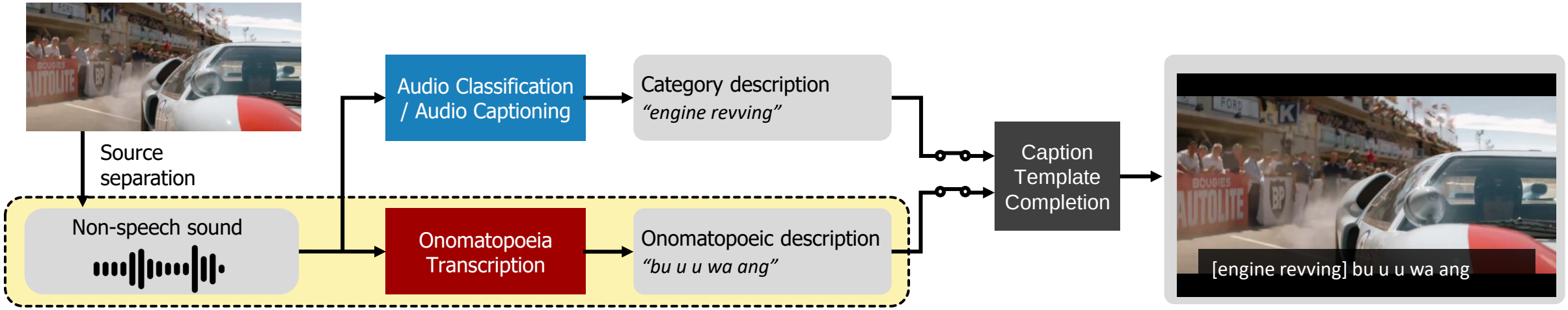
CHI 2025

OnomaCap

Making Non-speech Sound Captions Accessible and Enjoyable through Onomatopoeic Sound Representation

