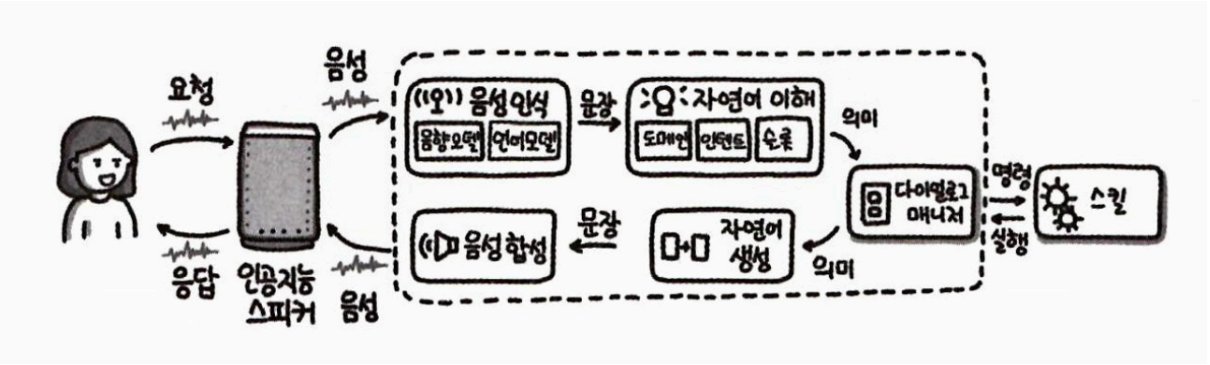
5장 스마트스피커

음성인식 트리거 → 음성 → 문장생성 → 다이얼로그 매니져 → 명령과 실행 → 결과문구 자연어 출력 → 문장을 음성합성 → 출력



음향 모델은 음성의 파형에서 단어를 인식

1. 음성인식

- 초기 : if(~word) then~
- 1차진화 : 은닉 마르코프 모델 ⇒ 확률을 통해 단어 산출
- 들어온 단어를 그대로 텍스트화 하고, 맥락을 추론해서 정상 단어로 변경하는 작업을 진행.
  예> 사가 → 사과

2. 자연어 이해

- 과거: 모든 단어를 동사와 명사로 구분하여 조건문을 작성

발화	도메인	인텐트	슬롯	슬롯 필링
“오늘 날씨 어때?”	날씨	조회	위치: 현재 위치	o
“최신 가요가 듣고싶어”	음악	재생	대상: 최신 가요	x
“레스토랑 예약해 줘”	예약	진행	장소: 빕스 시간: 오늘 오후 7시 인원: 3명	o

- 발화 → 도메인, 인텐트 정의
- + 구체적인 세부사항 정의 요청

  추가 대화를 통해서 세부사항을 추가하는것을 “멀티 턴” 이라고 한다.

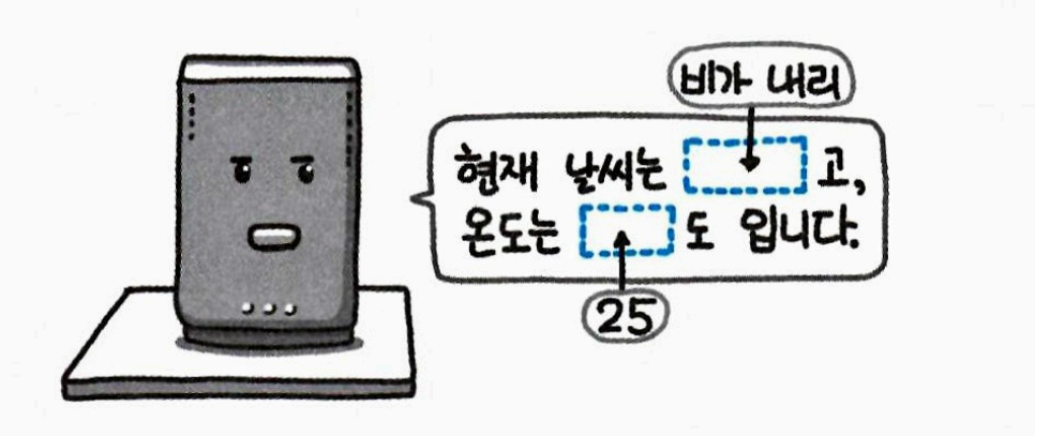
  멀리턴을 통해서 추가 정보를 획득하여 목적의 “빈 슬롯” 을 채워나간다.
- 발화 : “오늘 날씨 알려줘 “
- 도메인정의 : 날씨
- 인텐트정의 : 조회
- 세부사항 요청 : 어디날씨 ? → 부산 or 현재위치
- 슬롯필링여부 : o / x

