

CHƯƠNG TRÌNH HỖ TRỢ TRẺ KHIẾM THÍNH LUYỆN ÂM QUA VIỆC TẬP HÁT

Lê Hoài Bắc, Nguyễn Đức Hoàng Hạ

Khoa Công nghệ Thông tin, Trường Đại học Khoa học Tự nhiên – ĐHQG-HCM
(Bài nhận ngày 30 tháng 11 năm 2004, hoàn chỉnh sửa chữa ngày 22 tháng 02 năm 2005)

TÓM TẮT: Phần lớn trẻ điếc đều có thể **LẮNG NGHE** và **NGHE ĐƯỢC** âm nhạc khi chúng đeo máy trợ thính và trẻ cảm nhận âm nhạc bằng một cách riêng nào đó. Âm nhạc luôn hấp dẫn con người trong mọi thời đại, mọi nền văn hóa và dĩ nhiên cũng rất hấp dẫn với trẻ điếc. Nhưng làm sao trẻ học nhạc được, làm sao trẻ biết hát? Kỹ thuật xử lý và nhận dạng âm thanh hiện nay đã có thể hỗ trợ trẻ trong vấn đề này. Từ khả năng nhận ra nốt nhạc trong bài hát, nhận ra lời trong bài hát (thu từ micro), chúng tôi xây dựng một chương trình hỗ trợ người khiếm thính (đặc biệt là trẻ) tập hát. Chương trình gồm 2 mục tiêu, phần luyện thanh và phần tập hát.

1. GIỚI THIỆU

Bài báo này nhằm giới thiệu một phương pháp hỗ trợ trẻ khiếm thính luyện giọng. Đó là luyện giọng qua việc tập hát. Mỗi người có một cách cảm nhận bài hát khác nhau. Ý tưởng quan trọng ở đây là giúp trẻ khiếm thính nhận thức ra được âm nhạc qua màn hình kết hợp với đồ họa sinh động của máy tính. Bài báo gồm 3 phần:

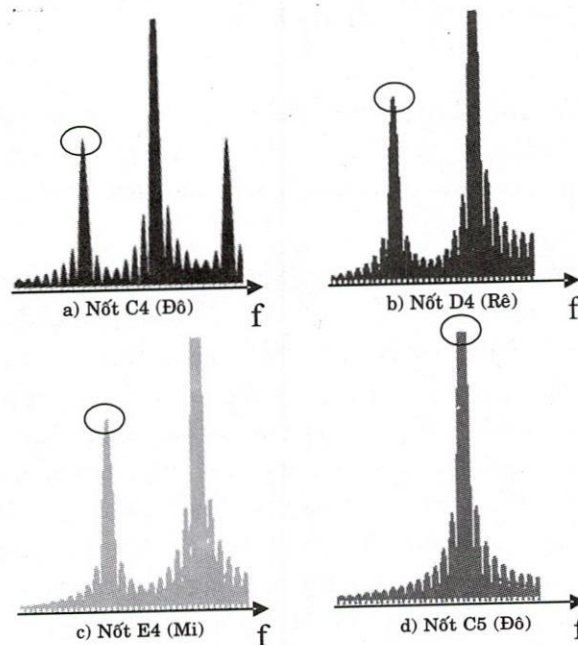
- Phần kỹ thuật sẽ trình bày một số kỹ thuật quan trọng trong chương trình như nhận dạng nốt nhạc từ âm thanh, nhận dạng lời bằng mạng nơ ron.
- Phần ứng dụng sẽ trình bày những lợi ích và giới thiệu chương trình.
- Phần cuối là đánh giá và kết luận

