공학석사 학위논문

음성 인터페이스 기반 음악 서비스에서의 쿼리 유형 별 특성 연구

– 사용자가 추천 스트림에 기대하는 요소를 중심으로 –

2023 년 2 월

서울대학교 융합과학기술대학원

지능정보융합학과

박 상 아

음성 인터페이스 기반
음악 서비스에서의
쿼리 유형 별 특성 연구

– 사용자가 추천 스트림에

기대하는 요소를 중심으로 –

지도 교수 이 중 식

이 논문을 공학 석사 학위논문으로 제출함

2023 년 2 월

서울대학교 융합과학기술대학원

지능정보융합학과

박 상 아

박상아의 공학석사 학위논문을 인준함

2023 년 2 월

위 원 장 _____ 서 봉 원 _____ (인)

부위원장 _____ 이 중 식 _____ (인)

위 원 _____ 이 교 구 _____ (인)

초 록

음성 인터페이스에서 음악 서비스는 핵심적인 도메인이다. 조사에 따르면, 음악은 스마트 스피커 사용자들이 가장 일상적으로 사용하는 서비스 중 하나이며 시간 당 사용 빈도 역시 가장 높게 나타난다. 지배적으로 사용되는 도메인인 만큼, 음악 경험은 음성 인터페이스의 입문 또는 이탈에도 영향을 줄 수 있다. 따라서 음성 인터페이스만의 음악 경험을 이해해야 한다.

음성 인터페이스에서의 음악은, 하나의 쿼리를 트리거하면 자동 생성된 음악 리스트가 연속으로 재생되는 형태이다. 기존의 모바일 인터페이스가 '검색 (쿼리 입력) > 탐색 (결과 리스트 탐색) > 재생 (곡 클릭)'의 순서로 이어지는 것과 달리, 음성 인터페이스에서는 탐색 단계가 존재하지 않으며 검색과 재생이 동시에 이루어진다. 즉 하나의 쿼리에 의해 상위 결과와 그 연관 곡들이 스트림 형태로 출력된다.

따라서 음성 인터페이스만의 음악 경험을 파악하기 위해서는 사용자가 입력하는 쿼리를 이해해야 한다. 쿼리 형태에 따라, 사용자가 결과에 대해 예상하는 바가 달라질 것이기 때문이다. (e.g., "신나는 재즈 틀어줘" VS "음악 틀어줘") 이때, 사용자가 결과에 대해 가지는 기대치와 실제 검색 결과의 간격이 크면, 경험에 영향을 미칠 수 있다. 긍정적인 경우 새로운 노래를 발굴하는 세렌디피티로 이어지기도 하지만, 부정적인 경우에는 추천에 대한 불신이나 음성 인터페이스 사용 저하로 이어질 가능성이 있다.

본 연구에서는 음성 인터페이스의 음악 도메인에서 사용되는 쿼리의 유형을 이해하고, 쿼리 유형 별로 사용자가 기대하는 추천 스트림을 파악하고자 한다. 이를 바탕으로, 음성 인터페이스 음악 추천 방식에의 디자인 함의점을 제시하고자 한다. 본 연구는 크게 두 개의 조사를 진행하였다.

1차 조사의 목적은 음성 인터페이스의 음악 도메인에서 사용되는 쿼리의 유형을 이해하는 것이다. 스마트 스피커 사용자 9명의 3개월 치 음악 관련 로그 2,723개를 수집한 후, 선행 연구를 기반으로 음악을

트리거 하는 쿼리를 유형화하였다. 그 결과 음악 쿼리는 크게 세 가지로 분류되었다: 1) SQ - Specific Query, 곡이나 아티스트로 요청, 2) NSQ - Non-Specific Query, 기준을 제시하지 않음, 3) DQ - Descriptive Query, 분위기나 장르를 묘사. 로그 분석 결과, 쿼리 유형 별로 재쿼리를 시도하는 횟수와 시점이 다르게 나타났다. 로그 기반 인터뷰 결과, 쿼리 유형에 따라 발화 의도와 만족도가 서로 다르게 나타났다.

2차 조사의 목적은 쿼리 유형에 따라 사용자가 추천 결과에 대해 가지는 기대를 심층적으로 파악하는 것이다. 5일 동안 27명의 참여자에게 음성 인터페이스로 음악을 트리거하는 ESM 태스크를 부여하여, 설문을 통해 음악 추천에 대한 기대와 인식을 5점 척도로 수집하였다. 총 쿼리 290개에 대한 기대와 인식 설문이 수집되었으며, 분석 결과 다음의 특성이 도출됐다: 1) SQ - 예상 내의 연관성 높은 곡들을 기대하며 만족도가 높음. 2) NSQ - 새로움, 다양성, 의외성 높은 곡들을 기대하며 만족도는 낮으나 결과에 관용적임. 3) DQ - 새로움, 다양성, 의외성 높은 곡들을 기대하며 만족도가 낮고 결과에 엄격함.

본 연구의 결과를 토대로, 다음과 같은 논의를 진행하였다. 첫째, 음성 인터페이스의 음악 경험이 기존과 크게 세 지점에서 다르며 (배경적 청취, 일람성 부재, 인식 오류 가능성), 이에 따라 사용자들의 쿼리 선택이 전략적으로 달라진다. 둘째, 사용자들이 기대하는 요소를 토대로 쿼리 유형 별 추천 스트림의 설계 방식을 제언한다.

본 연구는 음성 인터페이스 기반 음악 도메인에서 사용되는 쿼리의 유형을 파악하고, 추천에 대한 사용자의 기대와 인식을 쿼리 유형 별로 확인하였다. 또한 음성 인터페이스 기반 음악 경험과 쿼리의 특성을 밝히고, 쿼리 유형 별 음악 추천 방식을 제언하였다.

주요어 : 음성 인터페이스, 음악 추천 스트림, 음악 쿼리.
학 번 : 2020-24842

목 차

제 1 장 서 론 7
제 1 절 연구의 배경 7
제 2 절 연구의 목적 9

제 2 장 관련 연구 10
제 1 절 음성 인터페이스의 특성 10
1.1 음성 인터페이스의 모달리티 특성
1.2 음성 인터페이스의 쿼리
제 2 절 음악 도메인의 사용자 니즈와 쿼리 14
2.1 스트리밍 서비스로의 변화
2.2 음악 서비스에서의 사용자 니즈와 행동
2.3 음악 서비스의 쿼리
제 3 절 사용자 중심의 음악 연구 방법 18
3.1 음악 도메인에서의 사용자 조사
3.2 사용자 중심의 음악 추천 평가 척도

제 3 장 연구 문제 21
제 1 절 연구 문제 21
제 2 절 연구 구조 22

제 4 장 1차 조사 23
제 1 절 연구 방법 23
1.1 수집 방법
1.2 연구 참여자 선정 기준 및 모집
1.3 분석 방법
1.3.1 쿼리 유형 분류
1.3.2 쿼리 유형 별 패턴 분석
제 2 절 연구 결과 29
2.1 쿼리 유형
2.2 쿼리 유형 별 재쿼리 패턴
2.3 쿼리 유형 별 발화 의도와 만족 여부
제 3 절 소결론 34

제 5 장 2차 조사 36
제 1 절 연구 방법 36
1.1 수집 방법
1.2 연구 참여자 선정 기준 및 모집
1.3 분석 방법
제 2 절 연구 결과 44
2.1 쿼리 유형에 따른 스트림 인식
2.2 쿼리 유형에 따른 스트림 기대
2.3 쿼리 하위 유형과 사용자 특성에 따른 스트림 기대
제 3 절 소결론 48

제 6 장 논의 49
제 1 절 음성 인터페이스의 음악 쿼리 특성 49
제 2 절 음성 인터페이스의 쿼리 별 추천 스트림 제언 53

제 7 장 결론 55
제 1 절 연구 요약 55
제 2 절 연구 한계 57
제 3 절 연구 의의 58

부록 59

참고문헌 63

표 목차

[표 1] 1차 조사 참여자의 음성 인터페이스 사용 행태..... 25
[표 2] 1차 조사로 도출된 쿼리 유형 29
[표 3] 쿼리 유형에 따른 재쿼리 패턴 30
[표 4] 쿼리 유형 별 발화 의도와 만족 여부..... 32
[표 5] 2차 조사의 태스크 중 수집되는 설문 문항..... 39
[표 6] 2차 조사 참여자 목록 41
[표 7] 2차 조사를 통해 수집된 데이터 예시 43
[표 8] 음성 인터페이스의 음악 쿼리 특성 요약..... 49

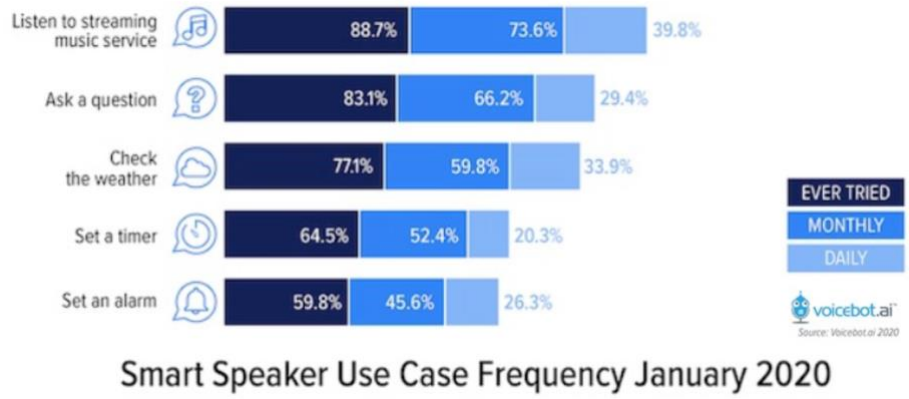
그림 목차

[그림 1] 2020년 1월 기준 스마트 스피커에서 자주 사용되는 서비스 목록 7
[그림 2] 기존 인터페이스와 음성 인터페이스의 음악 경험 8
[그림 3] Hosey et al. (2019)가 제시한 음악 검색에서의 사용자 태도(mindset) 15
[그림 4] 본 연구의 전반적인 구조 22
[그림 5] 1차 조사에서 수집된 원시 데이터..... 24
[그림 6] 재쿼리의 예시 27
[그림 7] 쿼리 유형에 따른 재쿼리 시점 31
[그림 8] 2차 조사의 ESM 태스크 구조..... 37
[그림 9] 2차 조사 참여자에게 제공된 태스크 안내문 38
[그림 10] 쿼리 유형에 따른 스트림 기대 요소 45
[그림 11] 쿼리 유형에 따른 스트림 인식 46
[그림 12] 헤비 유저와 라이트 유저의 쿼리 유형 별 만족도.. 47
[그림 13] 쿼리 유형 별 추천 스트림에 기대 차이 53

제 1 장 서 론

제 1 절 연구의 배경

음성 인터페이스에서 음악 서비스는 핵심적인 도메인이다. Bentley et al. (2018)가 구글 홈 사용자 88명의 사용 기록을 분석한 연구에 따르면, 음악은 스마트 스피커에서 시간 당 사용 빈도가 가장 높으며 전체 명령어 중 40% 이상을 차지할 만큼 지배적이다. 뿐만 아니라 Ammari et al.(2018)의 조사에서는 스마트 스피커 장기 사용자들이 가장 일상적으로 사용하는 도메인 중 하나로 꼽히기도 했다 (음악, 검색, IoT). [그림 1]의 Voicebot.ai의 조사에서도 나타난 것처럼, 음악 스트리밍 서비스는 스마트 스피커 사용자들이 최소 한 번 이상 사용했거나 (88.7%) 월에 한 번 이상 (73.6%), 혹은 하루에 한 번 이상 (39.8%) 사용하는 핵심적인 도메인이다. [①]



[그림 1] 2020년 1월 기준
스마트 스피커에서 자주 사용되는 서비스 목록

① BRET KINSELLA, “Streaming Music, Questions, Weather, Timers and Alarms Remain Smart Speaker Killer Apps, Third-Party Voice App Usage Not Growing”. <Voicebot.ai>, 2020.05.03.

초기 사용자와 장기 사용자가 모두 사용하는 서비스인 만큼, 음악 경험은 음성 인터페이스 자체의 입문과 이탈에도 영향을 줄 수 있다. 따라서 음성 인터페이스만의 음악 경험을 이해하는 것이 필요하다.

음성 인터페이스 기반 음악 서비스는 하나의 쿼리를 트리거 하면 자동 생성된 음악 리스트가 연속으로 재생되는 형태이다. 즉, 음악을 트리거 할 때 조작할 수 있는 요소가 '쿼리' 뿐이라는 것이다. 이는 기존 인터페이스의 음악 트리거 과정보다 단축된 형태이다.

예컨대 [그림 2]처럼 모바일 환경에서는, 사용자가 검색을 위해 쿼리를 입력하면 검색 결과 리스트가 출력된다. 사용자는 클릭이나 스크롤을 통해 해당 리스트를 탐색한 다음, 원하는 곡을 최종적으로 재생하게 된다.

하지만 보이스 인터페이스에서는 사용자가 검색 결과를 탐색할 수 없다. 검색과 재생이 분리되어 있지 않으므로, 쿼리를 입력함과 동시에 상위 결과와 그 연관 곡들이 즉각적으로 출력된다. 이로 인해 개별 곡에 대한 검색보다는 음악 스트림의 재생에 의존하게 된다.

[그림 2] 기존 인터페이스와 음성 인터페이스의 음악 경험

그렇기 때문에 음성 인터페이스에서는 사용자가 입력하는 쿼리를 이해하는 것이 중요하다. 쿼리의 유형에 따라 사용자가 결과에 대해 예상하는 바가 달라질 것이기 때문이다. 예컨대 사용자가 "신나는 재즈 틀어줘"라고 발화했을 때에는, "음악 틀어줘"라고 발화했을 때보다 더욱 특정한 곡의 흐름을 기대한다.

이때 사용자가 결과에 대해 가지는 기대치와 실제 검색 결과 사이의 갭이 크면 경험에 영향을 줄 수 있다. 긍정적인 경험으로 이어질 경우, 사용자는 예상하지 않았던 새로운 노래를 발굴하는 세렌디피티(Serendipity)를 느낄 수 있다 (Jin et al., 2018). 하지만 본 연구에서 진행한 인터뷰에 따르면, 예상과 다른 추천 결과는 종종 부정적인 경험으로 이어지기도 한다. 모든 인터뷰 참여자들이 음성 인터페이스에서 음악을 전환하는 이유 중 하나로 '기대했던 곡이 나오지 않음'을 꼽았으며, 특정 참여자들의 경우에는 추천에 대한 불신이나 음성 인터페이스 사용 저하로 이어지기도 했다. 따라서 사용자의 기대치를 고려해 보다 섬세한 추천을 제공할 필요가 있다.

한편 음성 쿼리는 in-situ의 사용자 니즈를 명확하게 포착할 수 있다 (Xiao et al., 2021; Lovine et al., 2021; Kostric et al. 2021). 이 특성은 음악 도메인에서 효과적으로 작동할 것이다. 음악 청취가 스트림화 되면서 사용자들은 1) 저마다 다양한 기준으로 새로운 노래를 발견하고자 하며, 2) 배경적 청취가 잦아져 사용자의 상황 맥락과 의도를 파악하는 것이 중요해졌기 때문이다 (Jin et al., 2018; Volokhin & Agichtein, 2018).

즉 다변적이고 복합적인 사용자의 음악 니즈를, 해상도 높은 음성 쿼리를 이용해 정교하게 포착하고자 하는 것이 본 연구의 동기이다.

제 2절 연구의 목적

위와 같은 맥락에서, 본 연구에서는 음성 인터페이스의 음악 도메인에서 사용되는 쿼리의 유형을 이해하고, 쿼리 유형 별로 추천 스트림에 기대하는 요소를 파악한다.

이를 통해, 음성 인터페이스 쿼리의 차별점을 밝히고 음악 추천 방식에의 디자인 함의점을 제시하고자 한다.

제 2 장 관련 연구

본 연구가 탐구하고자 하는 것은 1) 음성 인터페이스에서 나타나는 음악 쿼리의 유형 2) 쿼리 유형에 따라 추천 스트림에 기대하는 요소이다.

따라서 음성 인터페이스의 특성과 기존 인터페이스와의 차별점을 파악해야 한다. 또한 음악 도메인에서 사용자의 니즈와 행동이 어떻게 나타나고, 쿼리에 대한 기존 연구는 어떻게 진행되었는지 살펴보아야 한다. 마지막으로 음악 도메인의 사용자 중심 연구 방법과 추천 평가 척도를 파악해, 본 연구에 반영하여야 한다.