3. 다이얼로그 매니저, 명령의 실행

- 실질적인 명령어의 실행을 구현한다.
- 발화를 통화 자연어의 이해를 위한 분석 및 “멀티턴” 방식의 세부사항 요청을 통해 등록된 다이얼로그를 실행한다.
- **지역 날씨 조회
  - 결과 : 서울의 강남, 25도 비

4. 자연어 생성 및 대화 디자인

- 자연어 텍스트 생성 : 비가 내리고 온도는 25도 입니다.
- 자연어 텍스트를 통한 음성인식 생성
  - 날씨 관련 템플릿을 통해 자연어 텍스트를 생성



5. 연결합성 USS(Unit Selection Synthesis)

- 성우가 녹음한 짧은 음원을 이어서 목소리를만들어낸다.
- 최근에는 딥러닝 기반의 합성 : 구글의 타코트론2(tacotron2) 을 통해서 자연스러운 목소리를 만들어 낸다. ⇒ TTS (text to speech)

6. 음성출력

- 성우가 만들어낸 음성을 스피커로 출력

스피커 핵심기술

- wake-up word : 특정 단어를 말하면 반응을 시작하는 트리거 단어를 통해 동작을 시작하게 한다.
  - “헤이 카카오, 기가지니” 등과 같은 특정단어를 등록하여 동작을 시작한다.
- 딥러닝도입:

        - 개인의 강세, 속도 등을 딥러닝을 통해서 보완해 나간다.
            - 음성의 잡음, 속도의 개인차에 따라 인식이 안될때가 많았다.
            - 현재는 인식률이 높다.
    - 언어모델의 보정
        - “개떡같이 말해도 찰떡같이 응답한다”
            - 예: 오늘 날씨 엿되? ⇒ 오늘 날씨 어때? 의 형식으로 단어를 보정하여 자연어의 이해를 하고, 다이얼로그 메니저를 통해 기능을 동작한다.
    - 슬롯필링(slot filling) - 자동 설정
        - 오늘 날씨 어때? 라고 했을 때 지역을 굳이 물어보지 않고 현재의 위치를 기준으로 데이터를 재설정하여 응답하도록 한다.
    - 멀티턴
        - 재질문을 통해서 조건을 완성해나간다.
        - 레스트랑 예약해줘 ⇒ 어떤 레스트랑인지 정보가 부족(레스트랑이 다수 존재)
            - 재질문 : 어느 레스토랑을 예약할까요? ⇒ 응답: xx 레스토랑, 가까운곳 ⇒ 빠진정보 획득
    - 음성의 연결 “USS”
        - 성우가 만든 단어들을 조합하여 문장을 완성
        - 구글의 타코트론2(tacotron2) 를 통해서 자연스러운 문장을 만들어내기도 한다.

결론 스마트 스피커의 미래

자연스러운 자유로운 대화보다는 정확한 정보 전달을 위해서 템플릿을 통한 대화응답 방식을 많이 사용한다.

그래서 대화가 많이 이어지지 않을 수도 있다.

부적절하고 부정적인 대화를 하지 않도록 통제하려는 의도가 있기도하다.