

TỔNG HỢP TIẾNG NÓI TRÊN THIẾT BỊ GIÁM SÁT HÀNH TRÌNH SPEECH SYNTHESIS ON THE TRACKING SURVEILLANCE DEVICE

Trần Thu Hà¹, Phạm Minh Dương²

¹ Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh

² Trung tâm Nghiên cứu và Đào tạo Thiết kế Vi mạch (ICDREC)

Ngày tòa soạn nhận bài 06/11/2015, ngày phản biện đánh giá 30/11/2015, ngày chấp nhận đăng 17/12/2015.

TÓM TẮT

Bài báo này trình bày quá trình thiết kế và thi công thiết bị giám sát hành trình có tích hợp giải thuật tổng hợp tiếng nói bằng phương pháp ghép nối để giúp cho người sử dụng có thể giám sát, quản lý phương tiện của mình một cách trực quan. Đồng thời, người sử dụng có thể truyền thông tin xuống lái xe để phát thông báo. Tổng hợp tiếng nói bằng phương pháp ghép nối-lựa chọn đơn vị các từ, cụm từ hoặc câu làm cho tiếng nói sau khi tổng hợp đạt được chất lượng về mức độ tự nhiên và mức độ dễ nghe. Văn bản đầu vào cần được chuẩn hóa thành dạng chữ hoa trước khi tiến hành tổng hợp ghép nối các đơn vị âm thanh bằng các ngữ cảnh thích hợp. Ngoài ra, thiết bị giám sát hành trình này cũng có đầy đủ các tính năng theo quy chuẩn QCVN 31:2014/BGTVT của Bộ Giao thông vận tải.

Từ khóa: Thiết bị giám sát hành trình; giải thuật tổng hợp tiếng nói; lựa chọn đơn vị; phương pháp ghép nối; Bộ Giao thông vận tải.

ABSTRACT

This paper presents the design and fabrication of the tracking surveillance device with speech synthesis algorithm by pairing method in which the users can monitor and manage their vehicles intuitively. At the same time, users can transmit the information to the driver to broadcast notifications. The speech by the unit selection-pairing synthetic method with words, phrases or sentences made speech after synthesis achieved high quality of natural level and easy listening level. The input text must be standardized into uppercase before processing the pairing synthesis with audio units by appropriate context. Besides, the tracking surveillance device also has fully complicant features of QCVN 31:2014 BGTVT, the standard of the Ministry of Transport.

Keywords: The tracking surveillance device; speech synthesis algorithm; unit selection; pairing method.

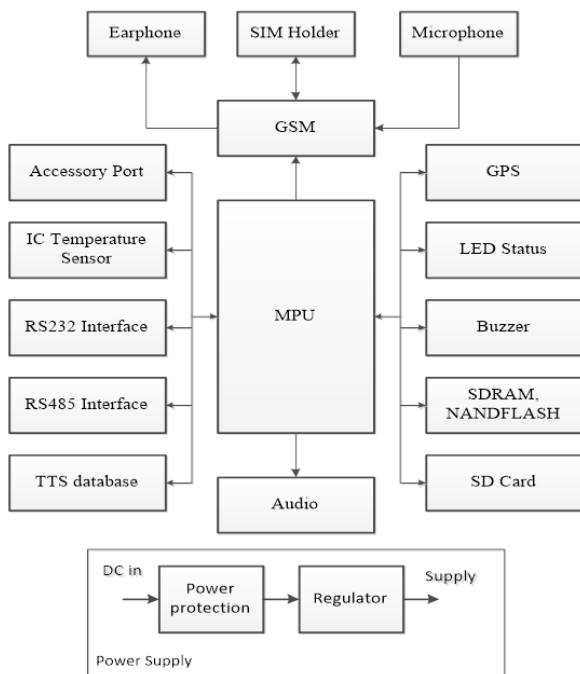
1. GIỚI THIỆU

Hiện nay thiết bị giám sát hành trình (TBGSHT) ở nước ta đang phát triển rất nhanh, nhất là từ khi Nghị định 08 của Bộ Giao thông vận tải số: 08/2011/TT-BGTVT ban hành. Ngoài các tính năng mà Nghị định đưa ra, TBGSHT còn có thêm một số tính năng mở rộng để đáp ứng nhu cầu sử dụng của khách hàng. Đặc biệt là các tính năng này có thể được tích hợp cùng với TBGSHT nhằm tăng tính tiện ích và giảm chi phí cho người sử dụng. Chẳng hạn các tính năng đó là: máy tính cước taxi, thông báo thông tin từ trung tâm điều hành, rao trạm xe buýt, thanh toán tiền tự động bằng RFID, v.v.