2. KỸ THUẬT

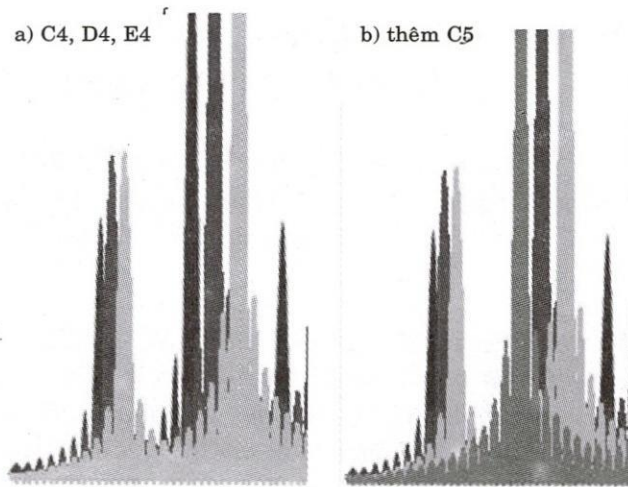
2.1. Nhận dạng nốt nhạc từ giọng hát

Âm thanh là tín hiệu được tổng hợp từ những sóng âm có tần số đơn giản dạng SIN hoặc COS. Dùng phép biến đổi Fourier, ta có thể chuyển từ tín hiệu tổng hợp ra lại thành các tín hiệu đơn vị (tín hiệu dạng SIN hoặc COS có tần số cố định).

Đây là hình ảnh một số nốt nhạc đã chuyển về dạng miền tần số



Hình 1: Minh họa hình phổ của các nốt Đô, Rê, Mi, Đố



Hình 2. Sự khác biệt các tần số tạo nên các nốt nhạc

Ta thấy rất rõ trong hình 2b, nốt Đố (C5) hoàn toàn trùng với tần số cộng hưởng kế tiếp của nốt Đô (C4). Vậy việc nhận ra các nốt nhạc là việc phát hiện các điểm cực trị đánh dấu trong hình 1. Thuật toán phát hiện cực đại như sau:

```

Input:  double *f: dãy tần số của tín hiệu
        int n;
Output:  f0: tần số của nốt nhạc
B1: min = 0;
B2: for (i=1; i<n; i++) {
    if (f[i]>f[i-1] && f[i]>f[i+1]) {
        if (f[i]<min) return f0;
        min = f[i];
        f0 = i;
    }
}
    
```

Ý tưởng chính là tìm cực đại đầu tiên không tăng thì cực đại trước sẽ là f_0 . Đặc tính này đúng với một giọng hát. Không đúng hoàn toàn với nhiều giọng hát và khi có nhạc đệm quá lớn ảnh hưởng vào. Phạm vi của bài báo này chỉ cần ở mức này là đủ.

2.2. Nhận dạng lời bài hát

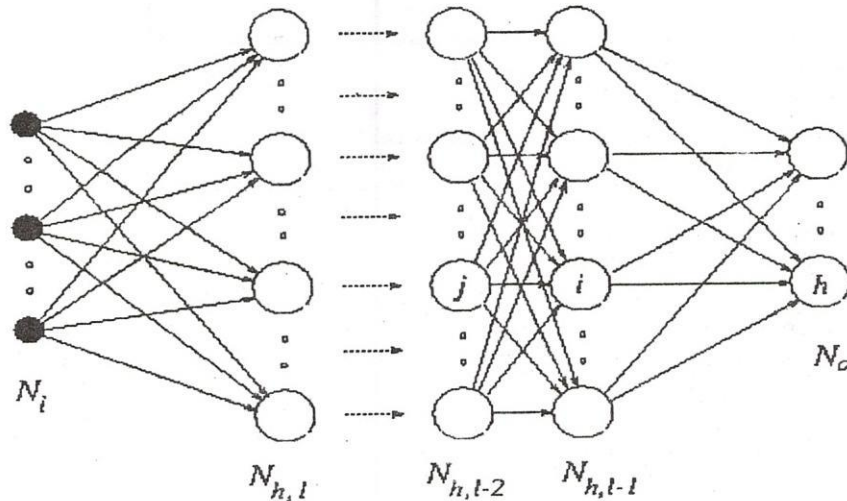
Lời bài hát được chia làm 2 loại, một là những âm đơn giản như a, e, o,...; hai là lời của bài hát thật sự. Nhận dạng lời bài hát tổng quát là một vấn đề khó, và là hướng nghiên cứu tiếp theo của đề tài. Hiện tại, đề tài chỉ mới giải quyết các âm đơn giản được hát trong bài hát.

