

음성 기반의 AI 서비스 디자인 가이드

목차

- I. Voice UI
- II. 타겟 사용자 분석 및 태스크 정의
- III. 예상 발화 작성 가이드라인
- IV. 대화 흐름 디자인하기
- V. Prompt 작성 가이드라인

I. Voice UI

I. Voice UI

- NUGU 플랫폼에서 제공될 Play 서비스는 **사용자와 음성 대화로 상호작용**을 한다는 점이 다른 일반적 서비스와 가장 큰 차이점



■ 장점

1. 구어 (Spoken language)를 기반으로 실행되므로, **디지털 디바이스 사용에 익숙하지 않은 사용자**도 쉽게 사용
2. 사용자가 다른 일을 하면서도 이용 가능
 - **운전 중 상황과 같이 손과 눈이 자유롭지 않은 상황**에서 음성 인터페이스의 유용성은 더 커짐
3. 계층 구조로 이루어진 메뉴를 가진 GUI에서는 원하는 기능에 접근하기 위해서는 여러 단계를 거쳐야 함
 - 반면, 음성 인터페이스는 **한 번의 발화로 바로 실행**이 가능
4. 음성은 감성적인 관계를 맺는데 중요한 수단으로 이용
 - 음성 인터페이스를 통해 사용자와 감성적인 관계 맺기가 용이
5. 소리는 사용자가 집중하고 있지 않더라도 **주의를 끌고, 콘텐츠 및 정보 제공**이 가능

- 한계
 1. 제공하고자 하는 기능이 음성 상호작용만으로 **사용자에게 충분한 가치를 제공할** 수 있는지 검토
 - 정보 탐색의 과정 및 결과가 복잡하거나, 음성보다는 화면을 통해 상호작용하는 것이 더 효과적인 기능의 경우, 도식화된 정보나 목록 없이 음성만으로 인터랙션을 제공하는 것은 적합하지 않음
 - 예) 영화표 예매를 위한 좌석 선택
 - 2. 간결하고 명확하게** 사용자에게 정보를 전달해야 함
 3. 사람의 인지적 특성상 단기적으로 기억할 수 있는 정보량은 한계가 있음
 - 특히 시각적인 정보 (Visual cue) 가 없는 상태에서 대화나 내용이 길어지면 이전 상황을 사용자가 기억하기 힘들 수 있음
 - 음성은 화면만큼 많은 양의 정보를 단시간에 전달하기 어려우므로 **중요한 정보만 전달**하는 것이 좋음

- 한계
 4. 사람과 사람 간의 대화와 같이 **정형화되지 않은 절차나 발화를 할 가능성을 고려**하여 예외 사례를 빠짐 없이 디자인해야 함
 5. 기기와 어느 정도 거리를 두고 음성으로 상호작용하기 때문에, 타인이 들을 수 있다는 점을 고려
 - 프라이버시 관련 내용이나 비밀번호와 같이 **민감한 개인정보를 음성으로 발화하는 것은 가급적 지양**
 6. 음성 명령 시에 주위의 잡음이나 다른 사람의 말소리가 들어가면 서비스 실패 확률이 높아지므로 파티장, 전시회장과 같은 환경에서는 좋은 인터페이스가 아님
- 결론적으로 음성 서비스 디자인은 스마트폰과 같이 디스플레이가 있는 기기에서 **제공하던 서비스를 그대로 음성으로 바꾸는 것이 아님**
 - 정보의 탐색 과정을 생략한 **간단한 기능 위주로 디자인**해야 하며, 상세한 오류 처리 (Error Handling)도 필요

- 음성 서비스 디자인 시 고려할 7가지 항목

1. 간결하고 명확하게 사용자에게 정보를 전달해야 함
2. 다른 사람이 들을 수 있다는 것을 감안하여 설계해야 함
3. 다른 행동을 하면서도 서비스 이용이 가능하다는 점을 활용 함
4. 인식이 어려운 발화를 유도하지 않아야 함
5. 서비스 전체를 한눈에 보기 어렵다는 것을 감안해야 함
6. 한 명과 대화하는 것처럼 일관성 있게 설계해야 함
7. 사람 간 대화 같이 자연스럽게 디자인해야 함

