# *Developing Music Technology for Health and Learning*

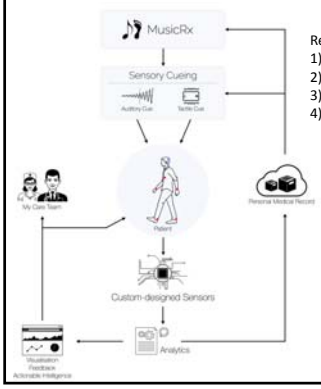
Ye WANG
School of Computing
NUS Graduate School for Integrative Sciences and Engineering
National University of Singapore

www.smcnus.org

## Outline

- **Music & Wearable Technology for Health** *(MusicRx)*
- **Music Technology to Motivate Foreign Language Learning** *(SLIONS related)*
  1. NUS-48E corpus
  2. Lexical Novelty Score (LNS)
  3. Intelligibility of Sung Lyrics (IoSL)
  4. Pronunciation Evaluation of Sung Lyrics
  5. Perceptual Evaluation of Singing Quality (PESnQ)

## Music & Wearable Technology for Health (MusicRx)




Research Problems:
1) iRACE (2014)
2) Auditory Tempo Stability (2014)
3) Domain specific music recommendation
4) Domain specific music composition


**iRACE**

**An iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease**

4

## Music Technology to Motivate Foreign Language Learning (SLIONS Karaoke)

Research Problems:
1) lyric complexity (ISMIR 2015)
2) Singing voice intelligibility (ISMIR 2017)
3) Singing-to-text transcription (ISMIR2017)
4) Domain specific song recommendation



*Subproject 1:*

The NUS Sung and Spoken Lyrics Corpus (NUS-48E): A Quantitative Comparison of Singing and Speech

Zhiyan Duan, Haotian Fang, Bo Li, Khe Chai Sim and Ye Wang

## Creation of an Annotated Database for Comparative Analysis of Singing and Speech

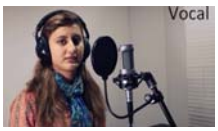
SUNG

SPOKEN

"Edelweiss, edelweiss
Every morning you greet me"

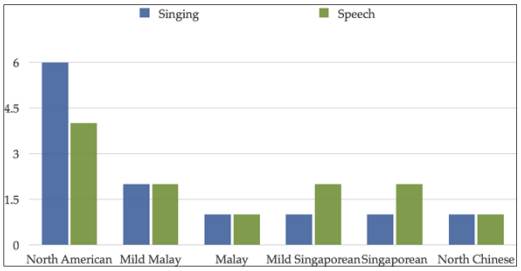
## Participants, and Song Selection

- 6 males, 6 females
- Varying levels of vocal training experience (0 – 10+ years)
- Soprano, alto, tenor, baritone and bass

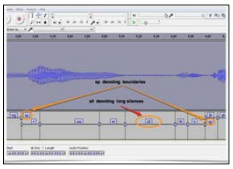



- Phonetic balance (140 ~ 980 phoneme per song)
- Tempo balance (68 ~ 150 bpm)
- Popularity
- Ease of learning

## Subjects – A Wide Range of Accents




## Annotation – Identifying Individual Phonemes, based on CMU's Dictionary




We use Audacity to annotate the sound files.

- Annotated spoken/sung tracks: 4 (4 tracks per subject)
- Total Length: 169 mins
- Phoneme Count: 25,474
- Spoken data: alignment of labels from sung data


*Subproject 2:*
Quantifying Lexical Novelty in Song Lyrics

Robert J Ellis, Zhe Xing, Jiakun Fang, & Ye Wang

## Motivation


- Second-language acquisition
  - The complexity of the lyric should be matched to the level of the learner
  - A search engine that enables finding lyrics based upon topic, mood, and lyric complexity could be used to facilitate L2 learning programs

## Quantify Lyric Novelty for Music Recommendation?


No more. No way. No more. I say.
You do it. I don't. You will. I won't.