2.2.1. Mạng nơ-ron nhân tạo

Các phương pháp xử lý tín hiệu kỹ thuật số (Digital Signal Processing - DSP) truyền thống thường dựa trên các thuật toán chuyển đổi dữ liệu từ dạng này sang dạng khác thông qua các thủ tục cài đặt sẵn và hầu hết các kỹ thuật này đòi hỏi phải có các tham số nào đó để thực thi. Các thuật toán này mô tả công việc phải thực thi trên dữ liệu, còn các tham số cung cấp các chuẩn để đánh giá chúng. Việc lựa chọn các tham số thích hợp trở nên quan trọng hơn chính các thuật toán.

Trên cơ sở mô phỏng cách thức hoạt động của bộ não người, mạng nơ-ron nhân tạo (Artificial Neural Network – ANN, gọi tắt là mạng nơ-ron) ra đời đã phát huy ý tưởng này lên đỉnh điểm, bằng cách sử dụng các thuật toán cực kỳ đơn giản kèm theo các tham số được tối ưu hóa. Mạng nơ-ron đã thay thế các phương pháp giải quyết vấn đề truyền thống bằng các chiến lược thử và sai, các giải pháp thực dụng, và phương pháp luận “kết quả này tốt hơn kết quả kia”.

Một mạng nơ-ron đa lớp có kiến trúc phân lớp, mỗi lớp bao gồm nhiều nơ-ron. Các nơ-ron này nhận tín hiệu vào từ các nơ-ron thuộc lớp ngay trước nó và truyền tín hiệu ra cho các nơ-ron ở các lớp sau. Không có kết nối giữa các nơ-ron trong cùng một lớp. Mặc dù một mạng nơ-ron có thể có rất nhiều lớp nhưng kiến trúc thông dụng nhất là một mạng nơ-ron với 3 lớp: lớp nhập (input layer), lớp ẩn (hidden layer), và lớp xuất (output layer).



Hình 3: Một mạng nơ-ron đa lớp với 1 lớp (1-2 lớp ẩn)

Mạng được huấn luyện bằng phương pháp lan truyền ngược. Quá trình lan truyền ngược của mạng nơ-ron chính là quá trình mạng tự cập nhật lại các trọng số của các nơ-ron nhằm đạt được giá trị của các tín hiệu ra (output) như mong muốn.

Độ lỗi của mạng được tính theo công thức sau:

$$E = \frac{1}{2} \times \sum_{i=1}^{N_o} (d_i - a_i)$$

trong đó a_i là giá trị mạng sinh ra, d_i là giá trị mà ta mong muốn tín hiệu ra sẽ đạt tới và N_o là số tín hiệu ra (output).

Sự thay đổi trọng lượng được tính theo công thức:

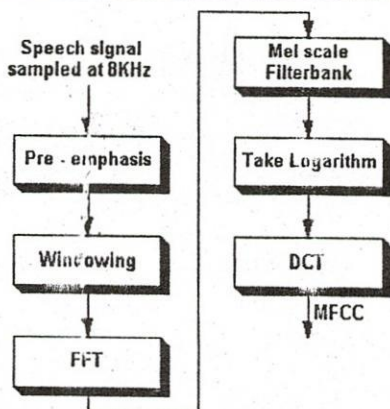
$$\Delta w_{ij} = -\gamma \times \frac{\delta E}{\delta w_{ij}}, \text{ với } \gamma \text{ là tốc độ học (learning rate).}$$

