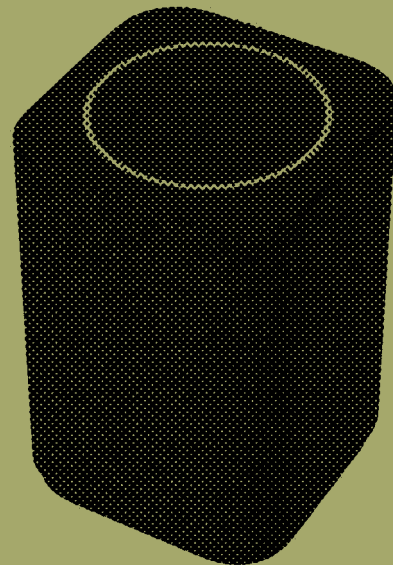


카카오미니

사람을 향하다



앞선 리포트에서는 카카오미니가 사람의 목소리를 어떻게 인식하고, 사람처럼 스무고개 게임을 하는 지에 대해 기술적으로 설명하는 글이 게재됐습니다. 이번 호에는 카카오미니 디자인의 탄생 비화를 들려달라고 카카오미니 프로젝트 팀에 요청했습니다. 두 명의 담당자가 바쁜 시간을 쪼개 글을 써줬습니다. 먼저, 외형 전반에 대해서는 카카오미니 디자인 작업을 진두지휘한 최경국 님이 답을 주셨습니다. 이 글이 카카오미니 생김새의 이유를 기술해 준 글이라면, 황병식 님은 카카오미니 몸체 구성구석의 기능과 그것들의 제작 배경에 대한 설명해 주셨습니다.

카카오미니 하드웨어 툭아보기

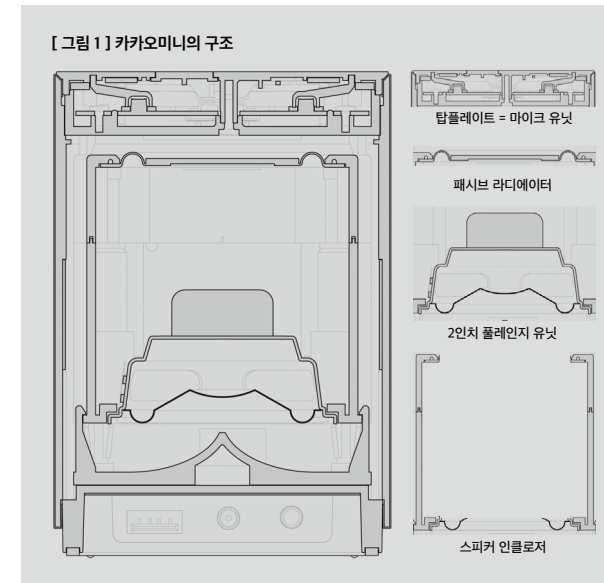
카카오미니의 디자인은 사용자의 목소리를 잘 들을 수 있는 마이크를 설계하는 작업에서부터 시작한다. 사용자의 음성을 잘 들을 수 있는 사용자와 마이크 사이의 최적의 거리를 계산한 뒤 이를 바탕으로 최적의 비율로 높이를 결정했다. 다시 말해 디자인적 요소와 2인치 플레인지 스피커가 필요로 하는 볼륨(공간) 두 가지를 모두 고려한 것이다.

그리고 사용자에게 좋은 품질의 음향을 제공하기 위해 카카오미니의 내부 공간 중 최대한 많은 부분을 스피커 유닛에 할당했다. 그 외의 나머지 공간에는 카카오미니를 동작시키는데 필요한 하드웨어를 배치했다. 이렇게 카카오미니의 하드웨어는 스피커로써 가져야 할 기본 역할에 충실하고자 노력했고, 열악한 구조 속에서도 음성인식 성능을 이끌어 내기 위해 많은 고민을 했으며 이를 하드웨어 디자인에 반영했다.