JooYeong Kim, Jin-Hyuk Hong

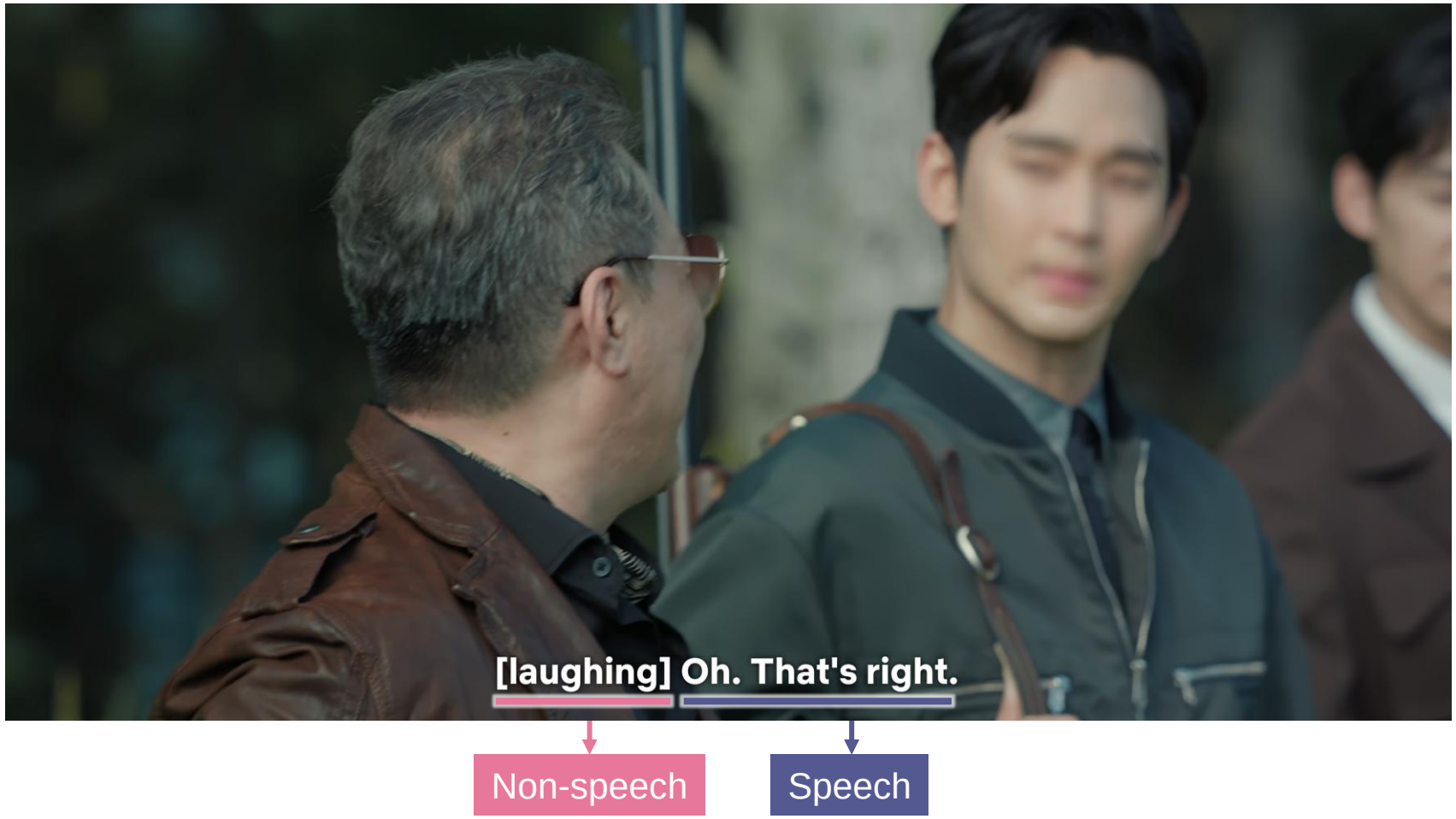
Soft Computing & Interaction Lab, GIST



Dynamic non-speech sounds

d/Deaf and Hard-of-Hearing (DHH) viewers

Non-speech sound captions for accessibility



[laughter]

Need for improvement in non-speech sound captions

1) Non-speech sound captions are still rare in most content

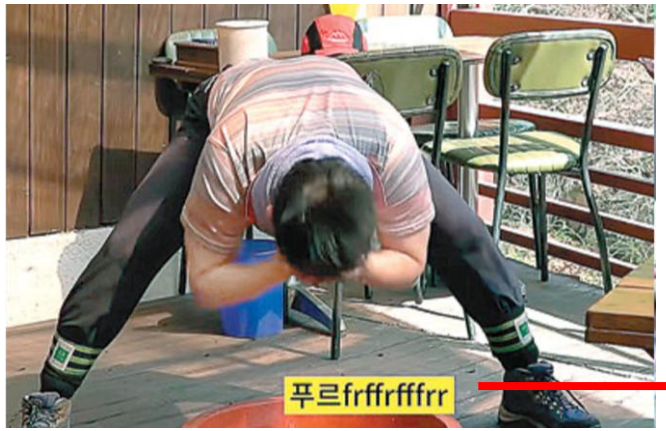
2) Captions focus on “category-based information”

Examples of Subtitles for the Deaf and Hard-of-Hearing (SDH)

Automatic non-speech captioning by YouTube

Onomatopoeia

a linguistic expression that imitates sound,
capturing how it is perceived through vocal imitation.



[phureu-frffrfffrr]

This onomatopoeic expression mimics the sound of splashing water and rubbing the face — like the sound of someone washing their face.

[ho-rop jap jju-wap jjok jjop nyam-nyam-jjop]

This onomatopoeic expression mimics various food-related sounds — like tasting, slurping, lip-smacking, and chewing — to express how someone enjoys eating in a fun and expressive way.

We can imagine the sound nuance from onomatopoeia

"Onomatopoeia is closely related to repetitive patterns *(image schemas)* arising from interactions with the physical world." [1]

Onomatopoeia		Sound source (category)		Image schema		
boom	➡	explosion	+			
				crash	smoke	negative feeling

[1] Maria Catrical and Annarita Guidi, Onomatopoeias: a new perspective around space, image schemas and phoneme clusters. Cognitive Processing 2015

Challenges

Lack of
empirical research

Lack of
dataset & models

Research questions

RQ1. What experience do DHH individuals have with onomatopoeic expressions?

RQ2. How can onomatopoeic expressions be automatically transcribed from sound?

RQ3. How do non-speech sound captions with onomatopoeia affect video viewing?

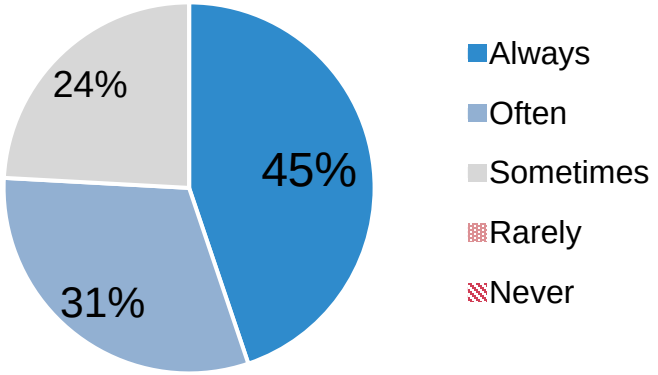
RQ1. What experience do DHH individuals have with onomatopoeic expressions?

