



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0153731
(43) 공개일자 2022년11월21일

(51) 국제특허분류(Int. Cl.)
G10L 13/08 (2006.01) G06F 40/40 (2020.01)
(52) CPC특허분류
G10L 13/08 (2013.01)
G06F 40/40 (2020.01)
(21) 출원번호 10-2021-0060914
(22) 출원일자 2021년05월11일
심사청구일자 2021년05월11일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암
동5가)
(72) 발명자
임희석
서울특별시 성북구 안암로 145 자연계캠퍼스애기
능생활관 311호 NLP&AI연구실
박찬준
서울특별시 송파구 중대로 24, 306동 1303호 (문
정동, 올림픽해밀리타운)
(74) 대리인
김홍석

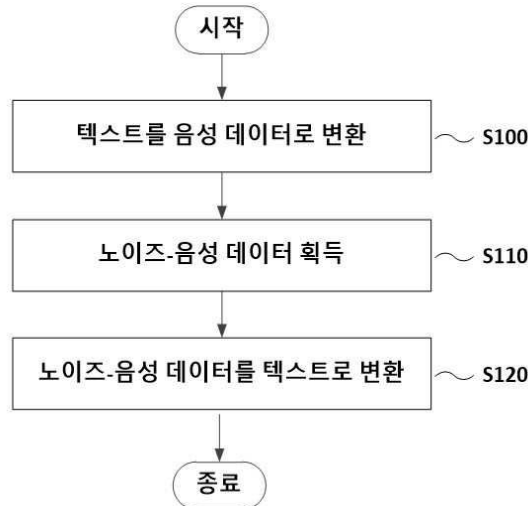
전체 청구항 수 : 총 6 항

(54) 발명의 명칭 학습 데이터 생성 방법 및 이를 이용한 음성 인식 후처리 방법

(57) 요약

학습 데이터 구축 방법 및 이를 이용한 음성 인식 방법이 개시된다. 상기 학습 데이터 구축 방법은 적어도 프로 세서를 포함하는 컴퓨팅 장치에 의해 수행되고, 복수의 문장들을 포함하는 제1 텍스트 데이터를 제1 음성 데이터 로 변환하는 단계, 상기 제1 음성 데이터에 노이즈를 부가하여 제2 음성 데이터를 획득하는 단계, 및 상기 제2 음성 데이터를 제2 텍스트 데이터로 변환하는 단계를 포함하고, 상기 제2 텍스트 데이터는 상기 제1 텍스트 데이 터에 포함된 복수의 문장들 각각에 대응하는 문장이 포함된다.

대표도 - 도1



(52) CPC특허분류

G06N 20/00 (2021.08)

G10L 15/26 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711116144
과제번호	2018-0-01405-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	Human-inspired 복합지능 원천기술 연구 및 인력양성
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

적어도 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되는 학습 데이터 생성 방법에 있어서,
복수의 문장들을 포함하는 제1 텍스트 데이터를 제1 음성 데이터로 변환하는 단계;
상기 제1 음성 데이터에 노이즈를 부가하여 제2 음성 데이터를 획득하는 단계; 및
상기 제2 음성 데이터를 제2 텍스트 데이터로 변환하는 단계를 포함하되,
상기 제2 텍스트 데이터는 상기 제1 텍스트 데이터에 포함된 복수의 문장들 각각에 대응하는 문장이 포함되는,
학습 데이터 생성 방법.

청구항 2

제1항에 있어서,
상기 노이즈는 비소리나 천둥소리를 포함하는 자연 소음 또는 구급차 소리나 경찰차 소리를 포함하는 공해 소음을 포함하는,
학습 데이터 생성 방법.

청구항 3

제1항에 있어서,
상기 제2 음성 데이터는 상기 제1 음성 데이터의 주파수 스펙트럼의 일부를 마스킹함으로써 생성되는,
학습 데이터 생성 방법.

청구항 4

적어도 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되는 음성 인식 방법에 있어서,
생성된 학습 데이터를 이용하여 시퀀스-투-시퀀스(seq2seq) 모델을 학습시킴으로써 음성 인식 후처리 모델을 생성하는 단계;
소정의 음성 인식 모델을 이용하여 입력 음성 데이터에 대응하는 텍스트 문장을 생성하는 단계; 및
상기 음성 인식 후처리 모델을 이용하여 상기 텍스트 문장을 교정하여 교정된 텍스트 문장을 출력하는 단계를 포함하는 음성 인식 방법.