— "Stop It" by Nomeansno (1985)

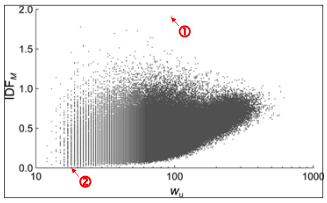
Herodias and I have led a phantom cavalcade
Through veiled and pagan history where superstitions reigned
And Christendom sought to pervert

— "Haunted Shores" by Cradle of Filth (1996)

## Approaches to Lyric Novelty Analysis

- A common text processing technique is to use the inverse document frequency (IDF).
- We explored this and came up with improvement relevant to lyrics.
- An issue we addressed was scaling.
- We observed data via visualization.

## First-pass LNS: $IDF_M$



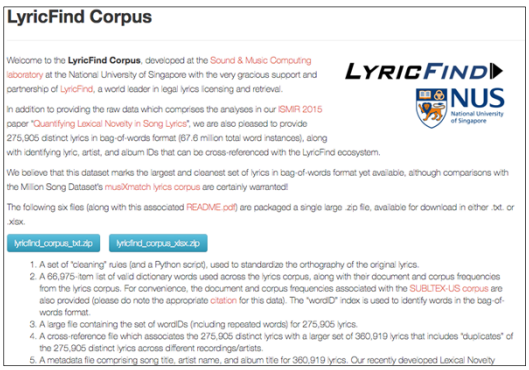

① There's Syria, Lebanon, Israel, Jordan
Both Yemens, Kuwait, and Bahrain
The Netherlands, Luxembourg, Belgium, and Portugal
France, England, Denmark, and Spain
— "Yakko's World" from *Anamaniacs* (1993)

② I can't think straight
Help me now before it's too late
Now what do I care?
'Cause we're going nowhere

— "Going Nowhere" by Cut Copy (2004)

## NUS-48E Lyric Corpus

http://www.smcnus.org/lyrics/



**LyricFind Corpus**

Welcome to the **LyricFind Corpus**, developed at the Sound & Music Computing laboratory at the National University of Singapore with the very gracious support and partnership of LyricFind, a world leader in legal lyrics licensing and retrieval.

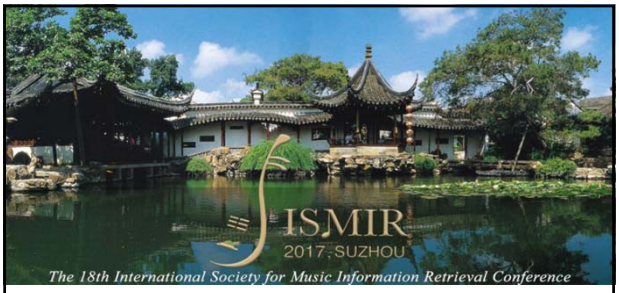
In addition to providing the raw data which comprises the analyses in our ISMIR 2015 paper "Quantifying Lexical Novelty in Song Lyrics", we are also pleased to provide 275,905 distinct lyrics in bag-of-words format (67.6 million total word instances), along with identifying lyric, artist, and album IDs that can be cross-referenced with the LyricFind ecosystem.

We believe that this dataset marks the largest and cleanest set of lyrics in bag-of-words format yet available, although comparisons with the Million Song Dataset's musiXmatch lyrics corpus are certainly warranted!

The following six files (along with this associated README.pdf) are packaged a single large .zip file, available for download in either .txt. or .xlsx.

lyricfind_corpus_txt.zip  lyricfind_corpus_xlsx.zip

1. A set of "cleaning" rules (and a Python script), used to standardize the orthography of the original lyrics.
2. A 66,975-item list of valid dictionary words used across the lyrics corpus, along with their document and corpus frequencies from the lyrics corpus. For convenience, the document and corpus frequencies associated with the SUBLTEX-US corpus are also provided (please do note the appropriate citation for this data). The "wordID" index is used to identify words in the bag-of-words format.
3. A large file containing the set of wordIDs (including repeated words) for 275,905 lyrics.
4. A cross-reference file which associates the 275,905 distinct lyrics with a larger set of 360,919 lyrics that includes "duplicates" of the 275,905 distinct lyrics across different recordings/artists.
5. A metadata file comprising song title, artist name, and album title for 360,919 lyrics. Our recently developed Lexical Novelty




