|  |
| --- |
| Text to Speech System |
| **Đánh giá chất lượng tiếng nói tổng hợp theo phương pháp MOS Test**  Version: 1.0   |  |  |  |  | | --- | --- | --- | --- | | **Chuẩn bị tài liệu**  **Written by** | **Họ và Tên**  **Full Name** | **Chữ kí**  **Signature** | **Ngày**  **Date** | | **VNPT Technology** | **Nguyễn Thị Lan** |  | **22/02/2016** | |
|  |

Nội dung

|  |
| --- |
|  |

[1. THIẾT KẾ MOST TEST 2](#_Toc444244003)

[1.1 Yêu cầu chung 2](#_Toc444244004)

[1.2 Câu hỏi đánh giá 2](#_Toc444244005)

[1.3 Môi trường test 2](#_Toc444244006)

[1.4 Tiến hành thí nghiệm 2](#_Toc444244007)

[1.5 Đánh giá kết quả 3](#_Toc444244008)

[2. THU THẬP DỮ LIỆU TEST 4](#_Toc444244009)

[2.1 Yêu cầu chung 4](#_Toc444244010)

[2.2 Danh sách câu được chọn 4](#_Toc444244011)

[2.3 Dữ liệu tiếng nói 6](#_Toc444244012)

[2.4 Quy tắc đặt tên file 6](#_Toc444244013)

[3. Danh SÁCH NGƯỜI THAM GIA TEST 6](#_Toc444244014)

[4. KẾT QUẢ BÀI TEST 7](#_Toc444244015)

# THIẾT KẾ MOST TEST

## Yêu cầu chung

Thí nghiệm MOS cho phép chúng ta cho điểm và so sánh chất lượng tổng quát của các hệ thống tổng hợp tiếng so với giọng thật tự nhiên. Người tham gia được yêu cầu đánh giá chất lượng các câu họ được nghe thông qua các câu hỏi và lựa chọn tương ứng. Điều kiện thí nghiệm phải đảm bảo giống nhau với tất cả người nghe.

## Câu hỏi đánh giá

Câu hỏi 1:

*Chất lượng cho tiếng nói mà bạn nghe thấy có thể được đánh giá là?*

Các lựa chọn:

*(5) Tuyệt vời, như giao tiếp thật*

*(4) Tốt, tương đối tự nhiên*

*(3) Khá, có thể giao tiếp được*

*(2) Kém, khó giao tiếp được*

*(1) Tồi tệ, không thể giao tiếp được*

Câu hỏi 2:

*Công sức bạn cần bỏ ra khi nghe để nghe và nắm bắt đoạn tiếng nói là?*

Các lựa chọn:

*(5) Hoàn toàn thư giãn*

*(4) Tương đối thư giãn*

*(3) Cần đôi chút tập trung*

*(2) Cần tương đối tập trung*

*(1) Cần tập trung hoàn toàn*

Các lựa chọn (1) – (5) tương ứng với số điểm sẽ được dùng để tính điểm MOS Test.

## Môi trường test

Bài test được thực hiện trong phòng kín, người nghe sử dụng headphones.

## Tiến hành thí nghiệm

Đặt tất cả các file *txt* và *wav* vào một thư mục đặt tên là *MOS*. Lưu ý tên các file như quy định trong phần 2.4 Quy tắc đặt tên file. Khi được yêu cầu chọn thư mục làm việc gốc thì browse đến thư mục chứa thư mục *MOS*.

Bước 1:

Người nghe bật công cụ thí nghiệm (file jar đính kèm, sử dụng lệnh *java –jar PerceptionTestVSSDC.jar*) và điền các thông tin cơ bản sau đó chuyển tới giao diện MOS Test như sau:



Các câu được phát theo thứ tự hoàn toàn ngẫu nhiên. Số lượng âm tiết của các câu từ 2 đến 30 âm tiết, do đó các câu có độ dài 1 đến 24 giây. Người nghe chỉ được nghe một lần duy nhất.