II. 타겟 사용자 분석 및 태스크 정의

II. 타겟 사용자 분석 및 태스크 정의

- 디자인의 첫 번째 단계

- 제공하고자 하는 **서비스의 타겟 사용자와 사용자의 상황을 분석**하고, 제공할 **주요 기능을 목록으로 정의**하는 단계
 - 예를 들어, 자동차에서 스타벅스 커피를 주문하고, 해당 매장에 도착해서 주문한 커피를 받아가는 서비스를 제공한다고 할 경우
 - 사용자의 운전 상황에 대한 이해와 사용자의 요구를 바탕으로 서비스 범위를 결정
 - 기존 커피 주문 시에는 샷이나 얼음 추가 등 여러 가지 옵션을 추가할 수 있지만, 운전 상황에서 여러 번의 사용자 선택을 거치는 것은 사고 발생의 위험이 있음
 - 따라서, 차량 내 커피 주문 서비스에서는 옵션 추가 기능을 제공하지 않을 수 있음
- 제공하고자 하는 기능 범위가 결정되면 이를 **단위 기능으로 분류하고 흐름을 구조화**

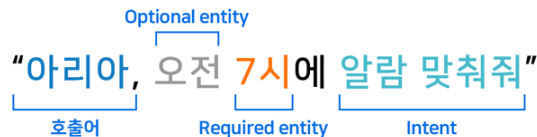
- **제공하고자 하는 기능 (Task) 목록을 정의**하고, 각 기능을 사용자가 예측 가능한 범위로 분리 또는 병합한 뒤 **핵심 기능과 주변 기능을 계층 또는 흐름으로 구조화**
- 예) 커피 주문 서비스 태스크
 - 사전 등록 : 계정 연동 / 결제 방식 등록
 - 커피 주문 : 주문 시작 / 메뉴 선택 / 매장 선택 / 주문 접수 / 주문 진행 상태 알림
 - 주문 관리 : 주문 조회 / 잔액 조회 / 주문 취소
 - 기타 : 도움말

- 사용자의 발화 패턴을 바탕으로 발생 가능한 경우 (Case)를 구분하여 세부 기능 단위로 정의
 - 예) 기능 : 커피 주문 (주문 시작 / 메뉴 선택 / 매장 선택)
 - Case 1 : 메뉴명, 지점명을 모두 발화한 경우
 - “스타벅스 소공동점에서 아메리카노 톨 사이즈 한 잔 주문해줘”
 - Case 2 : 메뉴명, 지점명을 발화하지 않은 경우
 - “스타벅스에서 주문해줘”
 - Case 3 : 메뉴명만 발화한 경우
 - “스타벅스에서 아메리카노 주문해줘”
 - Case 4 : 지점명만 발화한 경우
 - “스타벅스 소공동점에서 주문해줘”

III. 예상 발화 작성 가이드라인

III. 예상 발화 작성 가이드라인

- 단위 기능 (Task) 정의가 완료되면 각 기능별로 **사용자의 예상 발화를 작성**
 - 예상 발화 디자인은 사용자와 NUGU의 대화 상황에서 실제 발화할 것 같은 명령을 자연어 문장으로 기술
- 사용자 발화 문장의 구성 요소는 기기를 **Wake-up 시키는 호출어** 외에 **Entity**와 **Intent**로 구분
 - 예) 알람을 설정하는 기능에 대한 발화는 아래와 같이 정의될 수 있음