Preliminary survey

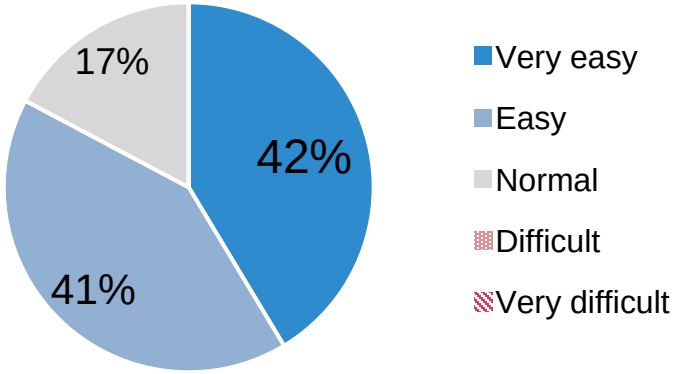
29 Korean DHH participants

- 22 – 49 years old (M = 31, SD = 6.31)
- d/Deaf (20), hard-of hearing (9)
- Profound (19), Severe (7) moderate (1) mild (2)

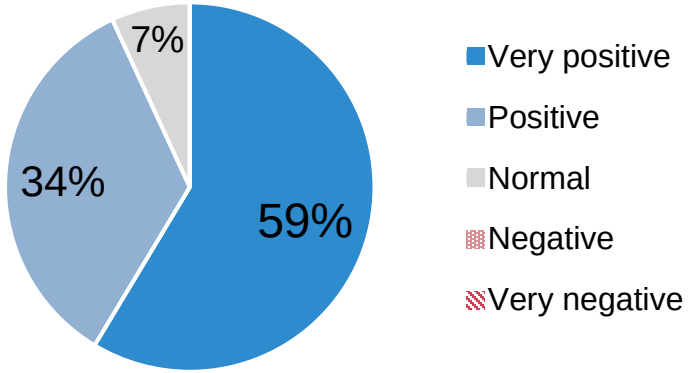
Q1. How often do you think you encounter onomatopoeia in the mediums?



Q2. Do you find it difficult to understand the sounds described when reading onomatopoeic text?



Q3. Do you think onomatopoeia positively impacts your experience when watching comics or videos?

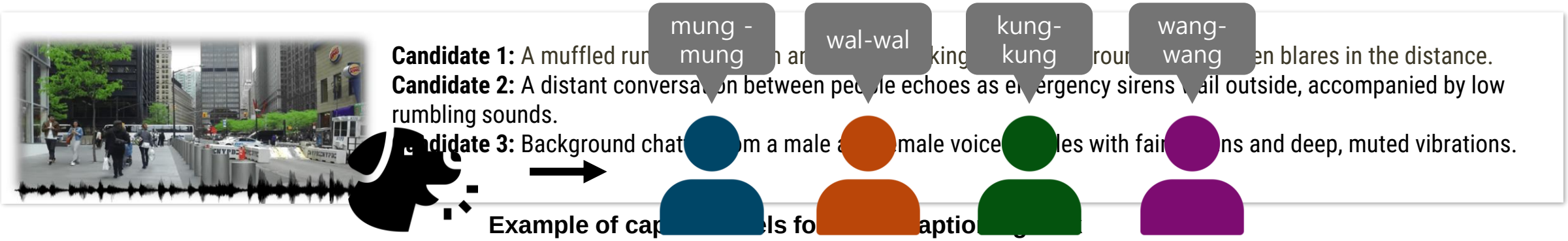


RQ2. How can onomatopoeic expressions be automatically transcribed from sound?