Sau khi nghe, người nghe tiến hành cho điểm theo hai câu hỏi đánh giá chất lượng và độ dễ nghe của câu vừa nghe bằng cách click vào lựa chọn tương ứng.

Bước 2:

Người nghe cảm thấy ok với các lựa chọn đã chọn, click **LƯU KẾT QUẢ VÀ NGHE CÂU TIẾP THEO**, ngược lại người nghe có thể click **Bỏ qua, quay lại sau** để đánh giá lại câu này sau khi nghe hết tất cả các câu.

Bước 3:

Khi kết thúc bài test, click **Đóng cửa sổ**.

## Đánh giá kết quả

Kết quả bài test được lưu trong file *MOSTest-[Loại giọng]* với cấu trúc như sau:

<?xml version="1.0" encoding="UTF-8"?>

<MOSTest>

<Utterance FileName="VNSP\_00016-ANA.wav" TextContent="Ông bà, cha mẹ cũng chuẩn bị một ít tiền để mừng tuổi con cháu trong nhà và hàng xóm bạn bè thân thích.">

<Result EffortScore="4" GeneralScore="4" NumberOfListen="1"

RecordTime="1455867871204" SubjectId="4"/>

</Utterance>

<Utterance FileName="VNSP\_00014-ANA.wav" TextContent="Trong những ngày vui này còn có tiết mục mừng tuổi chúc Tết.">

<Result EffortScore="3" GeneralScore="3" NumberOfListen="1"

RecordTime="1455867889630" SubjectId="4"/>

</Utterance>

</MOSTest>

Điểm đánh giá độ tự nhiên là điểm trung bình cộng EffortScore của của câu hỏi 1 trên tất cả người nghe.

Điểm đánh giá độ dễ nghe là điểm trung bình cộng GeneralScore của câu hỏi 2 trên tất cả người nghe.

# THU THẬP DỮ LIỆU TEST

## Yêu cầu chung

Dữ liệu phải bao phủ được đầy đủ các trường hợp tổng hợp như: chứa các ký tự đặc biệt về đơn vị đo, ngày tháng, số, chữ viết tắt. Độ dài các câu phải đa dạng.