카카오미니의 구조

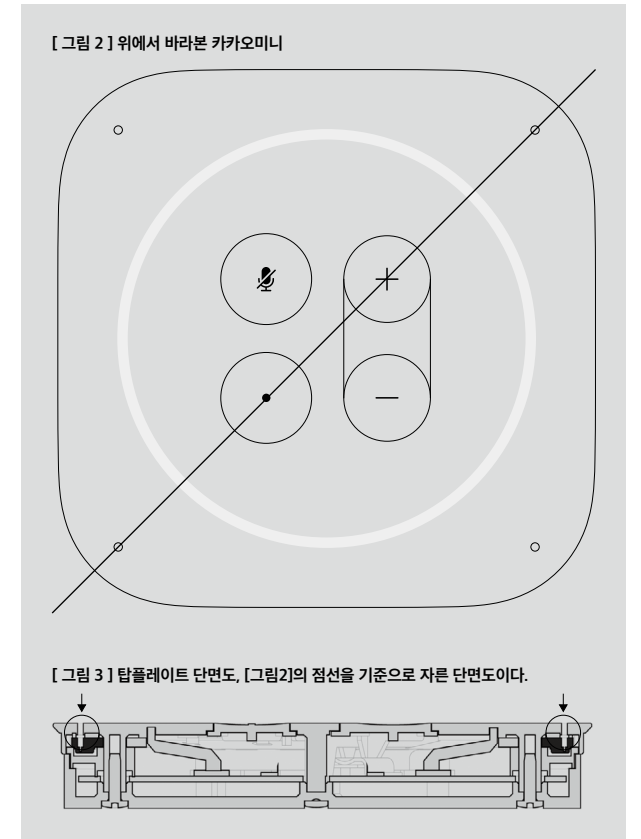
카카오미니는 크게 음성인식을 담당하는 마이크 유닛과 스피커를 담당하는 스피커 유닛으로 구성되어 있다. [그림 1]을 보면 카카오미니의 상단부에 마이크 유닛이 있고, 하단부에 스피커 유닛이 위치한다. 스피커 유닛에 많은 부분이 할당된 것을 확인할 수 있다.

스피커 유닛을 조금 더 자세히 살펴보면, 아래쪽에 2인치 플레인지 유닛(full range unit)과 중저음 보강을 위해 상단의 패시브 라디에이터(passive radiator)로 구성이 된다. 플레인지 싱글 유닛이란, 하나의 유닛(진동판)으로 모든 대역을 재생하는 방식이다. 패시브 라디에이터란 진동판을 움직이는 보이스 코일과 영구자석이 없는 유닛으로, 인클로저(스피커를 담은 박스) 내부의 공기 압력에 대응하여 동작되며 오디오의 저역대를 보강하게 된다.



음성 입력

카카오미니는 사용자의 음성을 정확히 듣기 위해 4개의 마이크가 상부에 배치되어 있다. 마이크 모듈(module)은 음성주파수 대역에서의 공진¹⁾을 방지하고, 입력되는 음성의 마이크 별 음량을 맞추기 위해 기구물(탑플레이트)²⁾, PCB(printed circuit board, 인쇄 회로 기판)³⁾ 디자인에 여러 기법을 적용했다.



예를 들어 카카오미니 마이크의 입력포트는 PCB와 기구의 홀(hole)을 통과하게 되는데 이 홀의 구조는 음성대역의 주파수에서 공진을 일으키지 않도록 디자인이 되었다. 공진이 발생할 경우 특정 주파수의 소리가 왜곡되어 음성인식에 악영향을 주거나 혹은 전혀 인식할 수 없는 음성으로 변형될 수 있다. 따라서 공진을 방지하는 설계 기술은 필수 요소라고 할 수 있다.

한편 마이크 모듈은 동일 음원에 대하여 PCB가 기구에 조립된 상태에서 마이크간 레벨 편차를 최소화하기 위하여, 다음 방법들을 적용했다.

- 4개의 마이크 부품을 싱글 보드에 탑재
- 마이크 포트 위치의 PCB 홀 특수 가공
- 카카오미니 내부의 영향을 억제하기 위한 2중 실링(sealing)⁴⁾구조

마이크 간 레벨 편차란 4개의 마이크에서 입력되는 마이크 별 음성 크기의 차이를 의미하며, 레벨 편차를 최소화하는 이유는 4개의 마이크에서 입력되는 오디오를 합성하는 과정에서 마이크 별 음성 크기가 달라지면 왜곡이 발생하여 음성인식에 어려움을 줄 수 있으므로 음압이 최대한 비슷해야 하기 때문이다.

사용자의 음성 왜곡을 억제하기 위하여 마이크 모듈의 품질 관리 또한 철저하게 거치고 있다. 일반 사용자의 환경에서는 사용자의 음성 이외의 많은 노이즈(noise)원(原)이 존재한다.