제1절에서는 음성 인터페이스의 특성을 알아본다. 세부적으로 음성 인터페이스의 모달리티 특성을 서술하고, 음성 인터페이스의 쿼리에 대한 기존 연구들을 살펴본다. 제2절에서는 음악 도메인의 사용자 니즈와 쿼리를 알아본다. 세부적으로 스트리밍 서비스의 등장과, 해당 서비스에서의 사용자 검색 니즈와 행동을 파악한다. 또한 음악 쿼리에 대한 기존 연구를 훑어본다. 제3절에서는 사용자 중심의 음악 연구 방법을 알아본다. 세부적으로 사용자 연구 방법과 음악 추천의 평가 척도를 살펴본다.

제 1 절 음성 인터페이스의 특성

1.1 음성 인터페이스의 모달리티 특성

음성 인터페이스(Voice interface) 상에서, 사용자는 구어를 통해 시스템과 자연스러운 방식으로 상호작용한다. 사용자는 손의 움직임 없이, 빠르고 직관적으로 기기를 조작할 수 있다 (Pearl, 2016).

음성 인터페이스의 가장 큰 특징은, 모바일이나 웹과 달리 시각적 큐(visual cue)가 존재하지 않는다는 점이다. 음성 인터페이스의 차별적인 경험은 이러한 모달리티 특성에서 파생된다. 시각적 큐가 존재하지 않기 때문에, 사용자는 시스템을 자유롭게 조작하거나 조작에 대한 결과를 예측하기 어렵다. (Myers et al., 2018; Corbett & Weber,

2016; Luger & Sellen, 2016) 선택 가능한 옵션을 화면에 보여주어 사용자가 빠르게 스키밍(skimming)할 수 있는 기존 인터페이스와 달리, 음성 인터페이스는 옵션에 대한 정보를 제공하기 어렵다. 이는 사용자가 무엇을 말해야 할지 모르거나, 가능한 기능을 쉽게 알지 못하는 'Learnability' 또는 'Discoverability' 문제로 이어지기도 한다 (Corbett & Weber, 2016).

하지만 자연어(Natural language)로 사용자와 소통하기 때문에, 인간에게 가장 자연스러운 방식으로 커맨드를 수집할 수 있다. 사용자는 자신의 경험이나 의견, 선호 등을 시스템에 명시적으로 입력할 수 있으며, 인터랙션이 일어나는 그 순간(in-situ)의 니즈를 표현할 수 있다 (Xiao et al., 2021; Lovine et al., 2021; Kostric et al., 2021). 이는 특히나 사용자의 피드백이 요구되는 추천 시스템에서 유용하게 작용한다. 이 때문에 대화형 추천 시스템(CRS, Conversational Recommender System) 분야에서는 '현재의 구체적인 선호를 수집하는 방안'에 대한 연구가 활발하게 진행되고 있기도 하다 (Jannach et al., 2022; Kostric et al., 2021; Kang et al., 2017; Grasch et al., 2013).

위와 같은 특성 때문에, 음성 인터페이스의 사용자 연구는 쿼리를 수집해 분석하는 방식으로 진행되는 경우가 많다. 예컨대 Guy(2018)은 웹 음성 검색 환경에서, Bentley et al. (2018), Beneteau et al. (2019), Mavrina et al. (2022)는 스마트 스피커 환경에서 사용자 로그를 수집하여 음성 인터페이스에서 사용되는 기능을 양적으로 분석하였다. 이때 Beneteau et al. (2019)는 사용자와 로그를 함께 살피며 발화 의도를 묻는 인터뷰를 병행하였다.

Trippas et al. (2018), Kiesel et al. (2018)은 음성 기반의 검색에서 나타나는 쿼리의 유형과 경험을 이해하기 위해 in-lab 환경에서 대화 기반 검색 실험을 진행하였다. 또한 Zhao et al. (2022)는 비디오를 음성 인터페이스로 조작할 때 어떤 쿼리가 사용되는지 관찰하기 위해 Wizard of Oz 방식으로 실험을 진행하였다. Myers et al. (2018)은 캘린더를 조작할 때 어떤 쿼리가 사용되는지 수집하기 위해 새로운 프로토타입을 제작하여 일정 기간 사용하도록 했다.

1.2 음성 인터페이스의 쿼리

음성 쿼리에 관한 이전 연구들은 대개 음성 쿼리의 ASR 인식 향상에 치중되어 있다 (Chelba & Schalkwyk, 2013, Li et al.,2009). 본 연구에 참고할 수 있는 선행 연구들에는 1) 대화형 검색의 특성과 2) 음성 쿼리의 특성 및 전략 연구들이 있다.

먼저 대화형 검색 측면에서, Trippas et al.(2018)은 음성으로만 정보 검색이 이루어질 때 나타나는 상호작용의 양상을 살펴보고자 in-lab 환경의 실험을 진행했다. 실제 검색 시스템을 이용하는 대신, 두 명의 참가자가 각각 정보 탐색자와 검색 엔진 역할을 맡아 대화를 나누었다.

음성 검색의 상호작용은 보다 복잡했다: 기존 텍스트 쿼리 형태 외에도 다양한 형태의 발화를 보였다 (query babbling, instructions 등). 또한 시스템과 사용자의 협업이 증가했다: 시스템의 결과에 사용자가 추가 정보나 기준을 요청하는 등 능동적으로 개입했다.

음성 쿼리의 특성 측면에서, Guy (2016), Guy et al. (2018), Xing et al.(2020)은 텍스트 쿼리와 음성 쿼리를 직접적으로 비교했다. 음성 쿼리가 더 자주 사용되는 맥락은 다음과 같다: 결과가 음성으로 출력되기 쉬운 오디오-비디오 콘텐츠/ 스크린과의 상호작용이 덜 필요한 정보 주제/ 발음하기 쉽지만 쓰기 어려운 단어.

또한 음성 쿼리는 텍스트 쿼리에 비해 평균적으로 문자의 길이와 지속 시간이 길게 나타나고, 언어적 표현이 자연스럽고 보다 설명적인 경향을 보인다. 그리고 음성 쿼리는 사용자의 주관적인 기준과 판단으로 형성되는 경향을 보인다 (e.g., “슬픈 영화 추천해줘”).

사용자 인식 측면에서, 음성 쿼리는 텍스트 쿼리에 비해 자연스럽고, 흥미로우며, 편리하다고 인식되었다. 사용하는 상황 맥락(e.g., 공공장소 vs 혼자 있음)이 경험에 중요한 영향을 끼쳤다. 대개의 문제는 ASR 오류로, 결과에 대한 신뢰도 하락으로까지 이어지기도 했다.

음성 쿼리 형성(formulation) 전략에 대한 연구는, 대부분 음성 인터페이스의 소통 오류 상황을 주제로 한다. Myers et al.(2018)은 사용자들이 음성 기반 프로토타입을 이용할 때 발생하는 오류와, 그것을 어떤 전략으로 극복하는지 분류했다. 교정, 단순화, 새로운 발화, 새 정보 요청, 종료 등의 음성 쿼리가 나타났다. 또한 Mavrina et

al.(2022)는 음성 인터페이스에서 나타나는 소통 오류를 파악하기 위해, 사용자의 대응 전략과 시스템의 반응 유형을 정리했다. 사용자의 대응 전략은 크게 reformulation, repetition, articulation으로 나타났다.

이외에도 특정 도메인의 음성 쿼리를 분류한 연구들이 있다. Zhao et al.(2022)는 비디오를 시청할 때 음성으로 조작하는 태스크를 제공하여, 조작에 사용하는 정보에 따라 두 가지의 커맨드를 발견했다 (Navigation command, Content-based command). 또한 Kang et al.(2017)에 따르면, 음성으로 영화를 검색하는 태스크에서 참가자들은 다양한 방식으로 니즈를 표현했다: 구조화되지 않고 모호한 정보(e.g., “결말이 열려 있거나 반전이 있는 영화”), 주관적인 감정이나 경험(e.g., “슬픈 영화”), 영화 퀄리티(e.g., “흥미로운 캐릭터, 기발한 스토리”), 다른 영화와의 유사성 등. 연구자들은 해당 쿼리를 크게 ‘objective’, ‘subjective’, ‘navigation’ 세 가지로 분류했다.

본 연구에서는 음성 인터페이스가 가지는 특성을 기반으로, 기존과 다른 음성 인터페이스 기반 음악 경험을 탐구해 보고자 한다. 이때 언급한 선행 연구들과 같이, 음성 쿼리를 이용해 사용자의 음악 니즈와 의도를 파악하고자 한다.

제 2 절 음악 도메인의 사용자 니즈와 쿼리

2.1 스트리밍 서비스로의 변화

음악 듣기의 수단은 물리적 장치(CD, 카세트 등)에서 디지털 저장소(모바일, 컴퓨터)로, 그리고 스트리밍 서비스로 변화하였다. Lee et al.(2017)에 따르면, 스트리밍 서비스의 등장으로 인해 1) 사용자들은 수많은 카탈로그에 접근하게 되었으며, 알고리즘으로 더 다양한 콘텐츠를 발견할 수도 있게 되었다. 실제로 새로운 곡을 학습하기 위해 검색이 이루어지기도 한다 (Jin et al., 2018; Hosey et al., 2019). 2) 내가 저장해 놓은 곡의 리스트뿐만 아니라, 알고리즘으로 자동 생성된 곡의 '스트림'을 듣게 되었다. 쿼리를 입력하는 행동이 '특정 곡 검색'을 넘어, '특정 스트림 검색'으로까지 이어지게 된 것이다.

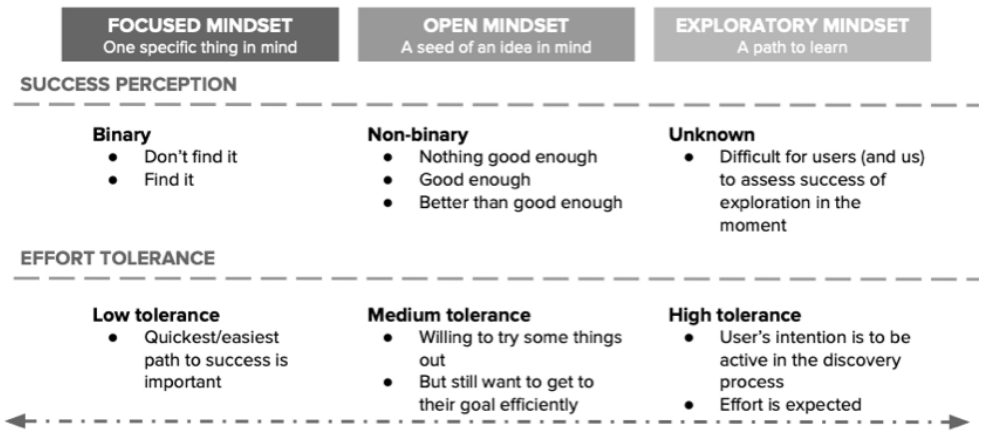
이로 인해, Schedl et al.(2018)은 음악 도메인에서 중요한 챌린지 중 하나로 'APG(Automatic Playlist Generation) 또는 APC(Automatic Playlist Continuation)'를 꼽았다. 이는, 플레이리스트의 타겟 특성에 걸맞은 트랙을 시스템이 자동적으로 추가하는 것을 의미한다. 실제로 사용자들에게 seed song(쿼리)으로 생성된 플레이리스트를 듣고 평가를 요청한 Lee et al.(2011)의 연구에서, 평가에 고려하는 기준이 다양하게 나타났다 (무드, 장르, 가사, 템포, 악기 등).

사용자가 음악—특히 스트림을 청취하는 목적은 상황에 따라 매우 달라진다. 음악 청취는 독립적으로 이루어지기보단 배경적으로 이루어지며 (운전, 운동 등 병행 활동 시), 이러한 활동이나 상황 맥락에 따라 서로 다른 타입의 음악을 원한다 (Volokhin & Agichtein, 2018). 따라서 다변적인 사용자의 의도를 파악하는 것이 무엇보다 중요하다.

제1절에서 서술하였듯, 음성 쿼리는 사용자의 니즈를 구체적으로 수집할 수 있다. 따라서 음성 인터페이스에서는 다변적인 사용자의 음악 니즈를 더 잘 포착할 수 있다는 잠재력이 있다.

2.2 음악 서비스에서의 사용자 니즈와 행동

사람들이 스트리밍 플랫폼에서 어떠한 음악 검색 니즈를 가지는지 탐구한 연구가 있다 (Hosey et al, 2019). 해당 연구에서는 사람들이 음악 스트리밍 플랫폼에서 검색을 어떻게 진행하며, 검색 결과를 어떻게 평가하는지 조사하기 위해 27명을 대상으로 반구조화 인터뷰를 진행하였다. 연구 결과, 사용자들은 음악 검색 경험을 크게 '성공'과 '노력' 측면에서 평가하며, 이것은 그들의 태도(mindset)에 따라 달라진다 [그림 3].



[그림 3] Hosey et al.(2019)가 제시한 음악 검색에서의 사용자 태도.

연구자들은 음악 검색에서의 사용자 태도(mindset)를 크게 세 가지로 분류했다. 첫 번째는 찾고자 하는 바가 명확한 'Focused mindset'이다. 이때의 성공 기준은 이분법적(찾았다–찾지 못했다)이며, 노력에 대한 관용 기준이 낮은 편으로 가장 빠르고 효율적으로 검색이 이루어졌을 때 경험을 좋게 평가했다.

두 번째는 명확한 대상은 없지만 원하는 경향성은 있는 'Open mindset'이다. 이때의 성공 기준은 'Good enough – Nothing good enough – Better than good enough'와 같은 형태로, 비교적 느슨하게 인식된다. 또한 노력에 대한 관용 기준은 상대적으로 높아, 좋은 결과를 찾을 수 있다면 노력을 감수할 수 있다.

세 번째는 새로운 장르의 음악을 배우고 싶어하는 'Exploratory mindset'으로, 원하는 바가 있으나 어떻게 검색할지 모르는 상태이다. 성공 기준은 사용자가 스스로 파악하기 어려우며, 노력에 대한 관용 기준은 가장 높은 편이다.

사용자들이 음악 검색 태도(mindset)에 따라 결과에 대한 기대치를 다르게 가진다는 점이 주목할 만하다. Focused mindset의 경우, 쿼리를 작성하는 과정에서 들이는 노력(타이핑, 수정)은 충분히 감수할 수 있다고 응답했다. 대신 정확한 결과에 대한 기대가 높게 나타났다. 반면 Open mindset의 경우, 사용자들은 쿼리를 작성하는 과정보다 결과 리스트를 탐색하는 과정에 더 노력을 들이고자 한다.

Li et al.(2019)은 후속 연구에서, 세 가지의 사용자 태도를 다시 둘로 나누었다 (Focused, Non-focused). 이때 클릭, 스트림, 쿼리 등의 행동 데이터는 사용자의 태도와 연관이 있었다. 예컨대 Focused mindset에서는 '검색' 행동이 자주 나타났으며 더 많은 노력(긴 쿼리, 클릭까지 오랜 시간)을 들였다. Non-focused mindset에서는 노력을 덜 들이고, 좋은 제안을 위해 시스템에 더 의존했다. 하지만 세션 초기의 적은 로그만으로, 사용자의 태도를 예측하는 것이 어렵다는 한계를 보였다.

2.3 음악 서비스의 쿼리

음악 쿼리의 구체적인 유형을 분석한 연구는, 주로 Google Answers, 네이버 지식iN 등의 QA 커뮤니티에 나타나는 음악 정보 니즈에 집중한다 (Downie & Cunningham, 2002; Bainbridge et al., 2003; Lee et al., 2005; Lee et al., 2010). QA 커뮤니티의 음악 도메인에서 사용자들이 올리는 질문의 형태는 다음과 같다: 특정 곡 확인하기, 곡의 출처 알기, 곡 추천받기, 곡 평가받기 등. 니즈의 주제(질문에 응답이 필요한 정보의 주제)는 다음과 같다: 가사, 번역, 의미, 악보, 관련 곡, 장르, 가수 등. 음악 니즈를 해결하고자 할 때 주로 사용하는 정보는 서지학적 데이터(bibliographic data, 곡 제목/ 가수/ 날짜 등)가 대부분이며, 이는 검색자의 마음 속에 원하는 특정 아이템이 있는 경우이다.

