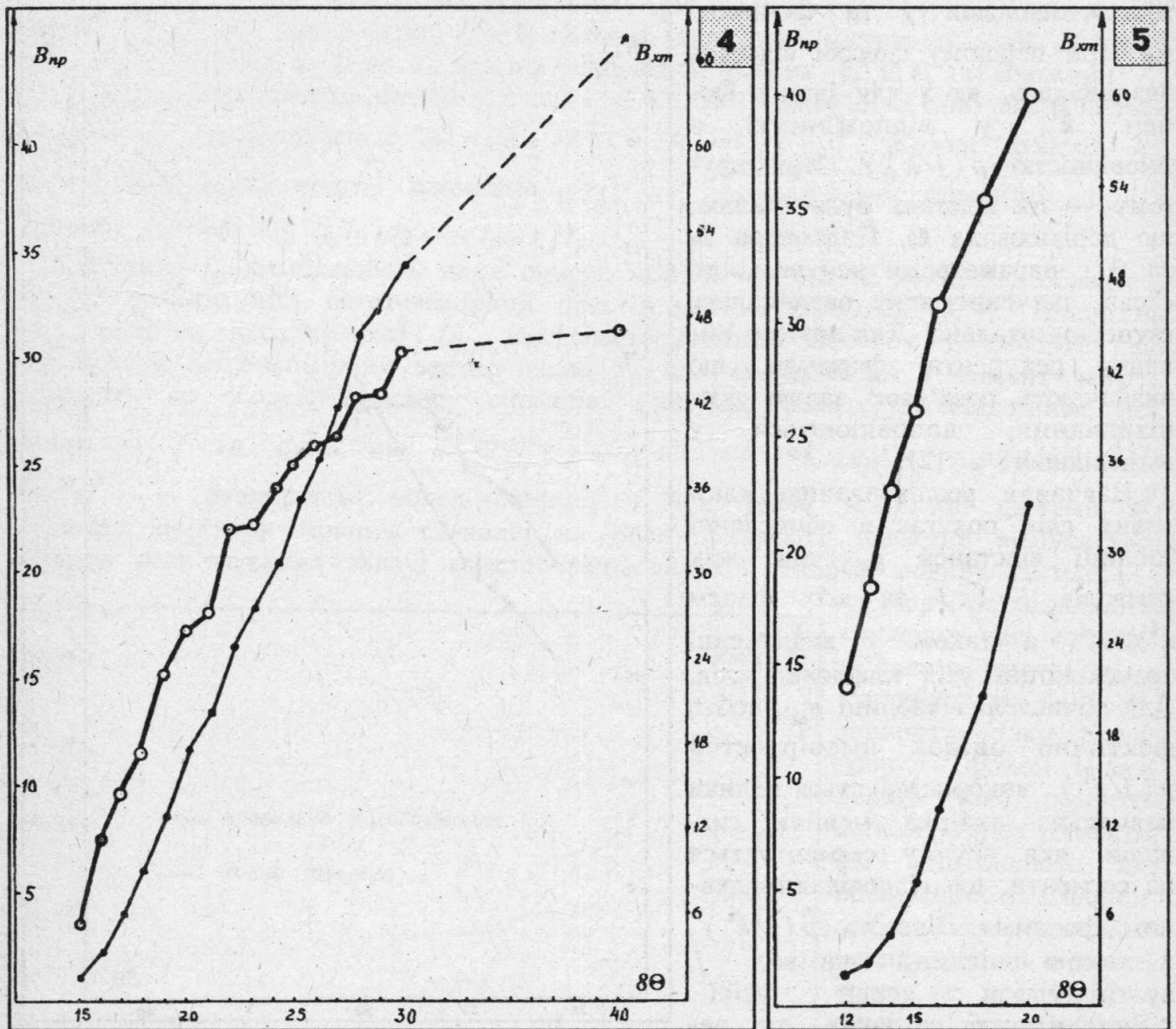
між символами j та фонемою k_0^1 . При першому способі відстань визначалась, як і для інших фонем k^1 , у відповідності з ймовірністю $p(j/k_0^1)$. При другому — ця відстань була сталою, що дорівнювала Θ . Параметри m та Θ є параметрами нейтрального слова, значення яких визначались експериментально. Для другого випадку рекурентні формули, що визначають розв'язок задачі розпізнавання, доповнюються у відповідності з [2].

Навчання розпізнаванню ключових слів полягає в обчисленні таблиці відстаней g_{jk}^1 для всіх символів $j=1:J$ та всіх фонем $k^1 \in K^1$, а також у визначенні транскрипцій усіх ключових слів. Для обчислення таблиці g_{jk}^1 , тобто фактично оцінок ймовірностей $p(j/k^1)$, використовується велика навчальна вибірка мовних сигналів, яка вручну сегментується на сегменти, що відповідають певним фонемам. Частота $\hat{p}(j/k^1)$, з якою певний символ j , зустрічається серед усієї різноманітності символів, що ре-

презентують фонему k^1 , приймається за оцінку ймовірності $p(j/k^1)$. Алфавіт фонем визначається експертом в залежності від мовлення й не обов'язково збігається з сукупністю фонем, яку визначають для даної мови фонетисти. Фонетичні транскрипції, тобто опис слів в алфавіті вибраних фонем, або задаються експертами, або можуть бути визначені автоматично за орфографічними текстами слів. Процес ручної сегментації та обчислення таблиці відстаней g_{jk}^1 , $j=1:J$, $k^1 \in K^1$, є трудомістким, але він виконується один раз під час настроювання системи. Після цього зміна та поповнення словника ключових слів відбувається шляхом введення в систему з клавіатури ПЕОМ фонетичних транскрипцій або орфографічних текстів слів, тобто навчання відбувається за текстами слів.

Експериментальні дослідження системи розпізнавання ключових слів були проведені з використанням фонограм зв'язного англійського мовлення. Попередньо був виконаний етап навчання розпізнаванню. Для цього за допомогою ручної сегментації мовних сигналів було одержано 1000 сегментів 56 фонем англійського мовлення. Весь простір мовних сигналів було розбито на 128 областей і для всіх символів $j=1:128$ і фонем $k^1=1:56$ були обчислені частоти $\hat{p}(j/k^1)$.





На підставі цих частот обчислювалась таблиця відстаней g_{jk}^1 , $j = 1 : 128$, $k^1 = 1 : 56$. Фонетичні транскрипції слів (по одній на кожне слово) були складені експертом.

