**<LEVERAGING LARGE TEXT CORPORA FOR END-TO-END SPEECH SUMMARIZATION >**

* How2 dataset에서 외부 텍스트 요약 데이터를 사용해서 E2E SSum 의 데이터 증강방법의효과 연구
* END-TO-END SPEECH SUMMARIZATION(E2E SSum)은 음성에서 직접 요약문장을 생성하는 방법
* 중간에 텍스트 출력을 포함하지 않아서 ASR 오류에서 자유롭고(적고?) 오직 음향 정보만 사용해서 요약문을 생성
* Cascade approach (ASR + 텍스트 요약 모델)보다 ASR 오류를 줄이고 비언어적 정보를 포함하며 전체 시스템 감소
* 대량의 (음성과 요약) 데이터 수집 어려움( 1. Tts를 활용해 합성음성을 생성하고 텍스트 요약과 함께 훈련, 2. 합성음성 대신 phoneme swquence를 직접 입력)
* How2 dataset에 외부 텍스트 요약 데이터를 사용해서 effectiveness of data augmentation
* E2E SSum의 imput으로 텍스트 요약 데이터만 사용

**< XNOR-FORMER: LEARNING ACCURATE APPROXIMATIONS IN LONG SPEECH TRANSFORMERS>**

* Self-attention내부의 key-query 연산의 특성을 조사하여 새로운 선형 transformer 모델 개발
* 음성 인식 & 음성요약에 최신 기술 능가 (librispeech-100 음성 인식 벤치마크, interview 음성 인식 벤치 마크 에서 1%의 wer 개선, how2 요약에서 rouge 5점 개선)

**< END-TO-END SPEECH SUMMARIZATION USING RESTRICTED SELF-ATTENTION >**

* **Cascade & end-to-end 모델 비교 , cascade는 error**
* **How2 dataset 사용 (2000h 음성,번역,요약 등)**
* **요약 평가(ROUGE, METEOR, BERTscore**
* **ASR은 ESPnet 사용(학습 CTC-attenction 사용)**