음성 인터페이스 상의 음악 쿼리에 대한 기존 연구가 존재한다. Thom et al. (2020)은 음성 인터페이스에서만 특징적으로 나타나는 'Non-Specific Query'(e.g., "Play music")에 집중해, 발화 의도와 기대치를 인터뷰로 탐구했다. 본 연구의 NSQ 정의는 다음과 같다: 음악 재생에 대한 요청이나 사용자 발화에 특정성이 없어, '어떠한 음악'을 틀어야 할지 모르겠는 요청. 또한 해당 연구는 NSQ와 대비되는 개념으로서 Descriptive query (e.g., "Play hiphop" or "Play something calming")와 Specific query (artists or albums)를 언급했다.

이 연구의 결과에 따르면, 사용자들이 NSQ(Non-specific query)를 발화하는 의도는 노력 없이 음악을 틀고 싶음, 개인화에 대한 기대 등이다.

제시된 세 가지 유형은 쿼리에 나타난 '니즈의 구체성'에 따라 분류된 것이다. 이는 사용자들이 어떠한 추천 스트림을 기대하는지 파악하기에 용이한 기준이므로, 본 연구에서는 Thom et al.(2020)이 제시한 개념—Specific query, Non-specific query, Descriptive query에 따라 음성 인터페이스의 음악 쿼리의 특성을 살펴볼 것이다.

제 3 절 사용자 중심의 음악 연구 방법

3.1 음악 도메인에서의 사용자 조사

사용자가 음악 스트리밍 서비스에서 가지는 태도·인식·평가를 수집하기 위해서는 데이터 수집 이상의 연구 방법을 이용해야 한다. 아래는 사용자 중심의 음악 조사를 진행한 선행 연구들이다.

Li et al.(2019)는 음악 검색에서 사용자 태도에 따라 행동이 어떻게 달라지는지 보고자 했다. 이를 위해, 먼저 최근에 경험한 검색 시나리오를 떠올리게 했다 (일반적인 검색, 구체적인 니즈가 있는 검색, 구체적인 니즈가 없는 검색). 각 시나리오에 대한 워크쓰루(walkthrough)를 진행하며, 검색마다 서베이 질문을 띄워 응답을 수집한 후 사후 인터뷰를 진행했다.

Thom et al.(2020)은 음성 인터페이스에서 나타나는 'NSQ(Non-specific query)'의 발화 의도와 기대하는 바를 알기 위해 반구조화 인터뷰를 진행했다. 먼저 음악을 듣는 시나리오(바로 지금 듣기, 운동, 파티, 아침, 퇴근, 집안일, 출퇴근, 하루 시작)를 제시하면서, NSQ 발화 의향과 발화 동기를 물었다. 이후 실제로 시스템에 NSQ를 발화하게 한 후, 결과에 대한 사용자의 기대치를 물었다.

Tang & Jhang(2020)은 음악 탐색 행동을 수집하기 위해 ESM(Experience Sampling Method)를 진행했다. ESM은 참가자에게 주어진 기간 동안 반복적으로 응답을 요청하는 방법으로, 일상적 상황에서 자연스럽게 데이터를 수집할 수 있다. 이 연구에서는 음악 탐색 행동을 수집하기 위해, 2주 동안 2-3시간 간격으로 참가자들에게 온라인 설문 링크를 전송하여, 데일리한 음악 감상 행동 데이터를 수집했다 (상황맥락/ 장르/ 친숙도/ 청취 경로/ 만족도).

Lee et al.(2011)은 실험 환경에서 사용자들이 플레이리스트를 평가하는 기준 요소를 파악했다. 참가자로 하여금 시드(seed) 쿼리로 서로 다른 세 장르를 트리거하여, 5곡을 30초씩 연속으로 듣게 했다. 이에 대한 만족도를 5점 척도로 수집하고, 이후 in-depth 인터뷰로

만족의 이유와 원하는 플레이리스트 특성을 질문하였다.

위 연구 방법을 참고하여, 1차 조사에서는 쿼리 유형을 파악하기 위해 로그를 수집한다. 이후 쿼리 유형 별 의도와 만족을 수집하기 위해, Li et al.(2019)와 Them et al.(2020)와 같이 로그 기반의 인터뷰를 진행한다.

2차 조사에서는 쿼리로부터 촉발된 플레이리스트에 대한 인식과 기대를 수집한다. 따라서 Lee et al.(2011)처럼 쿼리로부터 트리거된 리스트를 특정 조건에 맞춰 듣게 한 다음 (본 연구는 15분) 설문을 수집한다. 다만 인식과 기대를 보다 자연스러운 맥락에서 수집하기 위해, Tang & Jhang(2020)처럼 ESM(Experience Sampling Method)을 진행하여, 5일 동안 하루 2회씩 수집한다.

3.2 사용자 중심의 음악 추천 평가 척도

다수의 음악 추천 평가 연구에서 정확도 기반의 정량 지표를 많이 사용하지만, 정확도가 높다고 해서 곧 사용자의 만족으로 이어지지는 않는다. 따라서 사용자 중심의 음악 추천 평가 연구들에서는 유저 중심의 'beyond accuracy' 지표들이 제시된다 (Jannach, 2022; Kim et al. 2020; Kamehkhosh & Jannach & Bonnin, 2018; Zhang et al. 2012).

사용자 중심의 척도로 가장 빈번하게 언급되는 것은 새로움(novelty), 다양성(diversity), 의외성(serendipity)이다.

새로움(novelty)은 이전에 접한 적 없는 새로운 것인가에 대한 척도이다. 다양성(diversity)은 추천된 아이템이 얼마나 다양한가에 대한 척도이다. 의외성(serendipity)은 추천된 아이템이 얼마나 예상 밖의 의외의 것인가에 대한 척도이다.

Jannach (2022)는 대화형 추천 시스템 평가 척도로 prediction accuracy, item coverage, novelty, diversity, serendipity를 제시했다. Kim et al. (2020)은 대인 간 추천과 시스템의 추천을 사용자 중심으로 평가하기 위해 relevance, novelty, diversity, serendipity를 기준으로 삼았다. Kamehkhosh & Jannach & Bonnin(2018)은 시스템의 자동적인

노래 추천 평가를 위해, relevance, novelty, accuracy, diversity, familiarity, popularity, freshness를 수집했다. Zhang et al. (2012)는 음악 추천의 평가 척도로 accuracy, diversity, novelty, serendipity를 사용했다.

선행 연구들을 반영하고 본 연구에 맞게 요소를 더 추가하여, 2차 조사에서 추천 스트림에 대한 기대와 인식을 수집할 때 최종적으로 다음 항목을 기준으로 삼고자 한다.

- History relevance: 평소 듣던 것과 유사한 것이 재생되길 바란다 혹은 재생되었다.
- Diversity: 장르·가수 측면에서 다양하게 재생되길 바란다 혹은 재생되었다.
- Novelty: 전에 알지 못했던 새로운 곡들이 재생되길 바란다 혹은 재생되었다.
- Serendipity: 예상 밖의 좋은 곡들을 발견하길 바란다 혹은 발견했다.
- Expectation: 재생된 음악은 대략 예상이 된다 혹은 예상과 유사했다.

제 3 장 연구 문제

제 1 절 연구 문제

본 장에서는 탐색하고자 하는 연구 문제를 설명한다. 제2장의 관련 연구들을 고려해 보았을 때, 음성 인터페이스의 음악 경험은 기존과 차이가 있음에도 불구하고, 이를 실제 로그 기반으로 탐색한 연구가 부족하다. 따라서 본 연구에서는 음성 인터페이스의 음악 도메인에서 사용되는 쿼리의 유형을 이해하고, 쿼리 유형에 따라 사용자가 추천 스트림에 기대하는 요소를 파악하고자 한다.

연구 문제 1. 음성 인터페이스 기반의 음악 쿼리는 어떤 양상을 보이는가

1.1 쿼리를 어떻게 유형화할 수 있는가
1.2 쿼리 유형에 따라 재쿼리 행동에 차이가 있는가
1.3 쿼리 유형에 따라 발화 의도가 어떻게 달라지는가

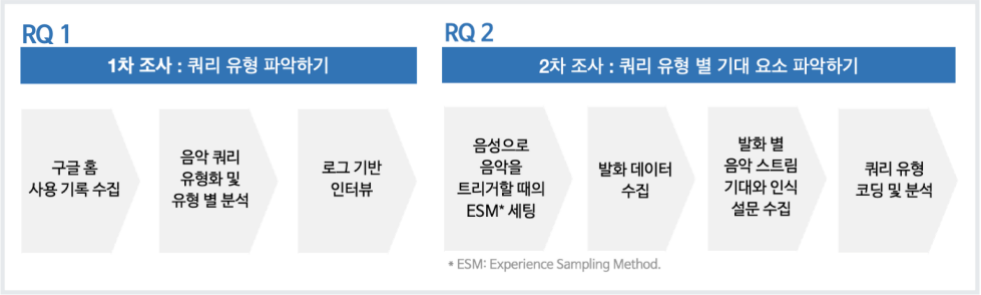
연구 문제 1에서는 실제 스마트 스피커 사용자들의 로그를 수집하고 세션 단위로 분석한 다음 인터뷰를 진행함으로써 음악 쿼리 유형을 파악하고자 한다.

연구 문제 2. 각 쿼리 유형에 따라 사용자의 추천 스트림에 대한 기대는 어떻게 달라지는가

2.1 쿼리 유형에 따라 어떠한 추천 스트림을 기대하는가
2.2 쿼리 유형에 따라 기대 대비 인식이 낮은 스트림은 무엇인가
2.3 쿼리의 하위 유형 및 사용자 특성에 따라 기대가 달라지는가

연구 문제 2에서는 앞서 도출한 쿼리의 분류를 기준으로, 사용자들이 추천 스트림에 기대하는 요소를 ESM으로 수집한다. 이를 바탕으로 음성 인터페이스의 음악 추천 방식에의 디자인 함의점을 제시하고자 한다.

제 2 절 연구 구조



[그림 4] 본 연구의 전반적인 연구 구조.

연구는 크게 1차 조사 2차 조사로 이루어져 있다. 1차 조사에서는 연구 문제 1을 탐색하며, 쿼리 유형 파악을 위해 구글 홈의 사용 기록을 수집한다. 선행 연구를 참고하여 음성 인터페이스의 음악 쿼리를 유형화한다. 이후 음악 관련 로그를 세션 단위로 나누어, 쿼리 유형에 따라 재쿼리 패턴이 어떻게 나타나는지 분석한다. 마지막으로 사용 기록을 기반으로 참여자와 인터뷰를 진행한다.

2차 조사에서는 연구 문제 2를 탐색하며, 쿼리 유형 별 기대 요소를 파악하기 위해 ESM(Experience Sampling Method)을 활용한다. 참여자가 음성 인터페이스로 음악을 재생할 때마다 추천 스트림에 대한 기대 및 인식을 묻는 설문을 수집한다. 설문 응답을 각각의 발화 데이터와 매칭한 뒤, 쿼리 유형 별로 나누어 통계 분석을 진행한다.

제 4 장 1차 조사

1차 조사의 목적은 음성 인터페이스에서 나타나는 쿼리 유형을 파악하기 위함이다. 이를 위해 1) 스마트 스피커 사용 기록을 수집하여 쿼리를 유형 별로 코딩하고 2) 쿼리 유형 별 행동을 분석한 후 3) 사용 기록을 기반으로 발화 의도와 만족 여부 등에 대한 인터뷰를 진행하였다.

제1절에서는 연구 방법을 서술하고, 제2절에서는 연구 결과로서 분류된 쿼리 유형과 행동 패턴, 인터뷰 결과를 차례로 서술한다.

제 1 절 연구 방법

1.1 수집 방법

일상에서의 자연스러운 인터랙션을 파악하기 위해, 대표적인 스마트 스피커 중 하나인 구글 홈(Google Home)의 최근 사용 기록을 수집한다. 이를 위해 총 9명의 참여자에게 최근 3개월 분의 사용 기록을 요청했다.

이때 1) 구글 홈 사용 기록 페이지[②]에서 직접 공유하고 싶지 않은 정보를 삭제한 다음 2) 3개월 분량의 json 기록 내역을 다운로드하도록 안내했다. 모든 연구 참여자가 이 과정을 명확히 인지하고 실행할 수 있도록, 민감한 정보 삭제 안내를 포함한 프로토콜 파일을 별도로 제공하였다.

수집된 스피커 사용 기록의 원시 데이터 형태는 아래 [그림 5]와 같다. 이는 연구자의 실제 구글 홈 사용 기록 중 일부이며, 타임 스탬프, 사용자의 발화, 시스템의 대응이 기록된다.

② https://myactivity.google.com/u/1/product/assistant/

2022. 6. 30. 오전 9:45:18 KST	음성으로 종료 음(를) 검색했습니다.	
2022. 6. 30. 오전 9:45:03 KST	음성으로 가사 없는 lofi jazz 틀어줘 음(를) 검색했습니다.	
2022. 6. 29. 오후 11:55:11 KST	음성으로 로파이 틀어 줘 음(를) 검색했습니다.	네, 알겠습니다.
2022. 6. 29. 오후 4:17:52 KST	음성으로 가사 없는 여름 노래 틀어 줘 음(를) 검색했습니다.	YouTube Music에서 그 시절의 여름 | 하사이시 조, 가사 없는 [PLAYLIST] 을 재생합니다..
2022. 6. 29. 오후 4:17:22 KST	음성으로 일할 때 듣는 노래 틀어 줘 음(를) 검색했습니다.	네, 알겠습니다.
2022. 6. 29. 오후 4:16:51 KST	음성으로 이럴 때 듣기 좋은 노래 틀어 줘 음(를) 검색했습니다.	알겠습니다.
2022. 6. 29. 오후 3:51:45 KST	음성으로 가사 없는 힙합 틀어 줘 음(를) 검색했습니다.	네, 알겠습니다.
2022. 6. 29. 오전 9:45:47 KST	음성으로 종료 음(를) 검색했습니다.	
2022. 6. 29. 오전 9:45:03 KST	음성으로 가사 없는 lofi jazz 틀어줘 음(를) 검색했습니다.	
2022. 6. 28. 오전 10:15:05 KST	음성으로 가사없는 힙합 재즈 틀어 줘 음(를) 검색했습니다.	알겠습니다.
2022. 6. 28. 오전 9:45:05 KST	음성으로 가사 없는 lofi jazz 틀어줘 음(를) 검색했습니다.	
2022. 6. 28. 오전 9:20:10 KST	음성으로 다이나믹듀오 MSG 틀어 줘 음(를) 검색했습니다.	다이나믹 듀오의 MSG 말씀이신가요?
2022. 6. 28. 오전 9:19:53 KST	음성으로 힙합 틀어 줘 음(를) 검색했습니다.	알겠습니다.
2022. 6. 28. 오전 9:15:56 KST	음성으로 9시 루틴 삭제해 줘 음(를) 검색했습니다.	취소합니다.
2022. 6. 28. 오전 9:00:23 KST	음성으로 종료 음(를) 검색했습니다.	
2022. 6. 28. 오전 9:00:05 KST	음성으로 가사 없는 lofi jazz 틀어줘 음(를) 검색했습니다.	
2022. 6. 27. 오후 8:57:42 KST	음성으로 이거 좋아요 눌러 줘 음(를) 검색했습니다.	
2022. 6. 27. 오후 3:51:20 KST	음성으로 힙합 lofi 재즈 틀어 줘 음(를) 검색했습니다.	알겠습니다.
2022. 6. 27. 오후 3:30:24 KST	음성으로 종료 음(를) 검색했습니다.	

[그림 5] 1차 조사에서 수집된 원시 데이터.

이후 연구자의 가공을 거친 사용 기록을 토대로 인터뷰를 진행했다. 사용 기록을 함께 보는 것은 인터뷰 중 참여자의 기억, 경험, 의견 등을 더욱 효과적으로 이끌어내기 위함이다. 인터뷰는 응답에 따라 질문을 유연하게 조정할 수 있는 반구조화(semi-structured) 형태로 진행되었다. 각 쿼리 유형에 해당하는 발화들을 보며 발화 의도, 상황 맥락, 만족 여부를 질문하였다. 인터뷰는 비대면 플랫폼(Zoom)을 통해 약 1시간 가량 진행되었다.

1.2 연구 참여자 선정 기준 및 모집

참여자 선정 기준은 다음과 같다: 1) 구글 홈(Google Home)에 외부 음악 서비스를 연동해 3개월 이상 이용 중인 사용자이며 2) 일주일에 한 번 이상 구글 홈으로 음악을 듣는 사용자.