Експерименти полягали в автоматичному розпізнаванні 200 реалізацій 50 ключових слів, що містилися серед більш як 3500 слів фонограми зв'язного англійського мовлення 11 дикторів тривалістю 40 хвилин. Сигнал фонограми безперервно надходив у систему розпізнавання, яка в реальному часі й розв'язувала задачу розпізнавання ключових слів. При цьому, на відміну від [1], заздалегідь не було відомо, скільки ключових слів у фразах та в якому порядку вони з'являються.

Було проведено чотири порівняльні експерименти. Перший експеримент проводився у точній відповідності з запропонованою моделлю. Рішення про нейтральне слово приймалось у відповідності з другим варіантом моделі нейтрального слова, а саме, коли відстань між символами j та узагальненою фонемою нейтрального слова була сталою та дорівнювала Θ . Результати розпізнавання в залежності від значення параметру Θ (параметр $m = 10$) наведені на рис. 2. Ліва вісь визначає відсоток правильних відповідей, коли ключове слово розпізнається у відповідний момент звучання фонограми. Права вісь визначає відсоток хибних

тривог, коли система «знаходить» ключове слово там, де в дійсності його немає. На інтервалі значень Θ між 30 та 40 експериментальні дані поки що не одержані, і передбачуваний характер поведінки кривих позначено пунктиром.

Другий експеримент був проведений за умов, коли рішення про нейтральне слово приймалося на підставі двох моделей нейтрального слова, а саме, коли хоча б одна з моделей вказувала на нейтральне слово. Результати наведені на рис. 3.

Третій експеримент проводився з деякими змінами в способі одержання символічного опису J_{0l} сигналу, а саме: спочатку сегментувався сигнал, представлений послідовністю багатовимірних векторів ознак $X_{0l} = (x_1, x_2, \dots, x_i, \dots, x_l)$, далі для кожного сегмента обчислювався вектор x_s , який найкраще апроксимував цей сегмент [3], і уже цей вектор замінювався на відповідний символ j . Подальша технологія експерименту збігалася з першою. Результати експерименту наведені на рис. 4.

Четвертий експеримент був проведений за умов, коли навчання розпізнаванню ключових слів проводилося на основі навчальної вибірки ключових слів, що була сформована з сигналу зв'язного мовлення для випадку, коли обчислювались так звані акустичні транскрипції слів [3]. Результати наведені на рис. 5. Аналізуючи експериментальні дані, можна дійти таких висновків.

1. При збільшенні Θ усі криві, що показують результати правильного розпізнавання, мають тенденцію до насичення, а криві, що відповідають хибним тривогам мають експоненціальний характер зростання.

2. Використання двох моделей нейтрального слова дещо зменшує відсоток хибних тривог.

3. Результати експериментів кращі при навчанні за навчальними вибірками.

Література

1. *Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models* / J.G. Wilpon, L.R. Rabiner, C.-H. Lee, E.R. Goldman // IEEE Trans. on ASSP. 1990. — 38, № 11. — P. 1870-1878.
2. Винцюк Т.К. Распознавание слов в потоке слитной речи // Автоматическое распознавание слуховых образов : Тез. докл. 16-го Всесоюз. семинара (APCO-16) (7-12 января 1991 г.). — М., 1991. — С. 24-25.
3. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. — Киев : Наук. думка, 1987. — 264 с.
4. Винцюк Т.К., Куляс А.И., Пилипенко В.В. Обучение распознаванию речи по текстам слов // Автоматическое распознавание слуховых образов : Тез. докл. 15-го Всесоюз. семинара (APCO-15) (13-17 марта 1989 г.). — Таллинн, 1989. — С. 79-81.



Оптимальний тональний і формантний аналіз сигналів мовлення

Олег Карпов

Дніпропетровський університет

Україна, 320625 Дніпропетровськ, МСП 10
просп. Гагаріна, 72
Тел.: (056) 239-1671

Тональний мовний сигнал має лінійчатий спектр, в котрому лінія з найнижчою частотою відповідає частоті основного тону голосу (ОТ). У випадку дзвінких звуків огинаюча лінійчатого спектру має максимуми на деяких частотах, які називаються формантними. Пошук формантних частот полягає в попередньому формуванні огинаючої спектру та знаходженні максимумів цієї огинаючої. Задачою даної роботи є побудова алгоритму оптимального пошуку частоти ОТ

і формантних частот як зв'язаних між собою параметрів на коротких інтервалах аналізу: відповідних тривалостей періоду основного тону T_{OT} для OT T_{F2} і T_{F2} для формантних частот $F1$, $F2$.

Класичне перетворення Фур'є (ДПФ та ШПФ) для коротких інтервалів аналізу не може з потрібною точністю виявити спектральну лінію, що відповідає частоті OT , оскільки коефіцієнти Фур'є будуть розміщені на частотах, кратних величині $F=1/T$, де T — інтервал аналізу. Тому для кожного інтервалу аналізу T необхідно здійснити деталізацію спектру в частотному діапазоні, близькому до частоти основного тону, для пошуку F_{OT} . Відповідно, для обчислення достатньо точних значень частот першої та другої формант на кожному періоді формант T_{F1} , T_{F2} необхідно деталізувати спектр в діапазонах частот, близьких до частот $F1$, $F2$. Деталізувати спектр (відновити спектр) в деякому частотному діапазоні можна, якщо обчислити проміжні спектральні коефіцієнти між коефіцієнтами Фур'є, які лежать на частотах $F_i = i * F$ для інтервалу аналізу T або аналогічних інтервалів для формантних частот, тобто спектральні коефіцієнти повинні бути обчислені на частотах

$$F_k = F_i / m = i * F / m, \quad k = i / m, \quad (1)$$

де m — коефіцієнт дробності номерів спектральних коефіцієнтів. Спектр буде обчислений в діапазоні частот $F + F_{max} / m$, і максимальний номер спектрального коефіцієнта буде дорівнювати $N / 2 / m$, де N — кількість відліків початкового сигналу.

Алгоритм розв'язання задачі. Основою алгоритму є косинусний спектральний аналіз [1], для якого початковий процес $f(t)$ може бути описаний як

$$s_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[\frac{1}{2} + \sum_{k=1}^n \cos k(t-x) \right] dt, \quad (2)$$

де

$$\sum_{i=1}^N f(t_i) \cos k(t_i - x) = s_k(x) \quad (3)$$

— k -а гармоніка в дискретному зображенні сигналу, яка задана виразом (1), N — число відліків сигналу $f(t)$. Спектральні коефіцієнти $S(k)$ являються амплітудними значеннями функцій $s_k(x)$:

$$S(k) = \max_x s_k(x). \quad (4)$$

На кожному періоді OT обчислюються динамічні спектри, для яких спектральні коефіцієнти будуть розміщені на частотах, кратних частоті основного тону.