## Danh sách câu được chọn

|  |  |  |
| --- | --- | --- |
| *STT* | *Tên file txt* | *Nội dung* |
| *1* | *VNSP\_00014.txt* | *Trong những ngày vui này còn có tiết mục mừng tuổi chúc Tết.* |
| *2* | *VNSP\_00016.txt* | *Ông bà, cha mẹ cũng chuẩn bị một ít tiền để mừng tuổi con cháu trong nhà và hàng xóm bạn bè thân thích.* |
| *3* | *VNSP\_00030.txt* | *Có nhiều thứ tôi được hoá thân vào số phận các nhân vật khác nhau.* |
| *4* | *VNSP\_00044.txt* | *Điều này tưởng chừng như đơn giản nhưng không phải dễ nhận ra nhất là khi mình thành công lúc còn quá trẻ.* |
| *5* | *VNSP\_00059.txt* | *Kiểu đi lắc lư ngộ nghĩnh của chim cánh cụt chính là một cách thức thông minh để đạt được điều đó.* |
| *6* | *VNSP\_00061.txt* | *Loài chim này thường phải đi bộ tới hơn 100km qua Nam Cực băng giá để tìm ra các vùng nước có cá, do vậy chúng cần tiết kiệm năng lượng tối đa.* |
| *7* | *VNSP\_00069.txt* | *Hết cam, mía lại đến tiêu, điều, cafe.* |
| *8* | *VNSP\_00087.txt* | *Cậu hay ăn hiếp những bạn yếu thế hơn mình và biết cách luồn lách sao cho có lợi về mình.* |
| *9* | *VNSP\_00096.txt* | *Trách nhiệm đó thuộc về ai?* |
| *10* | *VNSP\_00106.txt* | *Có trải qua mới biết nhiều cô gái thời nay dễ dãi trên mức bình thường.* |
| *11* | *VNSP\_00109.txt* | *Tiến sĩ Nguyễn Đức Nghĩa, trưởng ban đào tạo Đại học Quốc Gia thành phố HCM cho biết Đại học Quốc Gia chưa đặt ra ngưỡng là do các biện pháp kỹ thuật (hồ sơ thí sinh, kiểm tra hồ sơ) không đảm bảo và phải có dư luận tốt.* |
| *12* | *VNSP\_00186.txt* | *Do vậy, muốn nâng được một trọng lượng nhất định thì chỉ còn cách là làm rộng hết cỡ có thể diện tích cánh máy bay để có được lực nâng cần thiết.* |
| *13* | *VNSP\_00260.txt* | *Bởi trước khi rời nước anh chị đã ký khế ước với quốc gia Việt Nam là sẽ về nước phục vụ sau khi tốt nghiệp.* |
| *14* | *VNSP\_00287.txt* | *Còn công đoàn thì cho rằng họ đơn giản chỉ phản đối cách làm của giới chủ vốn đang phớt lờ mọi nguyên tắc an toàn trong công việc.* |
| *15* | *VNSP\_00292.txt* | *Hồi 16h chiều ngày 8/12, vị trí tâm bão ở vào khoảng 13,4 độ Vĩ Bắc; 112 độ Kinh Đông, trên vùng biển phía Bắc quần đảo Trường Sa; cách bờ biển Nam Trung Bộ khoảng 300 km về phía Đông.* |
| *16* | *VNSP\_00318.txt* | *Khi nào em không yêu anh nữa em sẽ nói thẳng với anh.* |
| *17* | *VNSP\_00375.txt* | *Nhưng mùa thu mưa nhiều quá, mùa đông thì lạnh còn mùa hè thì trời lại quá nóng.* |
| *18* | *VNSP\_00396.txt* | *Cám ơn ông.* |
| *19* | *VNSP\_00400.txt* | *Xin lỗi ông, cái này là cái gì?* |
| *20* | *VNSP\_00401.txt* | *Bà không nhìn ra à?* |
| *21* | *VNSP\_00406.txt* | *Đàn ông gì mà giống khúc củi quá!* |
| *22* | *VNSP\_00445.txt* | *Chà, đẹp quá.* |
| *23* | *VNSP\_00446.txt* | *Nhà này rộng bao nhiêu mét?* |
| *24* | *VNSP\_00457.txt* | *Xa quá!* |
| *25* | *VNSP\_00462.txt* | *Dạ, hay đấy anh ạ.* |
| *26* | *VNSP\_00464.txt* | *Mình rủ thêm Thảo và Nghĩa đi!* |
| *27* | *VNSP\_00471.txt* | *À, anh muốn đến nhà em chơi.* |
| *28* | *VNSP\_00496.txt* | *Ngon lắm, đầu mùa nhưng rất ngọt.* |
| *29* | *VNSP\_00499.txt* | *Lần sau chị mua nữa nhé.* |
| *30* | *VNSP\_00502.txt* | *Thế em đã biết nấu chưa?* |
| *31* | *VNSP\_00523.txt* | *Nước cốt dừa có cần không?* |

## Dữ liệu tiếng nói

Dữ liệu để test bao gồm hai loại:

*Tiếng nói tổng hợp*: tiếng nói đầu ra thu được từ hệ thống tổng hợp tiếng, gồm có Ailab và VBEE.

*Tiếng nói tự nhiên*: tiếng nói thu bởi giọng thật được dùng làm tham chiếu khi đánh giá kết quả test.