글 | 황병식 bruce.h@kakaocorp.com

카카오에서 10년 가까이 하드웨어 개발자로서 은둔한 끝에, 큰 빛을 보게 되었습니다. 카카오가 하드웨어 개발 역량을 보유하고 있다는걸 모르는 분들과 앞으로 많은 이야기를 나눌수 있게 되기를 희망합니다.

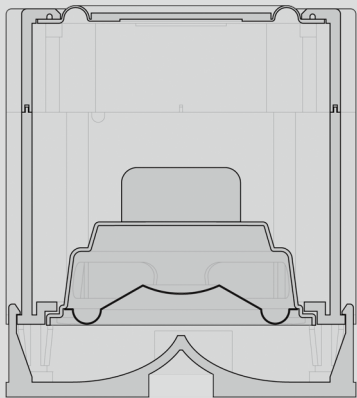
사용자의 음성 명령 이외의 다른 모든 소리 및 카카오미니에서 출력되는 소리(음악/음성 멘트)도 노이즈로 규정된다.

그 중에서도 마이크로 입력되는 노이즈 중에 가장 큰 것은 바로 카카오미니에서 출력되는 소리이다. 카카오미니는 스피커와 마이크가 근접해 있으므로 음악이나 음성 멘트가 출력 될 때에 사용자의 음성 명령을 정확히 받기 위해서는 카카오미니의 출력과 사용자의 음성 명령을 구분하여 음성엔진으로 전송해야 한다. 카카오미니에는 사용자의 음성 명령을 정확히 인식하기 위하여 AEC(acoustic echo cancellation) 기술이 적용되었다. AEC 기술은 카카오미니의 출력이 사용자의 음성 명령과 함께 마이크에 입력 될 때, 사용자의 음성 명령만을 추출한다.

오디오 출력

카카오미니의 스피커는 풀레인지 싱글 유닛(full-range single unit)과 패시브 라디에이터(passive radiator)로 구성되었다. 앞서 설명했듯이 풀레인지 싱글 유닛은 하나의 유닛을 사용하기 때문에 모든 대역의 소리를 만족스럽게 낼 수 없다. 이 때문에 일반적으로 저역대의 소리, 웅장한 소리 등을 보강하고자 우퍼(woofer)⁵나 다른 영역대의 소리에 최적화된 유닛을 이용하곤 한다. 하지만 카카오미니는 크기가 작기 때문에 우퍼나 다른 유닛을 이용하는 것은 사실상 불가능했다. 패시브 라디에이터를 이용해 싱글유닛의 저역대를 보강하는 것이 가장 적절한 설계기법이었다.

[그림 4] 카카오미니 스피커유닛(하단부)



카카오 미니의 작은 크기에 좋은 소리를 담기 위해 최대한의 공간을 스피커 유닛에 할당하였다. 좋은 소리를 위해 정격 7W(8옴⁶ 유닛 기준) 급의 Class-D AMP를 사용하였으며, 3옴 스피커 유닛을 사용하여 약 13.5W(최대)의 출력을 낼 수 있도록 설계되었다. 풀레인지 싱글 유닛의 성능을 아사히카세이 마이크디바이스(AKM)의 고음질 코덱의 5-band Parametric EQ

기반 튜닝을 통해 최대한 이끌어 내고 있다.

카카오미니는 자체 모노(mono) 사운드⁷ 보다 더 나은 출력을 사용자가 선택할 수 있도록 고급 사용자를 위한 AUX 출력단자를 내장하고 있다. 따라서 사용자는 이미 보유한 오디오 시스템 혹은 스테레오 스피커 등을 연결하여 사용할 수 있다.

AUX 단자를 통해 출력되는 사운드 또한 AEC 과정을 거치도록 설계되어 있으므로 카카오미니의 음성 명령을 처리하는데 영향을 주지 않는다. 단, AUX 단자에 연결된 스피커의 출력이 사용자의 음성보다 현저하게 높을 경우 영향을 줄 수 있다.