На першому кроці визначається частота основного тону. Для цього визначається положення першої лінії лінійчатого спектру на деякому інтервалі T , більшому за період основного тону T_{OT} , і на інтервалі T , заданому N відліками, обчислюється спектр за формулами (3), (4) для номерів спектральних коефіцієнтів вигляду (1) при $i = m + N / 2$. В цьому випадку визначається детальний спектр в діапазоні частот $F + F_{max} / m$ і перший максимум спектру знаходиться на частоті основного тону F_{OT} . Для поліпшення достовірності пошуку частоти OT можна визначити положення другої лінії спектру, вона буде розміщуватися на частоті, кратній частоті OT .

На другому кроці визначається інтервал аналізу T_{OT} , що дорівнює періодові основного тону з числом відліків N_{OT}

$$N_{OT} = N * m / (i - m), \quad (5)$$

де i — номер коефіцієнта в виразі (1). На інтервалі T_{OT} обчислюється спектр $S(i)$ Фур'є або детальний спектр за формулами (1), (3), (4) при $m=1$ (при $m=1$ спектр Фур'є і детальний спектр збігаються). Спектр буде оптимальним по відношенню до частоти ОТ. Для обчисленого спектру визначається глобальний максимум

$$S_{\max} = \max_i S(i), \quad (6)$$

де $i = 1 + N_{OT} / 2$. Значення S_{\max} відповідає першій форманті, а його положення j визначає номер спектрального коефіцієнта, котрий знаходиться в околі дійсного положення першої форманти.

На *третьому кроці* обчислюються формантні параметри для кожного періоду першої форманти T_{F1} . Для цього по знайденому на попередньому кроці значенню j знаходиться усереднений період першої форманти τ_{F1} і число відліків n_{F1} , необхідних для представлення сигналу в межах цього періоду

$$n_{F1} = N_{OT} / j. \quad (7)$$

Оскільки j знаходиться в околах дійсного формантного максимуму, для наступних обчислень береться

$$n_{F1}^* = 2 * n_{F1}. \quad (8)$$

Для заданого n_{F1}^* числа точок обчислюється детальний спектр $S_F(i)$ за формулами (1), (3), (4) для $m_1 \geq j$. При цьому m_1 вибирається таким чином, щоб при звуженні спектру справа $F_{cp1} = F_{\max} / m_1$ друга форманта знаходилась в діапазоні $F^* + F_{cp1}$, де ліва межа спектру починається з частоти F^* і відповідає числу точок n_{F1}^* . Для спектру $S_F(i)$ знаходиться глобальний максимум, який визначає дійсне положення першої форманти для поточного формантного періоду

$$S_{F\max} = \max_i S_F(i). \quad (9)$$

Для положення j цього максимуму знаходиться період першої форманти T_{F1} і число точок N_{F1} , необхідних для опису сигналу на цьому періоді. Другий максимум визначає дійсне положення другої форманти. Третій крок повторюється стільки разів, скільки періодів першої форманти T_{F1} розміщується на періоді основного тону T_{OT} , потім процес продовжується з першого кроку для кожного періоду ОТ доти, поки не буде оброблено весь тональний сигнал.

Відзначимо, що алгоритм має додатковий позитивний ефект, за яким зменшення інтервалу аналізу (часового вікна) по ланцюжку N, N_{OT}, N_{F1} і обчислення спектрів здійснює одночасно фільтрацію спектра, полегшуючи режим пошуку відповідних максимумів і їх положення.

Оскільки спектри $S_F(i)$ являються оптимальними по відношенню до частоти основного тону, то спектри і сигнал голосового джерела ОТ можна вважати відновленими точно в межах точності обчислення частоти F_{OT} і їх можна описати математичними моделями—аналітичними функціями виду [2]

$$S_{OT}(t) = \sum_{k=1}^{\infty} p^k \sin kt = \frac{p \sin x}{1 - 2p \cos x + p^2}, \quad (10)$$

$$S_{OT}(t) = \sum_{k=1}^{\infty} p^k \cos kt = \frac{1 - p \cos x}{1 - 2p \cos x + p^2}, \quad (11)$$

$$S_{OT}(k) = p^k. \quad (12)$$

Функції (10), (11) описують сигнал ОТ, а формула (12) описує спектр ОТ. В загальному випадку функції (10), (11), (12) задають сімейства кривих по параметру p , де p відображає амплітудні властивості тонального сигналу.

Література

1. Толстов Г.П. Ряды Фурье. — М.: Наука, 1980. — 381 с.
2. Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм, рядов и произведений. — М.: 1962. — 1100 с.

Synthesis of the patterns for speaker independent speech recognition system

Pylyp Korkmazky

Chernivtsi Institute of Automatics

4-9, Bozhenko Side-Street, Chernivtsi
274029, Ukraine
Phone: (037) 223-90-12

Let's gather some set of the speech training samples

$$X = \{ \bar{X}_{n,v}^{(f)} \} (n = \overline{1, N}; v = \overline{1, V}; f = \overline{1, F_{n,v}};),$$

uttered by V different speakers for the vocabulary, consisting of N words. Let's determine patterns for each of N words and each of V speakers. Let's create code books for each of the speakers and replace all patterns of the word segments with the appropriate code vectors. Let's mark the word patterns, obtained in such a way, as $\bar{E}_{n,v}$. Let's determine values S_v :

$$S_v = \sum_{n=1}^N \sum_{z=1}^V \sum_{f=1}^{F_{n,z}} \rho(\bar{E}_{n,v}; \bar{X}_{n,z}^{(f)}), \quad v = \overline{1, V}. \quad (1)$$

Here $\rho(\bar{E}_{n,v}; \bar{X}_{n,z}^{(f)})$ is the distance between the f -th training sample of the n -th word and pattern $\bar{E}_{n,v}$ of the n -th word of the v -th speaker. Let's determine the speaker v^* according to formula:

$$v^* = \arg \min_v S_v. \quad (2)$$

Let's make a segmentation of all samples of the n -th word ($n = \overline{1, N}$) of the speaker v relatively to the patterns $\bar{E}_{n,v}$ (e.g. with the help of the dynamic program method). This segmentation sets a correspondence between frames of the training samples of the speaker v and code vectors of the speaker v^* . All frames of the training samples of the speaker v are marked with the corresponding numbers of code vectors of the speaker v^* . The centers of the clusters of those frames, gotten the same numbers, become code vectors of a new code book of the speaker v . Performing of such a procedure for all V speakers allows to create V new code books. Code vectors in different code books correspond to each other by means of common numbers. The last circumstance allows to determine a distance between the code books and to perform a clusterization procedure for V code books. The centres of the obtained W code books clusters ($W < V$) we shall call the base code books.

