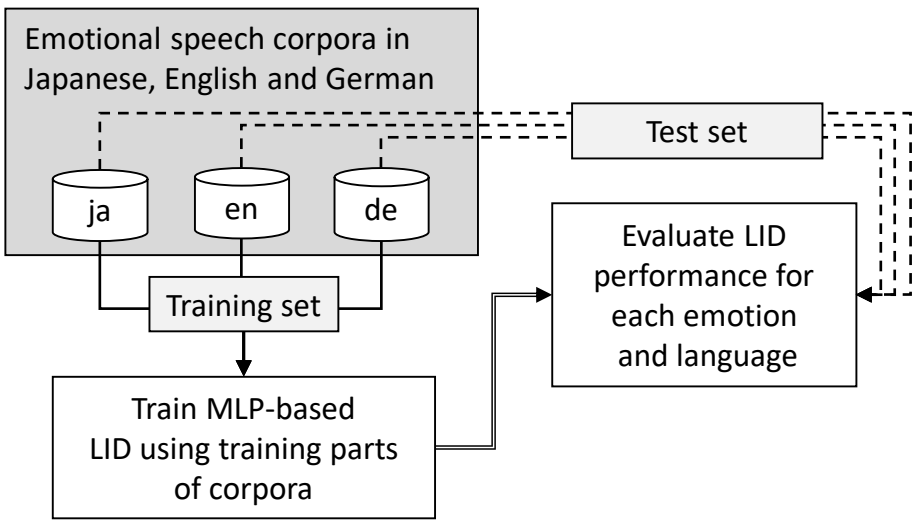
Automatic Spoken Language Identification Using Emotional Speech

Panikos Heracleous[†], Akio Yoneyama[†], **Kohichi Takai**[†,‡], Keiji Yasuda[†,‡]
[†]KDDI Research, Inc. , Japan
[‡]Nara Institute of Science and Technology, Japan

Introduction

- **Spoken Language Identification(LID)** is the process of **automatically recognizing the language from the uttered speech of an unknown speaker**.
- LID is of vital importance in both human-to-machine and human-to-human interaction in multilingual environment including dialogue system and multilingual speech translation system.
- Conventional LID studies focus only on neutral emotion speech. However, in real applications and for comprehensive research investigations, the use of emotional speech in LID is crucial.
- The current study aims at **investigating the effectiveness and performance of LID system** when **emotional speech is used for both training and evaluation**.

Analysis Flow



Step1: Divide emotional speech corpora in Japanese, English and Germany into training set and test set.

Step 2: Extract features using i-vectors and train a Multi-Layer Perceptron (MLP) based language classifier using training set.

Step 3: Evaluate LID performance for each emotion and language.

Experiments

Experimental Settings

Emotion speech corpus used for the experiments

Language	Training set size	Test set size	# of emotion labels
en	1000 utterances	200 utterances	7
de	280 utterances	152 utterances	7
ja	512 utterances	256 utterances	4

Experimental Results

Spoken language identification recalls [%] using English, German, and Japanese speech data

Emotion Label	Language			
	en	de	ja	Average
Normal	100.0	95.5	97.0	97.5
Emotional	97.4	87.6	96.5	93.8

Conclusion

- **LID using emotional speech**, the average recall was **93.8%**, slightly lower than the average recall when using normal speech.
- However, the recalls were **closely comparable** and the differences between normal and emotional recalls were **not statistically significant**.