Đề tài sẽ thực hiện tổng hợp tiếng nói tiếng Việt [2, 4, 10] và lập trình nhúng vào TBGSHT để phát thông báo bằng cách truyền thông tin văn bản từ máy chủ xuống TBGSHT qua giao thức TCP/IP hoặc bằng tin nhắn SMS, đề tài này có thể ứng dụng thay thế cho bộ đàm trong taxi, rao trạm xe buýt, v.v. tùy vào ứng dụng cụ thể mà phát triển thêm.

2. MÔ TẢ THIẾT BỊ

Hình 1 mô tả toàn bộ sơ đồ khối của TBGSHT có tích hợp thuật toán tổng hợp tiếng nói bằng phương pháp ghép nối – lựa chọn đơn vị.



Hình 1. Sơ đồ khối của TBGSHT có tích hợp TTS

Chức năng mỗi khối được mô tả như sau:

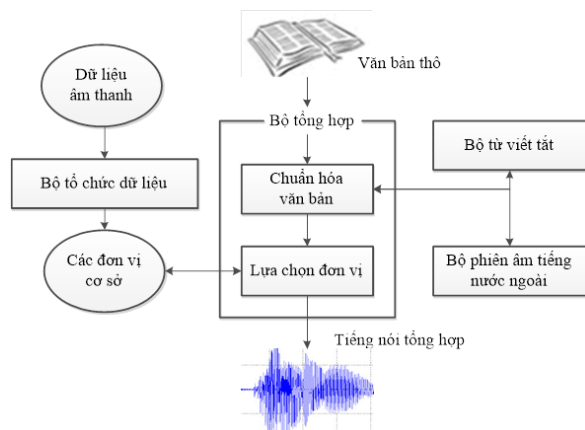
- Khối MPU: vi xử lý AT91SAM9260, tốc độ 180 MHz.
- Khối SDRAM, NAND FLASH: 64 MB SDRAM và 256 MB NAND Flash.
- Khối Power Supply: hỗ trợ nguồn cấp DC từ 8V – 36V, cung cấp nguồn cho các khối chức năng hoạt động.
- Khối GSM: giao tiếp với MPU để truyền nhận dữ liệu lên trung tâm điều khiển qua giao thức TCP/IP, giao thức HTTP hoặc giao thức FTP.
- Khối GPS: phân tích các thành phần dữ liệu được nhận từ module GPS, ta có được các thông số như: kinh độ, vĩ độ, vận tốc di chuyển, hướng di chuyển, thời gian thực, v.v.
- Khối LED trạng thái và Buzzer: hiển thị và thông báo các trạng thái của TBGSHT.
- Khối SD Card: lưu trữ cơ sở dữ liệu âm thanh và các thông tin hoạt động của TBGSHT.
- Khối Accessory Port: khối này dùng để

giao tiếp với các tín hiệu của xe như: khóa xe, cửa xe, máy điều hòa, vận tốc cơ của xe, nhiên liệu, nút nhấn SOS, v.v.

- Khối IC Temperature Sensor: dùng IC số để cảm biến nhiệt độ xe, hoặc các ứng dụng có nhiệt độ từ 0 đến 100 độ C.
- Khối RS232 Interface: giao tiếp với phần mềm phân tích dữ liệu TBGSHT của Tổng cục Đường bộ Việt Nam theo QCVN 31:2014/BGTVT của Bộ Giao thông vận tải.
- Khối RS485 Interface: dùng để mở rộng giao tiếp với các ngoại vi khác như Camera, RFID, và một số ngoại vi theo yêu cầu của khách hàng.
- Khối TTS database: lưu trữ thư viện âm thanh để ứng dụng đọc văn bản.
- Khối Audio: giải mã và phát âm thanh cho ứng dụng TTS.