*Subproject 3:*
Intelligibility of Sung Lyrics: A Pilot Study

Karim M. Ibrahim, David Grunberg, Kat Agres, Chitralekha Gupta and Ye Wang

## Automatic Assessment of Intelligibility for Language Learning

## Approach

Dataset collection → Dataset Labelling → Features Selection → Model Training → Evaluation

1. Collect a dataset using five genres
2. Estimate intelligibility of the songs according to human perception
3. Extract feature set that reflects clarity of song
4. Train Support Vector Machine
5. Correlating model estimation to users' ratings.

Dataset collection → Dataset Labelling → Features Selection → Model Training → Evaluation

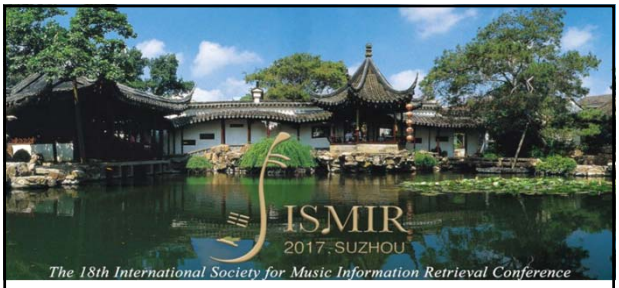
- 3 classes of interpretability
- SVM Classification accuracy: 66%

Confusion matrix of SVM output

| | High | Moderate | Low |
|---|---|---|---|
| High | 33 | 9 | 1 |
| Moderate | 10 | 30 | 2 |
| Low | 4 | 8 | 3 |

## Applications

- Language immersion is important for learning a foreign language
- Recommending music based on intelligibility for learning purposes may aid in motivation
- We intend to make our dataset available to the research community


*Subproject 4:*

## Towards automatic mispronunciation detection in singing

Chitralekha Gupta, David Grunberg, Preeti Rao, Ye Wang

## Overview

- **Learning a second language (L2) through singing** is shown to be effective and is used in pedagogy
- Automatic pronunciation evaluation of singing is desirable for L2 learning
- But finding training data is challenging


## Problem Statement

- **Automatic pronunciation error detection in South-East Asian English accents singing (Malaysian: M, Indonesian: I, Singaporean: S)** :

➢ What are the error patterns observed in non-native singing compared to non-native speech?

➢ If only native English speech models are available, can we detect pronunciation errors in non-native English singing, given that we know the singer's L1 (native language)?

24

## Error patterns in South-East Asian English accents

From speech analysis literature

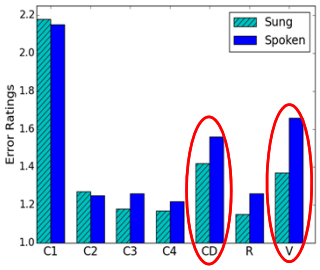
| ID | Error | Examples |
|---|---|---|
| C1 | /dh/→/d/ | thy → die; mother → moder |
| C2 | /th/→ /t/ | thought → taught; nothing → noting |
| C3 | /t/→ /th/ | to→thu; sitting → sithing |
| C4 | /d/→ /dh/ | dear → dhear |
| CD | Word-end consonant deletion | moment → momen |
| R | Rolling /r/ | ray → rray |
| V | vowel error | fool→full; sleeping→slipping |

➢ **Are all of these error patterns also observed in singing?**

## Subjective analysis

**Dataset**

- 26 sung and 26 spoken songs by 8 unique subjects (4M, 4F) - 3 Indonesian, 3 Singaporean, and 2 Malaysian
- All of the error patterns were subjectively rated by 3 English speaking judges