- 사용자 예상 발화 문장을 디자인할 때 고려 사항
 - 범용성
 - 일관성
 - 독립성
 - 확장성

III. 예상 발화 작성 가이드라인

- 발화의 범용성

- **보편적으로 사용되는 단어 및 구어체에서 사용되는 형식의 문장으로 디자인 함**

- 일반적인 사용자가 서비스 기능을 사용할 때 말하게 될 문장 그대로, 구어체 형식의 문장으로 작성
 - 보편적인 발화인지? 전문 용어를 사용하지 않았는지? 기계적으로 느껴지지 않는지? 살펴보아야 함
 - 개인의 발화 습관에 편향될 수 있으므로 여러 사람과 교차 확인을 하는 것이 좋음

- 발화의 일관성

- **서비스 내에서 혹은 유사 서비스와 발화 패턴이 일관적이어야 함**

- 서비스 내에서, 유사한 기능에 대한 발화 패턴은 일관적이어야 함
 - A 정보를 조회하는 기능의 예상 발화로 ‘알려줘’, ‘찾아줘’가 정의되었다면, B 정보를 조회하는 기능에 대한 예상 발화도 동일하게 정의하는 것이 좋음
 - 한 기능을 익힌 사용자가 다른 서비스에서 유사한 기능을 쉽게 이용할 수 있도록, 다른 서비스에서 제공하는 유사한 기능이 없는지 파악하고, 이를 참조하여 발화 패턴을 일관되게 적용하는 것도 중요함

III. 예상 발화 작성 가이드라인

■ 발화의 독립성

■ 각 기능별 예상 발화는 서로 구분되어야 하며, 의미가 명확히 인식될 수 있어야 함

- 다른 기능 실행에 영향을 주지 않는 발화인지를 고려하여 디자인해야 함
- 사용자의 요청을 정확히 파악하기 힘든 모호하거나 중의적인 표현은 배제하는 것이 좋음
- 발음이 어려운 경우 오인식 될 수 있음
 - 거센소리 (ㄱ, ㅋ, ㆁ, ㆅ), 된소리 (ㄲ, ㄸ, ㅃ, ㅆ, ㅉ) 는 인식률이 높아 이를 활용하면 오인식의 가능성이 줄어듦

■ 발화의 확장성

■ 하나의 의도에 대하여 다양한 표현을 작성하도록 함

- 사용자의 발화는 여러 가지 표현으로 나타낼 수 있으므로, 다양한 서술어와 Entity의 조합으로 예상 발화들을 작성
 - 조사나 어미는 자연어 처리를 통해 자동으로 학습되므로 이에 대한 확장은 정의하지 않아도 됨
- 사용자가 특정 시점 (예 : 어제, 오늘 등) 이나 조건 (예 : 서울, 부산 등) 을 한정하여 발화하거나 Entity 없이 발화하는 경우에는 별도 기능으로 구분하여 예상 발화를 작성

IV. 대화 흐름 디자인하기

IV. 대화 흐름 디자인하기

- 사용자와 NUGU 에이전트 간 대화는 **사용자의 발화에 대하여 NUGU 에이전트가 동작 (Action)을 실행하기** 까지의 흐름을 의미
- 대화의 유형
 - **Single-turn 대화**
 - 사용자가 발화한 요청에 대해 NUGU 에이전트가 답변한 대화가 종료되는 대화 유형
 - **Multi-turn 대화**
 - 사용자와 NUGU 에이전트가 대화를 2회 이상 주고받는 대화 유형

■ 명령 실행

- 사용자의 의도에 맞춰 서비스를 지원할 수 있는 경우

- 예) 사용자 : 아리아, 핫한 음악 틀어줘

NUGU : 멜론 Top 100을 들려드릴게요. (음악 재생)

- 명령 실행 시 실행의 결과가 사용자에게 명확하게 인지될 수 있고, 답변이 오히려 번거로울 수 있는 경우 Prompt 없이 해당 동작만 실행할 수 도 있음

- 예) 사용자 : (음악 재생 중) 아리아, 음악 그만

NUGU : (음악 재생 종료)