3. PHƯƠNG PHÁP TỔNG HỢP TIẾNG NÓI

Kiến trúc tổng quát của hệ thống tổng hợp tiếng nói tiếng Việt bằng phương pháp ghép nối – lựa chọn đơn vị được mô tả như hình 2, gồm 4 thành phần chính, mỗi thành phần đóng vai trò xử lý riêng:



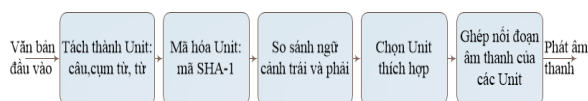
Hình 2. Kiến trúc tổng quát của hệ thống tổng hợp tiếng nói

- **Bộ tổ chức dữ liệu:** thực hiện quản lý dữ liệu âm thanh, và phân đoạn chúng thành các đơn vị cơ sở.
- **Bộ tổng hợp:** thực hiện chuyển văn bản sang tiếng nói thông qua hai bước chính

là chuẩn hóa văn bản và tổng hợp ghép nối. Trong bước chuẩn hóa văn bản, các dạng chữ thường, chữ hoa, số, từ viết tắt, tên địa danh tiếng nước ngoài, các ký hiệu đặc biệt v.v. sẽ được chuyển thành dạng chuẩn chữ hoa mà phần tổng hợp ghép nối có thể sử dụng để tạo âm thanh tương ứng.

- *Bộ từ viết tắt*: chuyển đổi các từ viết tắt thông dụng thành dạng chữ hoàn chỉnh của tiếng Việt.
- *Bộ phiên âm từ nước ngoài*: chuyển từ nước ngoài sang cách đọc tiếng Việt chính xác nhất có thể.

Tổng hợp tiếng nói bằng phương pháp ghép nối – lựa chọn đơn vị được mô tả như hình 3, sau khi văn bản đầu vào được chuẩn hóa, ta tiến hành tách văn bản thành các đơn vị (Unit) như: câu, cụm từ hoặc từ, rồi mã hóa các đơn vị này thành từng mã SHA-1 [1] tương ứng, thực hiện so sánh ngữ cảnh trái và phải của từng mã SHA-1 và chọn mã SHA-1 thích hợp nhất trong cơ sở dữ liệu, ghép nối các đoạn âm thanh của từng đơn vị lại với nhau, sau đó phát âm thanh được tổng hợp ra hệ thống.



Hình 3. Quy trình thực hiện phương pháp ghép nối – lựa chọn đơn vị

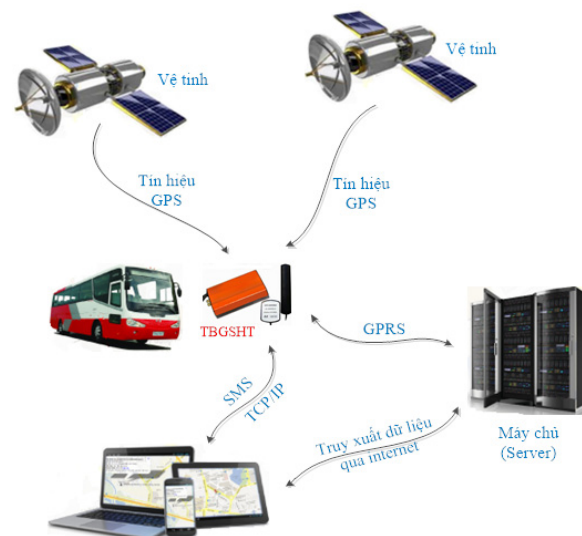
4. HOẠT ĐỘNG HỆ THỐNG

Trong hình 4, TBGSHT có hai chức năng chính cần xử lý: chức năng thứ nhất là TBGSHT phải thực hiện đầy đủ các tính năng theo Nghị định của Bộ Giao thông vận tải, chức năng thứ hai là tổng hợp tiếng nói bằng phương pháp ghép nối – lựa chọn đơn vị.

Trong chức năng thứ nhất, TBGSHT thu thập dữ liệu thời gian, vận tốc, tọa độ, v.v. từ vệ tinh, thu thập các tín hiệu từ xe, xử lý thông tin cần thiết để truyền đến trung tâm giám sát qua GPRS.

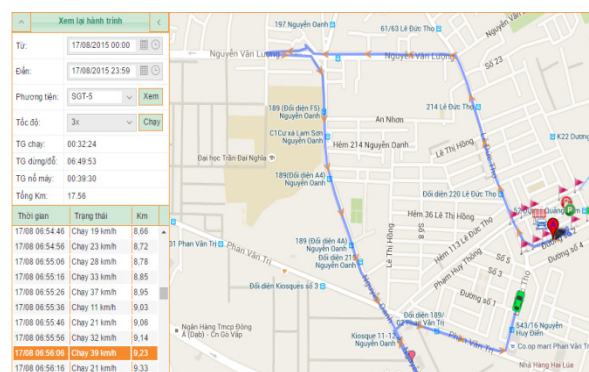
Trong chức năng thứ hai, TBGSHT nhận văn bản từ SMS hoặc từ máy chủ qua giao thức TCP/IP để tổng hợp tiếng nói. Văn

bản cần tổng hợp được lưu vào bộ nhớ của thiết bị, sau đó văn bản sẽ được chuẩn hóa và đưa vào cơ sở dữ liệu để so sánh, tổng hợp và phát ra âm thanh.



