

디플리 한국어 감성 발화 말뭉치

Summary

화자 쌍이 3 종류의 text sentiment를 가진 대본을 3 종류의 audio sentiment로 읽은 음성이 녹음되었습니다. 녹음은 잔향의 수준이 다른 무향실, 원룸, 댄스 스튜디오 등 3 가지 장소에서 이루어졌으며, 음원으로 부터의 거리 및 기기의 효과를 알아볼 수 있도록, 모든 실험은 3 가지의 다른 거리에서 2 종류의 스마트 폰을 이용해 기록했습니다.

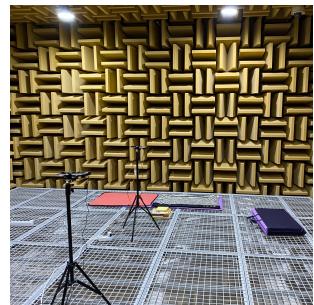
Recording contents	부정, 중립, 긍정의 세 가지 text sentiment를 갖는 대본을 세 가지 audio sentiment(부정, 중립, 긍정)으로 읽은 음성, 대본 (스크립트: 영화 리뷰(긍정적, 부정적), 일상적 대화(중립적))
Recording environment	무향실(잔향 없음), 원룸(중간 잔향), 댄스 스튜디오(높은 잔향)
Device	<u>iPhone X</u> (iOS), <u>Samsung Galaxy S7</u> (Android)
Distance from the source	0.4m, 2.0m, 4.0m
Volume	~ 290 hours, ~ 190,000 utterances, ~ 107 GB
Format	wav(44100Hz, 16-bit, mono), or h5(16000Hz, 16-bit, mono)
Language	한국어
Demographics	34 명의 한국인으로, 26% 남성, 74% 여성으로 구성되어 있습니다. 이들 중 47%는 20대, 20.5%는 30대, 17.5%는 40대, 6%는 50대, 그리고 9%는 60대입니다.



<Fig 1. Studio apartment>



<Fig 2. Dance studio>



<Fig 3. Anechoic chamber>

What's inside the Deeply Korean Read Speech Corpus?

읽기 음성 데이터 세트는 3 종류의 text sentiment를 가진 대본을 3 종류의 audio sentiment로 읽은 것이 녹음된 289.9 시간 분량의 데이터 세트입니다. 녹음은 3 종류의 공간에서 진행되었으며 (원룸, 댄스 스튜디오, 무향실), 매 녹음은 화자로부터 다양한 거리에서 (0.4m, 2.0m, 4.0m), 2 종류의 스마트폰을 사용하여 진행됐습니다 (iPhone X, Galaxy S7) .

Text sentiment와 audio sentiment는 다음과 같이 분류됩니다:

Negative text sentiment는 부정적인 내용을 담고 있는 대본, neutral text sentiment는 중립적인 내용을 담고 있는 대본, positive text sentiment는 긍정적인 내용을 담고 있는 대본을 나타냅니다. Negative, positive text sentiment에는 각각 부정적인/긍정적인 내용의 영화 리뷰가 이용되었으며, neutral text sentiment에는 어떤 감정이 포함되어 있지 않은 일상 대화가 이용되었습니다.

Negative voice sentiment는 화자가 대본을 부정적인 톤으로 읽는 것을 의미하는데, 일관성을 위해 화가 난 것처럼 읽어달라고 부탁드렸습니다. Neutral voice sentiment는 화자가 대본을 어떤 감정도 느껴지지 않게 중립적인

톤으로 읽는 것을 의미합니다. Positive voice sentiment는 화자가 대본을 기쁜 것처럼 읽는 것을 의미합니다. 각 종류의 voice sentiment는 대본의 내용(text sentiment)와 별개로, 모든 종류의 대본에 대해서 발화되었습니다. 예를 들어, 부정적인 내용을 부정적인 톤으로만 읽는 것이 아니라, 중립/긍정적인 톤으로도 읽도록 진행됐습니다.

데이터 세트에는 대본(speech-to-text aligned), 스피커, 연령, 성별, 잡음, 장소 유형, 거리 및 녹음기기와 같은 메타 데이터도 포함됩니다.