Sound-to-onomatopoeia transcription model

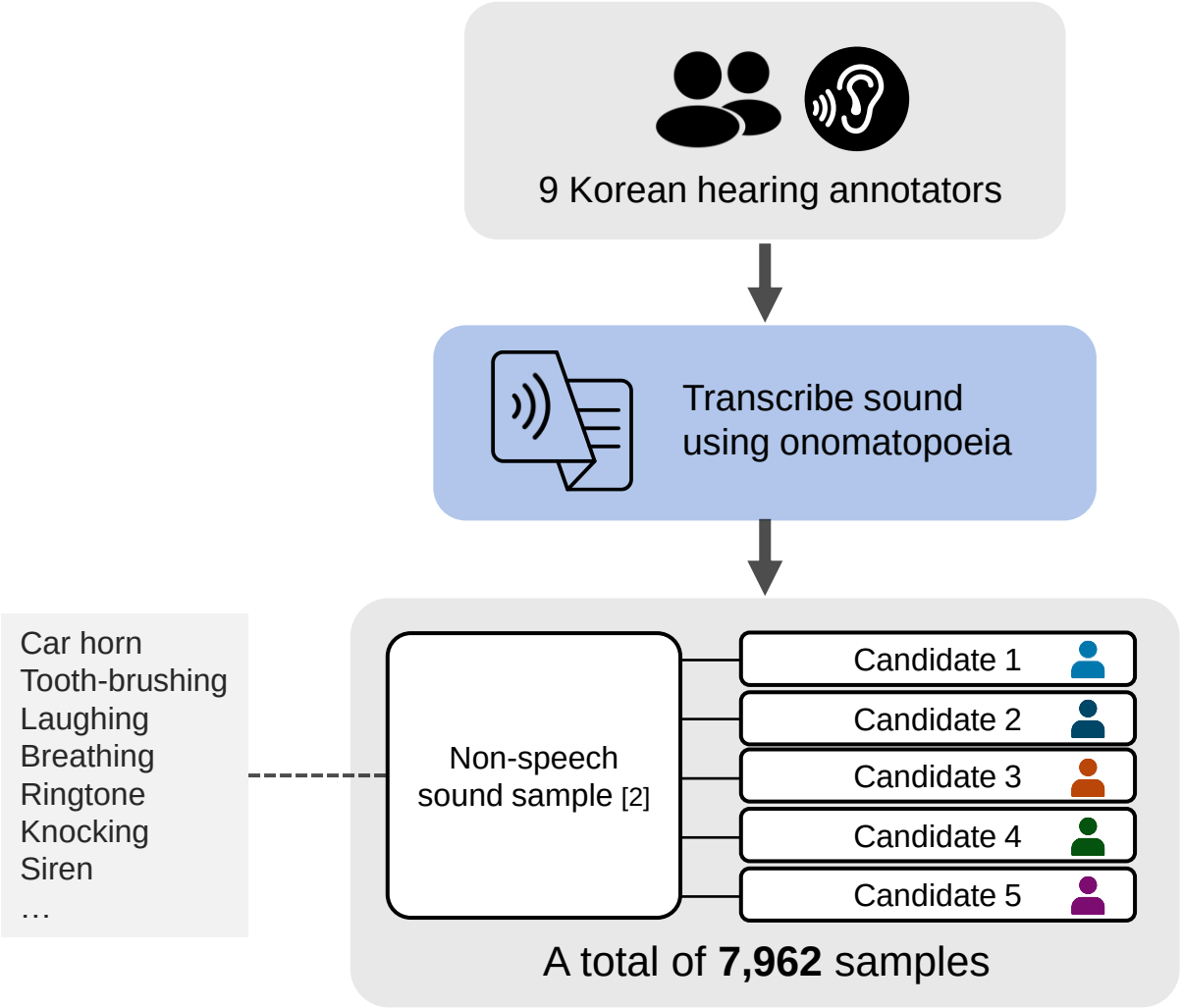
subjectivity & one-to-many mapping

Approach of Audio Captioning



RQ2. How can onomatopoeic expressions be automatically transcribed from sound?

