

# 서비스 플로우 설계 (Supertone API 기반)

## 개요

Supertone API 는 기성 음성과 사용자 정의 음성을 제공하여 텍스트를 실제 음성처럼 합성하는 REST API 이다. 2025 년 7 월 31 일 업데이트에서 사전제공 음성(프리셋)과 사용자 정의 음성이 분리되었으며, v1/custom-voices 에서 클론 음성을 직접 등록 · 검색 · 편집할 수 있게 되었다[1]. 또한 텍스트를 스트리밍 방식으로 변환하는 TTS API 가 베타 출시되어 지연 시간을 줄일 수 있다[2]. 본 서비스는 홍보 문구나 공공 안내를 작성한 뒤 TTS 음원으로 변환하고, 지정 시간에 전송하는 것이 목적이다. 도지사·시장 같은 지자체장 목소리를 클론해 메시지에 활용하고, 필요하면 배경 음악과 믹싱하는 기능도 제공한다.

## 시스템 구성

- **사용자 인터페이스:** 문구 작성 및 발송 시간을 지정하는 관리자 화면. 메시지 작성 시 300 자 제한과 언어·감정 스타일을 안내한다.
- **음성 선택/등록 모듈:**
  - **프리셋 음성 목록:** GET /v1/voices 로 Supertone 이 제공하는 음성 목록을 불러와 이름, 언어, 스타일, 샘플 등을 확인한다[3].
  - **프리셋 음성 검색:** GET /v1/voices/search 에서 이름·언어·감정 스타일 등으로 필터링해 적합한 음성을 찾을 수 있다[4].
  - **클론 음성 등록:** 지자체장 등 특정 인물의 음성을 등록하려면 POST /v1/custom-voices/cloned-voice 를 호출하여 3 MB 이하의 WAV/MP3 파일과 이름을 업로드한다. files 는 필수이며, 이름 길이는 최대 100 자로 제한되고 Free 티어에서는 지원되지 않는다[5].
  - **클론 음성 관리:** 등록된 음성은 GET /v1/custom-voices 로 목록화하고, GET /v1/custom-voices/{voice\_id}로 단일 음성 정보를 조회하며, 필요 시 PATCH /v1/custom-voices/{voice\_id}로 이름/설명을 수정한다.
- **TTS 생성 모듈:**

- **길이 예측:** 실제 음원을 생성하기 전에 `POST /v1/predict-duration/{voice_id}`로 예측 길이를 확인한다. 이 API 는 텍스트에 대한 예상 음성 길이(초)를 반환하며 음성이 생성되지 않아 크레딧을 소비하지 않는다[6]. 예측 결과를 통해 문구를 조정하거나 예약 시간을 계산한다.
- **TTS 생성:** `POST /v1/text-to-speech/{voice_id}`로 텍스트를 음성으로 변환한다. 요청 본문에는 최대 300 자 텍스트, 언어(en/ko/ja), 감정 스타일 등을 설정하며[7], 추가로 `pitch_shift`, `speed`, `similarity` 등 세부 음성 설정을 지정할 수 있다[8]. 출력 형식은 WAV 또는 MP3 를 선택한다. `POST /v1/text-to-speech/{voice_id}/stream` 를 사용하면 음원이 스트리밍 형태로 즉시 반환되어 실시간 서비스에 적합하다.
- **믹싱 모듈:** TTS 로 생성된 음원을 오디오 편집 라이브러리(FFmpeg 등)로 배경 음악 또는 다른 음원과 믹싱한다. 예를 들어 도지사 음성으로 인사한 뒤 배경 음악과 지역 안내를 결합할 수 있다. Supertone API 는 믹싱 기능을 제공하지 않으므로 서비스 서버에서 처리한다.
- **발송/스케줄링 모듈:** 생성된 최종 음원을 파일 시스템이나 클라우드에 저장하고 예약 시간에 맞춰 문자·푸시·전화 송출 서비스에 전달한다. 예약 시스템은 시간대(Asia/Seoul)를 고려해 동작하며, 요청 실패 시 재시도 로직을 포함한다.
- **사용량/크레딧 관리:**
  - `GET /v1/usage` 는 지정 기간의 API 사용 시간을 시간별/일별 버킷으로 분석할 수 있다[9].
  - `GET /v1/voice-usage` 는 음성별 TTS 사용 기록(날짜, 음성 ID, 스타일, 사용 시간 등)을 반환해 어떤 음성이 얼마나 사용됐는지 파악할 수 있다[10].
  - `GET /v1/credits` 는 남은 크레딧을 조회하며, 크레딧은 Play 와 API 가 공유되고 초 단위로 차감된다는 점을 안내한다[11].