## Quy tắc đặt tên file

Các file tiếng nói tổng hợp của VBEE đặt tên theo định dạng: *[Tên file text]-[VSY].wav*. Ví dụ: file text đầu vào *VNSP\_00523.txt* thì file tổng hợp của hệ thống của VBEE sẽ tên là *VNSP\_00523-VSY.*

Các file tiếng nói tổng hợp của Ailab đặt tên theo định dạng: *[Tên file text]-[ASY].wav*. Ví dụ: file text đầu vào *VNSP\_00523.txt* thì file tổng hợp của hệ thống của Ailab sẽ tên là *VNSP\_00523-ASY.*

Các file tiếng nói tự nhiên dùng để tham khảo đặt tên theo định dạng: *[Tên file text]-[NAT].wav*. Ví dụ: file text đầu vào *VNSP\_00523.txt* thì file sẽ tên là *VNSP\_00523-NAT.wav.*

# Danh SÁCH NGƯỜI THAM GIA TEST

Số lượng người tham gia: 20 người, trong đó 10 nam và 10 nữ.

|  |  |  |
| --- | --- | --- |
| *STT* | *Họ tên* | *Thông tin* |
| *1* | *Đặng Trung Hiếu* | *Nam, 24 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *2* | *Đào Hải Đăng* | *Nam, 24 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *3* | *Đinh Hoàng Hải* | *Nam, 24 tuổi, Nam Định, Đại học, Kỹ sư* |
| *4* | *Lê Đại Cát* | *Nam, 24 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *5* | *Lê Văn Trung* | *Nam, 25 tuổi, Hà Tĩnh, Đại học, Kỹ sư* |
| *6* | *Mai Quốc Khánh* | *Nam, 25 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *7* | *Ngô Duy Tiệm* | *Nam, 27 tuổi, Nam Định, Đại học, Kỹ sư* |
| *8* | *Ngô Hải Linh* | *Nam, 25 tuổi, Hải Dương, Trên Đại học, Kỹ sư* |
| *9* | *Nguyễn Văn Lương* | *Nam, 23 tuổi, Vĩnh Phúc, Đại học, Kỹ sư* |
| *10* | *Nguyễn Văn Thoan* | *Nam, 25 tuổi, Hưng Yên, Đại học, Kỹ sư* |
| *11* | *Trần Xuân Hiếu* | *Nam, 28 tuổi, Quảng Ninh, Trên Đại học, Kỹ sư* |
| *12* | *Nguyễn Hồng Hạnh* | *Nữ, 37 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *13* | *Nguyễn Thị Hường* | *Nữ, 23 tuổi, Vĩnh Phúc, Đại học, Kỹ sư* |
| *14* | *Phạm Minh Huệ* | *Nữ, 27 tuổi, Vĩnh Phúc, Đại học, Kỹ sư* |
| *15* | *Phạm Thái Linh* | *Nữ, 35 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *16* | *Phạm Thị Mai Hoa* | *Nữ, 27 tuổi, Hà Nội, Đại học, Kỹ sư* |
| *17* | *Văn Thị Thu Hà* | *Nữ, 24 tuổi, Vĩnh Phúc, Đại học, Kỹ sư* |
| *18* |  |  |
| *19* |  |  |
| *20* |  |  |

# KẾT QUẢ BÀI TEST

Lưu trữ:

Dữ liệu kết quả của 17 người tham gia test được lưu trong thư mục *Results*. Điểm test của các giọng được lưu tương ứng trong các file text dưới thư mục của mỗi người, nội dung file theo format: *GeneralScore\_EffortScore*

Thống kê:

Điểm MOS trung bình cho độ tự nhiên của các giọng chênh lệch khá lớn, giọng thật tự nhiên với số điểm 4.24 cao hơn so với giọng tổng hợp của VBEE ~ 0.3 điểm và so với giọng tổng hợp của Ailab là ~1.6 điểm.

Điểm MOS trung bình cho độ dễ nghe của các giọng chênh lệch khá lớn, giọng thật tự nhiên với số điểm 4.43 cao hơn so với giọng tổng hợp của VBEE ~ 0.3 điểm và so với giọng tổng hợp của Ailab là ~1.4 điểm.