참여자 모집을 위해 세 곳의 온라인 대학 커뮤니티에 모집 글을 기재하였다. 최종적으로 연구에 참여한 인원은 총 9명이었다. 참여자의 음성 인터페이스 사용 행태는 [표 1]과 같다. 사용 기록 제출과 인터뷰를 모두 완료한 참여자에게는 5만 원의 참여비가 지급되었다.

	스피커 사용 기간	스피커 이용 주기	스피커 기반 음악 청취 주기	발화 기록 수
P01	1년 이상	하루에 한 번 이상	하루에 한 번 이상	977
P02	6개월 이상	하루에 한 번 이상	일주일에 한 번 이상	263
P03	2년 이상	하루에 한 번 이상	하루에 한 번 이상	471
P04	6개월 이상	하루에 한 번 이상	하루에 한 번 이상	1298
P05	1년 이상	하루에 한 번 이상	2~3일에 한 번 이상	759
P06	2년 이상	2~3일에 한 번 이상	2~3일에 한 번 이상	276
P07	2년 이상	하루에 한 번 이상	하루에 한 번 이상	551
P08	1년 이상	하루에 한 번 이상	하루에 한 번 이상	462
P09	1년 이상	하루에 한 번 이상	2~3일에 한 번 이상	1175

[표 1] 1차 조사 참여자의 음성 인터페이스 사용 행태.

1.3 분석 방법

1.3.1 쿼리 유형 분류

9명의 참가자로부터 전체 구글 홈 사용 기록 6,232개가 수집되었다. 분석을 위해, 원시 데이터는 크게 두 단계에 거쳐 가공되었다.

첫째, 전체 발화 가운데 음악 관련 발화를 구분한 다음, 음악 관련 발화에서 음악 쿼리를 구분했다. 구분 기준은 아래와 같다.

- 음악 관련 발화: 사용자의 음악 관련 발화, 또는 시스템의 음악 관련 대응, 또는 동시간대에 생성된 미디어 컨트롤, 또는 음악과 관련된 루틴.

- 음악 쿼리: 음악 관련 발화 가운데 음악을 재생하는 발화. (*"…틀어 줘", "…재생"* 의 형태)

총 음악 관련 발화 2,723개와 음악 쿼리 1,585개를 추려냈다.

둘째, 음악 쿼리를 니즈의 구체성에 따라 유형화했다. 쿼리 유형의 기준은 2장에서 언급한 바와 같이, 음성 인터페이스에서 차별적으로 나타나는 음악 쿼리를 탐구한 Thom et al.(2020)의 연구를 참고했다.

해당 연구에서는 '음악 재생에 대한 요청이나 사용자 발화에 특정성이 없어 어떠한 음악을 틀어야 할지 모르겠는 요청'을 'NSQ(Non-Specific Query)'라고 정의했다. 또한 아티스트나 앨범 등을 요청하는 쿼리를 'Specific Query'라고, "Play hiphop", "Play something calming"처럼 묘사가 담긴 쿼리를 'Descriptive Query'라고 언급하였다.

선행 연구의 개념을 참고하여, 본 연구에서는 음악 니즈의 구체성에 따라 쿼리를 크게 세 가지를 분류 기준으로 삼았다.

- Specific Query: 곡이나 아티스트가 명확한 쿼리
- Non-Specific Query: 기준을 제시하지 않는 쿼리
- Descriptive Query: 원하는 장르나 분위기를 묘사하는 쿼리

위 기준을 고려하여, 앞에서 수집한 쿼리 기록 1,585개를 비교하고 유사한 쿼리를 같은 범주로 구성하는 순환적 과정을 거쳐 분류하였다.

이외에도 세 개의 기준에 포함되지 않는 기타 발화들이 존재했다. 그중 가장 많이 나타난 것은, 기존에 이미 생성된 리스트를 부르는 쿼리였다 (e.g., “내 ‘운동’ 플레이리스트 틀어줘”, “내가 좋아요 한 노래 틀어줘”). 이 쿼리는 상술한 세 쿼리와 다르게, 첫 곡은 물론 이후 곡까지 명확하여 사용자가 완전히 예측 가능하다는 특성이 있다. 하지만 본 연구는 ‘예측되지 않는 스트림’에 대한 기대를 파악하는 것이 목적이므로, 주요 분석에서 제외하였다.

1.3.2 쿼리 유형 별 패턴 분석

쿼리 유형 별 행동을 분석하기 위해, 사용 기록을 세션으로 나눈 후 세션 내 쿼리의 패턴을 분석했다. 세션의 분리 기준은 다음과 같다: 1) '음악 종료' 관련 발화가 존재하면 세션 종료, 2) 마지막 발화 후 100분 이상 다음 발화가 없으면 세션 종료.

쿼리의 패턴을 보기 위해, 세션 내 재쿼리 양상을 분석했다. 재쿼리란, 동일한 세션 내에서 초기 쿼리로 음악이 재생된 이후 새로운 쿼리를 다시 시도하는 행동을 말한다. 재쿼리의 예시는 아래 [그림 6]과 같다.

재쿼리가 시도됨

세션	time	user command	type	system response
1	19일 15:09	음악 틀어줘	NSQ	네, 나만의 추천 믹스 들려드릴게요.
1	19일 15:11	음악 꺼줘		
2	19일 17:00	가사 없는 재즈 틀어줘	DQ	네, 유튜브에서 jazz 스테이션을 재생합니다.
2	19일 17:03	쳇 베이커 틀어줘	SQ	네, 쳇 베이커 스테이션을 재생합니다.
2	19일 17:15	볼륨 30%		
3	19일 20:00	아이브 러브다이브	SQ	네, 유튜브에서 러브 다이브를 재생합니다.

[그림 6] 재쿼리의 예시.
동일 세션 내에서 새로운 쿼리를 다시 시도함.

이때 에러로 인한 재쿼리는 분석에서 제외되었다. 에러로 간주한 경우는 다음과 같다: 인식 오류로 시스템 반응이 부적절(e.g., “죄송하지만, 잘 이해하지 못했습니다”)하여 재쿼리, 또는 특정 결과에 정상적으로 닿기 위한 (e.g., 시스템이 유사한 제목의 다른 노래를 재생) 재쿼리.

이를 기반으로, 쿼리 유형에 따른 재쿼리 패턴을 두 측면에서 분석했다: 1) 쿼리 유형 별로, 어떤 유형으로 재쿼리를 시도하는가? 2) 쿼리 유형 별로, 어느 시점에서 재쿼리가 시도되는가?

1)을 위해 세션 내 재쿼리 패턴(초기 쿼리 유형-재쿼리 유형)을 파악했다. 2)를 위해 세션 내 재쿼리가 이루어지는 모든 경우에 대해 (이전 쿼리 - 다음 쿼리)의 시간 간격을 연산하였다. 이전 쿼리의 유형(SQ, NSQ, DQ)에 따라 집단을 나누어 이분산 가정 ANOVA 및 gameshowell 사후 검정을 진행했다.

제 2 절 연구 결과

1.1 쿼리 유형

수집한 1,585개의 쿼리를 니즈 구체성에 따라 분류한 결과, 쿼리 유형은 아래 표와 같이 나타났다.

쿼리 유형	기준 및 예시	총 발화 비율 (전체 1,585)
SQ (Specific Query)	곡 또는 아티스트가 명확함 "뉴진스의 Hype boy 틀어줘"	70% (1,110)
DQ (Descriptive Query)	원하는 분위기나 장르를 묘사함 "공부할 때 듣기 좋은 노래 틀어줘", "재즈 틀어줘"	13% (205)
NSQ (Non-Specific Query)	기준을 제시하지 않음 "음악 틀어줘"	14% (223)

[표 2] 1차 조사로 도출된 쿼리 유형.

SQ(Specific Query)는 곡이나 아티스트가 명확한 경우로 (e.g., "뉴진스의 Hype boy 틀어줘"), 수집된 전체 로그에서 가장 많이 시도되었다. DQ(Descriptive Query)는 원하는 분위기나 장르를 묘사하는 경우이다 (e.g., "공부할 때 듣기 좋은 노래 틀어줘", "재즈 틀어줘"). NSQ(Non-Specific Query)는 기준을 제시하지 않는 경우이다 (e.g., "음악 틀어줘"). DQ와 NSQ의 발화 빈도는 유사하게 나타났다.

1.2 쿼리 유형 별 재쿼리 패턴

1.1에서 도출된 쿼리 유형에 따라 재쿼리 패턴을 분석한 결과는 아래와 같다. 상술하였듯, 재쿼리란 동일한 세션 내에서 초기 쿼리로 음악이 재생된 이후 새로운 쿼리를 다시 시도하는 행동이다. 사용자들은 곡을 전환하고자 할 때, "다음 곡"이라고 발화(전체 음악 관련 로그 2,723개 중 59개)하기보다는 재쿼리를 시도(전체 음악 관련 로그 중 628개)하는 경향을 보였다.

초기 쿼리	다음 쿼리	빈도 (에러로 인한 재쿼리는 제외)
SQ 에서	재쿼리 X	58%
	SQ로 재쿼리	36%
	DQ로 재쿼리	3%
	NSQ로 재쿼리	3%
DQ 에서	재쿼리 X	63%
	SQ로 재쿼리	16%
	DQ로 재쿼리	20%
	NSQ로 재쿼리	1%
NSQ 에서	재쿼리 X	71%
	SQ로 재쿼리	24%
	DQ로 재쿼리	3%
	NSQ로 재쿼리	3%

[표 3] 쿼리 유형에 따른 재쿼리 패턴

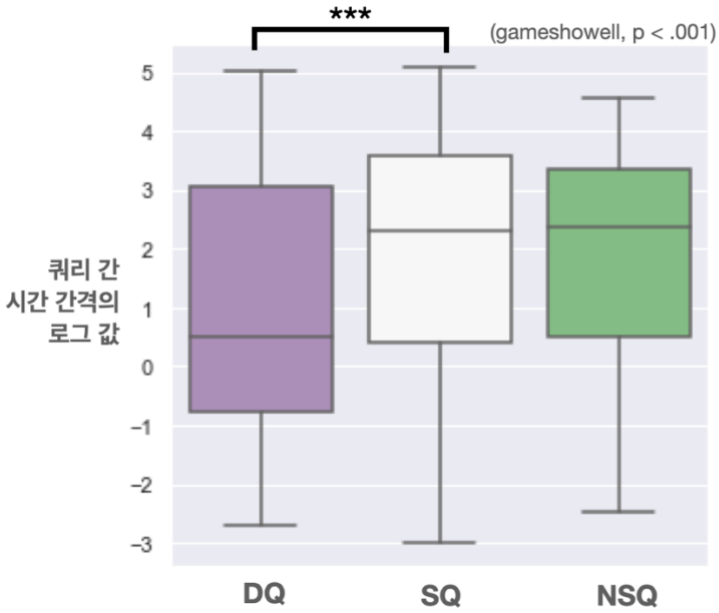
1) 쿼리 유형 별로, 어떤 유형으로 재쿼리를 시도하는가 [표 3] :

SQ(Specific Query)로 음악을 재생한 상황 중 58%에서 재쿼리 시도가 이루어지지 않았다. 이는 세션이 종료될 때까지 SQ 이후에 추가적인 쿼리가 시도되지 않은 경우다. 동일 세션에서 SQ 이후에 재쿼리가 시도된 경우, 대부분 SQ로 이루어졌다 (e.g., “Hype boy 틀어줘” 이후 “사건의 지평선 틀어줘” 시도).

DQ(Descriptive Query)로 음악을 재생한 상황 중 63%에서 재쿼리 시도가 이루어지지 않았다. 동일 세션에서 DQ 이후에 재쿼리가 시도된 경우, 주로 SQ (e.g., “신나는 노래 틀어줘” 이후 “Hype boy 틀어줘” 시도) 또는 DQ(e.g., “신나는 노래 틀어줘” 이후 “K-pop 틀어줘”)로 이루어졌다.

NSQ(Non-Specific Query)로 음악을 재생한 상황 중 71%에서 재쿼리 시도가 이루어지지 않았다. 동일 세션에서 NSQ 이후에 재쿼리가 시도된 경우, 주로 SQ로 이루어졌다 (e.g., “음악 틀어줘” 이후 “Hype boy 틀어줘”).

정리하자면, NSQ로 노래를 재생한 경우에는 대부분 세션이 종료될 때까지 재쿼리 시도 없이 음악을 들었다 (전체 NSQ의 71%). 또한 모든 쿼리 유형에서, 재쿼리를 시도하는 경우 주로 SQ가 그 수단으로 사용되는 모습을 보였다.



[그림 7] 쿼리 유형에 따른 재쿼리 시점

2) 쿼리 유형 별로, 어느 시점에서 재쿼리를 시도하는가 [그림 7]:

Gameshowell 사후 검정 결과, DQ의 재쿼리 시점은 SQ의 재쿼리 시점보다 유의미하게 일렀다 (ps < 0.001). 다시 말해, DQ로 음악을 재생한 후 얼마 지나지 않아 다른 쿼리로 곡을 전환하는 경향을 보였다. 유의미하지는 않으나, DQ의 재쿼리 시점은 NSQ의 재쿼리 시점보다도 이르게 나타났다.

1.3 쿼리 유형 별 발화 의도와 만족 여부

사용 기록 기반 인터뷰 내용을 녹음 후 전사하여 주제 분석(thematic analysis)으로 정리하였다. 이를 위해, 인터뷰 내용 중 유사한 내용을 가진 의미 단위를 범주로 구성하는 과정을 순환적으로 거쳤다.

각 쿼리 유형에 따라 도출된 주제와 세부 내용은 아래 [표 4]와 같다. 표의 괄호 안 숫자는 언급된 횟수이다.

쿼리 유형	주제	발화 의도	만족 여부
SQ (Specific Query)	명확한 니즈	- 최근 자주 듣는 노래 (4) - (다른 활동 중) 해당 곡 니즈가 생겨서 (4)	
	스트림 컨트롤	- 필요한 무드에 맞는 대표 곡으로서 (3) - 재생 중인 곡이 마음에 들지 않을 때 (2)	- 원하는 미세한 무드가 맞춰짐 - 이후 곡들에 집중하지 않기도 함
	인식 오류	- 인식 오류로 인한 재쿼리 - 음성 인식에 대한 실험적 발화이기도 함	- 인식 오류로 인한 불만 (2)
NSQ (Non-Specific Query)	추천 기대	- 추천에 대한 궁금증 혹은 기대 (4)	- 취향을 잘 반영함 (4) - 새롭지 않거나 엉뚱한 노래가 나옴
	모호한 니즈	- (다른 활동 중) 아무거나 듣고 싶음 (2)	- 적당한 만족감 - 결과에 대한 예측이 불가능함
	인식 오류 방지	- 인식 오류를 피하기 위함 (2)	
	기능 시도	- 기능에 대한 호기심	
DQ (Descriptive Query)	모호한 니즈	- 분위기 형성을 위함 (2) - 듣고 싶은 곡이 없거나 장르 지식이 없음	
	스트림 컨트롤	- 특정 장르 내 취향인 곡들을 듣고 싶어서 - 재생 중인 곡이 마음에 들지 않아서	- 제시한 기준에는 맞으나 취향이 아님 (3) - 무엇이 나올지 예상 가능 (2) - 구체적인 요구대로 재생됨 - 유사한 결과의 반복

[표 4] 쿼리 유형 별 발화 의도와 만족 여부

SQ의 발화 의도는 크게 '명확한 니즈', '스트림 컨트롤', '인식 오류'로 나타났다: 이는 각각 듣고 싶은 곡이 명확하거나, 원하는 무드의 대표 곡으로 발화하거나, 인식 오류로 인한 재발화하는 경우이다.

SQ의 만족 여부는 크게 '스트림 컨트롤', '인식 오류' 측면에서 거론되었다: '스트림 컨트롤' 측면에서 원하는 미세한 무드를 조정할 수 있어 만족하나, 이후 곡에 별로 집중하지 않기도 한다. '인식 오류' 측면에서 음성 인식의 실패로 인한 불만을 보인다.

NSQ의 발화 의도는 크게 ‘추천 기대’, ‘모호한 니즈’, ‘인식 오류 방지’, ‘기능 시도’로 나타났다: 이는 각각 추천에 대한 기대로, 아무거나 듣고 싶어서, 인식 오류를 피하고 싶어서, 기능에 대한 호기심으로 발화하는 경우이다.

NSQ의 만족 여부는 크게 ‘추천 기대’, ‘모호한 니즈’ 측면에서 거론되었다: ‘추천 기대’ 측면에서 취향이 잘 반영된다는 의견이 대부분이지만 새롭지 않거나 엉뚱한 노래를 맞닥뜨리기도 한다. ‘모호한 니즈’ 측면에서 결과에 대한 예측이 불가능하여 불만을 보인다.

