

Deeply Parent-Child Vocal Interaction Dataset

Summary

The interaction of 24 pairs of parent and child(total 48 speakers), such as reading fairy tales, singing children's songs, conversing, and others, is recorded. The recordings took place in 3 different types of places, which are an anechoic chamber, studio apartment, and dance studio, of which the level of reverberation differs. And in order to examine the effect of the distance of mic from the source and device, every experiment is recorded at 3 distinct distances with 2 types of smartphone, iPhone X, and Galaxy S7.

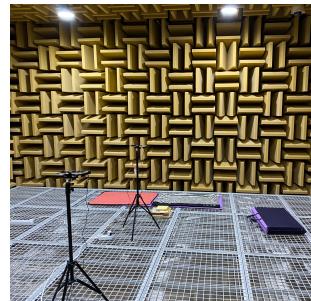
Recording contents	A parent and his/her child interacting(singing children's song, reading fairytales, conversing, ...)
Recording environment	Anechoic Chamber (no reverb), Studio apartment (moderate reverb), Dance studio (high reverb)
Device	<u>iPhone X</u> (iOS), <u>Samsung Galaxy S7</u> (Android)
Distance 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	Korean
Demographics	24 parents, with 17% males and 83% females, and 12.5% parents are in 20s, 62.5% in 30s, and 25% in 40s. 24 children, with 46% males and 54% females, and 21% children are in the age group of 1-2, 54% in 3-4, and 25% in 5-6



<Fig 1. Studio apartment>



<Fig 2. Dance studio>



<Fig 3. Anechoic chamber>

What's inside the children interaction dataset?

The Parent-Child Vocal Interaction Dataset consists of 281.3 hours of audio clips of the different types of interactions between a parent and his/her child. The participants are encouraged to record repetitively in all 3 types of place (anechoic chamber, studio apartment, dance studio), and every recording is conducted systematically at 3 ordinal distances(0.4m, 2.0m, 4.0m) with 2 types of device(iPhone X and Galaxy S7).

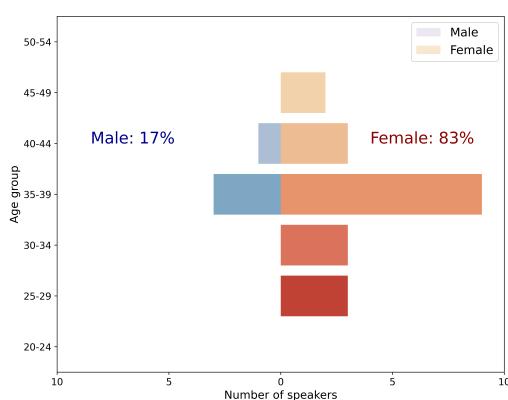
The dataset also includes metadata such as a speaker, age, sex, noise, type of place, distance, and device. The type of utterances is categorized as follow: that of the parents is categorized into 3 groups, 'singing', 'reading', and 'other utterances', the singing category includes singing a children's song and a lullaby, and other utterances literally contain all the other utterances than reading and singing, mostly spontaneous speech with his/her child. However, that of children is categorized into 5 groups, 'singing', 'reading', 'crying', 'refusing', and 'other utterances'. Children's singing category also includes singing a children's song and lullaby, refusing is an utterance caused when the children didn't like the subject, wanted to do different stuff, or so, and other utterances are same as the parents'.

Information & Statistics

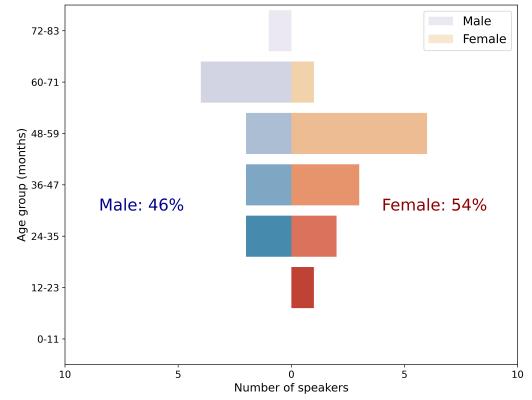
Figures 4 and 5 demonstrate the demographic information of the parents and the children respectively. The parent-participants comprise 17% of males and 83% of females, and 75% of males and 60% of females are in their 30s. And the child-participants comprise 46% of males and 54% of females, and males are distributed almost evenly throughout the age group, and 46% of females are 4 years old and so on.