Điểm MOS độ tự nhiên của VBEE lớn hơn Ailab ~1.2 điểm, độ dễ nghe của VBEE lớn hơn Ailab ~1.2 điểm. Độ chênh lệch khá lớn một phần có thể do ảnh hưởng bởi yếu tố vùng miền, tuy nhiên dựa vào độ chênh lệch khá nhỏ giữa giọng thật tự nhiên và giọng tổng hợp của VBEE thì có thể khẳng định chất lượng của VBEE là tương đối tốt.

Bảng 1 Điểm MOS test của tất cả người tham gia

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Name | VNA | | VSY | | ASY | |
|  |  | GS | ES | GS | ES | GS | ES |
| 1 | DangTrungHieu | 3.9 | 4.06 | 4 | 4.06 | 2.48 | 3.1 |
| 2 | DaoHaiDang | 4.23 | 4.32 | 3.52 | 3.61 | 1.97 | 1.77 |
| 3 | DinhHoangHai | 4.42 | 3.77 | 4.13 | 3.74 | 3.87 | 3.52 |
| 4 | LeDaiCat | 3.84 | 3.97 | 3.42 | 3.39 | 2.26 | 2.32 |
| 5 | LeVanTrung | 4.32 | 4.65 | 4.16 | 4.45 | 2.03 | 2.45 |
| 6 | MaiQuocKhanh | 4.65 | 4.94 | 4.19 | 4.74 | 3.71 | 4.1 |
| 7 | NgoDuyTiem | 4.06 | 4.71 | 3.9 | 4.68 | 3.32 | 4.19 |
| 8 | NgoHaiLinh | 4.06 | 4.74 | 4.06 | 4.68 | 2.65 | 2.71 |
| 9 | NguyenVanLuong | 4.77 | 4.26 | 4.32 | 4.1 | 3.06 | 3.06 |
| 10 | NguyenVanThoan | 4.16 | 4.1 | 3.9 | 3.68 | 1.71 | 1.74 |
| 11 | TranXuanHieu | 3.87 | 3.68 | 3.55 | 3.45 | 2.03 | 2 |
| 12 | NguyenHongHanh | 4.16 | 4.48 | 4 | 4.39 | 2.77 | 2.65 |
| 13 | NguyenThiHuong | 3.84 | 4.58 | 3.55 | 4.52 | 2.74 | 3.23 |
| 14 | PhamMinhHue | 4.81 | 5 | 4.23 | 4.97 | 2.1 | 3.26 |
| 15 | PhamThaiLinh | 4.13 | 4.55 | 3.94 | 4.45 | 3.39 | 3.87 |
| 16 | PhamThiMaiHoa | 4.81 | 5 | 4.48 | 4.84 | 2.74 | 4.61 |
| 17 | VanThiThuHa | 4.13 | 4.55 | 2.94 | 3.35 | 1.94 | 1.74 |
|  | **Average** | **4.244706** | **4.432941** | **3.899412** | **4.182353** | **2.633529** | **2.96** |
|  |  |  |  |  |  |  |  |
|  | VNA: Giọng thật tự nhiên | |  |  |  |  |  |
|  | VSY: Tổng hợp từ VBEE | |  |  |  |  |  |
|  | ASY: Tổng hợp từ Ailab | |  |  |  |  |  |
|  | GS: Điểm đo chất lượng độ tự nhiên | | |  |  |  |  |
|  | ES: Điểm đo chất lượng độ dễ nghe | | |  |  |  |  |

Hình 1 Điểm MOS cho độ tự nhiên của các giọng

Hình 2 Điểm MOS độ tự nhiên trung bình của các giọng

Hình 3 Điểm MOS cho độ dễ nghe của các giọng

Hình 4 Điểm MOS trung bình độ dễ nghe của các giọng