Hình 4. Sơ đồ hoạt động tổng quan của TBGSHT

Các thông tin của TBGSHT được gửi lên máy chủ và được xử lý hiển thị trên bản đồ số giúp cho người quản lý có thể quan sát phương tiện trực quan. Hình 5 biểu diễn hành trình của xe trên bản đồ số.



Hình 5. Bản đồ số giám sát TBGSHT

Nội dung văn bản cần tổng hợp tiếng nói có thể được truyền bằng tin nhắn SMS hoặc bằng máy chủ qua giao thức TCP/IP được mô tả như hình 6.



Hình 6. Văn bản tổng hợp được truyền bằng SMS hoặc TCP/IP

Bảng 1 mô tả một vài ví dụ đánh giá mức độ tổng hợp tiếng nói về mức độ tự nhiên và mức độ dễ nghe.

Bảng 1. Đánh giá một số ví dụ về mức độ tổng hợp tiếng nói

| Nội dung | Số ký tự | Đánh giá mức độ tự nhiên và mức độ dễ nghe (%) |
|--|----------|--|
| Máy chủ (Server) | | |
| Cú pháp: \$MTTS_Nội_dung Cấu trúc có tối đa 1024 byte, mỗi ký tự có dấu thay đổi từ 1 đến 3 byte. Vậy trung bình nội dung có tối đa khoảng 512 ký tự. | | |
| Xin chào bạn. | 13 | 97 |
| Chào mừng các bạn tân sinh viên năm 2015. | 41 | 96 |
| Chào mừng các bạn sinh viên trường ĐH Sư Phạm Kỹ Thuật TPHCM. | 61 | 98 |
| Chương trình tổng hợp tiếng nói trên TBGSHT. | 44 | 98 |
| Theo nhiều chuyên gia, Khoa học Máy tính là một chuyên ngành khó vì nó thiên về lý thuyết, học thuật. | 101 | 95 |
| Facebook công bố kế hoạch phóng vệ tinh cung cấp đường truyền Internet cho các khu vực hẻo lánh của châu Phi. | 109 | 97 |
| Tin nhắn SMS | | |
| Cú pháp: C4321VMTTS_Nội_dung Cấu trúc có tối đa 134 byte, mỗi ký tự có dấu là 2 byte, nội dung có tối đa 67 ký tự. | | |
| Xin chào bạn. | 13 | 97 |
| Chào mừng các bạn tân sinh viên năm 2015. | 41 | 96 |
| Chương trình tổng hợp tiếng nói trên TBGSHT. | 44 | 98 |

5. KẾT LUẬN

Đề tài đã hoàn thành các vấn đề đặt ra và được đánh giá như sau:

Tổng hợp tiếng nói tiếng Việt và lập trình nhúng vào TBGSHT:

- Chuẩn hóa văn bản đầu vào, tổng hợp tiếng nói bằng phương pháp ghép nối – lựa chọn đơn vị.
- Viết chương trình tổng hợp tiếng nói trên máy tính có hệ điều hành Linux và trên thiết bị giám sát hành trình.
- Viết chương trình cho TBGSHT đáp ứng đầy đủ các tính năng theo Nghị định của Bộ Giao thông vận tải.
- Tích hợp hệ thống tổng hợp tiếng nói tiếng Việt vào TBGSHT.
- Truyền nội dung văn bản cần tổng hợp bằng tin nhắn SMS hoặc bằng máy chủ qua giao thức TCP/IP.

- So sánh kết quả tổng hợp tiếng nói tiếng Việt trên thiết bị TBGSHT với trên máy tính, thì trên TBGSHT tốc độ chậm hơn trong việc chuyển văn bản thành tệp tin âm thanh, nhưng âm thanh được đọc phát ra là tương đương nhau. Phần cơ sở dữ liệu có thể được đánh giá là lớn hơn so với các phương pháp tổng hợp Formant hay phương pháp ghép nối dùng các diphone.
- Phương pháp này cho chất lượng tiếng nói tốt hơn, với lại kho dữ liệu âm thanh được lưu trong SD card (4GB), và thuận lợi ứng dụng trên thiết bị giám sát hành trình xe ô tô. Tổng hợp tiếng nói được ứng dụng thực tế trên thiết bị giám sát hành trình, xử lý được các từ viết tắt, số, tên riêng và một số địa danh tiếng nước ngoài mà các thuật toán tổng hợp trước đây còn hạn chế [5, 9].