Figures 6 and 7 illustrate the length distribution of each utterance by each class and the average length of each utterance by each class. As shown in figure 6, most of the parents' utterances are shorter than 10 seconds and most of the children's utterances are shorter than 5 seconds. Figure 7 shows that parents vocalize longer than their children comparing the same types of vocalization.

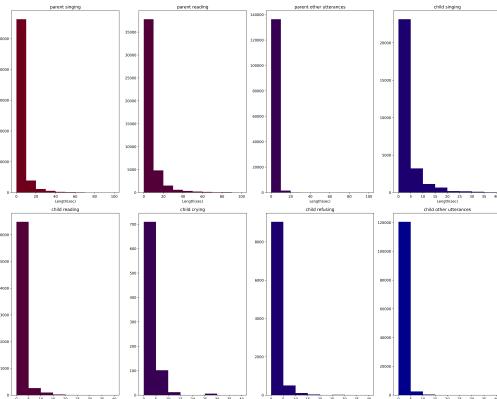
Figure 8 describes the total length(in hours) of each utterance type. The different colors consisting of each bar indicate whether the utterance is monophonic or polyphonic, in other words, whether the utterance involves a single speaker or more than two speakers(in our case, two). As shown below, you can observe that most type of the utterance type is mostly monophonic besides child singing and that the parents have led the interaction in the real world given that we didn't give any instructions on how should the interaction be done.



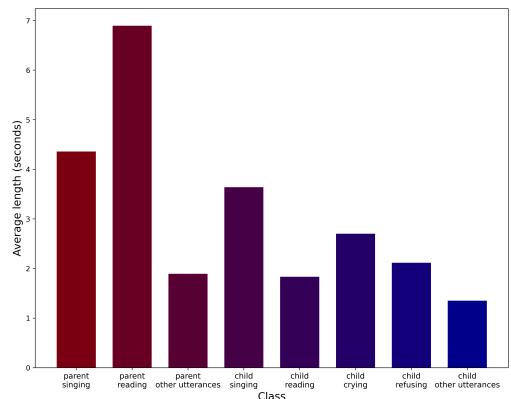
<Fig 4. Age distribution of parents by sex>



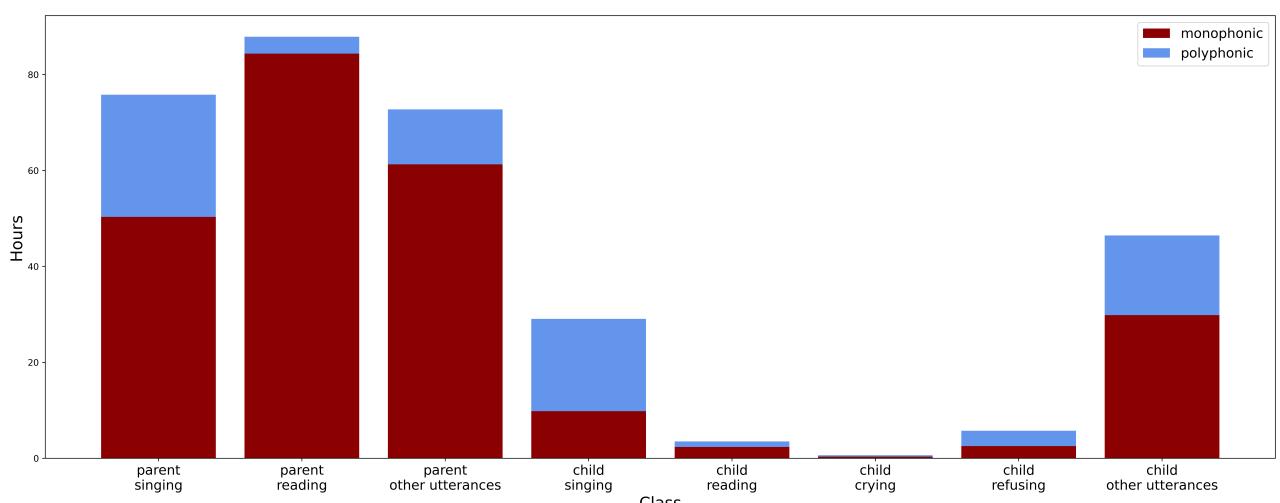
<Fig 5. Age distribution of children by sex>



<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