DQ의 발화 의도는 크게 ‘모호한 니즈’, ‘스트림 컨트롤’로 나타났다: 각각 분위기 형성 또는 원하는 곡이 없어서, 특정 기준 내에서 취향인 곡을 듣고 싶어서 발화하는 경우이다.

DQ의 만족 여부는 ‘스트림 컨트롤’ 측면에서, 구체적 요구를 따르는 결과를 예상 가능하나 취향에 맞는 결과가 아니라는 의견이 거론되었다.

또한, 음성 인터페이스의 음악 경험이 전반적으로 모바일과 어떤 차이가 있는지 질문하였다. 그 결과, ‘더욱 배경적으로 들으며 분위기 형성 용도’와 ‘듣는 노래의 폭이 넓거나 새로운 노래를 기대함’이 가장 많이 언급되었다. 그 외에도 ‘곡을 보다 연속적으로 듣게 됨’이라는 응답이 있었다.

제 3 절 소결론

연구 문제 1. 음성 인터페이스 기반의 음악 쿼리는 어떤 양상을 보이는가

1.1 쿼리를 어떻게 유형화할 수 있는가
1.2 쿼리 유형에 따라 재쿼리 행동에 차이가 있는가
1.3 쿼리 유형에 따라 발화 의도가 어떻게 달라지는가

1차 조사에서는 연구 문제 1. '음성 인터페이스의 음악 쿼리가 어떤 양상을 보이는가' 를 살펴보았다. 구글 홈 사용 기록을 수집해 유형과 행동 패턴을 분석하고 기록 기반의 인터뷰를 진행하였다. 1차 조사의 결과를 바탕으로 정리한 소결론은 다음과 같다.

1-1. 쿼리를 어떻게 유형화할 수 있는가. 음성 인터페이스의 음악 쿼리는 표현된 니즈의 구체성에 따라 SQ(Specific Query), NSQ(Non-Specific Query), DQ(Descriptive Query)로 나눌 수 있다. 전체 사용 기록 중 SQ가 가장 많이 사용되었으며, NSQ와 DQ는 비슷하게 사용되었다.

1.2. 쿼리 유형에 따라 재쿼리 행동에 차이가 있는가. 음악 쿼리 유형에 따라 재쿼리 패턴이 다르게 나타났다. NSQ로 음악을 재생했을 때, 동일 세션에서 재쿼리 없이 세션을 종료하는 비율이 가장 높게 나타났다 (71%).

모든 쿼리에서, 재쿼리를 시도하는 경우 SQ를 이용하는 경향을 보였다. 또한 세 쿼리 가운데 DQ에서 재쿼리 시점이 가장 이르게 나타났다. 이는 쿼리 유형에 따라 결과에 대한 관용도가 다르기 때문일 수 있다.

1.3. 쿼리 유형에 따라 발화 의도가 어떻게 달라지는가. 음악 쿼리 유형에 따라 사용자들의 발화 의도와 만족도가 다르게 나타났다. SQ는 주로 곡에 대한 명확한 니즈가 있거나, 스트림을 곡으로 컨트롤하고자 할 때 사용되었다. 원하는 무드를 미세하게 맞출 수 있다는 점에서

만족하나, 인식 오류로 인한 불만이 나타났다.

NSQ는 니즈가 없을 때 추천 곡을 기대하거나 인식 오류를 피하기 위해 사용되었다. 취향이 잘 반영되어 만족하나, 엉뚱한 노래가 나오는 등 결과 예측이 불가능하다는 불만이 나타났다.

DQ는 니즈가 명확하지 않을 때 전반적인 스트림을 컨트롤하기 위해 사용되었다. 구체적 기준에 따라 예상 가능한 결과가 나와 만족하나, 취향과는 거리가 멀다는 불만이 나타났다.

1차 조사를 통해, 쿼리 유형에 따라 사람들이 서로 다른 스트림을 기대할 것임을 파악했다. 따라서 2차 조사에서는 쿼리 유형에 따라 구체적으로 어떤 요소를 기대하는지 살펴볼 것이다.

제 5 장 2차 조사

2차 조사의 목적은 음성 인터페이스의 음악 쿼리 유형에 따라, 추천 스트림에 대한 사용자의 인식과 기대를 파악하는 것이다. 이를 위해 ESM(Experience Sampling Method) 방식을 이용해, 음성 인터페이스로 음악을 재생할 때마다 기대와 인식에 관한 설문을 수집하였다.

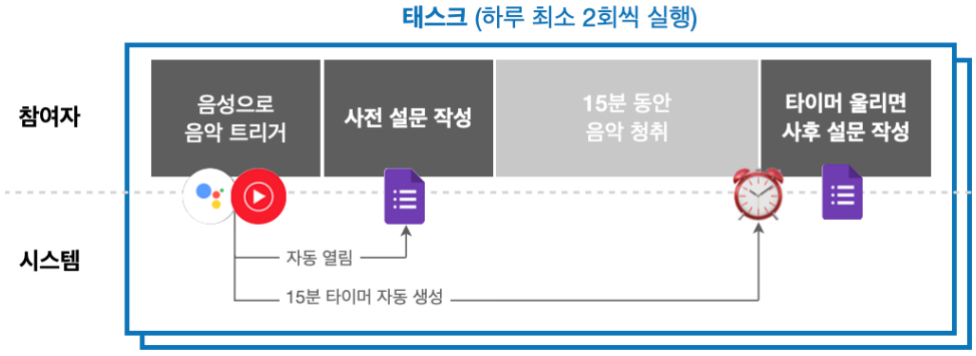
제1절에서는 연구 방법을 서술하고, 제2절에서는 연구 결과로서 쿼리 유형에 따른 인식과 기대의 차이를 서술한다.

제 1 절 연구 방법

1.1 수집 방법

2차 조사에서는 추천 스트림에 대한 기대와 인식을 수집하는 방법으로서 ESM(Experience Sampling Method)을 진행하였다. 제2절에서 언급하였듯, ESM은 참가자에게 주어진 기간 동안 반복적으로 응답하도록 요청하는 데이터 수집 방법이다. 일상적인 상황에서 태스크를 수행함으로써 비교적 자연스럽고 통제되지 않은 데이터를 수집할 수 있다는 장점이 있다. 본 연구에서는, 평소 음악을 듣는 상황 맥락 내에서 가지는 기대와 인식을 수집하기 위해 ESM 방법을 채택하였다.

본 연구가 제공한 ESM 태스크는 다음과 같다: 일상 생활에서 음성 인터페이스로 음악을 재생하고 청취하는 동안 설문을 작성 [그림 8].



[그림 8] 2차 조사의 ESM 태스크 구조.

참여자는 5일의 참여 기간 동안 태스크를 하루 최소 2회 이상 수행하여야 한다. 하나의 태스크는 네 단계로 이루어져 있다: 1) 구글 어시스턴트(Google Assistant) [③] 를 이용해, 유튜브 뮤직(Youtube Music) [④] 으로 음악 재생을 요청. 2) 사전 설문(재생될 추천 스트림에 대한 기대) 제출. 3) 15분 동안 음악 청취. 4) 사후 설문(재생된 추천 스트림에 대한 인식) 제출.

참여자에게 태스크 안내는 [그림 9]로 제공됐다. 이때 두 가지 주의사항을 전달했다: 1) 세 가지 방식의 음성 요청을 골고루 시도할 것. 2) 음악을 평소처럼 듣되 앱에 접속하지 않을 것.

이때, in-situ 데이터를 수집하기 위해 아이폰 단축어 [⑤] 기능을 활용했다. 단축어 기능은 복수의 작업을 자동화할 수 있다. 연구자는 두 개의 단축어를 생성해, 연구 참여자에게 공유했다: 1) 유튜브 뮤직(Youtube Music) 앱을 열면, '사전 설문' 링크가 자동으로 열리고 15분 타이머가 생성됨. 2) '사후 설문' 링크가 열림 (타이머 종료 후 참여자가 누를 단축어).

③ https://assistant.google.com/intl/ko_kr/
④ https://www.youtube.com/musicpremium
⑤ https://support.apple.com/ko-kr/guide/shortcuts/welcome/ios

[그림 9] 2차 조사 참여자에게 제공된 태스크 안내문.

전체 단계에서 수집되는 데이터는 크게 세 가지이다:

1. 초기 설문:
   - 평소 음악 성향 (Relevance, Diversity, Novelty, Serendipity 측면에서 5점 척도)
   - 음성 인터페이스 사용 빈도

2. 태스크 중 설문 [표 5]:
- 사전 설문: 추천 스트림 기대 (Relevance, Diversity, Novelty, Serendipity, Expectation) 5점 척도
- 사후 설문: 추천 스트림 인식 (Relevance, Diversity, Novelty, Serendipity, Expectation, Satisfaction) 5점 척도, 곡 스킵 횟수

3. 구글 어시스턴트 발화 데이터.

초기 설문에서 수집하는 '평소 음악 성향'과 '음성 인터페이스 사용 빈도'는, 사용자 특성이 결과에 영향을 미치는지 확인하기 위함이다.

사전 설문은 사용자가 쿼리를 발화한 후 음악이 본격적으로 재생되기 전에 작성된다. 따라서, 이후 재생될 음악 스트림에 대해 사용자가 바라는 요소를 수집할 수 있다 ("나는 …한 곡들이 재생되길 바란다"에 대한 5점 척도). 여기서 수집된 응답은, 두 번째 연구 문제 "사용자가 쿼리 별로 어떤 추천 스트림을 기대하는가?"에 대한 직접적인 답을 줄 수 있다.

사후 설문은 15분간 음악 재생이 끝난 후에 작성된다. 따라서, 재생된 음악 스트림에 대해 사용자가 인식한 요소—즉, 재생된 스트림의 요소를 수집할 수 있다 ("15분 동안 …한 곡들이 재생되었다"에 대한 5점 척도). 사후 설문에서 수집된 응답을 사전 설문의 응답과 비교하면, '기대 대비 인식'—즉 현재 추천 스트림에서 개선되어야 할 요소를 도출해낼 수 있다.

추천 스트림에 대한 기대와 인식을 평가하는 기준은, 제2장에서 상술한 선행 연구의 척도를 반영해 History relevance, Novelty, Diversity, Serendipity, Expectation 다섯 개의 요소로 확정하였다.

사후 설문에서는 이에 더해 전반적인 만족도와 스킵 횟수를 추가적으로 수집했다. 스킵 횟수를 제외한 모든 항목은 5점 리커트 척도로 수집되었다 (1 매우 그렇지 않다 ~ 5 매우 그렇다). 사전 설문과 사후 설문의 문항은 [표 5]와 같다.

척도	사전 설문 (청취 전 기대)	사후 설문 (청취 후 인식)
(History) **Relevance**	평소 듣던 것과 유사한 것이 재생되길 바란다.	평소 듣던 것과 유사한 것이 재생됐다.
Diversity	장르·가수 측면에서 다양하게 재생되길 바란다.	장르·가수 측면에서 다양하게 재생됐다.
Novelty	전에 알지 못했던 새로운 곡들이 재생되길 바란다.	전에 알지 못했던 새로운 곡들이 재생됐다.
Serendipity	예상 밖의 좋은 곡들을 발견하길 바란다.	예상 밖의 좋은 곡들을 발견했다.
Expectation	어떤 노래가 나올지 대략적으로 예상된다.	재생된 음악은 대략 예상했던 것과 유사하다.
Satisfaction		재생된 음악은 전반적으로 만족스러웠다.
Skip		음악을 몇 번 스킵했나요?

[표 5] 2차 조사의 태스크 중 수집되는 설문 문항.

1.2 연구 참여자 선정 기준 및 모집

참여자 선정 기준은 다음과 같다: 1) 유튜브 뮤직(Youtube Music)을 반 년 이상 사용 중이며 2) 음성 인터페이스 사용 경험이 있으며 3) 아이폰을 사용 중인 사람.

추천 스트림에 대한 기대와 인식이 서비스 별 알고리즘 차이에 영향을 받지 않도록, 참여자의 이용 서비스를 유튜브 뮤직(Youtube Music)으로만 한정하였다. 또한 충분히 개인화된 추천 스트림을 토대로 평가하기 위해, 해당 서비스를 6개월 이상 이용한 사람으로 한정하였다. ESM 태스크를 아이폰 단축어 기능으로 제공하기 때문에, 아이폰 사용자만을 모집하였다.

참여자 모집을 위해 세 곳의 온라인 대학 커뮤니티에 모집 글을 기재하였다. 최종적으로 연구에 참여한 인원은 총 28명이었으며, 참여가 불성실하다고 판단된 1명을 제외해 총 27명의 데이터로 분석을 진행하였다. 참여자의 음성 인터페이스 사용 행태는 [표 6]와 같다. 태스크를 모두 완료한 참여자에게는 3만 원의 참여비가 지급되었다.

P	음악 서비스 이용 주기 (1 번에 15 분 이상)	음성 인터페이스 사용 주기
1	일주일에 5 번 이상 7 번 미만	거의 매일 사용
2	매일	거의 사용하지 않음
3	일주일에 5 번 이상 7 번 미만	한 달에 한 번 이상 사용
4	매일	한 달에 한 번 이상 사용
5	매일	한 달에 한 번 이상 사용
6	일주일에 3 번 이상 5 번 미만	한 달에 한 번 이상 사용
7	일주일에 3 번 이상 5 번 미만	한 달에 한 번 이상 사용
8	매일	거의 매일 사용
9	매일	거의 매일 사용
10	매일	거의 매일 사용
11	일주일에 5 번 이상 7 번 미만	한 달에 한 번 이상 사용
12	매일	한 달에 한 번 이상 사용
13	매일	일주일에 한 번 이상 사용
14	매일	거의 사용하지 않음
15	매일	한 달에 한 번 이상 사용
16	매일	거의 매일 사용
17	매일	일주일에 한 번 이상 사용
18	매일	거의 매일 사용
19	일주일에 5 번 이상 7 번 미만	거의 사용하지 않음
20	매일	일주일에 한 번 이상 사용
21	매일	일주일에 한 번 이상 사용
22	매일	일주일에 한 번 이상 사용
23	매일	거의 사용하지 않음
24	매일	일주일에 한 번 이상 사용
25	매일	일주일에 한 번 이상 사용
26	매일	일주일에 한 번 이상 사용
27	매일	일주일에 한 번 이상 사용
28	매일	일주일에 한 번 이상 사용

[표 6] 2차 조사 참여자 목록

1.3 분석 방법

ESM 참여가 종료된 후, 연구자는 각각의 발화 데이터에 쿼리 유형을 마킹하고 설문 응답과 매칭하였다. 이때, 시스템의 응답이 사용자 발화와 관련이 없거나 오류로 인식되는 경우(e.g., 사용자가 동일한 발화를 반복함)는 분석에서 모두 제외하였다.

태스크를 통해 수집된 데이터의 예시는 [표 7]와 같으며, 이를 기반으로 통계 분석을 진행했다.

P	발화 데이터	구분	설문 구분	(History) Relevance	Diversity	Novelty	Serendipity	Expectation	Satisfaction	스킵 수
P1-1	Hype boy 틀어줘	SQ	사전	2	2	4	1	5		4
P1-1			사후	2	5	3	2	4	4	
P1-2	신나는 노래 틀어줘	DQ	사전	4	3	1	1	3		13
P1-2			사후	3	1	3	1	2	3	

[표 7] 2차 조사를 통해 수집된 데이터 예시.

분석은 크게 세 가지 차원에서 진행되었다: 첫째, 쿼리 유형에 따라 추천 스트림이 어떻게 인식되는가? 둘째, 쿼리 유형에 따라 어떤 추천 스트림을 기대하는가? 셋째, 쿼리 내 하위 유형 또는 사용자 특성에 따라서도 기대가 달라지는가?

첫째, 쿼리 유형에 따라 추천 스트림이 어떻게 인식되는지 알아본다. 독립 요인은 쿼리 유형(SQ, NSQ, DQ)이다. 종속 요인은 사후 설문으로 응답된 History relevance, Novelty, Diversity, Serendipity, Expectation, Satisfaction, 스킵 횟수이다. 각각의 종속 요인에 대해 독립 요인 간 차이를 보이는지 알아보는 것이 목표이다.

표본의 크기는 충분하지만 설문 문항이 서열 척도로 수집되었고 모든 분포가 정규성을 위반하므로, 비모수 검정인 크루스칼 왈리스 검정(Kruskal-Wallis test)을 진행하였다. 해당 검정이 유의미할 경우 본페로니 교정(Bonferroni Correction)으로 사후 검정을 진행했다.