■ 명령 대안 실행

- 사용자의 의도에 100% 부합하는 결과를 낼 수는 없지만 요청을 받아들일 수 있는 대안이 준비된 경우

- 예) 사용자 : 아리아, 아이유 재즈 음악 들려줘

NUGU : 재즈 음악을 들려드릴게요

■ 미지원 안내

- 사용자의 의도는 파악되었으나 서비스의 지원 범위에 벗어난 경우
- 서비스 지원 여부를 인지하지 못한 사용자는 지원되지 않는 기능에 대한 시도를 계속할 수 있으므로 지원할 수 없는 서비스에 대해 명확히 안내

- 예) 사용자 : 어제 날씨 알려줘

NUGU : 저는 오늘 이전의 날씨 정보는 가지고 있지 않아요

- 지원 가능한 다른 기능에 대해 추가적으로 안내하는 것도 좋음

- 예) 사용자 : 코끼리가 중국어로 뭐야?

NUGU : 지금은 한영사전만 찾아보실 수 있어요. 영어 단어를 찾아보시려면, “아리아, 영어로 코끼리가 뭐야?” 라고 말씀해보세요.

- 연결오류 안내

- 네트워크의 연결 상태나 서버에 문제가 발생한 경우

- 오류 (Error) 상황에 대해 지나치게 자세히 설명할 필요는 없으나 원인에 대해 안내를 하는 것이 좋음

- 예) 사용자 : 아리아, 날씨 알려줘

NUGU : (날씨 서버 연결 오류 상태) 지금 날씨 정보를 가져올 수 없어요. 잠시 후 다시 말씀해주세요.

- 응답 없이 상황을 종료하거나 거짓으로 안내해서는 안 됨

- 간결하고 직관적인 오류에 대해 안내하도록 하며, 가능하다면 사용자가 시도할 수 있는 다음 단계를 안내

- 예) 사용자 : 아리아, 멜론 틀어줘

NUGU : (멜론 서버 연결 오류 상태) 지금 멜론 서비스에서 음악을 가져올 수 없어요. 잠시 후 다시 말씀해주세요.

■ 서비스 도움말

- 사용자는 Play안에 다른 어떤 기능들이 있는지 알기가 어려움
- 사용자가 다양한 기능을 활용할 수 있도록, “도움말” 이나 “사용방법 알려줘” 와 같은 발화에 대하여 서비스 도움말을 제공할 수 있음
- 도움말로 제공할 안내가 많을 경우, 한번에 모든 정보를 제공하기 보다는 주요 기능별로 분리하여 제공하는 것이 좋음
- 사용자가 도움말을 직접 요청하지 않더라도, 필요에 따라 자연스러운 흐름을 방해하지 않는 선에서 첫 진입 시 혹은 응답 중에 필요한 도움말
(예 : 다른 기능을 소개) 을 제공하는 것도 고려해 볼 수 있음
 - 로그인이 필요한 서비스에서 로그인하지 않고 해당 서비스 명령을 한 경우, 로그인이 필요하다는 안내와 함께 로그인 방법에 대한 가이드를 제공하는 것이 좋음

- Slot-filing

- 사용자의 의도는 파악하였으나 추가 정보가 필요해 다시 묻는 경우
 - 요청 이행을 위한 Entity가 누락되었을 경우에는 사용자에게 부족한 Entity 정보를 더 얻은 후 서비스를 제공
 - 예) 사용자 : 아리아, 알람 설정해줘

NUGU : 몇 시로 설정할까요? ← Slot-filling

사용자 : 7시

NUGU : 내일 오전 7시로 알람을 설정했어요.