청구항 5

제4항에 있어서,
상기 음성 인식 후처리 모델을 생성하는 단계 이전에, 학습 데이터를 생성하는 단계를 더 포함하고,
상기 학습 데이터를 생성하는 단계는,
복수의 문장들을 포함하는 제1 텍스트 데이터를 제1 음성 데이터로 변환하는 단계;
상기 제1 음성 데이터의 주파수 스펙트럼의 일부를 마스킹하여 제2 음성 데이터를 생성하는 단계; 및
상기 제2 음성 데이터를 제2 텍스트 데이터로 변환하는 단계를 포함하는,
음성 인식 방법.

청구항 6

제5항에 있어서,
상기 학습 데이터는 상기 제1 텍스트 데이터와 상기 제2 텍스트 데이터를 포함하는,
음성 인식 방법.

발명의 설명

기술 분야

[0001] 본 발명은 학습 데이터 생성 방법 및 이를 이용한 음성 인식 후처리 방법 에 관한 것이다.

배경 기술

[0002] GEC(grammar error correction)은 주어진 문장에서 문법적인 오류들을 올바르게 교정하는 것으로, GEC 태스크(task)는 대부분 규칙 기반(rule-based) 또는 통계 기반(statistical-based)으로 교정이 수행된다. 그러나, 규칙 기반의 경우 규칙을 구축하는데 상당한 시간과 비용이 투자되어야 하며, 통계 기반의 경우 대용량 데이터가 필요하다.

[0003] 최근, 딥-러닝(deep-learning)의 등장과 함께, GEC 태스크를 저-원천 언어(low-resource language)를 번역하는 시스템의 관점으로 접근하는 많은 연구가 이루어지고 있다. 기계 번역(machine translation, MT)은 소스 언어(source language)를 타겟 언어(target language)로 번역을 진행하는 것을 의미한다. GEC를 기계 번역의 관점으로 보면, 문법적으로 오류가 있는 소스 문장을 문법적인 오류를 해소한 타겟 문장으로 번역을 진행하는 것으로 해석할 수 있다. 한편, 딥러닝 연구 방향들은 대량의 모델의 크기와 파라미터를 요구하는 사전 학습(pre-train) 및 미세 조정(fine-tuning)을 이용하기 때문에, 속도가 느리고 메모리가 부족하다. 이러한 한계점을 해결하기 위하여, 모델의 변경 없이 각종 후처리(post-processing)를 통해 모델의 성능을 향상시키는 연구가 활발히 이루어지고 있으나 이에 따른 한계점 역시 존재한다.

[0004] 자동 음성인식은 음성을 문자로 변환해주는 것으로, 딥-러닝의 등장으로 급격한 인식을 향상을 이뤘다. 음성인식기의 경우, 띄어쓰기가 이상하고 문맥에 맞지 않은 결과를 내보낼 수 있고, 숫자 처리를 하지 못하는 경우가 있다. 더 나아가 기호 부착(셈표, 마침표, 느낌표, 또는 물음표)을 하지 않고 철자 오류를 발생시킬 수 있다.

[0005] 또한, 음성 인식기는 음향 모델과 언어 모델로 구성되며 해당 모델을 다루기 위해서는 엄청난 시간과 비용이 필요하며 진입장벽이 높고 개발 및 유지보수 비용이 많이 들어간다.

발명의 내용

해결하려는 과제

[0006] 본원 발명이 해결하고자 하는 과제는 속도 및 성능이 우수하고 시간 및 비용을 단축할 수 있는 학습 데이터 생성 방법 및 이를 이용한 음성 인식 후처리 방법을 제공하는 것이다.

과제의 해결 수단

[0007] 본 발명의 일 실시예에 따른 학습 데이터 생성 방법은 적어도 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되고, 복수의 문장들을 포함하는 제1 텍스트 데이터를 제1 음성 데이터로 변환하는 단계, 상기 제1 음성 데이터에 노이즈를 부가하여 제2 음성 데이터를 획득하는 단계, 및 상기 제2 음성 데이터를 제2 텍스트 데이터로 변환하는 단계를 포함하되, 상기 제2 텍스트 데이터는 상기 제1 텍스트 데이터에 포함된 복수의 문장들 각각에 대응하는 문장이 포함된다.