둘째, 쿼리 유형에 따라 추천 스트림이 어떻게 기대되는지 알아본다. 독립 요인은 쿼리 유형(SQ, NSQ, DQ)이다. 종속 요인은 사전 설문으로 응답된 History relevance, Novelty, Diversity, Serendipity, Expectation이다.

이 역시 비모수 검정인 크루스칼 왈리스 검정(Kruskal-Wallis test)을 진행하였고, 해당 검정이 유의미할 경우 본페로니 교정(Bonferroni Correction)으로 사후 검정을 진행했다.

추가적으로, 추천 스트림에 대한 기대 대비 인식이 얼마나 차이나는지 알아본다. 각 쿼리 별 종속 요인의 사전 기대와 사후 인식을 비교해야 하므로, 대응표본 t-검정의 비모수 검정인 윌콕슨 부호순위 검정(Wilcoxon’s signed-rank test)를 진행했다.

셋째, 쿼리 내 하위 유형 또는 사용자 특성에 따라서도 기대가 달라지는지 알아본다.

1) 쿼리 내 하위 유형: 정보의 종류에 따라 SQ를 ‘곡’(e.g., “밤편지 틀어줘”)과 ‘가수’(e.g., “아이유 틀어줘”)의 하위 유형으로 나누었다. 또한 DQ를 ‘음악적 분위기’(e.g., “신나는 노래 틀어줘”), ‘상황이나 활동’(e.g., “운동할 때 듣는 노래”, “우울할 때 듣는 노래”), ‘장르’(e.g., “클래식 음악”)의 하위 유형으로 나누었다. 각각의 하위 유형을 독립 요인으로, 사전 설문의 응답을 종속 요인으로 설정하였다.

2) 사용자 특성: 음성 인터페이스 사용 빈도에 따라 사용자를 ‘헤비 유저’와 ‘라이트 유저’로 나누었다. 일주일에 한 번 이상 사용하는 경우 헤비 유저로 분류하였다. 독립 요인은 쿼리 유형과 사용자 집단, 종속 요인은 사전 설문의 응답이다. 이원분산분석(two-way ANOVA)의 비모수 검정 방식이 존재하지 않아, 6개의 그룹(쿼리 유형 3종류 X 사용자 집단 2종류)에 대한 본페로니 교정(Bonferroni Correction)을 진행했다.

모든 분석은 Python과 pandas, numpy, scipy, statsmodels 패키지를 이용하였으며, 통계적으로 유의미한 결과 위주로 보고한다 ($p < 0.05$).

제 2 절 연구 결과

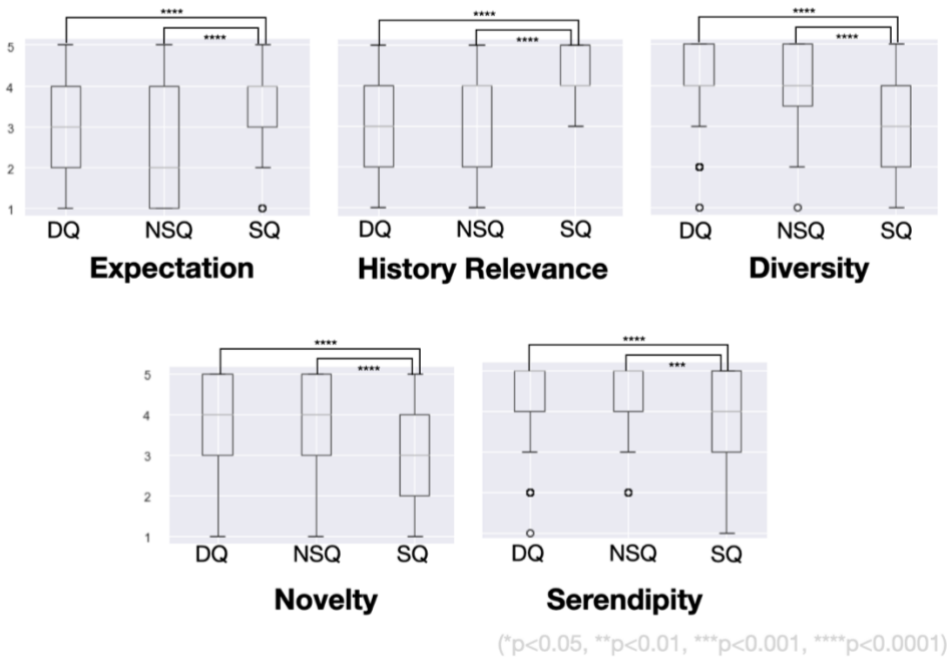
27명의 참여자로부터 총 290개의 쿼리 및 사전-사후 설문 쌍이 수집되었다. 이중 SQ는 96개, NSQ는 79개, DQ는 115개로 나타났다.

음성 인터페이스 사용 빈도로 사용자를 나누었을 때, 일주일에 한 번 이상 사용하는 '헤비 유저'가 15명, 일주일에 한 번 미만 사용하는 '라이트 유저'가 11명으로 나타났다. 참여 신청 시 수집한 5점 리커트 척도 설문에서, 참여자들은 전반적으로 평소에 새롭고 다양한 음악을 듣고자 하는 것으로 나타났다 (M=3.7, SD=0.7).

1.1 쿼리 유형에 따른 스트림 기대

[그림 10]은 사전 설문에서 받은 추천 스트림에 대한 기대를 쿼리 유형에 따라 분석한 결과이다. SQ는 재생될 스트림을 대략적으로 예상할 수 있으며 (Expectation, $p < 0.0001$), 평소에 듣던 것과 유사한 곡이 재생되길 기대한다 (History Relevance, $p < 0.0001$).

DQ와 NSQ는 재생될 스트림이 예상되지 않는 편이지만, 동시에 다양하고 새롭고 의외의 곡들이 재생되길 기대한다 (Diversity, $p < 0.0001$ / Novelty, $p < 0.0001$ / Serendipity, $p < 0.001$).

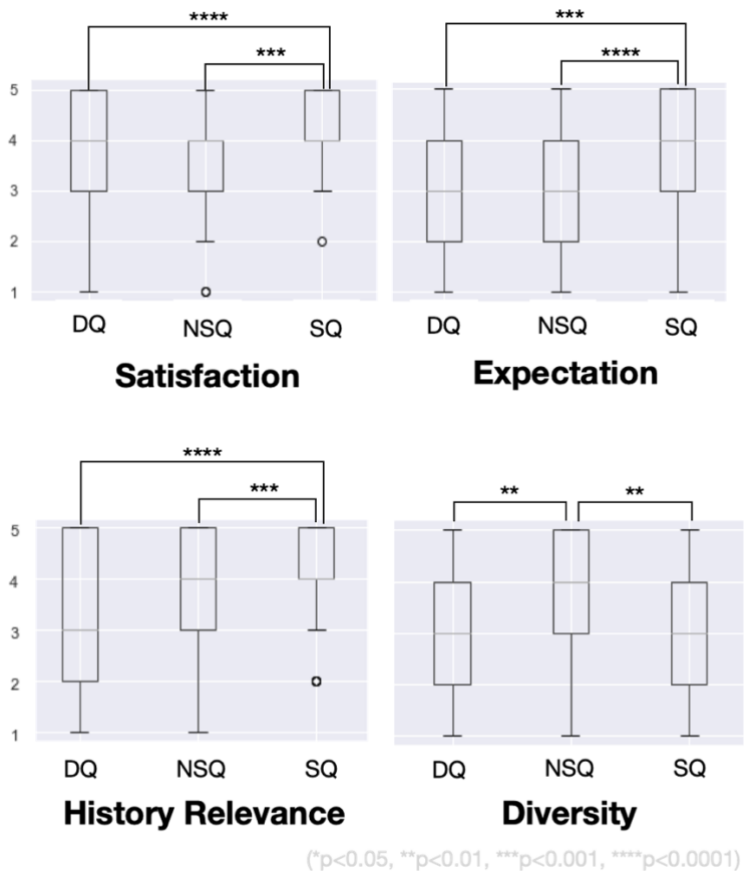


[그림 10] 쿼리 유형에 따른 스트림 기대 요소.

1.2 쿼리 유형에 따른 스트림 인식

[그림 11]은 사후 설문에서 받은 추천 스트림에 대한 인식을 쿼리 유형에 따라 분석한 결과이다. 다른 쿼리들과 비교했을 때, SQ는 예상했던 결과와 일치하는 편이고 (Expectation, $p < 0.001$), 평소에 듣던 것과 유사한 곡들이 재생된다 (History relevance, $p < 0.001$). 전반적인 만족도는 가장 높게 나타났다 (Satisfaction, $p < 0.001$).

DQ와 NSQ는 예상했던 결과대로 나오지 않는 편이다 (Expectation, $p < 0.001$), 특히 NSQ는 다양한 결과가 재생되지만 (Diversity, $p < 0.01$) 전반적인 만족도가 가장 낮다 (Satisfaction, $p < 0.001$).



[그림 11] 쿼리 유형에 따른 스트림 인식.

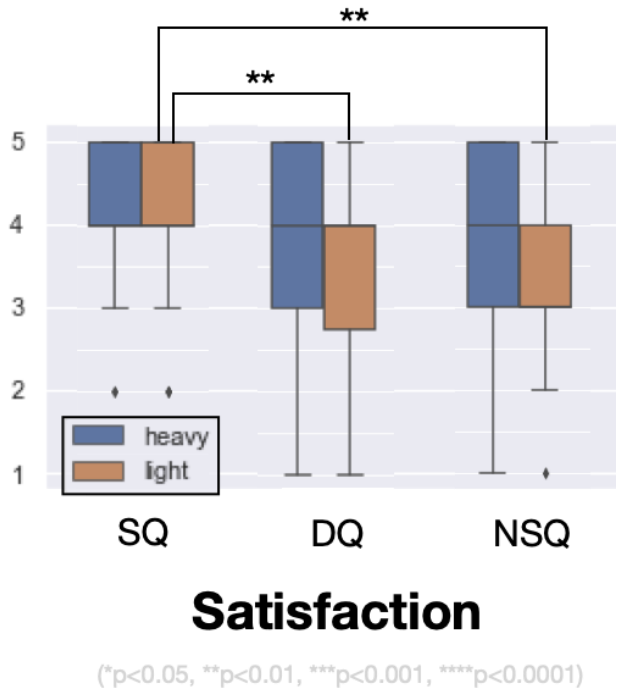
추가적으로, 쿼리 유형에 따른 기대 대비 인식을 살폈다 (사전 기대 설문과 사후 인식 설문에 대한 대응 표본 분석). 세 쿼리 모두에서, 기대에 비해 의외성에 대한 인식이 현저히 떨어졌다 (Serendipity, $p < 0.0001$). DQ에서는 다양성이 (Diversity, $p < 0.0001$), NSQ에서는 새로움이 (Novelty, $p < 0.001$) 기대에 비해 인식이 떨어졌다.

1.3 쿼리의 하위 유형과 사용자 특성에 따른 스트림 기대

SQ를 '곡'과 '가수' 유형으로 나누어, 추천 스트림을 다르게 기대하는지 살펴보았다. 곡을 요청하는 쿼리가 81개, 가수를 요청하는 쿼리가 15개로 나타났다. 분석 결과, '곡'으로 음악을 요청했을 때 새롭고 다양하고 의외의 스트림을 원했다 (Novelty, $p < 0.05$ / Diversity, $p < 0.05$ / Serendipity, $p < 0.01$). 동일한 SQ로 묶이더라도 '곡'은 비교적 DQ와 NSQ와 가까운 모습을 보였다.

DQ를 다시 '음악적 분위기', '상황이나 활동', '장르'로 나누어 분석했으나 유의미한 차이는 나타나지 않았다.

사용자를 음성 인터페이스 사용 빈도에 따라 '헤비 유저'와 '라이트 유저'로 나눈 결과, 쿼리 별 만족도의 차이는 '라이트 유저'에게만 나타나고 '헤비 유저'에게는 나타나지 않았다 [그림 12].



[그림 12] 헤비 유저와 라이트 유저의 쿼리 유형 별 만족도.

제 3 절 소결론

연구 문제 2. 각 쿼리 유형에 따라 사용자의 추천 스트림에 대한 기대는 어떻게 달라지는가

2.1 쿼리 유형에 따라 어떠한 추천 스트림을 기대하는가
2.2 쿼리 유형에 따라 기대 대비 인식이 낮은 스트림은 무엇인가
2.3 쿼리의 하위 유형 및 사용자 특성에 따라 기대가 달라지는가

2차 조사에서는 연구 문제 2. '음성 인터페이스 음악 쿼리의 유형에 따라 어떤 추천 스트림을 기대하는가' 를 심층적으로 살펴 보았다. 이를 위해 ESM 형식으로 참여자들의 일상적인 음성 음악 쿼리와 설문 응답을 수집했다. 2차 조사의 결과를 바탕으로 정리한 소결론은 다음과 같다.

2-1. 쿼리 유형에 따라 어떠한 추천 스트림을 기대하는가. 쿼리 유형에 따라 서로 다른 추천 스트림을 기대한다. SQ는 결과 예상이 잘되면서 평소 듣던 것과 유사한 곡들이 나오길 바란다. DQ와 NSQ는 결과 예상이 안 되지만, 보다 다양하고 새롭고 의외의 곡들이 나오길 바란다.

2-2. 쿼리 유형에 따라 기대 대비 인식이 낮은 스트림은 무엇인가. 참여자들의 기대와 실제 인식을 비교하여, 현재 추천 스트림에서 개선이 필요한 요소를 발견했다. 세 쿼리 모두 의외성(Serendipity)이 기대에 비해 충족되지 않고 있다. 또한 DQ에서는 다양성(Diversity)이, NSQ에서는 새로움(Novelty)이 기대에 비해 충족되지 않고 있다.

2-3. 쿼리의 하위 유형 및 사용자 특성에 따라 기대가 달라지는가. 동일한 쿼리이더라도 세부적인 요청 정보에 따라 기대하는 바가 달라짐을 확인했다. SQ 내에서도 '곡' 을 요청하는 쿼리는 '가수' 를 요청하는 쿼리보다 더욱 새롭고 다양한 스트림을 요구한다. 또한 음성 인터페이스 사용 빈도가 적은 '라이트 유저' 에서 DQ와 NSQ의 만족도가 떨어졌다.

1차 조사의 결과와 2차 조사의 결과에 대한 종합적인 논의는 제6장에서 서술한다.

제 6 장 연구 논의

본 장에서는 첫 번째 연구 문제에 해당하는 음성 인터페이스의 음악 경험과 쿼리 특성, 두 번째 연구 문제에 해당하는 쿼리 별 기대하는 추천 스트림에 대해 논의한다.

제 1 절 음성 인터페이스의 음악 경험과 쿼리 특성

먼저, 본 연구에서 밝힌 쿼리 유형과 특성을 요약하여 설명한다. 니즈 구체성에 따라, 음성 인터페이스의 음악 쿼리는 세 가지로 분류됐다 [표 8]. 1차 조사의 결과를 반영해, 요청하는 정보/ 발화 의도/ 재쿼리 행동의 인사이트를 기준으로 쿼리 간 차이를 살펴보았다.

	SQ (Specific Query)	DQ (Descriptive Query)	NSQ (Non-Specific Query)
쿼리로 요청하는 정보	- 곡 "Hype boy 틀어줘" - 가수 "뉴진스 노래 틀어줘"	- 음악적 분위기 "잔잔한 노래" - 상황·활동 "운동할 때 듣는 노래" - 장르 "클래식 음악 틀어줘"	-
발화 의도	원하는 대상이 명확	명확한 곡이 안 떠오름, 구체적인 스트림 컨트롤 원함	니즈 없이 추천에 대한 기대, 인식 오류 방지
재쿼리 행동의 인사이트	재쿼리 수단으로 가장 많이 사용됨	결과에 대한 기대가 높은 편	결과에 대한 기대가 낮은 편

[표 8] 음성 인터페이스의 음악 쿼리 특성 요약.

- SQ (Specific Query): 요청하는 정보 측면에서, 곡(e.g., “Hype boy 틀어줘”)과 가수(e.g., “뉴진스 노래 틀어줘”)로 나뉜다. 발화 의도 측면에서, 사용자는 원하는 대상이 명확하다. 그렇기 때문에 SQ는 재쿼리의 수단으로도 많이 사용된다. 인터뷰에서 곡을 전환하는 이유로 자주 언급된 것이 ‘처음에는 니즈가 없었으나 점점 생기는 상황’인데, 이때 사람들은 SQ를 이용했다.

