

# 063

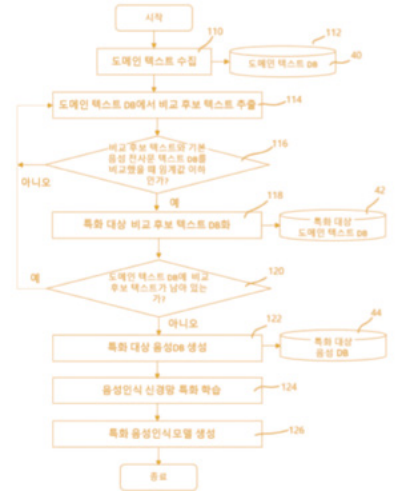
기술분류\_ 인공지능

## 실시간 스트리밍 종단형 음성인식기술(한국어, 다국어 및 발성 평가용) 및 화자 분리 기술(Ver.1.3)

### 01 기술 개요

#### 실시간 종단형(End-to-End) 음성인식기술(한국어, 다국어 및 발성 평가용) 및 화자 분리 기술

- 심층신경망기반 심층학습 기술의 고도화 및 대용량 언어음성 데이터에 기반하여 실현되는 고난이도 융복합 기술
- 원어민 및 비원어민의 대규모 한국어 및 다국어 발성을 수집하고 가공하여, ESPnet과 같은 오픈소스를 활용한 종단형 음성인식기의 훈련 리소스로 사용
- 음향모델과 언어모델, 발음사전 등을 별도로 학습하는 기존의 음성인식 모델을 종단형 방식으로 통합 학습하도록 개선함으로써 음성인식 성능을 대폭 향상시키고, 실시간 스트리밍 음성인식을 가능하게 함으로써 활용 범위를 대폭 향상



[대표도면]

### 02 기술 차별성

#### 종단형 음성인식 기술

- 기존의 언어모델, 음향모델, 발음사전 등을 별도로 학습하는 방식에서 벗어나 종단형으로 통합 학습하여 음성인식에 적용의

#### 실시간 스트리밍 종단형 음성인식 기술

- 음성이 입력되기 시작하면 연산을 시작하여 스트리밍 입력에 대해 발성 종료 후 1초 내외 결과 출력하며, 블록 단위로 판단하여 연속적으로 결과 출력 가능

#### 화자분리 기술

- 다수의 화자가 같은 장소에서 발생하는 경우 각 화자별로 발생한 구간을 추정하는 화자분리 기술

### 03 기술 키워드

#### 종단형 음성인식, 실시간 스트리밍, 화자분리

### 04 기술의 TRL 단계



필수기반

# 063

기술분류\_ 인공지능

## 실시간 스트리밍 종단형 음성인식기술(한국어, 다국어 및 발성 평가용) 및 화자 분리 기술(Ver.1.3)

### 05 사업화 포인트

### 06 활용 분야 및 시장 규모

### 07 지식재산권 현황

국내외 종단형 음성인식 기술을 서비스하고 있는 기업은 많지 않으며, 제공하는 경우에도 API를 통해 사용량을 제한하거나 일정용량 이상 유료과금하며 튜닝이 어려움. 실시간 스트리밍 음성인식 기술은 기술 시장 진출 극초기 단계로 시장 선점 필요

#### 활용 분야

실시간 방송 자막, 실시간 회의록 작성

#### 시장 규모 및 전망

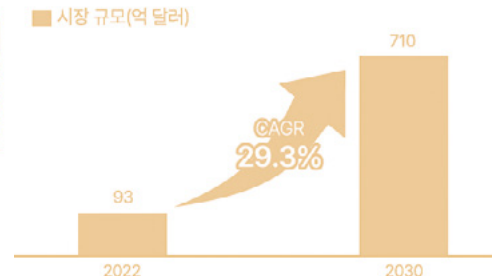
Fortune business insight에 따르면 글로벌 음성인식 시장 규모는 연평균 24.8% 성장해 2023년 126.2억 달러에서 2030년 710억 달러에 도달할 것으로 전망

[국내 자연어처리 시장]

구분	2018년	2019년	2020년(3)	2021년(3)	2022년(3)	2023년(3)	2024년(3)	2025년(3)	CAGR(%) (18~25)
시장규모	1.07	1.50	2.07	2.87	3.87	5.48	7.58	10.51	36.4%
인식률	0.42	0.58	0.81	1.13	1.57	2.13	3.04	4.24	30.1%
정확도(%)	0.26	0.36	0.40	0.60	0.97	1.38	1.89	2.64	30.9%
처리량(만)	0.27	0.37	0.51	0.70	0.96	1.30	1.81	2.49	37.3%
서비스량(만)	0.12	0.20	0.26	0.35	0.47	0.63	0.85	1.14	36.4%
성장률(%)	-	39.2%	38.4%	36.4%	38.4%	38.4%	39.3%	38.5%	-

(출처: 한국신용정보원, AI 기술 시장동향)

[글로벌 음성인식 시장 규모]



(출처: Fortune business insight)

#### 권리현황

특허명	도메인특화 음성인식 모델 구성 방법 및 장치와 이를 이용한 종단형 음성인식기
출원번호	10-2022-0001723
권리자	한국전자통신연구원
관리기관	한국전자통신연구원
담당자	이세희
문의처	042-860-6841