Let's use the dynamic programming formulas [1] for the recognition of a speech signal $\bar{x} = \{ \bar{x}_1, \dots, \bar{x}_p, \dots, \bar{x}_L \}$, uttered by some known speaker:

$$\left\{ \begin{array}{l} F_{n,i}(0) = F_{n,(i-1)}(0) + g(\bar{x}_i; \bar{e}_0) \\ F_{n,i}(r) = \min_{m_{n,r} \leq u \leq M_{n,r}} (F_{n,(i-u)}(r-1) + G_{n,i}(u, r)), \quad r = \overline{1, R_n} \\ G_{n,i}(u, r) = \sum_{v=i-u+1}^i g(\bar{x}_v; \bar{e}_{n,r}) \\ F_{n,i}(R_n+1) = \min (F_{n,i}(R_n); F_{n,(i-1)}(R_n+1) + g(\bar{x}_i; \bar{e}_0)). \end{array} \right. \quad (3)$$

Here $F_{n,i}(r)$ is the minimum integral distance between the initial part $\bar{x}^{(i)} = \{\bar{x}_1, \dots, \bar{x}_i\}$ of the speech signal \bar{x} and the initial part of the pattern $\bar{E}^{(r)} = \{\bar{e}_0, \dots, \bar{e}_{n,r}\}$; $g\{\bar{x}_v; \bar{e}_{n,r}\}$ is the distance between the speech frame \bar{x}_v and the pattern $\bar{e}_{n,r}$ of the r -th segment; $m_{n,r}$ and $M_{n,r}$ are minimum and maximum lengths of the r -th segment, \bar{e}_0 is the pattern of the background noise.

While performing of the recognition of the speech of an unknown speaker we shall synthesize his patterns with the help of the base code books. First, let's set a correspondence between base code vectors $\bar{Y}_{n,r}^{(v)}$ ($v = \overline{1, W}$), gotten the same numbers, and pattern segments $\bar{e}_{n,r}$ ($r = \overline{1, R_n}$). It is suggested to evaluate the values of the minimum integral distances $F_{n,i}^{(v)}(r)$ independently for the each base code book v ($v = \overline{1, W}$):

$$\left\{ \begin{array}{l} F_{n,i}^{(v)}(0) = F_{n,i}^{(v)}(0) + g(\bar{x}_i; \bar{e}_0) \\ F_{n,i}^{(v)}(r) = \min_{m_{n,r} \leq u \leq M_{n,r}} (F_{n,i}^{(v)}(i-u)(r-1) + G_{n,i}^{(v)}(u, r)), \quad r = \overline{1, R_n} \\ G_{n,i}^{(v)}(u, r) = \sum_{v=i-u+1}^i g(\bar{x}_v; \bar{Y}_{n,r}^{(v)}) \\ F_{n,i}^{(v)}(R_n+1) = \min (F_{n,i}^{(v)}(R_n); F_{n,i}^{(v)}(i-1)(R_n+1) + g(\bar{x}_i; \bar{e}_0)). \end{array} \right. \quad (4)$$

Here $g(\bar{x}_v; \bar{Y}_{n,r}^{(v)})$ is the distance between the frame \bar{x}_v and the base code vector $\bar{Y}_{n,r}^{(v)}$.

Let's evaluate values $\rho_{n,i}^{(v)}(r)$ for some \bar{x}_i fixed :

$$\left\{ \begin{array}{l} \rho_{n,i}^{(v)}(r) = \frac{1}{u_{n,i}^{(v)}(r)} G_{n,i}^{(v)}(u_{n,i}^{(v)}(r), r) \\ u_{n,i}^{(v)}(r) = \arg \min_{m_{n,r} \leq u \leq M_{n,r}} (F_{n,i}^{(v)}(i-u)(r-1) + G_{n,i}^{(v)}(u, r)) \\ r = \overline{1, R_n}; \quad v = \overline{1, W}. \end{array} \right. \quad (5)$$

Let's determine current estimates of the segment patterns for an unknown speaker by means of synthesis them with the help of the base code vectors:

$$\left\{ \begin{array}{l} \bar{e}_{n,i}^{(\gamma)}(r) = \sum_{v=1}^W \lambda_{n,i}^{(\gamma)}(r) \cdot \bar{Y}_{n,r}^{(v)}; \quad r = \overline{1, R_n} \\ \lambda_{n,i}^{(\gamma)}(r) = \left(\frac{1}{\rho_{n,i}^{(\gamma)}(r)} \right)^\gamma \cdot \left[\sum_{v=1}^W \left(\frac{1}{\rho_{n,i}^{(\gamma)}(r)} \right)^\gamma \right]^{-1}, \end{array} \right. \quad (6)$$

where γ is some parameter ($0 < \gamma < \infty$).

Evaluating the values $\bar{e}_{n,i}^*(r)$ for each of the sequential frame \bar{x}_i ($i = \overline{1, L}$), we get a possibility to synthesize an optimal pattern signal of the n -th word for an unknown speaker by means of calculating of the appropriate integral distances $F_{n,i}^*(r)$:

$$\left\{ \begin{array}{l} F_{n,i}^*(0) = F_{n,(i-1)}^*(0) + g(\bar{x}_i; \bar{e}_0) \\ F_{n,i}^*(r) = \min_{m_{n,r} \leq u \leq M_{n,r}} \left(F_{n,(i-u)}^*(r-1) + G_{n,i}^*(u,r) \right); \quad r = \overline{1, R_n} \\ G_{n,i}^*(u,r) = \sum_{v=i-u+1}^i g(\bar{x}_v; \bar{e}_{n,i}^*(r)) \\ F_{n,i}^*(R_n+1) = \min \left(F_{n,i}^*(R_n); F_{n,(i-1)}^*(R_n+1) + g(\bar{x}_i; \bar{e}_0) \right). \end{array} \right. \quad (7)$$

It is possible to create a code book for an unknown speaker on the base of the values $\bar{e}_{n,i}^*(r)$ in accordance with the results of the recognition. If it is necessary, this code book can add the set of the base code books, when a special threshold criterion indicates the necessity of such actions.