## 서비스 플로우 단계

1. 인증 및 준비
2. Supertone 개발자 콘솔에서 API Key 를 발급받는다. 모든 요청에는 `x-sup-api-key` 헤더를 포함해야 한다.

3. 프리셋 음성 또는 이미 등록된 클론 음성의 `voice_id` 를 확인해 두거나, 이후 단계에서 새로 생성한다.

#### 4. 문구 목적과 시나리오 정의

5. 발송할 문구의 목적(안전 안내, 지역 이벤트 홍보, 긴급 재난 알림 등)을 결정한다.
6. 청취자가 누구인지, 전달해야 할 감정(중립적, 따뜻한, 긴박한 등)을 설정한다. 이 정보는 TTS 요청의 `style` 파라미터 선택과 음성 선택에 영향을 준다[7].

#### 7. 음성 선택 또는 클론 생성

8. 프리셋 음성: GET `/v1/voices` 로 목록을 받아 언어·연령·성별·감정 스타일을 확인하고 적합한 음성을 선택한다[3].
9. 프리셋 음성 검색: 이름, 언어, 성별, 감정 스타일 등 다양한 조건으로 음성을 필터링하려면 GET `/v1/voices/search` 를 사용한다[4]. 예를 들어 한국어를 지원하고 밝은 스타일의 음성을 찾을 수 있다.

#### 10. 지자체장 음성 클론:

- a. 도지사·시장 음성 샘플(WAV/MP3)을 수집한다. 음성 파일은 3 MB 이하로 준비해야 한다[12].
- b. POST `/v1/custom-voices/cloned-voice` API 를 호출하여 파일과 `name` 을 업로드하여 클론 음성을 생성한다. 응답의 `voice_id` 를 기억한다.
- c. 필요 시 GET `/v1/custom-voices/{voice_id}`로 정보를 확인하고, PATCH `/v1/custom-voices/{voice_id}`로 이름이나 설명을 수정한다.

#### 11. 문구 작성 및 길이 확인

12. 관리자 화면에서 최대 300 자의 문구를 작성한다.
13. POST `/v1/predict-duration/{voice_id}`를 호출해 예상 음성 길이를 파악한다. 이 API 는 실제 음원을 생성하지 않으며 크레딧도 차감되지 않는다[13].
14. 예측 결과를 바탕으로 문구 길이 또는 발송 시간표를 조정한다.

#### 15. 음성 합성 (TTS)

16. POST /v1/text-to-speech/{voice\_id} (또는 실시간 응답이 필요할 경우 POST /v1/text-to-speech/{voice\_id}/stream)를 호출한다.
17. 요청 본문에는 작성한 text, language, 선택한 style 과 model 을 지정하고, output\_format 을 WAV 또는 MP3 로 설정한다[7].
18. 필요에 따라 voice\_settings 속성에서 피치 변화(pitch\_shift), 속도(speed), 발음 유사도(similarity) 등을 조정해 도지사나 시장의 발성 특징을 세밀하게 맞춘다[8].
19. 응답으로 반환되는 오디오 데이터를 저장하거나 스트리밍 결과를 받아 파일로 변환한다.

## 20. 음원 믹싱 및 후처리 (선택 사항)

21. 안내 멘트와 함께 배경 음악, 효과음 등을 삽입하려면 음성 합성 후 오디오 편집 라이브러리를 사용해 여러 트랙을 믹싱한다. 예를 들어 도지사 클론 음성으로 인사하는 부분과, 이후 청각적인 브랜드 효과를 겹쳐 재생할 수 있다.
22. 믹싱 시 TTS 음원의 길이에 맞춰 배경음 볼륨을 조절하고 시작/끝 포인트를 동기화한다.

## 23. 전송 일정 관리 및 발송

24. 최종 음원을 저장소(예: S3 버킷)에 업로드한다.
25. 예약 시스템(크론 잡 또는 메시지 큐)을 이용해 사용자 지정 시간에 발송하도록 스케줄링 한다. 서비스의 타임존 설정을 Asia/Seoul 로 고정해 국내 사용자에게 올바른 시간에 전송한다.
26. 발송 채널(앱 푸시 알림, 전화 안내, 방송 시스템 등)을 통해 음원을 전송하며, 성공/실패 로그를 기록한다.

## 27. 사용량 및 크레딧 모니터링

28. GET /v1/voice-usage 를 사용해 기간별 음성별 사용 시간과 스타일, 언어 등을 확인하고 월간 예산을 관리한다[10].
29. GET /v1/usage 를 통해 시간 단위 또는 음성 이름/ID 로 세분화된 사용 통계를 받을 수 있으며, API 호출수와 사용 시간을 파악한다[9].