[0008] 또한, 본 발명의 일 실시예에 따른 음성 인식 방법은 적어도 프로세서를 포함하는 컴퓨팅 장치에 의해 수행되고, 생성된 학습 데이터를 이용하여 시퀀스-투-시퀀스(seq2seq) 모델을 학습시킴으로써 음성 인식 후처리 모델을 생성하는 단계, 소정의 음성 인식 모델을 이용하여 입력 음성 데이터에 대응하는 텍스트 문장을 생성하는 단계, 및 상기 음성 인식 후처리 모델을 이용하여 상기 텍스트 문장을 교정하여 교정된 텍스트 문장을 출력하는 단계를 포함한다.

발명의 효과

[0009] 본 발명의 실시예들에 따르면, 학습 데이터의 무한한 생성이 가능하며 철자 오류, 띄어쓰기 오류, 숫자 변환,

기호 부착 등의 문제를 해결할 수 있다. 시퀀스-투-시퀀스 모델을 재사용 가능하여 개발비용을 줄일 수 있고, 전사자 없이 데이터를 자동으로 생성할 수 있어, 시간과 비용을 크게 단축시킬 수 있다.

[0010] 또한, 일반 상용화된 음성인식 시스템을 도메인 특화 후처리기 적용을 통해 도메인 특화 어휘 적용이 가능하다.

도면의 간단한 설명

[0011] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.

도 1은 본 발명의 일 실시예에 따른 학습 데이터 생성 방법을 설명하기 위한 순서도이다.

도 2는 도 1에 도시된 학습 데이터 생성 방법을 설명하기 위한 모식도이다.

도 3은 본 발명의 일 실시예에 따른 음성 인식 후처리 모델 생성 방법 또는 음성 인식 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0012] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.

[0013] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0014] 제1 또는 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.

[0015] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0016] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0017] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0019] 이하에서는, 본 발명의 일 실시예에 따른 학습 데이터 생성 방법 및 이를 이용한 음성 인식 후처리 방법을 첨부된 도면과 함께 상세하게 설명하기로 한다. 후술되는 학습 데이터 생성 방법은 적어도 프로세서 및/또는 메모리를 포함하는 컴퓨팅 장치에 의해 수행될 수 있다.

[0020] 도 1은 본 발명의 일 실시예에 따른 학습 데이터 생성 방법을 설명하기 위한 순서도이고, 도 2는 도 1에 도시된

학습 데이터 생성 방법을 설명하기 위한 모식도이다.