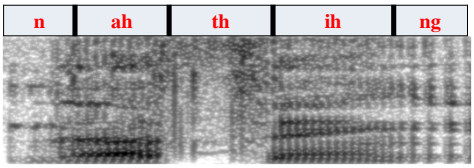



**Findings:**

- **Consonant Deletion** and **Vowel errors** are significantly **lower in singing** than in speech
- **Key Insight:** Only a subset of the error patterns that occur in speech occur in singing - **suggests a possible learning strategy**

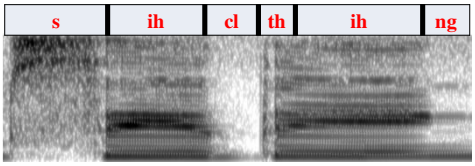
## Mispronunciation detection with limited data

### Sub-phonetic Modeling




**American**
"Nothing"

Dental Fricative /th/



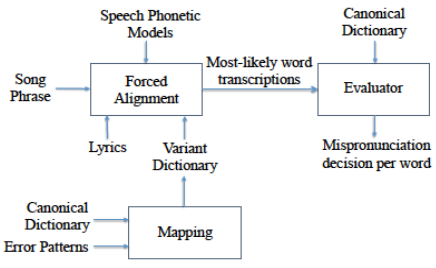

**Indonesian**
"Sitting" → "sithing"

Dental Stop /closure/+/th/

27

## System Overview

Converted all pronunciation patterns into a dictionary of words with acceptable and unacceptable pronunciation variants for LEX method



## Results

| | Dictionary A | Dictionary B |
|---|---|---|
| Definition | only American English phones (L2) | American phones+modified (L1-adapted) phone |
| Example | /th/ | /th/, /cl/+/th/ |
| F-score for M & S | 0.63 | **0.67** |
| F-score for I | 0.33 | **0.47** |

29

## Contributions

- We derive the error patterns in singing compared to speech in South-East Asian English accents and obtain mispronunciation rules for singing
- Combine acoustic models of sub-phonetic segments to represent missing L1 phone models
- Incorporate the above two methods in ASR framework to detect mispronunciation in singing

## Application

- Automated pronunciation analysis alongside singing may be useful for language learning


*Subproject 5:*

## Perceptual Evaluation of Singing Quality

(PESnQ)

Chitralekha Gupta, Haizhou Li, Ye Wang

## Goal

To develop a **perceptually-valid score** for evaluating singing quality

## Motivation

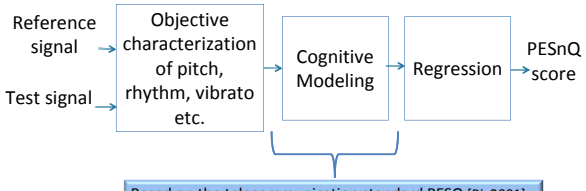
Such a score could serve as
- a complement to singing lessons
- make singing training more accessible to learners

## How do experts perceptually evaluate singing quality?

Reference Good Bad

- Rhythm Consistency
- Intonation Accuracy
- Appropriate Vibrato
- Voice Quality
- Pitch Dynamic Range
- Pronunciation

## PESnQ Formulation

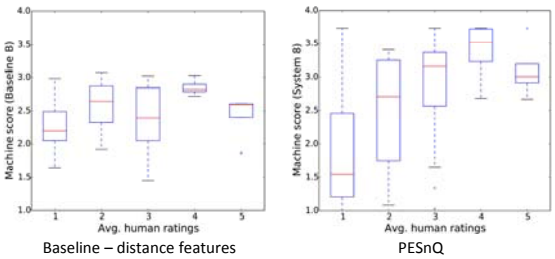



Based on the telecommunication standard PESQ [Rix2001]:

***a localized error in time has a larger subjective impact than a distributed error***

35

## Results




Baseline – distance features PESnQ

Adopting **the cognitive modeling theory** of **PESQ** to design a **PESnQ** score shows **96%** improvement over baseline scores in correlating with the music-expert human judges

