

# Deeply Parent-Child Vocal Interaction Dataset

## Summary

동화책 읽기, 동요 부르기, 대화 등 부모와 자녀 24쌍(총 48명)의 상호작용이 기록되어 있습니다. 녹음은 잔향 수준이 다른 무향실, 원룸, 댄스 스튜디오 등 3가지 장소에서 이루어졌다. 그리고 음원으로 부터의 마이크의 거리와 녹음 기기의 영향을 알아볼 수 있도록, 모든 실험은 3개의 다른 거리에서, 2종류의 스마트폰을 이용해서 진행되었다.

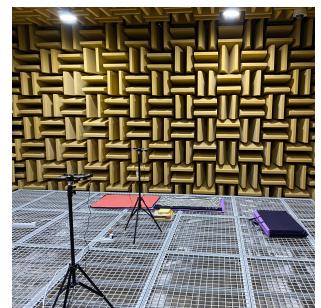
Recording contents	부모와 자녀의 상호 작용(동요 부르기, 동화 읽기, 대화 등)
Recording environment	무향실(잔향 없음), 원룸(중간 잔향), 댄스 스튜디오(높은 잔향)
Device	<u>iPhone X</u> (iOS), <u>Samsung Galaxy S7</u> (Android)
Distance from the source	0.4m, 2.0m, 4.0m
Volume	~ 282 hours, ~ 360,000 utterances, ~ 110 GB
Format	wav(44100Hz, 16-bit, mono), or h5(16000Hz, 16-bit, mono)
Language	한국어
Demographics	24명의 부모는, 남성이 17%, 여성이 83%이며, 12.5%가 20대, 62.5%가 30대, 25%가 40대입니다. 24명의 자녀는, 남성이 46%, 여성이 54%이며, 21%가 만 1~2세, 54%가 만 3~4세, 25%가 만 5~6세입니다.



<Fig 1. Studio apartment>



<Fig 2. Dance studio>



<Fig 3. Anechoic chamber>

## What's inside the children interaction dataset?

부모-자녀 음성 상호작용 데이터 세트는 부모와 자식 간의 서로 다른 유형의 상호작용을 보여주는 281.3시간의 오디오 클립으로 구성됩니다. 참가자들은 3 종류의 장소(무반향실, 원룸, 댄스 스튜디오)에서 반복 녹음할 수 있도록 권장됐으며, 모든 녹음은 3 종류의 거리(0.4m, 2.0m, 4.0m)에서 2 종류의 기기(iPhone X, 갤럭시 S7)를 통해 진행됐다.

말하기의 종류는 다음과 같이 분류된다: 부모가 '노래 하기' '동화책 읽기' '기타 말하기'으로 3개 그룹으로 나뉘며, 노래부르는 범주에는 동요와 자장가를 부르는 것이 포함되며, 기타 말하기는 문자 그대로 동화책 읽기와 노래 이외의 모든 다른 말들을 포함하고 있으며, 대부분 자녀와 즉흥적인 대화이다. 그러나 어린이의 경우는 '노래 하기', '동화책 읽기', '울기', '거부하기', '기타 말하기' 등 5개 그룹으로 분류된다. 노래 하기, 동화책 읽기, 기타 말하기는 부모의 그것과 같으며, '울기'는 아이가 녹음 도중에 울음을 터뜨린 경우, '거부하기'는 아이들이 특정 주제나 상호작용대신에 다른 것을 하고 싶어서 칭얼대는 등의 경우입니다.

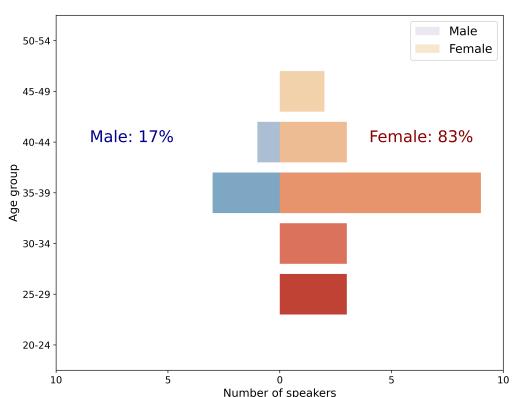
데이터 세트에는 화자, 연령, 성별, 노이즈, 장소 유형, 거리 및 장치와 같은 메타데이터도 포함됩니다.