- [0021] 도 1과 도 2를 참조하면, 단일 코퍼스(mono corpus, 단일 말뭉치)의 초기 텍스트(제1 텍스트 데이터로 명명될 수 있음)를 음성 데이터로 변환된다(S100). 단일 코퍼스는 복수의 문장들을 포함할 수 있다. 즉, 제1 텍스트 데이터는 복수의 문장들을 포함할 수 있다. S100 단계는 복수의 문장들 각각을 음성 데이터로 변환하는 단계를 의미할 수 있다. 즉, S100 단계의 수행 결과로써, 복수의 문장들 각각에 대응하는 복수의 음성 데이터들(제1 음성 데이터들이라 명명될 수 있음)이 생성될 수 있다. 음성 데이터를 생성하기 위해 TTS(Text To Speech) 기법이 사용될 수 있다.
- [0022] 이후, 변환된 음성 데이터 각각에 노이즈를 부가하여, 노이즈-음성 데이터를 획득할 수 있다(S110). 노이즈-음성 데이터는 제1 음성 데이터에 노이즈를 부가한 음성 데이터로써, 노이즈가 부가된 음성 데이터 또는 제2 음성 데이터로 명명될 수도 있다.
- [0023] 부가되는 노이즈는 상황과 환경에 따라 상이할 수 있으며, 복수의 노이즈를 부가하거나 단일의 노이즈를 부가하는 것도 가능하다. 예컨대, 노이즈는 자연 소음을 포함할 수 있다. 자연 소음은 비 소리, 천둥 소리, 번개 소리, 새 소리, 동물 소리, 폭포 소리, 시냇물 소리 등과 같이 자연에서 발생하는 소리를 의미할 수 있다. 또한, 노이즈는 공해 소음을 포함할 수도 있다. 공해 소음은 구급차 소리, 경찰차 소리, 자동차 경적 소리, 공사 현장에서 발생하는 소음, 교실이나 광장 등과 같이 특정 장소에서 발생하는 소음 등과 같은 소음을 의미할 수 있다.
- [0024] 또한, 제1 음성 데이터의 스펙트럼의 일부(예컨대, $a\%$ (a 는 0보다 크고 100보다 작은 실수로써, 예시적인 값은 15임))를 마스킹(masking)함으로써 제2 음성 데이터를 생성할 수도 있다. 이를 위해, 컴퓨팅 장치는 임의의 주파수 변환 기법(예컨대, 푸리에 변환(Fourier Transform, FT), 단기 푸리에 변환(Short-Time Fourier Transform, STFT), 웨이블릿 변환(Wavelet Transform) 등)을 이용하여 제1 음성 데이터를 주파수 변환한 후 음성 스펙트럼의 $a\%$ 를 마스킹할 수 있다. 여기서, 마스킹 동작은 미리 정해진 주파수 대역 또는 임의의 주파수 대역의 신호를 삭제(또는 제거)하는 것을 의미할 수 있다. 이후, 컴퓨팅 장치는 시간 도메인(time domain)으로 재변환함으로써 제2 음성 데이터를 생성할 수 있다. 일 예로, 제1 음성 데이터의 주파수 범위가 300 ~ 3400Hz일 때, 컴퓨팅 장치는 제1 음성 데이터의 주파수 스펙트럼의 일부, 예컨대 320~330, 490~500, 700~710, 950~960, 1500~1510, 2300~2310, 및 3100~3110Hz의 주파수 성분을 제거함으로써 제2 음성 데이터를 생성할 수 있다.
- [0025] 상술한 바와 같이, 제1 음성 데이터에 노이즈를 부가하여 제2 음성 데이터를 생성하기 위한 노이즈 데이터는 컴퓨팅 장치에 포함되는 저장부(저장 장치)에 미리 저장되어 있을 수 있다.
- [0026] 노이즈-음성 데이터는 최종 텍스트로 변환될 수 있다(S120). 최종 텍스트는 제2 텍스트 데이터로 명명될 수 있으며, 제2 텍스트 데이터는 제1 텍스트 데이터에 포함되는 복수의 문장들 각각과 대응되는 복수의 문장들을 포함할 수 있다. 제2 텍스트 데이터를 생성하기 위해 STT(Speech To Text) 기법이 사용될 수 있다.
- [0027] 초기 텍스트(제1 텍스트 데이터)와 최종 텍스트(제2 텍스트 데이터)는 병렬 코퍼스(parallel corpus, 병렬 말뭉치)를 이룬다. 즉, 제2 텍스트 데이터는 제2 음성 데이터에 대응되는 텍스트 데이터이고 제1 텍스트 데이터는 제1 음성 데이터의 기초가 되는 텍스트 데이터이므로, 제1 텍스트 데이터와 제2 텍스트 데이터(즉, 병렬 코퍼스)를 이용하여 소정의 모델을 학습할 경우, 음향 모델의 결과로 생성되는 텍스트의 오타자, 오류 등을 교정할 수 있는 음성 인식 후처리기 모델을 생성할 수 있다.
- [0028] 이와 같이, 텍스트를 음성 데이터로 변환하고, 음성 데이터에 노이즈를 추가하고 다시 텍스트로 변환시킴으로써 학습 데이터의 무한한 생성이 가능하다.
- [0030] 도 3은 본 발명의 일 실시예에 따른 음성 인식 후처리 모델 생성 방법 또는 음성 인식 방법을 설명하기 위한 흐름도이다. 도 3에 도시된 방법 또한 도 1에 도시된 방법과 마찬가지로 프로세서 및/또는 메모리(저장 장치)를 포함하는 컴퓨팅 장치에 의해 수행될 수 있다. 이때, 도 1의 방법을 실행하는 컴퓨팅 장치와 도 3의 방법을 실행하는 컴퓨팅 장치는 동일하거나 상이할 수 있다.
- [0031] 도 3을 참조하면, 우선 학습 데이터가 생성된다(S210). 학습 데이터는 제1 텍스트 데이터와 제2 텍스트 데이터를 포함할 수 있다. 제1 텍스트 데이터와 제2 텍스트 데이터는 각각 복수의 텍스트 문장들을 포함할 수 있다. 여기서, 제2 텍스트 데이터는 교정 대상 문장들을 의미할 수 있고, 제1 텍스트 데이터는 교정이 완료된 목표 문장들을 의미할 수 있다. 즉, 제2 텍스트 데이터의 문법적 오류 등이 교정된 후에는 제1 텍스트 데이터와 동일할