36

## Putting all together: SLIONS Project

**S**inging and
**L**istening to
**I**mprove
**O**ur
**N**atural
**S**peaking

---

**The efficacy of singing in foreign-language learning**

Psychology of Music
2015, Vol. 43(5) 627–640

DOI: 10.1177/0305735614528833
pom.sagepub.com
SAGE

**Arla J. Good[1], Frank A. Russo[1] and Jennifer Sullivan[2]**

**Abstract**

This study extends the popular notion that memory for text can be supported by song to foreign-language learning. Singing can be intrinsically motivating, attention focusing, and simply enjoyable for learners of all ages. The melodic and rhythmic context of song enhances recall of native text; however, there is limited evidence that these benefits extend to foreign text. In this study, Spanish-speaking Ecuadorian children learned a novel English passage for 2 weeks. Children in a sung condition learned the passage as a song and children in the spoken condition learned the passage as an oral poem. Children were tested on their ability to recall the passage verbatim, pronounce English vowel sounds, and translate target terms from English to Spanish. As predicted, children in the sung condition outperformed children in the spoken condition in all three domains. The song advantage persevered after a 6-month delay. Findings have important implications for foreign language instruction.

---

**The efficacy of singing in foreign-language learning**

Psychology of Music
2015, Vol. 43(5) 627–640

DOI: 10.1177/0305735614528833
pom.sagepub.com
SAGE

Singing can be intrinsically **motivating**, **attention focusing**, and simply **enjoyable** for learners of all ages.

**Abstract**

This study extends the popular notion that memory for text can be supported by song to foreign-language learning. Singing can be intrinsically motivating, attention focusing, and simply enjoyable for learners of all ages. The melodic and rhythmic context of song enhances recall of native text; however, there is limited evidence that these benefits extend to foreign text. In this study, Spanish-speaking Ecuadorian children learned a novel English passage for 2 weeks. Children in a sung condition learned the passage as a song and children in the spoken condition learned the passage as an oral poem. Children were tested on their ability to recall the passage verbatim, pronounce English vowel sounds, and translate target terms from English to Spanish. As predicted, children in the sung condition outperformed children in the spoken condition in all three domains. The song advantage persevered after a 6-month delay. Findings have important implications for foreign language instruction.

---

**The efficacy of singing in foreign-language learning**

Psychology of Music
2015, Vol. 43(5) 627–640

DOI: 10.1177/0305735614528833
pom.sagepub.com
SAGE

Children were tested on their abilities to **recall** the passage verbatim, **pronounce** English vowel sounds, and **translate** target terms from English to Spanish. As predicted, children in the sung condition outperformed children in the spoken condition.

The song advantage preserved after a **6-month delay**.

**Abstract**

This study extends the popular notion that memory for text can be supported by song to foreign-language learning. Singing can be intrinsically motivating, attention focusing, and simply enjoyable for learners of all ages. The melodic and rhythmic context of song enhances recall of native text; however, there is limited evidence that these benefits extend to foreign text. In this study, Spanish-speaking Ecuadorian children learned a novel English passage for 2 weeks. Children in a sung condition learned the passage as a song and children in the spoken condition learned the passage as an oral poem. Children were tested on their ability to recall the passage verbatim, pronounce English vowel sounds, and translate target terms from English to Spanish. As predicted, children in the sung condition outperformed children in the spoken condition in all three domains. The song advantage persevered after a 6-month delay. Findings have important implications for foreign language instruction.

## SLIONS Karaoke Prototyping


## Acknowledgements


- Zhiyan Duan
- Zhe Xing
- Jiakun Fang
- Karim M. Ibrahim
- Chitralekha Gupta
- Dania Murad
- Michael Barone
- Rob Ellis
- David Grunberg
- Kat Agres
- Douglas Turnbull


**Thank you!**

ISMIR
2017, SUZHOU, CHINA


Ye Wang
wangye@comp.nus.edu.sg
www.smcnus.org