Thiết bị có chức năng giám sát hoạt động của xe:

Thông báo trạng thái hoạt động của thiết bị bằng đèn, còi và có cơ chế tự kiểm tra các trạng thái xe; Nhận dạng, thay đổi lái xe sử dụng công nghệ thẻ RFID, chuẩn ISO/IEC 15693; Cảnh báo lái xe khi xe đang chạy: quá tốc độ, đăng nhập, đăng xuất; Giám sát thời gian lái xe liên tục, lái xe trong ngày và cảnh báo khi lái xe vượt quá thời gian qui định; Đếm số lần, thời gian đóng mở cửa; Đếm số lần, thời gian dừng đỗ; Giao tiếp với phần mềm phân tích dữ liệu của BGTVT qua cổng COM; Hiện thị trạng thái tín hiệu GPS, GSM, kết nối server, thẻ SD-card, lái xe và nguồn cấp; Hỗ trợ đo số km đi được (sai số đơn vị: 0.001km).

Thiết bị có chức năng hỗ trợ cho người quản lý và người sử dụng: SOS khi

cần sự trợ giúp nhanh; Chế độ chống trộm khi xe đỗ/ngủ; Kiểm tra tài khoản trong SIM và tự động cảnh báo khi tài khoản sắp hết; Kiểm tra trạng thái máy điều hòa của xe; Thay đổi các thông số hoạt động của TBGSHT bằng SMS theo cú pháp định sẵn.

LỜI CẢM ƠN

Xin chân thành gửi lời cảm ơn đến Ban lãnh đạo Trung tâm Nghiên cứu và Đào tạo Thiết kế Vi mạch (ICDREC) đã tạo điều kiện cho nhóm tác giả hoàn thành tốt luận văn thạc sĩ này.

Xin chân thành gửi lời cảm ơn đến toàn thể quý thầy cô Trường Đại học Sư phạm Kỹ thuật TP. Hồ Chí Minh đã giảng dạy, hướng dẫn, truyền đạt kiến thức, kinh nghiệm quý báu và tạo mọi điều kiện học tập tốt cho tác giả.

TÀI LIỆU THAM KHẢO

- [1] D. Eastlake, P. Jones, *US Secure Hash Algorithm 1 (SHA1)*, RFC, 2001.
- [2] DO Van Thao, TRAN Do Dat, NGUYEN Thi Thu Trang, *Non-uniform unit selection in Vietnamese Speech Synthesis*, Proceedings of the 2nd SoICT 2011, tr. 165-171, 2011.
- [3] Nguyễn Thị Thanh Mai, *Nghiên cứu các phương pháp nâng cao chất lượng tổng hợp tiếng Việt và thử nghiệm cho phần mềm VnVoice*, LVThS Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội, 2007.
- [4] Phan Thanh Sơn, Phùng Trung Nghĩa, *Một số vấn đề về tổng hợp tiếng nói tiếng Việt*, Hội thảo Quốc gia về Điện tử, Truyền thông và Công nghệ Thông tin, REV-ECIT2014, 2014.
- [5] Phan Thanh Tài, *Tổng hợp tiếng nói tiếng Việt sử dụng mô hình tổng hợp tiếng nói Formant*, LVThS Trường Đại học Sư phạm Kỹ thuật TP.HCM, 2009.
- [6] Qian, Yao và các cộng sự, *A fast table lookup based, statistical model driven non-uniform unit selection TTS*, Proc. in ICASSP2013, Vancouver, Canada, 2013.
- [7] Sakti, Sakriani và các cộng sự, *The Asian Network-based Speech-to-Speech Translation System*, Proc. in Automatic Speech Recognition & Understanding (ASRU), Merano, Italy, tr. 507-512, 2009.
- [8] Trung-Nghia Phung; Chi-Mai Luong và Masato Akagi, *A Hybrid TTS between Unit Selection and HMM-based TTS under limited data conditions*, Proc. in 8th ISCA Speech Synthesis Workshop, Barcelona, Spain, 2013.
- [9] Võ Văn Nguyên, *Tổng hợp tiếng nói sử dụng giải thuật TD_PSOLA*, LVThS Trường Đại học Sư phạm Kỹ thuật TP.HCM, 2013.
- [10] Vũ Hải Quân và Cao Xuân Nam, *Tổng hợp tiếng nói tiếng Việt theo phương pháp ghép nối cụm từ*, Tạp chí CNTT và TT, Tập V-1(1), tr. 70-76, 2009.