Reference

1. Vintsiuk T.K. Analyses, recognition and interpretation of speech signals. — Kiev : Nauk. Dumka, 1987. — 264p.



Instantaneous Spectrum Digital Analyzers

Volodymyr Krakovsky

Glushkov Institute of Cybernetics
of the AS of Ukraine

40, Academician Glushkov Avenue
Kyjiv, 252207, Ukraine
Phone : (044) 267-6962
Fax : (044) 266-1570

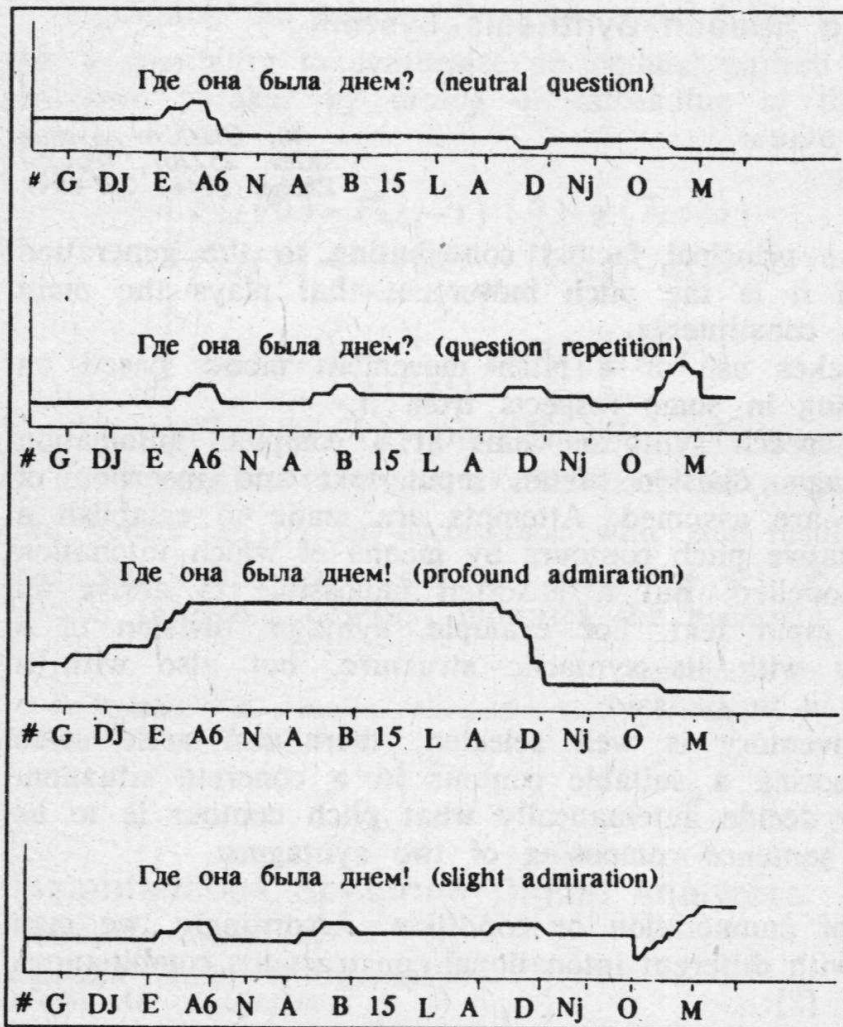
E-mail address vintsiuk%golos.kiev.ua@relay.ussr.eu.net

The dynamic spectral analysis is widely used in signal processing. The hardware for such analysis consists of so-called digital spectrum analyzers [1]. The computing power of those analyzers is limited by the number of complex multiplications required for processing samples of the input signal. The known recurrent algorithm reduce the number of the multiplications to one for computing each spectrum harmonic during each sliding step. Author improved the recurrent algorithm so, that it became possible to use one multiplication for computing two, four or eight spectrum harmonics simultaneously [2]. In the paper it will be considered digital analyzer structures implementing the above mentioned algorithms.

References

1. Цифровые анализаторы спектра / В.Н. Плотников, А.В. Белинский, В.А. Суханов, Ю.Н. Жигулевцев. — М., Радио и связь, 1990. — 184 с.
2. Краковский В.Я. Алгоритмы повышения быстродействия цифровых анализаторов мгновенного спектра—Кибернетика. — 1990. — № 5. — С. 113–115.





The 20 pitch contours differ in voice range (low, middle and high) and in pitch glides and jumps. The figure presents schematised pitch movement in a one and the same sentence uttered with different pitch contours.

Pitch contours are described by sequences of numerical values corresponding to fundamental frequency values. Although pitch contours representation is an adaptation of Lobanov's prosodemes description, the algorithm calculating F0 values for each 10 msec of an utterance is rather different.

Software was developed to facilitate the experimental part of the work. It includes automatic transcription module (for Ukrainian and Russian languages), graphical presentation of the phonetic knowledge, possibility of changing and expanding it readily and, at last, audio output.

References

1. Lobanov B.M. The Phonemophone text-to-speech system // Proc. 11th ICPHS. The Eleventh Intern. Congress of Phonetic Sciences (August 1-7, 1987). — Tallinn, 1987. — Vol. 1. — P. 120-124.
2. Bryzgunova E.A. Zvuki i intonacija russkoj reci. — M., 1969. — P. 251