Data collection

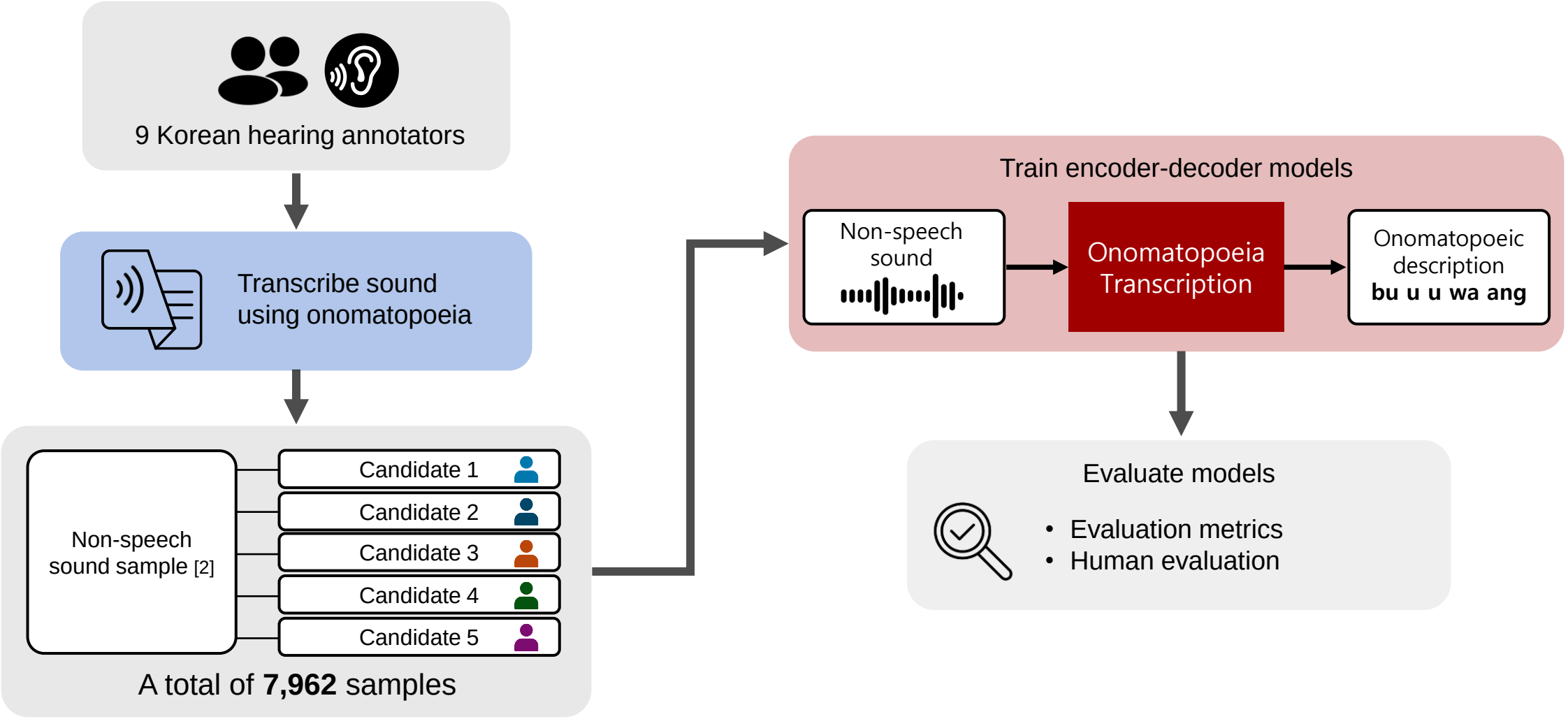


Dataset link

[2] Audio source: Firat ESC50, https://www.kaggle.com/datasets/buraktaci/firat-esc50

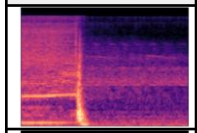
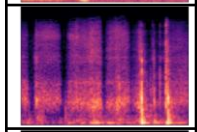
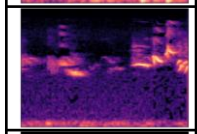
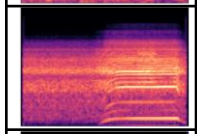
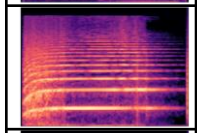
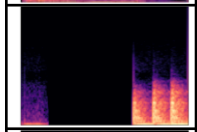
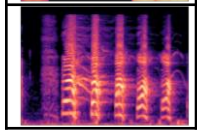
RQ2. How can onomatopoeic expressions be automatically transcribed from sound?

Training



RQ2. How can onomatopoeic expressions be automatically transcribed from sound?