2.2.2. Nhận dạng các âm đơn giản

Để giải quyết bài toán nhận dạng này, ta có một nhận xét quan trọng sau: Âm thanh được phát ra là do luồng không khí đi qua bộ máy phát âm của trẻ. Bộ máy phát âm chỉ duy trì ở một trạng thái và phát ra tiếng nhờ dây thanh quản rung và các cộng hưởng bên trong bộ máy. Vì vậy, tín hiệu âm thanh là tuần hoàn. Ta đã biết, tín hiệu tuần hoàn thì sẽ có phổ vạch. Một tính chất quan trọng ở đây là tại mỗi thời điểm, hình ảnh phổ là như nhau. Từ tính chất này, chúng tôi đã chọn mạng nơ-ron nhân tạo (Artificial Neural Network – ANN) để nhận dạng âm vị tại một thời điểm. Tổ chức mạng nơ-ron gồm 3 lớp: 12 nơ-ron vào tương ứng với vectơ đặc trưng, 120 nơ-ron ẩn, 9 nơ-ron ra tương ứng với số âm đơn giản (đặc trưng được chọn để nhận dạng là MFCC (Mel-Frequency Cepstral Coefficients)).

Ở đây, ta sử dụng phương trình sai phân tuyến tính hệ số hằng $y = \sum_{r=0}^1 b_r x(n-r)$ với $b_0=1, b_1=$

0.97 để Pre-emphasis tín hiệu. Sử dụng Hamming-Window để Windowing tín hiệu sau khi đã Pre-emphasis. Sau đó, dùng phép biến đổi Fourier để chuyển tín hiệu từ miền thời gian sang miền tần số. Dãy bộ lọc được thiết kế là dãy bộ lọc tam giác có tần số giữa đều nhau trên thang Mel.



Hình 4: Các bước tính đặc trưng MFCC

Danh sách các âm đơn giản:

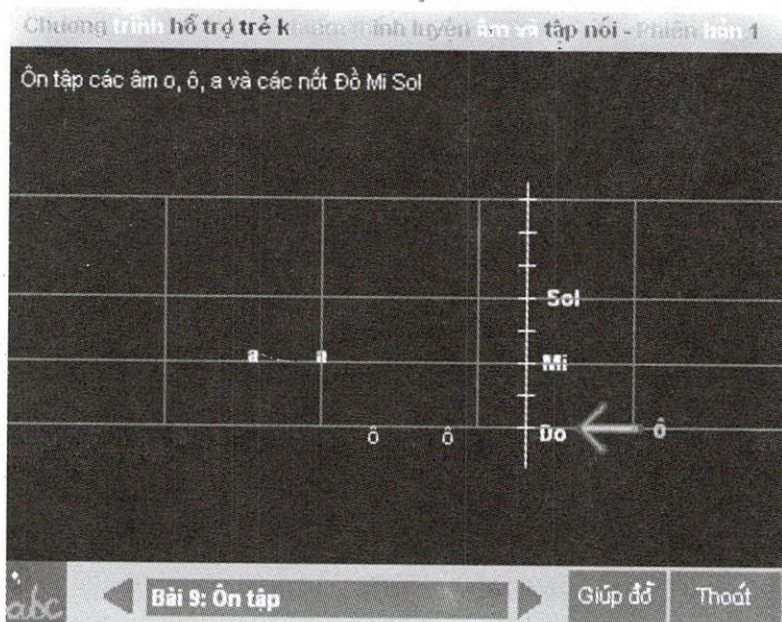
STT	Âm	Từ	STT	Âm	Từ
1	i	i	6	a:	a
2	e	ê	7	u	u
3	ɛ	ɛ	8	o	ô
4	ɔ	ư	9	ɔ	o
5	ɜ:	ơ			

3. ỨNG DỤNG

Những kỹ năng trẻ học được qua chương trình:

- *Kỹ năng sáng tạo:* Trẻ tự dựng lên giai điệu và tiết tấu.
- *Kỹ năng vận động thô:* Trẻ tự vận động theo nhạc (như vỗ tay, lắc mình...).
- *Kỹ năng về cảm xúc:* Trẻ biểu hiện nhiều cảm xúc qua âm nhạc như vui, buồn, tức giận; âm nhạc còn có thể làm thư giãn đầu óc.
- *Kỹ năng giao tiếp:* lắng nghe giai điệu, nhịp điệu, lời của bài hát; lắng nghe sự khác nhau trong âm nhạc (như cao-trầm, to-nhỏ...)