## Information & Statistics

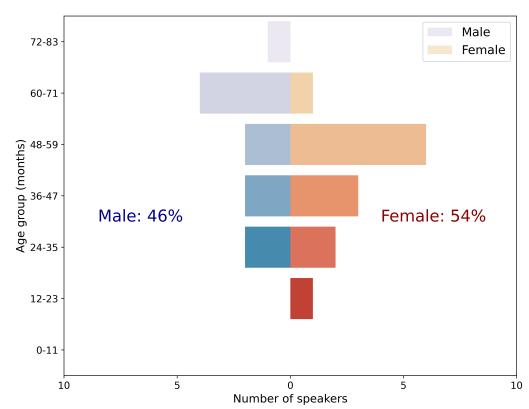
Figures 4 and 5는 각각 부모와 아이들의 인구통계학적 정보를 보여준다. 부모 참여자는 17%가 남성, 83%가 여성으로 구성되며, 남성의 75%, 여성의 60%가 30대이다. 그리고 어린이 참여자는 46%가 남성, 54%가 여성으로 구성되며, 남성은 전 연령대에 거의 고르게 분포하고 있으며, 여성의 46%는 4세 등으로 구성되어 있다.

Figures 6 and 7은 각 클래스에 따른 발화의 길이 분포와 각 클래스에 따른 각 발화의 평균 길이를 보여줍니다. Figure 6에서 볼 수 있듯이, 대부분의 부모의 발언 시간은 10초 미만이고 대부분의 아이들의 발언 시간은 5초 미만이다. Figure 7은 부모가 자녀보다 평균적으로 더 길게 발화하는 것을 보여준다.

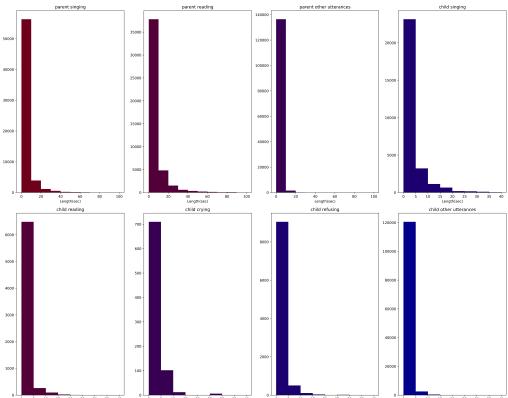
Figure 8은 각 발화 유형의 총 길이(시간)를 보여줍니다. 각 막대를 구성하는 다른 색은은 발화가 monophonic 한지 polyphonic 한지, 다시 말해서, 한 명의 화자가 발화하고 있는지, 아니면 두 명 이상의 화자가 발화하고 있는지 여부를 나타냅니다(본 데이터셋에선 최대 두 명). 아래 그림과 같이, 아이가 노래하는 경우를 제외하면 대부분 monophopic한 것을 확인할 수 있고, 부모의 발화량이 더 많은 것으로 미루어보아 실제 상황에서 부모가 대부분 대화 혹은 상호작용을 주도한다는 것을 알 수 있다.



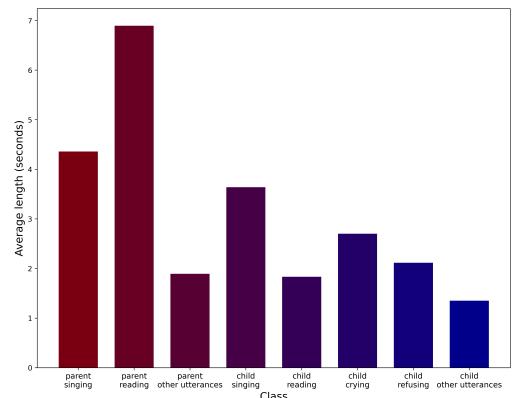
<Fig 4. Age distribution of parents by sex>