Моделювання мовних сигналів методом гомоморфної обробки

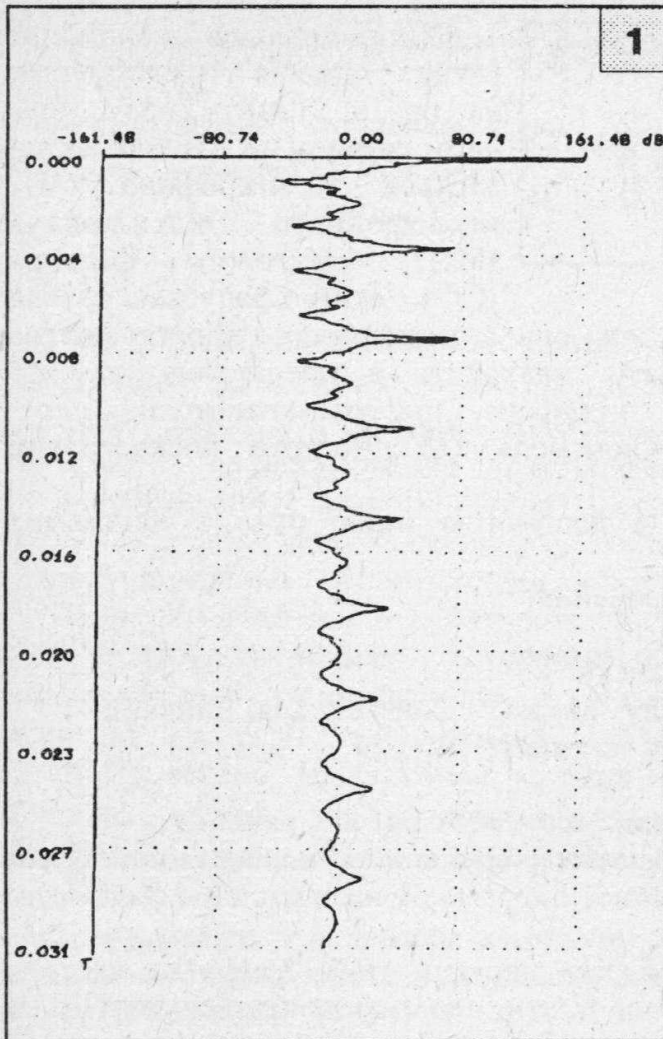
Володимир Мазур, Іван Маслій

Львівський університет

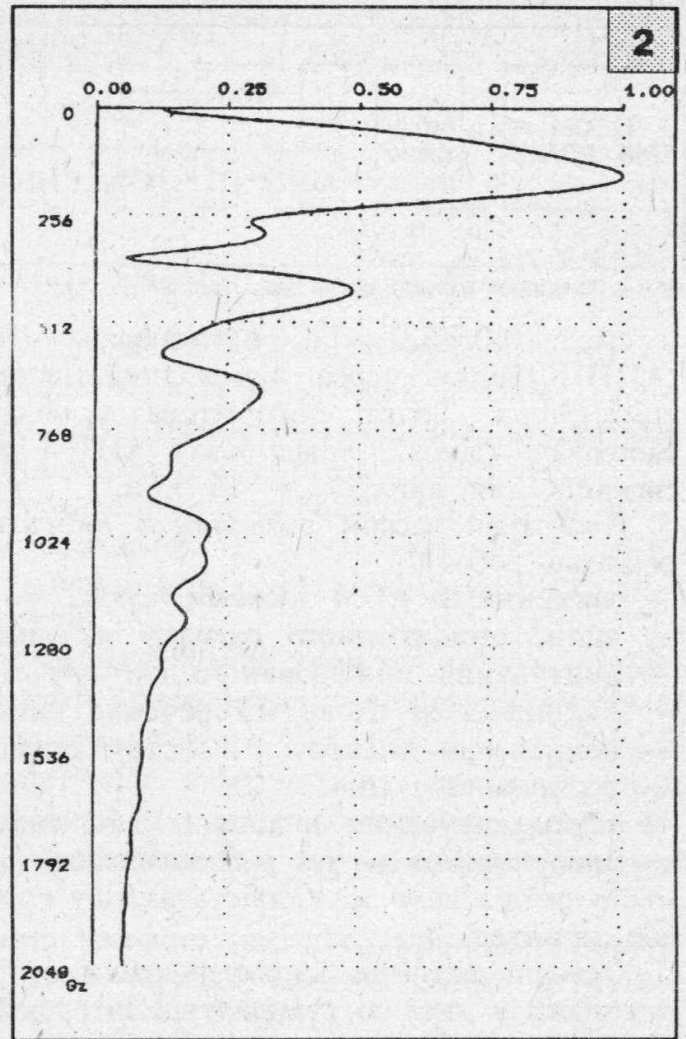
Україна, 290005 Львів
вул. Грушевського, 4
Тел.: (032) 79-556

В даній роботі описані програмно-апаратні засоби, що дозволяють на ПЕОМ IBM PC AT вводити, виводити та аналізувати мовний сигнал. Для первинної обробки мовлення пропонується метод гомоморфної обробки, який, поряд з іншими відомими методами, може використовуватись на етапі первинного опису. Особливу ефективність метод виявляє при аналізі функціонального і емоційного стану людини, оскільки він дозволяє виділити акустико-фонетичні параметри мовлення, які несуть інформацію про функціональний стан людини. На рис. 1 зображено кепстр голосного звуку «а», де час поданий в секундах, а амплітуда в децибелах.

Піки сигналу відповідають імпульсам збудження голосового тракту. Використовуючи кепстр сигналу можна отримати параметри, які широко використовуються різними



Кепстр голосного звуку «а»



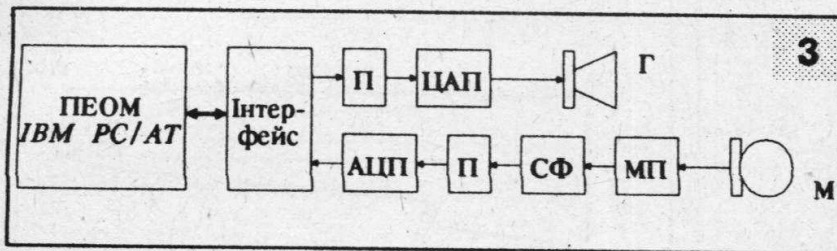
Спектр імпульсної характеристики мовного тракту для голосного «а»

авторами для аналізу функціонального і емоційного стану людини [1, 2], а саме частоту основного тону (ЧОТ), середнє значення ЧОТ, дисперсію ЧОТ, інтервал зміни ЧОТ і т.п. Окрім цього для розв'язку вище наведених задач пропонується використовувати імпульсну характеристику мовного тракту. На рис. 2 представлено спектр імпульсної характеристики мовного тракту для голосного звуку «а».

Якщо припустити, що на часових інтервалах 10 мс голосовий тракт має постійну форму і є відносно незалежним від джерела збудження, то його можна представити лінійною, інваріантною в часі, системою [3, 4], вихідний сигнал якої $s(t)$ є згорткою імпульсної характеристики голосового тракту зі збуджуючими коливаннями

$$s(t) = [p(t) * g(t) * v(t)]w(t), \quad (1)$$

де $p(t)$ — послідовність імпульсів, $g(t)$ — імпульсна характеристика гортані, $v(t)$ — імпульсна характеристика мовного тракту, $w(t)$ — функція вікна. У випадку гладкості функції $w(t)$ на протязі ефективної довжини імпульсних характеристик мовного тракту і гортані рівняння (1) можна переписати $s(t) = [p(t) * w(t)] * g(t) * v(t)$. Для розділення імпульсної характеристики голосового тракту $v(t)$ і імпульсів збудження $p(t)$ використовується клас гомоморфних систем, для яких справедливий узагальнений принцип суперпозиції відносно згортки [5].



Функціональна схема апаратних засобів

Функціональна схема апаратних засобів представлена на рис. 3. Тракт вводу мовного сигналу в ПЕОМ складається з мікрофона (М), мікрофонного підсилювача (МП), смугового фільтра (СФ), який пропускає сигнал

у смузі 0,05–8,0 кГц, підсилювача (П) та аналого-цифрового перетворювача (АЦП), звідки через відповідний інтерфейс надходить в оперативну пам'ять комп'ютера. Тракт синтезатора включає інтерфейс, цифрово-аналоговий перетворювач (ЦАП), підсилювач (П) і гучномовець (Г). Частота дискретизації сигналів для аналізу — 16 кГц.