Việc tập hát là một phần trong những bài học của trẻ. Trong chương trình dạy trẻ, tập hát đã được lồng vào một cách thích hợp. Sau đây là bài học tập hát được trích ra trong bộ giáo trình luyện âm tập nói tiếng Việt và phát triển tư duy.



Hình 4: Hình minh họa ứng dụng luyện tập hát.

4. KẾT LUẬN

Kết quả thực nghiệm về nhận dạng ra nốt nhạc khá chính xác (>99%). Về nhận dạng các âm đơn giản dùng ANN cũng cho kết quả rất cao và cả hai đều nhận đúng trên nhiều nhóm người. Sau đây là bảng đánh giá thử nghiệm.

STT	Loại thử nghiệm	Người thử nghiệm	Kết quả	Ghi chú
1	Nhận dạng nốt nhạc	2 Nam +2 Nữ đọc 10 lần các nốt C4(Đồ) tới B5(Si)	99.7%	Một số lỗi do thiết bị hoạt động không ổn định lúc mới bắt đầu và sau một thời gian mở micro quá lâu. Những lỗi này được bỏ qua.
2	Nhận dạng âm đơn giản	2 Nam +2 Nữ đọc 50 lần các âm cơ bản	99.4%	nt

THE PROGRAM OF SUPPORTING DEAF CHILDREN IN SOUND TRAINING THROUGH LEARNING TO SING

Le Hoai Bac, Nguyen Duc Hoang Ha

Faculty of Information Technology, University of Natural Sciences – VNU-HCM

ABSTRACT: *Most of the deaf children can listen to music and songs with the help of special devices, and they can feel the music with their own special way. Music and songs are always very interesting to everyone all over the world in every culture. Of course, music is also attractive to deaf children. The problems are how deaf children can learn music and learn how to sing. Nowadays audio processing and recognizing techniques can help deaf children to overcome obstacles to learn music and to sing. Utilizing the capability to recognize notes and lyrics from songs recorded by microphones, we developed the application to support the deaf, especially young children, in sound training. Our application includes two objectives: sound training and learning to sing.*

TÀI LIỆU THAM KHẢO

- [1]. Lê Hoài Bắc, Nguyễn Đức Hoàng Hạ, *Xây dựng chương trình hỗ trợ trẻ khiếm thính luyện âm tập nói*, Hội thảo Quốc gia về Công nghệ thông tin – 6/2002 Nha Trang, Việt Nam.
- [2]. Lê Hoài Bắc, Nguyễn Đức Hoàng Hạ, *Sử dụng mạng nơ-ron phát hiện âm vận trong tiếng nói tiếng Việt*, Hội nghị khoa học lần thứ III, Trường ĐHKH Tự nhiên, 10/2002, Thành phố Hồ Chí Minh, Việt nam.
- [3]. Quách Tuấn Ngọc, *Xử lý tín hiệu số*, Nhà xuất bản giáo dục, 1999.
- [4]. Carl G. Looney, *Pattern Recognition Using Neural Networks*, Oxford University Press, 1997.
- [5]. Phạm Thị Sửu, *Chương trình 36 buổi cho lớp mẫu giáo 5 tuổi*, NXBGD, 2001.
- [6]. Cao Xuân Hạo, *Tiếng Việt mấy vấn đề ngữ âm ngữ pháp ngữ nghĩa*, NXBGD, 1998.
- [7]. Một số tài liệu tham khảo trong các khóa đào tạo giáo viên dạy trẻ điếc tại Bệnh viện Tai-Mũi-Họng Tp. HCM, 2002.