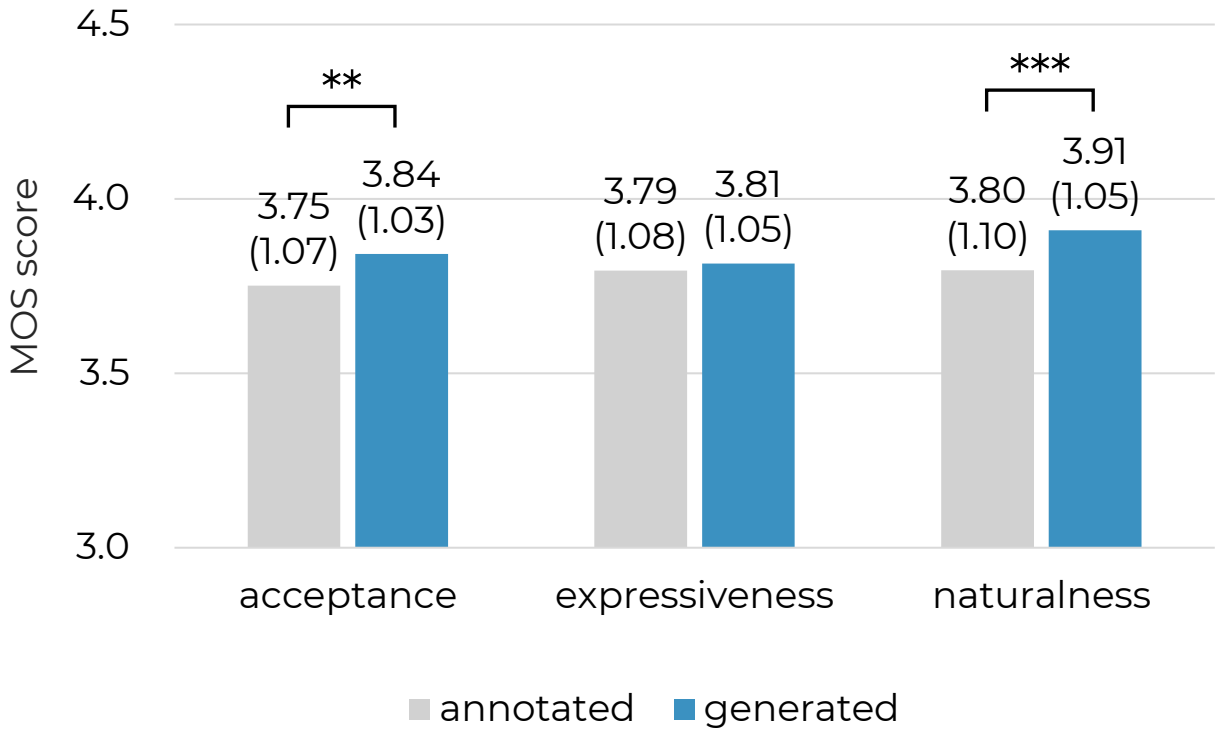
Model evaluation

Mel-Spectrogram	Category	Annotation 1	Annotation 2	Annotation 3	Annotation 4	Annotation 5	Prediction
	Car windows	ji i i ing tak 지이이잉탁	jjeu eu eu eung tak 쯔으으응탁	i i i ung keu 이이이웅크	wi i i ing tak 위이이잉탁	ji i i ing deub 지이이잉듭	wi i i i ing teo geok 위이이이잉터걱
	Chopping food	seo eo seu eu seu geu teok 서어스으스그턱	sseu euk sseuk seo reok teok teo reok 쓰윽쓰서럭턱터럭	seu eu euk seu eu euk s eu euk teuk deu deuk 스으윽스으윽스윽특드득	seuk seu eut seu euk te ok teo eok 슥스읏스윽턱터억	eu sak eu sa sak 으삭으사삭	seu geu euk seu geuk 스그윽스극
	Songbird	hwi i hwi o hwi i i l 휘이휘오휘이이이	hwi o hwi o ho ho jji ji j yeo 휘오휘오호호찌지져	hwi yu yu hwi yu jjae ae k jjyaek jjaek 휘유유휘유째액쨱짹	ppi ik ppi ppi it jjaek jjae aek jjaek 삐익삐삐잇짹째액짹	hwik ppo e o hwik hwik 휙뽀에오휙휙	jjaek jjaek hwi o o o 짹짹휘오오오
	Cellphone vibrating	ji i i i ing 지이이이잉	ki i jji i i i ing 키이찌이이이잉	u wi i i i i i eung 우위이이이이이응	u u ung jui i i i ing 우우웅즤이이이잉	swa a a a ppu wae ae a e aeng 솨아아아뿌왜애애앵	seu eu euk jui i i ing 스으윽쥐이이잉
	Siren	ppi yo o o o o o ong 삐요오오오오오옹	ppwi i i i ing 쀠이이이잉	ppu wae e e eng 뿌왜에에엥	ppi wi i i ing 삐위이이잉	ppae e e e e e e eng 빼에에에에에에엥	ppi e e e eng 삐에에에엥
	Door knocking	ttok ttok ttok 똑똑똑	ku ku kung 쿠쿠쿵	ttok ttok ttok 똑똑똑	ttok ttok ttok 똑똑똑	dok dok dok 독독독	dok dok dok 독독독
	Laughing	a ha ha ha ha 아하하하하	sseu eo a ha ha ha ha 쓰어아하하하하	ha ha ha ha ha 하하하하하	eu hat hat hat hat hat h a 으핳핳핳핳핳하	eu ha ha ha ha ha 으하하하하하	ha ha ha ha 하하하하