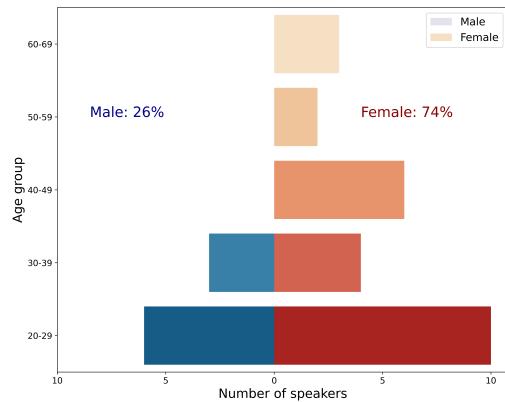
Information & Statistics

Figures 4는 화자들의 인구통계학적 정보를 보여줍니다. 참가자는 26%가 남성, 74%가 여성이고, 남성의 67%, 여성의 40%가 20대입니다.

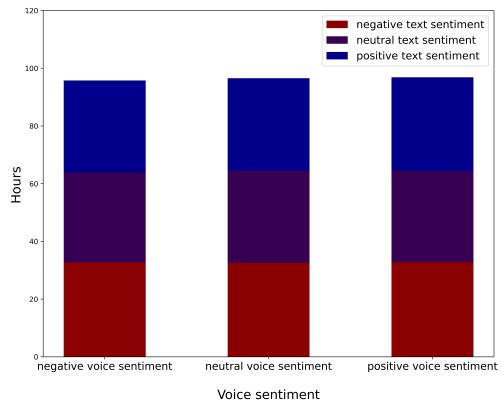
Figures 5는 특정 text sentiment, voice sentiment에 해당하는 발화의 총 길이를 보여줍니다. 가로로 늘어선 막대는 각각 다른 voice sentiment를 이용한 발화들의 총 길이가 서로 같다는 것을 보여줍니다.. 또한, 막대가 수직으로 쌓여 있는 것을 보면, 서로 다른 text sentiment를 가진 대본을 읽은 발화의 총 길이가 서로 매우 유사함을 알 수 있습니다.

Figure 6은 text sentiment와 audio sentiment에 따른 발화 당 평균 길이를 보여줍니다. Negative/positive text sentiment에는 영화 리뷰가 이용되었기 때문에, neutral text sentiment에 사용된 일상 대화보다는 평균적으로 그 길이가 긴 것을 확인할 수 있다.

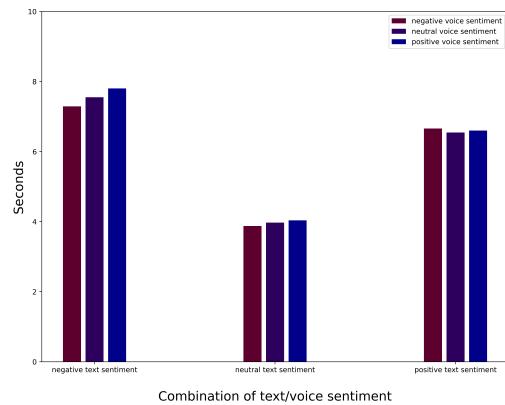
마지막으로, Figure 7은 text sentiment와 audio sentiment에 따른 발화 길이의 분포를 보여줍니다. 같은 text sentiment의 대본을 audio sentiment를 바꾸어가며 발화해도 발화의 길이의 분포는 변하지 않는 것처럼 보이며, 모든 분포는 positively-skewed 되어 있습니다.



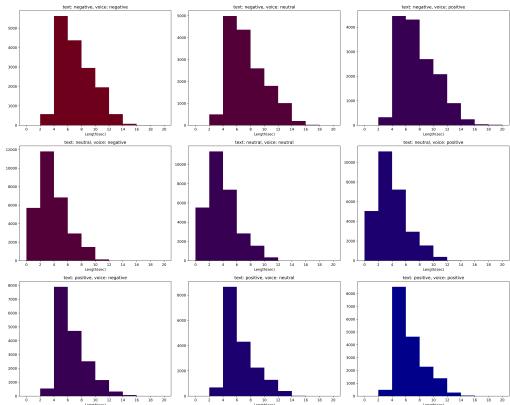
<Fig 4. Age distribution of the speakers by sex>



<Fig 5. Total length(hours) by text, voice sentiment>



<Fig 6. Average length(seconds) by text, voice sentiment>



<Fig 7. Length distribution by text, voice sentiment>

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}_{voice_sentiment}.wav

Example: sub2001_2020_11_29_00_22_0_0_0_-1.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

Text sentiment

- 1: ‘negative’
- 0: ‘neutral’
- 1: ‘positive’

Voice sentiment

- 1: ‘negative’
- 0: ‘neutral’
- 1: ‘positive’

Speaker

- a: speaker a
- b: speaker b

Age

First digit of the age (real age in h5 attributes, metadata.json)

Sex

- {0: ‘Female’, 1: ‘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