- DQ (Descriptive Query) : 요청하는 정보 측면에서, 음악적 분위기(e.g., “잔잔한 노래”)와 상황·활동(e.g., “운동할 때 듣는 노래”)과 장르(e.g., “클래식 음악 틀어줘”)로 나뉜다. 발화 의도 측면에서, 사용자는 명확한 곡이 떠오르지 않거나 구체적으로 스트림을 컨트롤하길 원한다. 재쿼리 행동을 보았을 때 DQ 이후에 재쿼리를 빠르게 시도하는 모습을 보여, 결과에 대한 기대가 높은 편으로 해석된다.

Thom et al.(2020)은 “Descriptive query를 시도하는 사람들은 개인화보다 결과에 대한 컨트롤을 원할 것이다.”라고 언급하였다. 하지만 인터뷰 결과, DQ가 ‘제시한 기준과 맞는 결과이지만 취향과 맞지 않아 불만족’한다는 의견이 상당수 수집되었다. 이는 DQ 사용자들이 개인화와 결과 컨트롤을 양자택일하기보다는, 둘 모두를 희망함을 의미한다. 하지만 2차 조사 결과, 사용자들은 DQ를 이용해 새롭고 다양한 노래를 원하기도 했다. 동일한 DQ더라도 추가적인 상황 맥락에 따라 기대하는 바가 다른 것으로 보여, 추후 연구가 필요하다.

- NSQ (Non-specific Query) : 아무런 정보도 요청하지 않는 쿼리로, 니즈 없이 시스템의 제안을 기대하거나 인식 오류를 피하기 위해 사용된다. NSQ를 한번 트리거하면 대개 재쿼리 없이 세션이 종료되는 모습을 보여, 결과에 대한 기대가 낮은 편으로 해석된다.

Thom et al. (2020)이 밝힌 것처럼, 사용자들은 발화의 간편함과 예측 가능성 사이의 trade-off가 존재한다는 것을 이해하고 있다 (e.g., “음악 틀어줘”: 발화가 간편하나 예측이 비교적 불가능함). 이런 이해를 기반으로, 자신이 원하는 스트림을 재생하기 위해 서로 다른 쿼리를 전략적으로 사용한다.

음성 인터페이스의 음악 경험은 기존과 크게 세 지점에서 다르며, 이에 따라 쿼리 사용이 전략적으로 달라진다

사용자의 쿼리 사용 전략은 음성 인터페이스의 차별적인 음악 경험과 긴밀하게 연결되어 있다. 본 연구에서 밝힌 음성 인터페이스의 음악 경험의 차별점은 1) 배경적 청취, 2) 일람성 부재, 3) 인식 오류 가능성이다. 이로 말미암아 나타나는 쿼리 사용 전략에 대해 논의하고자 한다.

첫째, **배경적 청취.** 1차조사의 인터뷰에서 모바일과의 음악 경험 차이점을 질문한 결과, 가장 많이 언급된 것이 '음악에 집중하기보단 배경 분위기를 형성하는 용도'라는 점, 그로 인해 '더욱 폭넓은 음악을 듣게'된다는 점이었다. 이는 음성 인터페이스의 콘텐츠형 서비스는 대개 병행 활동 중 배경적으로 소비된다는 고병휘(2020)의 연구 결과와도 일치한다. 곡에 크게 집중하지 않고 배경적으로 청취하기에 전반적으로 모바일에 비해 곡에 수용적이며, 이로 인해 폭넓은 음악을 듣게 되는 것이다.

쿼리 유형 별 발화 의도에서 나타난 분위기 형성 방식은 세 가지였다. 첫째, 원하는 스트림이 명확히 존재하는 경우 SQ(Specific query)를 이용해 그것의 대표 곡을 명령하기, 둘째, 원하는 스트림이 비교적 명확하지 않은 경우 나의 상황 또는 필요한 음악적 분위기를 DQ(Descriptive query)를 이용해 묘사하기, 셋째, 단순히 백그라운드 사운드 형성 용도로 (Non-specific query)를 이용해 아무 음악을 재생하기.

둘째, **일람성 부재**. 하지만 음성 인터페이스는 근본적으로 시각적 큐가 부재하기에 브라우징이 불가능하고, 결과 예측성이 떨어진다. 이는 기존 연구들에서도 지속적으로 논의된 바이다 (Myers et al., 2018; Corbett & Weber, 2016; Luger & Sellen, 2016). 이로 인해 음악 도메인에서는, 검색 결과나 이후 재생될 리스트를 확인하지 못하는 문제가 발생한다. 2차조사 결과, 참여자들은 특히 DQ(Descriptive query)와 NSQ(Non-specific query)에서 이후 곡들을 예상하지 못하는 모습을 보였다.

명확한 형태의 SQ(Specific query)는, 이러한 결과를 컨트롤하는 전략으로서 사용된다. 1차조사 결과, 초기 검색 결과가 마음에 들지 않아

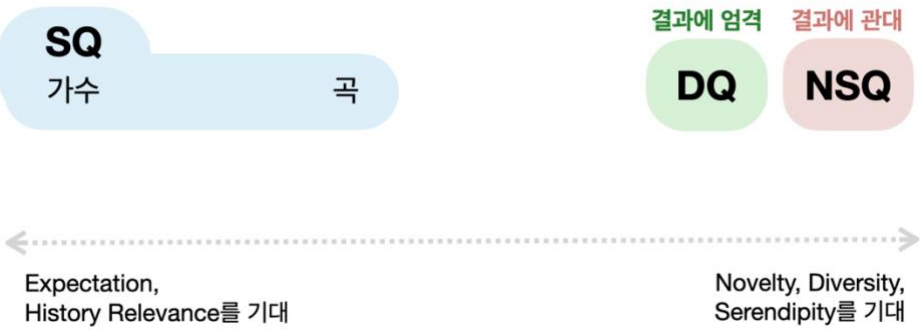
재쿼리를 시도하고자 할 때, SQ를 이용한 경우가 78%로 압도적으로 많았다. 실제 인터뷰 결과, *"'잔잔한 노래 틀어줘' 같은 요청(descriptive query)조차 기대했던 것과 다른 결과가 나와서, 구체적인 곡으로 무드를 조정하기도 한다 (P03, P04, P07)"* 는 내용이 언급되었다. 즉 예측성이 떨어지고 수용하기 어려운 결과에 대해, 예측이 용이하고 구체적인 쿼리를 사용하는 모습이다.

셋째, **인식 오류 가능성.** 또 다른 음성 인터페이스의 근본적인 문제는 인식 오류의 가능성이다. 인식 오류는 음악 도메인에서 더욱 피하기 어려운데, 창작자에 의해 제목이 일반적이지 않은 문자 또는 문법으로 표현되거나 (Springer & Cramer, 2018), 동일한 곡 제목을 다양한 가수가 발매하는 경우가 비일비재하기 때문이다 (Xiao et al., 2021).

본 연구의 인터뷰에서도 음성 인터페이스 음악 경험의 불만사항으로 가장 많이 언급된 것이 바로 음성 인식 오류였다. 이는 특히 SQ(Specific query)에서 지배적으로 나타났다. 1차조사 결과, 전체 오류 로그의 89%가 SQ에서 발생했다. 사용자들은 이러한 오류를 방지하기 위해, DQ(Descriptive query)나 NSQ(Non-specific query)를 사용하기도 한다고 답했다. 이는 Thom et al.(2020)이 논의한 대로, 결과의 미세한 컨트롤을 포기하고 적은 노력(effort)으로 분위기 형성을 채택하는 경우이다. 이렇게 인식 오류의 가능성은 다시 배경적 청취의 원인이 되기도 한다.

제 2 절 음성 인터페이스의 쿼리 별 추천 스트림 제언

또한 본 연구에서는 각 쿼리 유형에 따라 사용자가 기대하는 스트림의 형태를 구체적으로 살펴보았다. 해당 결과를 토대로, 음성 인터페이스의 쿼리 별 추천 스트림 설계 방식에 대해 제언하고자 한다.



[그림 13] 쿼리 유형 별 추천 스트림에 기대 차이.

- NSQ : 새롭고 다양하고 의외의 곡들이 재생되길 기대한다. 결과에 대한 수용이 관대하기 때문에, 개인화된 스트림이 인식되는 선에서 비교적 적극적인 추천을 시도해봄 직하다.

- DQ : 구체적인 기준을 제시하면서도 새롭고 의외의 곡들이 재생되길 기대한다. 하지만 응답에 따라 편차가 존재하는 것을 보아, 사용자 특성이나 상황 맥락에 따라 기대가 달라질 수도 있다. 본 연구에서 DQ를 세 가지 하위 유형으로 나누었으나, 기대하는 스트림 요소에 유의미한 차이가 없었다. 이에 대해서는 추후 연구에서 더 조사될 필요가 있다.

- SQ : 결과를 충분히 예상 가능하며 개인화된 추천 스트림이 재생되기를 기대한다. 실제로 재생 후 만족도도 가장 높게 나타났다. 하위 유형을 살펴보면, 곡으로 요청했을 때 DQ와 NSQ처럼 새롭고 다양한 추천을 기대하는 모습이었다.

또한 추천 스트림에 대한 기대와 인식을 비교 분석한 결과, 세 쿼리 모두 기대에 비해 '의외의 곡'을 발견하기가 어려웠다. 음성 인터페이스의 음악은 보다 배경적이고 탐색적인 목적으로 이용되는 경향이 있기에, 사용자가 의외의 좋은 곡을 발견할 수 있도록 시스템이 도와야 한다. 또한 NSQ에서 새로운 곡이 기대에 비해 덜 재생되었으며, DQ에서 다양한 곡이 기대에 비해 덜 재생되었다.

마지막으로, 음성 인터페이스 사용 빈도에 따라 '헤비 유저'와 '라이트 유저'로 나누어 분석했다. 그 결과 추천 스트림에 기대하는 요소는 두 집단에서 동일하게 나타났다. 하지만, 라이트 유저에서 DQ, NSQ의 만족도가 SQ보다 유의미하게 낮았다. 헤비 유저는 이러한 차이를 보이지 않았다. 두 집단이 쿼리를 사용하는 목적은 동일하지만, 예상과 다른 결과가 나왔을 때 라이트 유저는 그것을 수용하거나 컨트롤하는 것이 낯설기 때문일 것이다.

제 7 장 결 론

제 1 절 연구 요약

본 연구는 음성 인터페이스의 음악 도메인에서 사용되는 쿼리의 유형을 이해하고, 쿼리 유형 별로 추천 스트림에 기대하는 요소를 파악한 후, 음성 인터페이스 쿼리의 차별점을 밝히고 음악 추천 방식에의 디자인 함의점을 제시하고자 하였다.

이를 위해 1차 조사에서는 쿼리 유형 파악을 위해 구글 홈의 사용 기록을 수집해, 선행 연구를 참고하여 음성 인터페이스의 음악 쿼리를 유형화했다. 이후 음악 관련 로그를 세션 단위로 나누어, 쿼리 유형에 따라 재쿼리 패턴이 어떻게 나타나는지 분석하였다. 마지막으로 사용 기록을 기반으로 참여자와 인터뷰를 진행했다.

2차 조사에서는 쿼리 유형 별 기대 요소를 파악하기 위해 ESM(Experience Sampling Method)을 활용하였다. 참여자가 음성 인터페이스로 음악을 재생할 때마다 추천 스트림에 대한 기대 및 인식을 묻는 설문을 수집했다. 설문 응답을 각각의 발화 데이터와 매칭한 뒤, 쿼리 유형 별로 나누어 통계 분석을 진행했다.

<연구 문제 1>, 음성 인터페이스의 음악 쿼리 양상을 파악한 결과, 표현된 니즈의 구체성에 따라 SQ(Specific Query), NSQ(Non-Specific Query), DQ(Descriptive Query)로 나눌 수 있었다.

전체 사용 기록 중 SQ가 가장 많이 사용되었으며, NSQ와 DQ는 비슷하게 사용되었다. NSQ로 음악을 재생했을 때, 동일 세션에서 재쿼리 없이 세션을 종료하는 비율이 가장 높게 나타났다 (71%).

모든 쿼리에서, 재쿼리를 시도하는 경우 SQ를 이용하는 경향을 보였다. 또한 세 쿼리 가운데 DQ에서 재쿼리 시점이 가장 이르게 나타났는데 이를 쿼리 유형에 따라 결과에 대한 관용도가 다르기 때문이라고 해석하였다.

음악 쿼리 유형에 따라 사용자들의 발화 의도와 만족도 역시 다르게 나타났다. SQ는 주로 곡에 대한 명확한 니즈가 있거나, 스트림을 곡으로 컨트롤하고자 할 때 사용되었다. 원하는 무드를 미세하게 맞출 수 있다는 점에서 만족하나, 인식 오류로 인한 불만이 나타났다. NSQ는

니즈가 없을 때 추천 곡을 기대하거나 인식 오류를 피하기 위해 사용되었다. 취향이 잘 반영되어 만족하나, 엉뚱한 노래가 나오는 등 결과 예측이 불가능하다는 불만이 나타났다. DQ는 니즈가 명확하지 않을 때 전반적인 스트림을 컨트롤하기 위해 사용되었다. 구체적 기준에 따라 예상 가능한 결과가 나와 만족하나, 취향과는 거리가 멀다는 불만이 나타났다.

<연구 문제 2>, 음성 인터페이스 음악 쿼리의 유형에 따라, 기대하는 추천 스트림이 다르게 나타났다.

SQ는 결과 예상이 잘되면서 평소 듣던 것과 유사한 곡들이 나오길 기대한다. DQ와 NSQ는 결과 예상이 안 되지만, 보다 다양하고 새롭고 의외의 곡들이 나오길 바란다. 동일한 쿼리이더라도 세부적인 요청 정보에 따라 기대하는 바가 달라졌다. SQ 내에서 '곡'을 요청하는 쿼리는 '가수'를 요청하는 쿼리보다 새롭고 다양한 곡들을 기대했다.

이때 기대 대비 인식을 분석하여, 현재의 추천 스트림에서 개선이 필요한 요소를 발견했다. 세 쿼리 모두 의외성(Serendipity)이 기대에 비해 충족되지 않고 있었다. DQ에서는 다양성(Diversity)이, NSQ에서는 새로움(Novelty)이 기대에 비해 충족되지 않고 있었다.

본 연구는 이전에 정리되지 않았던 음성 인터페이스의 음악 쿼리를 로그 기반으로 유형화하였다. 그리고 쿼리 유형 별로 음악 추천 스트림을 설계할 때 고려해야 할 요소를 제언했다. 마지막으로, 음성 인터페이스의 음악 쿼리가 갖는 차별점을 기존의 음악 쿼리 연구와 비교하여 제시했다.

제 2 절 연구 한계

본 연구에서 세 가지 쿼리 유형을 밝혀냈지만, 쿼리 별 세부적인 하위 유형에 대해서는 (e.g., 'DQ(Descriptive Query)'의 음악적 분위기/ 상황이나 활동/ 장르.) 인터뷰나 통계 분석을 면밀하게 진행하지 못하였다. 특히 'SQ'의 경우, 하위 유형에 따라 사용자가 기대하는 추천 스트림이 다르게 나타났다. 'DQ'의 경우, 사용자가 스트림을 구체적으로 컨트롤하려는 모습과 새롭고 다양한 곡들을 기대하는 모습이 공존했다. 후속 연구에서는 쿼리 유형을 더 세밀하게 분류해, 첫 곡에 대한 기대와 이후 곡들에 대한 기대를 나누어 살펴볼 필요가 있다.

참여자와 태스크 측면에서, 1차 조사의 참여자가 비교적 적게 모집되었다 (9명). 음성 인터페이스 장기 사용자를 추가적으로 모집해 사용 기록을 수집한다면, 본 연구에서 밝힌 발화 예시 외에도 새로운 예시가 발견될 수 있다. 또한 2차 조사에서 자연스러운 데이터 수집을 위해 ESM을 진행했으나, 음악 청취 시간에 제한(15분)을 두고 설문을 수집했기에 완전히 자연스러운 데이터라고 말할 수 없다.

마지막으로, 유튜브뮤직(Youtube Music)만을 연구 대상으로 삼고 다른 음악 스트리밍 플랫폼은 살펴보지 않았다. 이는 음악 추천 알고리즘 간 차이를 배제하기 위한 선택이었으나, 다른 플랫폼에서 연구를 진행할 경우 본 연구 결과와 다르게 나올 가능성도 배제할 수 없다. 다양한 스트리밍 플랫폼 사용자로부터 종합적인 추천 스트림 기대 요소를 파악할 필요가 있다.