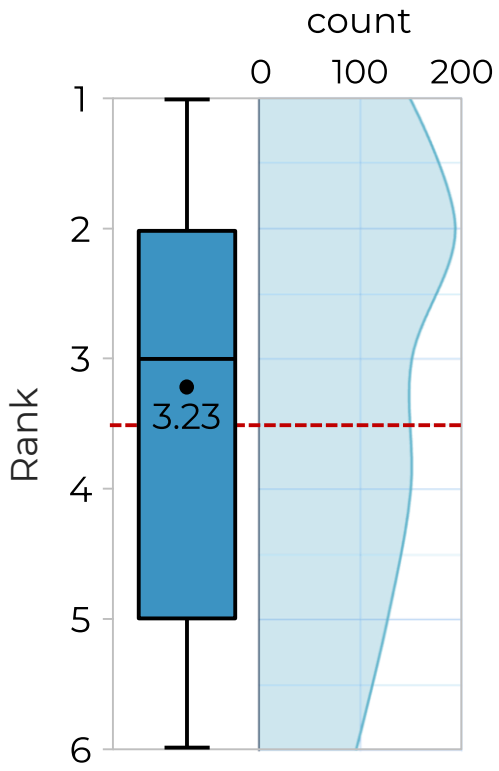
RQ2. How can onomatopoeic expressions be automatically transcribed from sound?

Model evaluation (human evaluation)

21 Korean hearing participants



Comparison of 5-point Mean Opinion Sores (MOS) between annotated and generated descriptions (Wilcoxon signed rank test; **$p < 0.01$, ***$p < 0.001$)



Rank distribution of generated descriptions for six onomatopoeic expressions (5 annotated + 1 generated)

RQ3. How do non-speech sound captions with onomatopoeia affect video viewing?

User study

25 Korean DHH participants

Diverse video genres (action, horror, comedy, documentary, …)

Questionnaire & interview

The Maze Runner (2014)

Category *(using Netflix captions)*

[그리버 울음 소리] [Griever chittering]

Onoma *(using the model-generated captions)*

[트르르릉] [teu reu reu reung]

Category + Onoma

[그리버 울음 소리] 트르르릉 [Griever chittering] teu reu reu reung

Example of Category + Onoma (translated to English)

RQ3. How do non-speech sound captions with onomatopoeia affect video viewing?