충전 포트

카카오미니는 스마트 기기 충전을 위한 CDP BC1.2⁸ 지원 USB-A 타입 포트를 제공한다. 시중에 유통되는 대부분의 스마트기기를 충전할 수 있으며, 최대 5V/1.5A의 출력을 낼 수 있다. 최근에 출시된 아이폰8(iPhone8)의 경우 약 70분 내외에 충전할 수 있다.⁹

상단 버튼 및 LED

카카오미니는 기본적으로 사용자의 음성 명령을 기반으로 동작하지만 주변 환경이 시끄러운 경우 웨이크업(wake up) 동작을 지원하기 위한 버튼과 볼륨 조정 및 다기능(multi-function) 버튼을 물리적으로 제공하고 있다.

웨이크업이란 음성인식을 깨운다는 의미로 카카오미니의 경우 '헤이 카카오'가 웨이크업 역할을 한다. 또한 사용자의 음성 명령에 대하여 직관적인 피드백을 제공하기 위하여 카카오미니의 상부에 풀컬러(full-color) 원형 LED를 배치하였으며, 이를 통해 사용자는 음성 명령에 대한 카카오미니의 즉각적인 피드백을 시각적으로 확인할 수 있다.

프로세서/OS

카카오미니는 서비스 확장성을 고려하여 안드로이드(android) 플랫폼에서 구동되며, 원활한 서비스 구동을 위하여 커스텀 드라이버(custom driver) 개발 및 BSP¹⁰(board support package) 최적화 과정을 진행하였다. 안드로이드의 원활한 구동을 위하여 카카오미니의 좁은 공간에는 ARM사의 쿼드코어 프로세서와 DDR/eMMC(embedded multi media card) 등의 많은 부품들을 탑재하였다.

그러나 좁은 공간과 디자인 등의 문제로 인해 AP(application processor) 보드에서 발생하는 열을 스피커 밖으로 배출하기

어려워졌고, 이로 인해 서비스가 구동될 때 열을 배출하는 것이 중요한 포인트가 되었다. 구조상 발생하는 이러한 발열을 줄이기 위해 추가적인 BSP 튜닝 작업이 여러 단계에 걸쳐 진행되었다.

글을 마치며

카카오미니는 아직까지도 조금은 생소한 음성 인터페이스 기반의 제품으로 사용자의 목소리를 어떻게 하면 더 잘 들을 수 있는지를 중점적으로 고민하고 설계됐으며, 이를 테스트하는데 많은 시간을 할애했다. 그 중에서도 AEC 즉 사용자의 음성 명령만을 추출하는 기능을 정상적으로 동작시키기 위해 많은 시행 착오를 겪었다.

이 기술의 기본 컨셉과 성능 측정, 기구 설계에 대한 노하우를 확보할 수 있었으며 앞으로 카카오가 만들어낼 많은 제품의 바닥을 다질수 있는 좋은 기회였다.

앞으로 등장할 음성인식 기반의 제품 및 서비스들이 더 많은 사용자의 소리를 듣고 만족할 만한 서비스를 제공하는데 있어, 카카오미니가 기반이 되기를 희망해본다.

^{*1} 출처 | <https://ko.wikipedia.org/wiki/공명> ^{*2} 참고 | 마이크간 입력 편차를 최소화하기 위한 여러 디자인 기법이 적용된 카카오미니의 헤드유닛 ^{*3} 참고 | 패닝/에폭시등의 절연판 위에 동박을 부착한 후 동박면에 회로를 형성시켜 그 위에 실장된 전자부품이 동작하도록 제작된 기판을 말한다. ^{*4} 참고 | 마이크 주변과 케이블이 연결되는 부분을 밀폐하는 것을 의미한다. ^{*5} 참고 | 우퍼란 일반 스피커의 부족한 저음을 보완해주기 위해 사용되는 저음 전용 스피커이다. ^{*6} 참고 | 옴은 스피커 보이스코일의 고유 저항값이다. ^{*7} 참고 | 1개의 채널로 출력되는 사운드이다. 반대로 스테레오는 2채널을 사용한다. 예를 들어 스테레오 방식을 이용하면 아이폰의 좌, 우에서 다른 소리가 출력되도록 할 수 있다. ^{*8} 참고 | http://www.usb.org/developers/docs/devclass_docs/USB_Battery_Charging_1.2.pdf ^{*9} 참고 | 계산상 수치이며 실제 충전시간은 달라질 수 있다 ^{*10} 참고 | 보드(board) 즉 하드웨어를 지원(support)해주는 소프트웨어의 묶음(package)을 의미한다.