제 3 절 연구 의의

본 연구는 1) 음성 인터페이스의 음악 쿼리를 로그 기반으로 유형화한 점, 2) 쿼리 유형 별로 음악 추천 설계 시 고려할 요소를 제언한 점, 3) 음성 인터페이스의 음악 쿼리가 갖는 차별점을 기존 연구와 비교해 제시한 점에서 의의가 있다.

먼저, 음성 인터페이스의 음악 쿼리를 로그 기반으로 유형화했다. 음성 인터페이스에서 음악 도메인이 핵심적임에도 불구하고, 전반적인 인터랙션 양상을 로그 기반으로 탐색한 연구는 기존에 없었다. 본 연구는 음성 쿼리에서 특징적으로 나타나는 'NSQ' 개념을 참고하여, 니즈 구체성에 따라 쿼리를 분류하였다. 해당 결과는 학술적 측면에서 음악 도메인에 나타나는 다양한 니즈를 확인했다는 데에 의의가 있다. 또한 산업적 측면에서, 시스템이 사용자 커맨드 유형을 자동으로 인식할 수 있는 기준을 마련했다는 데에 의의가 있다.

또한 쿼리 유형에 따라 음악 추천을 설계할 때 고려해야 할 요소를 제언했다. 두 차례의 조사를 통해, 각 쿼리에서 사용자들이 원하는 추천 스트림이 다르다는 점을 확인할 수 있었다. 음성 인터페이스 상 음악 추천을 설계하고자 하는 연구자 또는 산업 종사자에게 가이드라인을 제공한다는 점에서 의의가 있다.

마지막으로, 음성 인터페이스의 음악 쿼리가 갖는 차별점을 기존 연구와 비교해 제시했다. 기존 연구에서 살펴본 음악 쿼리는, 니즈가 명확한 상황에서만 입력되는 것들이기에 쿼리 내 정보가 서지 데이터(e.g., 가수 이름, 곡 제목, 날짜 등)인 것이 대부분이었다. 하지만 음성 인터페이스에서는 니즈가 명확한 상황뿐만 아니라 니즈가 존재하지 않는 상황에서도 쿼리가 입력된다 (e.g., "신나는 가사 없는 lofi 틀어줘"부터 "음악 틀어줘"까지 나타날 수 있음). 따라서 쿼리를 통해 표현되는 니즈가 훨씬 다층적이다. 뿐만 아니라 기존 인터페이스에서는 행동 데이터로 사용자의 니즈를 예측하고자 했으나, 세션 초기에 누적되는 행동 데이터가 적어 실패하였다. 반면 음성 인터페이스는 세션 시작 단계에서 사용자가 쿼리를 명확히 입력하므로 니즈 포착이 용이하다. 이를 토대로, 니즈 별 적응적인(adaptive) 추천 스트림을 제공할 수 있음을 암시한다.

부 록

사전 설문 문항

어떠한 말로 음악을 재생했나요? *

- ◯ 특정한 곡으로 (ex. "Hype boy 틀어줘")
- ◯ 분위기나 장르를 묘사해서 (ex. "운동할 때 듣기 좋은 노래" / "재즈 틀어줘")
- ◯ 아무 기준 없이 (ex. "음악 틀어줘")
- ◯ 기타: _____

유사함 (Relevance) - 내가 평소에 듣던 곡과 유사한 것들이 재생되길 바란다. *
(5: 매우 그렇다 ~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

다양성 (Diversity) - 장르나 가수 측면에서 다양한 곡들이 재생되길 바란다. *
(5: 매우 그렇다 ~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

새로움 (Novelty) - 전에 알지 못했던 새로운 곡들이 재생되길 바란다. *
(5: 매우 그렇다 ~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

의외성 (Serendipity) - 예상 밖의 좋은 곡들을 발견하길 바란다. (5: 매우 그렇다 ~ 1: 매우 그 *
렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

예상 정도 (Expectation) - 어떤 곡들이 재생될지 대략적으로 예상된다. (5: 매우 그렇다 ~ 1: *
매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

이제 15분 동안 음악을 들어 주세요.
!! 재생되는 동안은 유튜브뮤직 앱 접속 금지 !!

곡을 **전환**하고 싶으면, **상단바의 ⏩ 버튼**으로만 조작해 주세요.

사후 설문 문항

만족도 (Satisfaction) - 15분 동안 재생된 음악은 전반적으로 만족스러웠다. (5: 매우 그렇다 ~ 1: 매우 그렇지 않다) *

	5	4	3	2	1
.	◯	◯	◯	◯	◯

예상 정도 (Expectation) - 15분 동안 재생된 음악은, 내가 대략적으로 예상했던 것과 유사하다. (5: 매우 그렇다 ~ 1: 매우 그렇지 않다) *

	5	4	3	2	1
.	◯	◯	◯	◯	◯

곡 전환(⏩ 버튼 클릭)을 몇 번이나 했나요? *
유튜브뮤직 앱에서 재생된 목록을 참고해,
답을 작성해 주세요.

내 답변

유사성 (Relevance) - 15분 동안, 내가 평소에 듣던 곡과 유사한 것들이 재생되었다. *
(5: 매우 그렇다 ~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

다양성 (Diversity) - 15분 동안, 장르나 가수 측면에서 다양한 곡들이 재생되었다. *
(5: 매우 그렇다 ~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

새로움 (Novelty) - 15분 동안, 전에 알지 못했던 새로운 곡들이 재생되었다. (5: 매우 그렇다 *
~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

의외성 (Serendipity) - 15분 동안, 예상 밖의 좋은 곡들을 발견했다. *
(5: 매우 그렇다 ~ 1: 매우 그렇지 않다)

	5	4	3	2	1
.	◯	◯	◯	◯	◯

원하는 노래를 틀 때, 세 가지 유형의 명령을 골고루 시도해 주세요!
- **기준 없이** *("음악 틀어줘")*
- **분위기나 장르로** *(ex. "운동할 때 듣는 노래 틀어줘", "재즈 스테이션 틀어줘")*
- **특정 곡으로** *(ex. "뉴진스의 hype boy 틀어줘")*

참고 문헌

고병휘. (2020). *콘텐츠형 Voice User Interface 에서의 무응답 대응방식에 대한 연구* (Doctoral dissertation, 서울대학교 대학원).

Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, Search, and IoT: How People (Really) Use Voice Assistants. ACM Trans. Comput. Hum. Interact., 26(3), 17-1.

D. Bainbridge, S.J. Cunningham, and J.S. Downie. Analysis of queries to a Wizard-of-Oz MIR system: Challenging assumptions about what people really want. In Proc. of the 4th Int. Society for Music Information Retrieval Conf., Baltimore, Maryland, USA, 2003.

Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 243, 1 – 13.

Chelba, C., & Schalkwyk, J. (2013). Empirical exploration of language modeling for the google. com query stream as applied to mobile voice search. In Mobile Speech and Advanced Natural Language Solutions (pp. 197-229). Springer, New York, NY.

Eric Corbett and Astrid Weber. 2016. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16). Association for Computing Machinery, New York, NY, USA, 72 – 82.

Downie, J. S., & Cunningham, S. J. (2002). Toward a theory of music information retrieval queries: System design implications. In Proceedings: Third International Conference on Music Information Retrieval. ISMIR

2002: 13–17 October 2002, Paris, France. (c) 2002 IRCAM Centre Pompidou.

Grasch, P., Felfernig, A., & Reinfrank, F. (2013, October). Recomment: Towards critiquing-based recommendation with speech interaction. In Proceedings of the 7th ACM Conference on Recommender Systems (pp. 157–164).

Ido Guy. 2016. Searching by Talking: Analysis of Voice Queries on Mobile Web Search. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR '16). Association for Computing Machinery, New York, NY, USA, 35 – 44.

Christine Hosey, Lara Vujović, Brian St. Thomas, Jean Garcia- Gathright, and Jennifer Thom. 2019. Just Give Me What I Want: How People Use and Evaluate Music Search. In CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4 – 9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 12 pages.

Jannach, D., & Chen, L. (2022). Conversational Recommendation: A Grand AI Challenge. arXiv preprint arXiv:2203.09126.

Iman Kamehkhosh, Dietmar Jannach, and Geoffray Bonnin. 2018. How automated recommendations affect the playlist creation behavior of users. In Proceedings of the 23rd ACM Conference on Intelligent User Interfaces Workshops: Intelligent Music Interfaces for Listening and Creation (MILC'18).

Kang,J.,Condiff,K.,Chang,S.,Konstan,J.A.,Terveen,L.,Harper,F.M.:Understa nding how people use natural language to ask for recommendations. In: Proceedings of the Eleventh ACM Conference on Recommender Systems – RecSys 2017, pp. 229 – 237. ACM Press, New York (2017)

Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias

Hagen. 2018. Toward Voice Query Clarification. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18). Association for Computing Machinery, New York, NY, USA, 1257 – 1260.

Kim, H. J., Park, S. Y., Park, M., & Lee, K. (2020, September). Do Channels Matter? Illuminating Interpersonal Influence on Music Recommendations. In Fourteenth ACM Conference on Recommender Systems (pp. 663–668).

Kostric, I., Balog, K., & Radlinski, F. (2021, September). Soliciting User Preferences in Conversational Recommender Systems via Usage-related Questions. In Fifteenth ACM Conference on Recommender Systems (pp. 724–729).

Lee, J. H., Bare, B., & Meek, G. (2011, October). How Similar Is Too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation. In ISMIR (Vol. 11, pp. 109–114).

Lee, J. H., Wishkoski, R., Aase, L., Meas, P., & Hubbles, C. (2017). Understanding users of cloud music services: selection factors, management and access behavior, and perceptions. Journal of the Association for Information Science and Technology, 68(5), 1186–1200.

J.H. Lee. Analysis of user needs and information features in natural language queries seeking music information. Journal of the American Society for Information Science and Technology, 61(5):1025 – 1045, 2010.

Lee, J. H., Bare, B., & Meek, G. (2011, October). How Similar Is Too Similar?: Exploring Users' Perceptions of Similarity in Playlist Evaluation. In ISMIR (Vol. 11, pp. 109–114).

Ang Li, Jennifer Thom, Praveen Chandar, Christine Hosey, Brian St. Thomas, and Jean Garcia-Gathright. 2019. Search Mindsets: Understanding Focused and Non-Focused Information Seeking in Music Search. In The

World Wide Web Conference (WWW '19). Association for Computing Machinery, New York, NY, USA, 2971 – 2977.

Li, X., Nguyen, P., Zweig, G., & Bohus, D. (2009, April). Leveraging multiple query logs to improve language models for spoken query recognition. In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 3713–3716). IEEE.

Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems – CHI '16(2016), 5286 – 5297.

Mavrina, L., Szczuka, J., Strathmann, C., Bohnenkamp, L. M., Krämer, N., & Kopp, S. (2022). "Alexa, You're Really Stupid": A Longitudinal Field Study on Communication Breakdowns Between Family Members and a Voice Assistant. Frontiers in Computer Science, 4, 791704.

Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 6, 1 – 7.

Pearl, C. (2016). Designing voice user interfaces: principles of conversational experiences. " O'Reilly Media, Inc.".

Schedl, M., Zamani, H., Chen, C. W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. International Journal of Multimedia Information Retrieval, 7(2), 95–116.

Bruno Sguerra, Marion Baranes, Romain Hennequin, and Manuel Moussallam. 2022. Navigational, Informational or Punk-Rock? An Exploration of

Search Intent in the Musical Domain. In Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22). Association for Computing Machinery, New York, NY, USA, 202–211.

Springer, A., & Cramer, H. (2018, April). "Play PRBLMS" Identifying and Correcting Less Accessible Content in Voice Interfaces. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–13).

Tang, M. C., & Jhang, P. S. (2020). Music discovery and revisiting behaviors of individuals with different preference characteristics: An experience sampling approach. Journal of the Association for Information Science and Technology, 71(5), 540–552.

Johanne R. Trippas, Damiano Spina, Lawrence Cavedon, Hideo Joho, and Mark Sanderson. 2018. Informing the Design of Spoken Conversational Search. In Proceedings of 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA, March 11–15, 2018 (CHIIR '18), 10 pages

J.Thom, A.Nazarian, R.Brillman, H. Cramer, S. Mennicken. ""Play Music": User Motivations and Expectations for Non-Specific Voice Queries", 21st International Society for Music Information Retrieval Conference, Montréal, Canada, 2020.

Trippas, J. R., Spina, D., Cavedon, L., Joho, H., & Sanderson, M. (2018, March). Informing the design of spoken conversational search: Perspective paper. In Proceedings of the 2018 conference on human information interaction & retrieval (pp. 32–41).

Sergey Volokhin and Eugene Agichtein. 2018. Understanding Music Listening Intents During Daily Activities with Implications for Contextual Music Recommendation. In Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18). Association for

Computing Machinery, New York, NY, USA, 313 – 316.

Ziang Xiao, Sarah Mennicken, Bernd Huber, Adam Shonkoff, and Jennifer Thom. 2021. Let Me Ask You This: How Can a Voice Assistant Elicit Explicit User Feedback?. Proc. ACM Hum.Comput. Interact. 5, CSCW2, Article 388 (October 2021), 24 pages.

Xing, Z., Yuan, X., Wu, D., Huang, Y., & Mostafa, J. (2020, July). Understanding voice search behavior: review and synthesis of research. In International Conference on Human-Computer Interaction (pp. 305-320). Springer, Cham.

Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: introducing serendipity into music recommendation. In Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12). Association for Computing Machinery, New York, NY, USA, 13 – 22.

Yaxi Zhao, Razan Jaber, Donald McMillan, and Cosmin Munteanu. 2022. "Rewind to the Jiggling Meat Part": Understanding Voice Control of Instructional Videos in Everyday Tasks. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 58, 1 – 11.

Abstract

Study on Query Types of VUI-based Music Streaming Services

- Focusing on users' expectation for the recommendation streams -

SangAh Park
Department of Intelligence and Information
The Graduate School
Seoul National University

Music service is a core domain in Voice User Interface (VUI). According to the survey, music is one of the most used services by smart speaker users, and the frequency of use is also the highest. Thus, the music experience can affect entry into or exit of the voice user interface. It is necessary to understand the music experience of the voice user interface only.

In voice user interface, a single query triggers an auto-generated music stream. Unlike mobile interface, search and playback occur simultaneously without an exploration step. That is,

by one query, the top results and related songs are played as a stream.

To understand the music experience of voice user interface, it is necessary to understand the query input by users. Depending on the type of query, the user expectations for the recommendation are different (e.g., “Play calm jazz” vs. “Play music”). The gap between user expectations and actual results can affect the experience. In a positive case, it may lead to serendipity to discover new songs. But in a negative case, it may lead to distrust of recommendations or deterioration in the use of voice user interface.

This study aims to understand the types of queries in the music domain of voice user interface, and to identify the elements that users expect from recommendation streams for each type of query. This study conducted two major investigations.

The purpose of the primary investigation is to understand the types of queries used in the music domain of voice user interface. 2,723 music-related logs were collected from 9 smart speaker users, and music queries were classified based on previous study. As a result, music queries were largely classified into three categories: 1) SQ (Specific Query), request by song or artist, 2) NSQ (Non-Specific Query), no criteria presented, 3) DQ (Descriptive Query), mood or

genre description. As a result of log analysis, the number and timing of re-queries were different for each query type. As a result of the log-based interview, intention and satisfaction were different depending on the query type.

The purpose of the secondary investigation is to find out what users expect from recommendation for each type of query. 27 participants were given an ESM(Experience Sampling Method) task that triggers music using voice user interface for five days, and expectations of music recommendation were collected on a 5-point scale through a questionnaire. Survey responses were collected for a total of 290 queries, and the following characteristics were derived: 1) SQ - Songs with high relevance within the expectation were desired, and satisfaction was high. 2) NSQ, DQ - Novel, diverse, serendipitous songs were desired. Satisfaction was low.

Based on the results of this study, the following was discussed. First, the music experience of voice user interface is significantly different in three points – background listening, absence of visibility, possibility of recognition error. This allows users to strategically select queries. Second, based on the user expectations, we propose a design method for recommendation streams for each query type.

This study identified the types of music queries used in voice user interface, and confirmed user expectations of recommendation by query type. In addition, we revealed the characteristics of music queries and experiences in voice user interface, and suggested music recommendation direction for each query type.

.....

keywords : Voice User Interface, Music query, Music recommendation, User expectation.

Student Number : 2020-24842