수 있다. 또한, 제1 텍스트 데이터와 제2 텍스트 데이터는 도 1의 방법과 동일한 방법을 이용하여 생성될 수 있다.

[0032] 음성 인식 후처리 모델이 생성될 수 있다(S220). 음성 인식 후처리 모델은 음성 인식의 결과물로 출력되는 문장의 문법적 오류 등을 교정하는 모델을 의미한다. 음성 인식 후처리 모델은 S210 단계에서 생성된 학습 데이터를 이용하여 소정의 모델을 학습시킴으로써 생성될 수 있다. 일 예로, 시퀀스-투-시퀀스(seq2seq) 모델이 이용될 수 있다. 생성된 음성 인식 후처리 모델을 컴퓨팅 장치의 저장부(저장 장치)에 저장될 수 있다.

[0033] 이후, 소정의 음성 데이터가 입력되면, 알려진 음성 인식 모델을 이용한 음성 인식 동작이 수행될 수 있다(S230). 음성 인식 모델은 기존에 알려진 모델 등이 사용될 수 있으나, 사용되는 음성 인식 모델의 종류가 본 발명의 권리범위에 제한되는 것이 아님은 명백하다. 음성 인식의 결과는 해당 음성 데이터에 대한 텍스트 문장을 의미할 수 있다.

[0034] 다음으로, S220 단계에서 생성된 음성 인식 후처리 모델을 이용하여 음성 인식 결과에 대한 교정이 수행될 수 있다. 교정의 결과로, 교정된 텍스트가 출력될 수 있다.

[0035] 음성 인식기(음성 인식 모델)의 경우, 띄어쓰기가 적절하지 않고 문맥에 맞지 않는 결과(텍스트 문장)를 출력하고, 숫자 처리를 하지 못하며, 기호(웹표, 마침표, 느낌표, 또는 물음표 등) 부착을 하지 않으며, 철자 오류를 발생시키는 경우가 있는데, 전술한 음성 인식 후처리 방법을 사용함으로써, 철자 오류, 띄어쓰기 오류, 숫자 변환, 및 기호 부착 등의 문제를 해결할 수 있다. 즉, 고품질의 음질과 명확한 양질의 음성 데이터베이스를 구축할 수 있으며, 이를 이용하면 저비용으로 고성능의 음성 인식 모델을 생성할 수 있게 된다.

[0036] 또한, 기존의 맞춤법 교정 후처리기는 규칙기반, 통계기반(N-gram 언어모델)으로 좋은 성능을 보이지 못하고 있다. 특히, 규칙기반의 경우 코퍼스 구축에 엄청난 비용과 시간이 요구되고 있다. 그러나, 본 발명의 일 실시예에 따른 음성 인식 후처리 방법을 사용하는 경우, 전사자(phonetic transcription)의 역할을 없애고 데이터를 자동으로 생성할 수 있어, 시간 및 비용을 단축시킬 수 있다.