▪ Slot-filing

- Slot-filling 시에는 누락된 Entity에 해당하는 사용자 발화만 받아들이며, 그 외 발화에 대해서는 잘못된 명령임을 안내

- 2회 연속 잘못된 명령이 입력되거나, 미발화 상태가 7초 간 유지될 경우 대화가 종료 (Play 종료)

- 예) 사용자 : 아리아, 알람 맞춰줘

NUGU : 몇시로 맞춰드릴까요? ← Slot filling

사용자 : 몹시 흥분

NUGU : 원하시는 알람 시간을 이해하지 못했습니다. 시간을 다시 말씀해주세요. ← 필요한 Entity로 인식을 1회 실패

사용자 : 서울시

NUGU : 원하시는 알람 시간을 이해하지 못했습니다. ‘아리아, 오전 또는 오후 몇시 알람 설정해줘’ 라고 말씀해주세요. ← 필요한 Entity 인식을 2회 실패, Slot-filling을 더 시도 하지 않고 Idle 상태가 됨

- Slot-filing
 - Slot-filling 문장 작성 시에는 아래의 항목들을 고려
 - 사용자가 질문을 확실히 인지할 수 있도록 문장을 구성
 - 실행 가능한 기능에 대해서만 질문
 - 사용자에게 얻고자 하는 답변이 '예/아니오' 인지, 추가 정보인지에 따라 질문을 구성
 - 부족한 필수 Entity (Required Entity) 에 대해서만 Slot-filling하며, 선택적인 Entity (Optional Entity) 에 대해서는 기본값으로 설정하여 Slot-filling의 횟수를 최소화

■ 다음 명령 요청

- 사용자 발화에 대해 동작 후 다시 마이크를 열고 사용자의 새 명령을 기다리는 경우
 - 앞서 수행한 Play의 컨텍스트를 유지하고, Listening-passive 상태로 진입
 - Slot-filling과 마찬가지로 2회 연속 잘못된 명령이 입력되거나 미발화 상태가 7초 간 유지될 경우 대화가 종료 (Play 종료)

- 예) 사용자 : 배송차량 도착시간 알려줘

NUGU : 상온 배송차량이 아직 물류센터를 출발하지 않았습니다. < 답변

NUGU : 무엇을 더 도와드릴까요? (Listening 상태) < 다음 명령 요청

- 수행할 수 없는 명령을 한 경우에는 한 번 더 사용자에게 묻는 것도 고려해 볼 수 있음

- 예) 사용자 : 모짜르트의 애국가 들려줘

NUGU : 제가 처리할 수 없는 조합이네요. < 답변

NUGU : 다시 한번 말씀해주시겠어요? (Listening 상태) < 다음 명령 요청

V. Prompt 작성 가이드라인

V. Prompt 작성 가이드라인

- Prompt 란?
 - NUGU 에이전트의 음성 응답을 의미
 - 음성으로 전달되는 정보를 사용자가 잘 이해할 수 있도록 다음과 같은 가이드라인을 고려하여 디자인하는 것이 좋음
 - Prompt 정확성
 - Prompt 간결성
 - Prompt 다양성
 - Prompt 일관성

- 다음 사항에서는 **정보의 누락 없이 정확하고 명확하게 Prompt를 작성**해야 함
 - 정보 전달이 목적인 경우
 - 사용자가 각 기능 실행에 대한 답변 (Prompt) 피드백만으로도 이해할 수 있도록 필요한 정보를 빠짐없이 모두 포함
 - 질문인 경우
 - 사용자가 Prompt를 들었을 때 질문형으로 이해 할 수 있도록 문장을 구성
 - 원활한 상호작용을 위해 사용자가 대답해야 하는 발화를 명확하게 제시
 - 오류 상황인 경우
 - 오류 발생의 이유나 해결 방법을 구체적이고 정확하게 제시
 - 띄어쓰기와 구두점
 - TTS가 문장을 출력할 때 공백과 쉼표, 마침표를 인식하여 출력하므로 문장 안에 필요한 공백이나, 쉼표, 마침표 등을 적절하게 사용하여 흐름이 자연스럽게 작성
 - 예) BBQ 황금 올리브치킨이 16,000원에 준비되어 있습니다. 주문이 접수된 후에는 변경이나 취소가 불가능해요. 서울시 중구 을지로 2가 T타워로 배달을 원하시면 “아리아, 주문 접수” 라고 꼭 말씀해주세요.