Програмні засоби дозволяють вибрати з допомогою меню один з наступних режимів роботи:

- введення в ЕОМ фонем, слова чи речення;
- запис оцифрованого сигналу на диск;
- зчитування оцифрованого сигналу з диску;
- виділення зі слова чи речення сигналу окремої фонем для аналізу;
- обчислення спектру і кепстру мовного сигналу;
- розділення $p(t)$ і $v(t)$;
- прослуховування виділених або введених фонем, слів чи речення.

Використання меню робить систему простою і зручною в користуванні. Крім цього реалізована можливість виводу графічної інформації на дисплей комп'ютера або принтер.

Основні технічні характеристики розробленої системи такі. Апаратна частина виконана у вигляді стандартної інтерфейсної плати, що підключається до ПЕОМ IBM PC AT. Програмні засоби складаються з 15 модулів, написаних на мові С, що виконують наведені вище режими роботи, які працюють під керуванням спеціалізованої мовної моніторної програми. Дослідження можна виконувати в реальному масштабі часу. Необхідний для моделювання об'єм ОЗП динамічно розподіляється в межах 300–500 К байт. Об'єм програмного забезпечення не перевищує 150 К байт. Розроблені апаратно-програмні засоби можуть використовуватись для дослідження мовних сигналів як на предмет функціонального і емоційного стану людини, так і для розпізнавання мови.

Література

1. Галунов В.И., Манеров В.Х. Связь между психофизиологическим состоянием и характеристиками речевого сигнала. — Львов : АРСО, 1974. — С. 46–49.
2. Никонов А.В., Попов В.А. Особенности структуры речи человека-оператора в стрессовых ситуациях // Речь и эмоции. — Л. : Наука, 1975. — С. 11–16.
3. Фант Г. Акустическая теория речеобразования. — М. : Мир, 1964. — 284 с.
4. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. — М. : Радио и связь, 1981. — 495 с.
5. Oppenheim A.V. Nonlinear Filtering of Convolved Signals // Quart. Progr. Rept. 80. — Cambridge (Mass.) : Research Laboratory of Electronics, MIT, 1966. — P. 168–175.

Регулювання темпу мовної інформації

Юрій Рашкевич, Ярослав Гнатів,
Роман Марцишин, Зореслава Шпак

Львівський політехнічний інститут

Україна, 290646, Львів
вул. Степана Бандери, 12

Тел.: (032) 274-41-43

Факс: (032) 274-43-00

E-mail address: rush@cscd.lviv.ua@relay.ussr.eu.net

Значна надлишковість мовного сигналу, а також потреби його прискореного чи сповільненого відтворення (мовна пошта, логопедія, ввід-вивід мови в ЕОМ і т.д.) сприяли тому, що з середини 70-х років почалися інтенсивні роботи по створенню відповідних спеціалізованих пристроїв-регуляторів темпу мови (РТМ) [1].

В львівському політехнічному інституті аналогічні роботи ведуться з 1982 року. Проведено ряд теоретичних, експериментальних та конструкторських робіт, основні напрямки та результати яких і доповідаються на конференції.

Основні напрямки теоретичних досліджень.

1. Розроблені методи і алгоритми регулювання темпу мови на основі методу вибіркової сегментації сигналів, які дозволяють прискорювати/сповільнювати темп мови до 2-х разів при надійності розпізнавання фраз—90% [2].

2. Проведено дослідження методів нелінійного перетворення масштабу часу мовного сигналу з використанням попереднього розбиття сигналу на квазістаціонарні ділянки і наступного диференційованого перетворення довжини кожної із них [3].

3. Досліджені і розроблені алгоритми зміни часового масштабу мовного сигналу, шляхом побудови систем аналізу-синтезу мови на базі короткочасного перетворення Фур'є [4].

Основні дослідно-конструкторські роботи.

1. Розроблена система для дослідження мовних сигналів на базі мікроЕОМ «Искра-226» [5], яка дозволяє обробляти сигнал довжиною до 10 сек і динамічним діапазоном 48 дБ.

2. Розроблений апаратно-програмний інтерфейс з ІВМ РС/АТ для вводу-виводу мови в ЕОМ, який побудований на базі десятирозрядних АЦП і ЦАП з максимальною частотою дискретизації—30 кГц.

3. Розроблена і виготовлена партія з восьми регуляторів темпу мови, які дозволяють змінювати темп з коефіцієнтами 0,5...2 із дискретністю 0.125 при надійності розпізнавання команд не менше 90% [2]. Ці РТМ використовувалися для розв'язування задач судової експертизи і логопедії.

4. Розроблено пристрій логопедичного впливу на мовотворення [6], який дозволяє: плавно сповільнювати чи прискорювати відтворювану мову (коефіцієнт зміну темпу 1...2,5), зберігаючи при цьому її основні просодичні характеристики; отримувати ефект «відсталої» мови, вносячи у вихідний мовний сигнал часові затримки; виконувати додаткову звукову та світлову стимуляцію; відтворювати «чужу» мову, змінюючи тембр дикторського голосу та ін.

Основні напрямки подальших досліджень.

1. Вивчення особливостей темпоральної структури української мови з метою розробки алгоритмів високоякісного регулювання темпу мовної інформації.

2. Розробка алгоритмів перетворення часового масштабу мовних сигналів на базі систем аналізу-синтезу, які дозволяють розширити діапазон регулювання і підвищити надійність розпізнавання команд і окремих слів.

3. Втілення теоретичних і експериментальних результатів в спеціалізовані пристрої, які будуть використані в мовній пошті, для корекції «гелієвої» мови при водолазних роботах та ін.

Література

1. Шиффман М. Регулятор темпа речи для воспроизводящего магнитофона // Электроника. — 1974. — № 17. — С. 24–35.
2. Рашкевич Ю.М., Шпак З.Я. Регулирование темпа подачи речевой информации // Распознавание и синтез звуковых образов: Сб. науч. тр. — Киев: Ин-т кибернетики им. В.М. Глушкова АН УССР, 1987. — С. 70–77.
3. Rashkevich Ju.M. Nonlinear Transforms of the Time Structure of Speech Signals // Proc. of Latvian Signal Processing International Conference, April 1990, vol. 1. — P. 291–294.
4. Гнатив А.Н., Рашкевич Ю.М. Устройство для изменения темпа речевой информации. — А.с. № 1572296 (СССР).
5. Система для исследования речевых сигналов на базе микроЭВМ «Искра-226» / Х.Гепперт, Я.Н.Гнатив, И.Дочекал, Ю.М.Рашкевич, З.Я.Шпак // Автоматическое распознавание звуковых образов. — Таллинн: ИК АН ЭССР, 1989. — 325 с.
6. Hnativ A.N., Rashkevich Ju.M., Shpak Z.Y. A Device for Correction of Rhythmical Disorders of Speech Functions // Proc. of the Eleventh International Congress of Phonetic Sciences, August 1987, vol. 1. — P. 206–209.