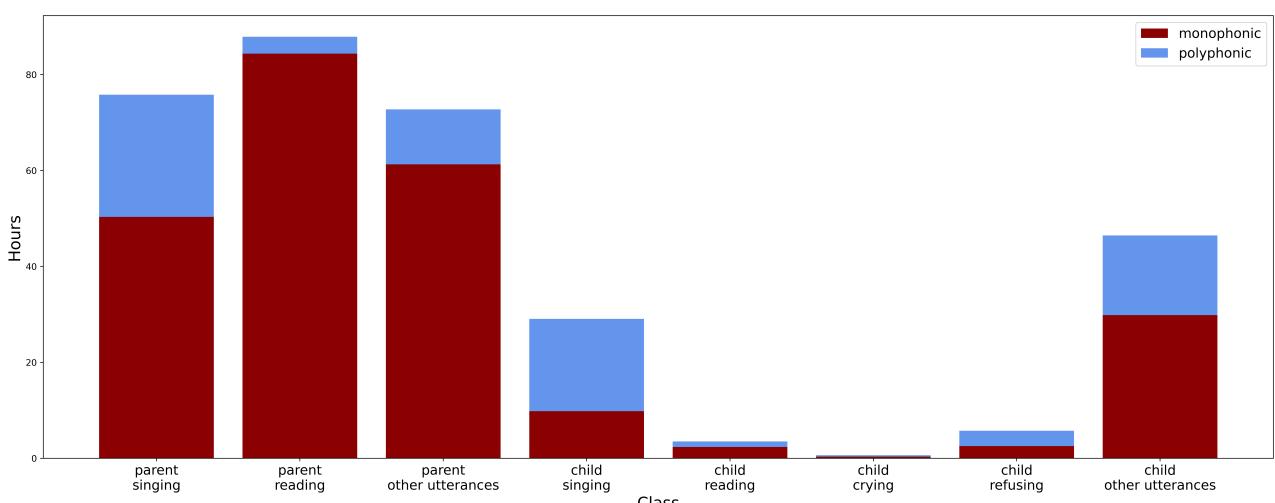
<Fig 5. Age distribution of children by sex>



<Fig 6. Length distribution of each utterance by class>



<Fig 7. Average length of each utterance by class>



<Fig 8. Total length(hours) by class>

## **Filename convention**

Recorded wav files are named under following format:

{subject ID}\_{yyyy}\_{mm}\_{dd}\_{sex\_a}{sex\_b}\_{age\_a}{age\_b}\_{location}  
\_{distance}\_{device}.wav

Example: sub2001\_2020\_11\_29\_00\_22\_0\_0.wav

**Subject ID** is a unique 4-digit alphanumeric code representing speaker pair, **{yyyy} {mm}**  
**{dd}** is a date of recording, **sex a** is a digitized code indicating the sex of speaker\_a(parent),  
**sex b** is a digitized code indicating the sex of speaker\_b(child), **age a** is indicating the first  
digit of the age group of speaker\_a(parent), **age b** is indicating the age of speaker\_b(child),  
**location** is a digitized code indicating where the recording took place, **distance** indicates the  
distance at which the recording was taken place from the source, and **device** is a digitized  
code indicating the device which was used to record.

## **How to decode?**

### **Class**

#### Parent(Speaker A)

- 0: 'singing'
- 1: 'reading'
- 2: 'other utterances'

#### Child(Speaker B)

- 0: 'singing'
- 1: 'reading'
- 2: 'crying'
- 3: 'refusing'
- 4: 'other utterances'

If polyphonic, {class\_a}{class\_b}. ex) 04 → speaker\_a: singing, speaker\_b: other utterances

### **Speaker**

- a: parent (monophonic)
- b: child (monophonic)
- ab: parent and child (polyphonic)

### **Age**

First digit of the age (real age in h5 attributes, if polyphonic, age is {age\_a}{age\_b})  
ex) 373 → age of speaker\_a: 37, age of speaker\_b: 3

### **Sex**

{0: 'Female', 1: 'Male'}, if polyphonic, sex is {sex\_a}{sex\_b}  
ex) 01 → speaker\_a is female, speaker\_b is male

### **Location**

{0: 'Studio apartment', 1: 'Dance studio', 2: 'Anechoic Chamber'}

### **Distance**

{0: 0.4 m, 1: 2.0 m, 2: 4.0m}

### **Device**

{0: iPhone, 1: Samsung Galaxy S7}

### **Noise**

{0: 'Noiseless', 1: 'Indoor noise', 2: 'Outdoor noise', 3: 'Both indoor and outdoor noise'}

## **License**

## **Contact & Purchase**

contact@deeplyinc.com