- 사용자와 음성으로만 상호작용하기 때문에 **효율적인 의사 전달을 위해 간결하게** 말해야 함
 - 불필요한 꾸밈말이나 조사를 사용하여 답변 (Prompt) 가 불필요하게 길어지지 않도록 주의
 - 기능 실행이 목적인 답변 (Prompt)인 경우에는 가능한 짧고 간결한 형태로 응답
 - 예) 사용자가 이미 알고 있는 주소를 안내할 경우, 시,군,구,동 등 전체 주소를 안내하기 보다는 동 이하 주소만 안내
 - 한 문장 안에는 기본 2개, 최대 4개의 정보만 배치
 - 전체 문장이 3문장 이상이 되어야 할 경우 중요한 정보를 포함하고 있을 경우에만 허용
 - 이 경우, 정보와 정보 사이를 구분하여 읽어 사용자의 이해를 도와야 함
 - 문장과 문장 사이에 동일한 단어 또는 술어를 반복 사용하여 중복되는 느낌을 주는 것을 지양
 - 예) 오전 날씨는 비가 오고, 오후에는 비가 와요 (X)
하루 종일 비가 와요 (O)

- 간결성 확보를 위한 정량적 가이드 사례
 - 작성된 답변 (Prompt)를 실제로 읽어보며 호흡을 체크
 - 한 번 이상 호흡해야 할 경우 문장의 길이를 줄이는 것이 좋음
 - 응답 하나당 약 500자 또는 90초 이하가 되도록 작성하는 것이 좋음
 - 긴 목록 항목이거나 사용자가 더 깊이 생각해야 하는 항목인 경우에는 시간을 더 확보
 - 목록의 형태로 안내해야 할 경우, 목록에 대한 제목을 먼저 제시해 앞으로 소개될 목록을 사용자가 예측할 수 있도록 하며 제목을 제시한 후에는 잠깐 쉬고 목록을 안내함
 - 항목이 많은 경우에는 모든 항목을 안내하는 것 보다 일부만 먼저 안내한 후 더 듣길 원하는지 사용자에게 의견을 묻도록 함
 - 이 경우, 먼저 안내되는 항목은 20초 이상 걸리지 않아야 함
 - 항목이 많은 경우에는 항목을 2개에서 5개 사이로 그룹화하여 안내
 - 문장 간 띄어 읽을 수 있도록 각 항목 뒤에 SSML (Speech Synthesis Markup Language) 을 사용

- 사용자가 자주 이용하는 서비스라면 매번 동일한 응답을 듣는 것이 지루할 수 있으므로 **다양한 응답으로 구성**

- 예) 사용자 : 아리아, 놀아줘

NUGU : 아, 저도요. 오후엔 졸리고 무료하고 좀 그렇네요. 제가 내는 퀴즈를 맞춰 보실래요?

사용자 : 아리아, 놀아줘

NUGU : 흠, 이럴 때는 뭘 해드려야 칭찬받을까요? 요즘 유행하는 신조어를 지금 배워보는 건 어때요?