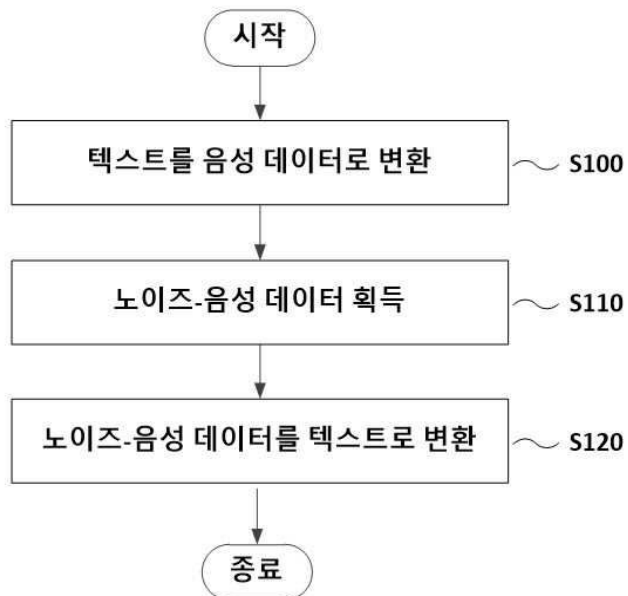
[0037] 더불어, 기구축된 음향 모델과 언어 모델 기반의 음성인식 시스템에 확장하여 적용 가능한 방법으로 음성인식을 훈련하기 위해서는 전사 데이터와 음성 데이터가 필요하다. 전사된 데이터를 G2P(grapheme to phoneme)를 진행하여 발음사전 구축 후 텍스트 정제 과정을 거쳐 언어 모델을 훈련할 수 있다. 음성 데이터를 이용해서는 음향 모델을 훈련하며, 이 음향 모델 훈련에서 필요한 음성 데이터베이스에 특정 도메인 음성 데이터베이스(비즈니스 회의, 주소 인식 등)를 TTS를 통해 구축하여 비용과 시간을 절감할 수 있는 방법으로 이를 통해 고품질의 음성 인식기를 개발할 수 있다.

[0038] 도메인 특화 환경에서, TTS를 통한 데이터를 구축하여 도메인 특화 방식을 이용하여 음성인식 모델을 만들 시 기존 제너릭(Generic) 모델보다 더 좋은 성능을 이루어낼 수 있다. 즉, 특정한 도메인에서의 음성 인식기 성능 향상에 도움을 줄 수 있다. 특히, 기존 상용화 시스템을 본 발명의 일 실시예에 따른 자기 지도 학습을 이용하면 도메인 특화 음성인식기로 변환시킬 수 있다. 원하는 도메인의 단일 말뭉치만을 이용해 TTS를 제작하고 구축된 병렬 말뭉치로 후처리기를 제작하면 상용화 시스템을 도메인 특화 음성인식기로 서비스 가능하다. 기업에서 상용화 시스템 보다 자체 음성인식기를 구축하는 이유는 도메인 특화 모델의 니즈 때문이다. 본 발명의 일 실시예에 따른 음성 인식 방법을 통해 상용화 시스템의 높은 인식률이라는 장점을 살려 도메인 특화 모델로 발전시킬 수 있다.

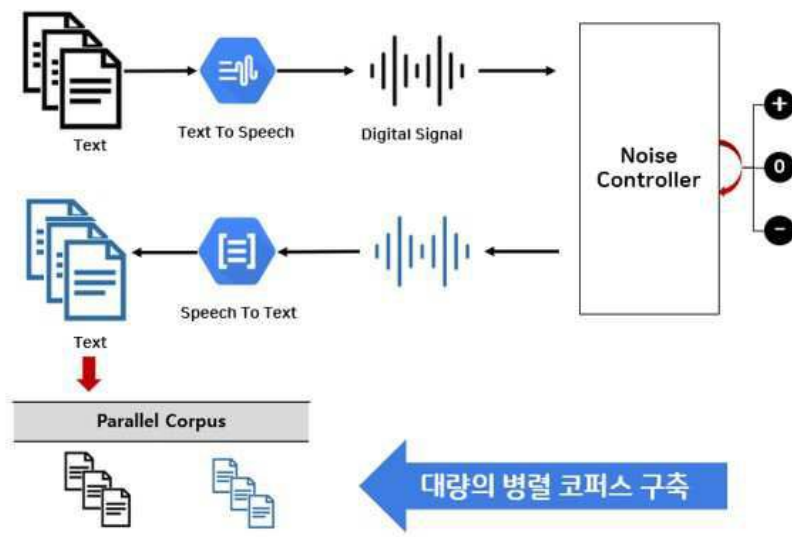
[0039] 이상, 첨부된 도면을 참조하여 본 발명의 실시예를 설명하였지만, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예에는 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

도면

도면1



도면2



도면3