Розробка методу дискретизації сигналів мовлення

Юрій Шабанов-Кушнарєнко, Валентина Чикіна

Харківський інститут радіоелектроніки

Україна, 310726 Харків
просп. Леніна, 14
Тел.: (057) 240-9446

У більшості сучасних систем обробки інформації найпершим перетворенням електричного сигналу є його дискретизація. Загальноприйнято у цьому випадку вживати аналого-цифровий перетворювач. Звичайно формується 8–12-розрядний двійковий код ординати решітчастої функції сигналу, що відповідає її вимірюванню приблизно з двома—трьома знаками точності. Інтервал між сусідніми точками береться, як правило, у межах від 100 мкс до 1 мс. У ряді випадків вживають значно точніші коди.

Звичайно для обґрунтування правомірності дискретизації сигналу посилаються на теорему дискретизації. Ця теорема стверджує, що у випадку, коли спектр $I(\omega)$ функції $f(t)$, заданий на всій дійсній осі, тотожно рівний нулю при частотах $\omega \geq \omega_0$, то за точними значеннями функції $f(t)$, знайденими в точках

$$t = 0, \pm \frac{2\pi}{2\omega_0}, \pm \frac{2\pi}{2\omega_0} \cdot 2, \dots, \text{ можна повністю відновити вид функції } f(t).$$

Оскільки будь-який прилад, що сприймає сигнал, має скінченну чутливість і скінченну смугу пропускання частот, для нього повинен існувати деякий скінченний поріг як щодо рівня сигналу, так і щодо часу. Отже, при виборі досить великої частоти дискретизації і досить високої точності при вимірюванні значень діаграми сигналу буде отримано код повідомлення, який містить в собі всю інформацію, яку сприймає приймач сигналу. В принципі, це, звичайно, вірне твердження, але воно носить надто загальний характер і не може бути достатнім при виборі оптимального способу дискретизації сигналу та встановленні її раціональних параметрів. Чи можливо змінити цей обсяг інформації без суттєвої шкоди для якості сигналу і якщо так, то наскільки і якими засобами? Чи впливає спосіб кодування сигналу на мінімально допустимий обсяг інформації,

що передається. На ці та на інші подібні запитання наука поки що не дає вичерпної відповіді. Їх розв'язання створило б наукову основу для суттєвого поліпшення технічних характеристик апаратури.

Аналіз літературних джерел показує, що до самого останнього часу розробки в цій галузі не були повністю успішними. У теорії зору вдалось зробити значно більше. Ще в 1834 році англієць Тальбот сформулював закон, який носить його ім'я, і який потім був застосований для розробки та обґрунтування практичних процедур дискретизації зорових сигналів. Згідно з законом Тальбота будь-який сигнал з незмінною яскравістю B_0 можна замінити таким спеціальним (у тому числі і дискретним) сигналом $B(t)$, який періодично змінюється на протязі часу (тобто мигтить), що спостерігач не помітить будь-яких тимчасових змін в своєму відчутті цього сигналу та сприйме його точно таким, як і сигнал B_0 . Для візуальної рівності сигналів $B(t)$ і B_0 необхідно, щоб середня яскравість сигналу $B(t)$ дорівнювала яскравості B_0 , тобто мала місце рівність

$$B_0 = \int_0^T B(t) dt / T. \quad (1)$$

Необхідно також, щоб період T сигналу $B(t)$ був досить малим.

Було отримано узагальнення закону Тальбота на випадок зорових картин, що повільно змінюються в просторі і часі. Тут нас буде цікавити узагальнення закону Тальбота лише на випадок зорових картин, які повільно змінюються на протязі часу. Нехай довільно вибрано зоровий стимул $B_0(t)$ та задано нескінченну послідовність зорових стимулів $\{B_N(t)\}_{N=1}^{\infty}$, яка задовольняє умові

$$\lim_{N \rightarrow \infty} \int_{t_1}^{t_2} B_N(t) dt = \int_{t_1}^{t_2} B(t) dt. \quad (2)$$

Тут t_1 і t_2 — довільні фіксовані моменти часу.

Узагальнений закон Тальбота свідчить, що відчуття, яке викликається зоровою картиною $B_N(t)$, при $N \rightarrow \infty$ не буде відрізнятися від зорового відчуття, яке викликається картиною $B(t)$. Умова (2) означає, що послідовність функцій $B_1(t), B_2(t), \dots$ збігається до функції $B(t)$ у середньому. Таким чином, відповідно узагальненому закону Тальбота, зоровий аналізатор ніби здійснює лінійне згладжування швидких коливань у діаграмі зорового сигналу.

Якщо по аналогії з зором під сигналом $B(t)$ розуміти енергетичний рівень сигналу, що змінюється на протязі часу, то закон Тальбота виявляється невірним. Якщо під сигналом $B(t)$ розуміти амплітудний рівень сигналу, то закон Тальбота у формі (1), хоча й підтверджується при випробуванні, однак являє собою цілком беззмістовне твердження. Сигнал просто не сприймається. Не сприймається, виявляється, також і будь-який періодичний звуковий сигнал, що задовольняє умові (1), якщо зробити досить малим його період (при $T \leq 50$ мкс).

Були проведені експерименти, в яких сигнали, що довільно вибираються, періодично переривалися з різною шпаруватістю λ та з періодом T . Одночасно з цим, для задоволення умові (2), переривчастий сигнал приймався як стимул $B_N(t)$, де під N розуміємо частоту переривання сигналу $N = 1/T$. Випробування показали, що при $N \geq N_0$ будь-який сигнал $B_0(t)$ не відрізняється від відповідного йому сигналу $B_N(t)$.

Важливо відзначити, що у формулюванні закону Тальбота має місце суттєва неточність. Справа в тому, що в ньому не береться до уваги поняття критичної частоти. Це закон граничний, ідеалізований, не пороговий. Тому узагальнений

закон Тальбота не дає відповіді на запитання про те, яким параметром оцінювати ступінь близькості сигналу $B_N(t)$ до сигналу $B_0(t)$ і при якому значенні цього параметра відчуття сигналів $B_N(t)$ і $B_0(t)$ інформаційно зійдуться. Цікаво як в теоретичному плані, так і з практичної точки зору спробувати так підсилити формулювання узагальненого закону Тальбота, щоб в ньому безпосередньо з'явився поріг. Це дало б змогу підійти до розробки наукових основ розрахунку раціональних параметрів дискретизації сигналу та вибору методу дискретизації. Це також дало б змогу підійти до вивчення експериментальними засобами конкретних значень величин порогів при рівних умовах прийому сигналу.