사용자 : 아리아, 놀아줘

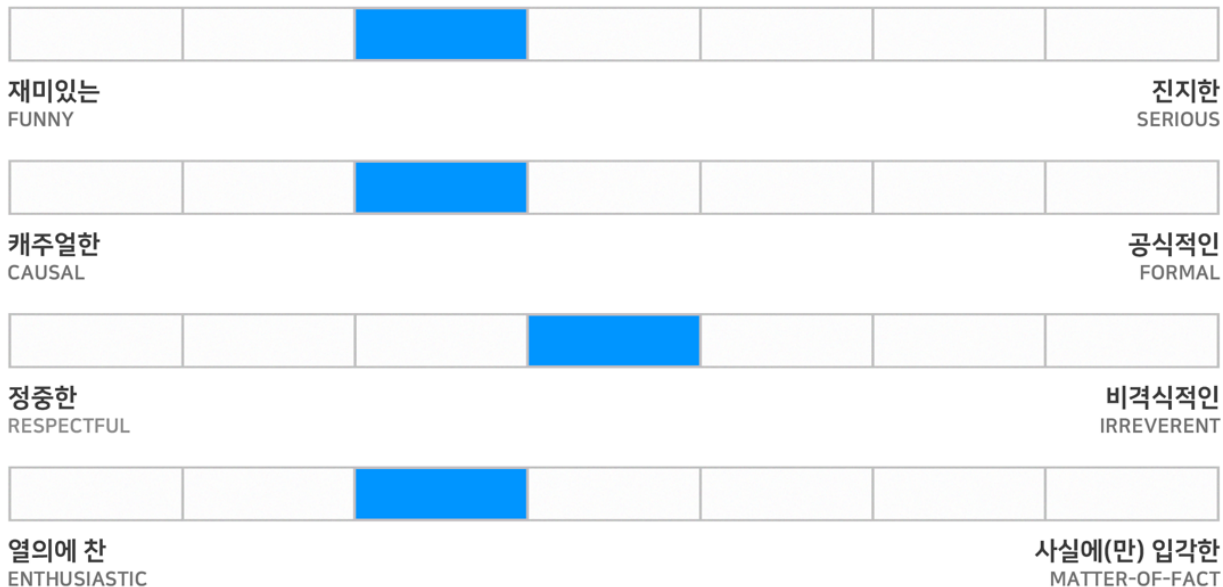
NUGU : 아, 이런 날은 뭔가 재미 있는 일이 생기면 좋을 텐데요. 음, 제가 준비한 무드등 개인기 한번 보실래요?

- NUGU 페르소나
 - 인격을 뜻하는 라틴어로서, NUGU 페르소나란 사용자가 NUGU 에이전트에 대해 가지는 인격에 대한 심성을 의미
 - 사람들을 다른 사람의 목소리를 들을 때 머릿속에 자동으로 (빠르고, 무의식적으로) 그 사람에 대한 캐릭터 (나이, 교육 정도, 감정의 상태, 얼마나 믿을 만한지, 친절한지 등)을 형성하게 됨
 - NUGU 사용자는 NUGU 에이전트가 제공하는 말투나 목소리 등으로부터 이러한 이미지를 갖게 됨
 - 서비스 전반에 걸쳐 일관된 페르소나를 유지하여 사용자가 NUGU 에이전트와 자연스럽게 상호작용할 수 있도록 VUX를 디자인하는 것이 필요
 - NUGU의 페르소나는 **20대 후반에서 30대 초반 여성**으로, **차분하고 편안함을 지향**하는 다정다감한 성격의 **친구 같은 비서**

- NUGU 페르소나

- 세부 대화 스타일은 다음과 같으며, 작성한 Prompt가 NUGU 페르소나와 일치하는지 검토하는 것을 권장

Kate Meyer, The Four Dimensions of Tone of Voice, Nielson Norman, 2016



■ 스타일

- 실제 대화에서도 언어가 동일한 사람들 간에 말투나 억양이 조금씩 다르며, 이는 각 개인의 습관이 반영
- 작성자 개인의 언어 습관이 Prompt에 미치는 영향을 줄이고, NUGU 페르소나에 맞는 어투를 적용하기 위한 기본 가이드는 다음과 같음
 - 구어체로 친근하게 응답
 - 가능한 구어체를 사용해 자연스러운 대화를 도출
 - 답변 (Prompt) 디자인 시, **글로 적다 보다 소설이나 에세이처럼 문어체로 작성될 우려**가 있음
 - 예) “아이유 밤편지를 재생합니다.” 보다는 “아이유의 밤편지, 들려드릴게요.”
 - 서술어 형태의 교차 배치
 - 여러 문장을 제공해야 한다면 문장 서술어 부분에 **‘~예요’ 와 ‘~입니다’ 표현을 교차 사용**하여 자연스러운 문장으로 구성
 - 예) 로그인 전에는 미리 듣기만 가능합니다. 전곡 듣기를 원하시면 누구 앱을 통해 로그인해 주세요.

End of Document