30. GET /v1/credits 를 호출하여 잔여 크레딧을 확인한다. 크레딧은 Play 와 API 에서 공유되며 초 단위로 차감된다[11]. 필요 시 Play 에서 충전해 API 사용량을 보충한다.

## 주요 API 엔드포인트 정리

섹션	메서드	엔드포인트 이름	요약/용도	문서 링크
<b>Introduction</b>	-	Introduction	Supertone API 기능, 업데이트 내용, 프리셋/커스텀 음성 분리 소개[1].	<a href="#">docs</a>
<b>Authentication</b>	-	Authentication	API Key 발급과 인증 헤더 설명.	<a href="#">docs</a>
<b>Voices</b>	GET	List voices	프리셋 음성 목록을 조회하며 이름, 언어, 스타일, 샘플 등을 반환한다[3].	<a href="#">docs</a>
	GET	Search voices	이름, 언어, 성별, 연령, 스타일 등으로 프리셋 음성을 필터링한다[4].	<a href="#">docs</a>
	GET	Get voice	특정 프리셋 음성의 자세한 정보를 얻는다.	<a href="#">docs</a>
<b>Custom Voices</b>	GET	List custom voices	사용자가 등록한 클론 음성 목록을 조회한다.	<a href="#">docs</a>
	GET	Search custom voices	이름/설명 등으로 사용자 정의 음성을 검색한다.	<a href="#">docs</a>
	GET	Get custom voice	특정 클론 음성 정보를 조회한다[14].	<a href="#">docs</a>
	PATCH	Edit custom voice	클론 음성의 이름·설명을 수정한다.	<a href="#">docs</a>
	POST	Create cloned voice	WAV/MP3 파일을 업로드해 새로운 클론 음성을 등록한다[5].	<a href="#">docs</a>
<b>Text to Speech</b>	POST	Create speech	지정한 음성으로 텍스트를 음성으로 변환한다. 언어·감정	<a href="#">docs</a>

섹션	메서드	엔드포인트 이름	요약/용도	문서 링크
Usage			스타일·음성 설정을 지원한다[15].	
	POST	Stream speech (Beta)	스트리밍 방식으로 텍스트를 음성으로 변환한다 (저지연 응답).	<a href="#">docs</a>
	POST	Predict duration	입력된 텍스트의 예상 음성 길이를 반환하며 크레딧을 소모하지 않는다[13].	<a href="#">docs</a>
	GET	Get usage	시간 버킷과 다차원 분해 방식으로 API 사용량을 분석한다[9].	<a href="#">docs</a>
	GET	Get voice usage	기간별 음성별 사용 시간과 스타일을 확인한다[10].	<a href="#">docs</a>
	GET	Get credit balance	남은 크레딧을 조회한다; 크레딧은 초 단위로 차감되며 Play 와 공유된다[11].	<a href="#">docs</a>

## 마무리 조언

- 클론 음성 사용 시 법적·윤리적 이슈를 고려해야 한다. 지자체장 음성을 사용할 때는 당사자의 허가를 받아야 하며, 정책 변화에 따라 서비스 이용이 제한될 수 있다.
- 음성 합성은 최대 300 자까지 지원하므로 길이가 긴 메시지는 여러 구간으로 나눠 각각 TTS 를 호출하고 편집 소프트웨어로 이어붙일 수 있다.
- 사용량과 크레딧을 정기적으로 모니터링하여 예산 초과를 방지하고, API 속도 제한(레이트 리밋)을 확인해 동시 호출수를 조절한다.

이와 같은 흐름으로 서비스 로직을 설계하면 텍스트 작성부터 TTS 합성, 클론 음성 관리, 믹싱 및 발송까지 체계적으로 구성할 수 있다.

---

[1] [2] Welcome - Supertone API Documentation

<https://docs.supertoneapi.com/en/user-guide/welcome>

[3] List voices - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/list-voices>

[4] Search voices - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/search-voices>

[5] [12] Create cloned voice - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/create-cloned-voice>

[6] [13] Predict duration - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/predict-duration>

[7] [8] [15] Create speech - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/text-to-speech>

[9] Get usage - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/get-usage>

[10] Get voice usage - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/get-voice-usage>

[11] Get credit balance - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/get-credit-balance>

[14] Get custom voice - Supertone API Documentation

<https://docs.supertoneapi.com/en/api-reference/endpoints/get-custom-voice>