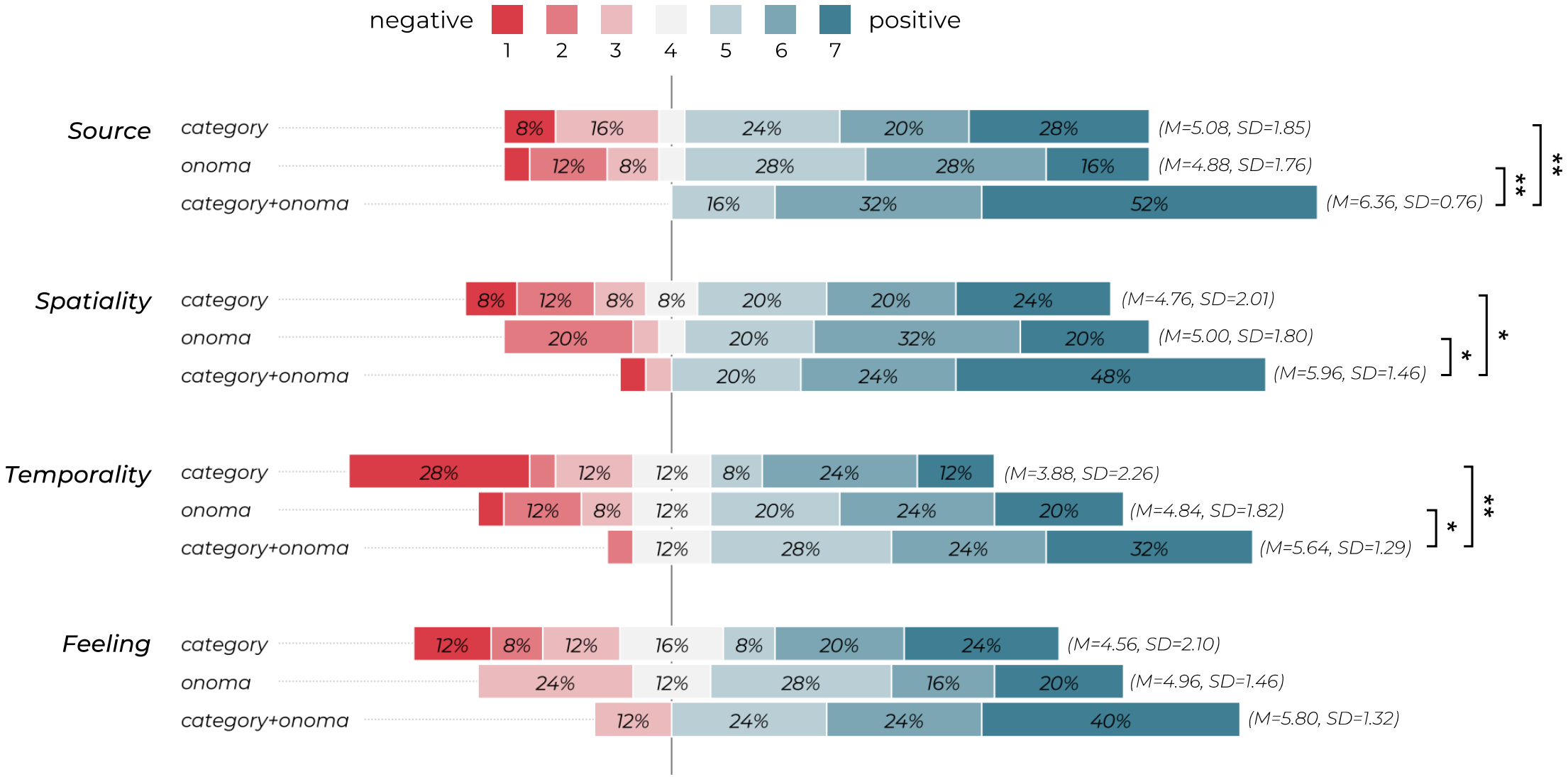
Results of the user study

- Perceived sound accessibility
- Perceived experience of non-speech sound captions
- Benefits of onomatopoeic expressions in captions
- Challenges of onomatopoeic expressions in captions
- Suitability across different genres
- Reactions to generated onomatopoeic expressions
- User preferences

For more details,
Please refer to the paper!

RQ3. How do non-speech sound captions with onomatopoeia affect video viewing?

Perceived sound accessibility



RQ3. How do non-speech sound captions with onomatopoeia affect video viewing?

Benefits of onomatopoeic expressions in captions

Differentiating between sounds of the same category

"Although the category descriptions were identical, the onomatopoeic expressions helped me understand the situation in the video." (P7)

Learning sounds through onomatopoeia

Deaf individuals are curious about all kinds of sounds. Through onomatopoeic expressions, I could learn about various sound information. It felt refreshing, like scratching an itch." (P25)

Enlivening video viewing

"Onomatopoeic captions are rare in horror movies, but when I tried it today, onomatopoeia was unique and scary. It was interesting to know about the sounds vividly. Previously, I couldn't relate when my hearing friend said sounds made movies scary, but after this experience, I now understand why." (P19)

RQ3. How do non-speech sound captions with onomatopoeia affect video viewing?

Challenge of onomatopoeic expressions in captions

Case 1) Unfamiliar sound

[teu reu reu reung]

The Maze Runner (2014)
Category caption: [monster chittering]

Case 2) Similar onomatopoeic expression

[ku gu u u u ung]

Predators, Lion (2023)
Category caption : [distant thunder]

Takeaways of OnomaCap

- OnomaCap transcribes non-speech sounds into onomatopoeic descriptions comparable to human annotations.
- Onomatopoeic descriptions improve non-speech sound accessibility and viewing experience compared to category-based captions.
- Onomatopoeic descriptions convey expressive sound nuances legibly and are adaptable across diverse genres.
- A sound-to-onomatopoeia transcription model will create new opportunities for captioning and sound-text interaction.

You can watch sample videos with English captions at this URL!

OnomaCap

Making Non-speech Sound Captions Accessible and Enjoyable through Onomatopoeic Sound Representation

JooYeong Kim Jin-Hyuk Hong

JooYeong Kim, Ph.D

Soft Computing & Interaction Lab.
Department of AI Convergence
GIST

jspirit01.github.io
jspirit01@gm.gist.ac.kr

Visit paper page! ☺

jspirit01.github.io/onomacap