

Supertone TTS 서비스 개선 설계서

📌 개선 요구사항 요약

사용자 요구에 따라 기존 설계에 아래 사항을 추가/보완한다:

- 목적 선택 독립화:** 사용자 메시지의 용도(공지, 축사, 광고 등)를 첫 단계에서 선택하고 이후 단계(예약 전송, 사용 로그)에서 이 값을 참조할 수 있도록 한다.
- 서버 검색 vs 클라이언트 필터링:** 음성 탐색 시 클라이언트 필터링을 기본으로 하되, 필요 시 GET /voices/search 엔드포인트를 호출하여 서버 측 검색 옵션을 제공할지 결정한다. 대규모 음성 라이브러리 도입 시 서버 검색이 유용하다[1].
- 보이스 클로닝 UI/스토리지/결과 관리:** 음성 샘플 업로드(3 MB 이하 WAV/MP3) → 스토리지 저장 → POST /custom-voices/create-cloned-voice 호출[2] → 클론 음성 ID 저장 및 상태 관리 인터페이스 추가.
- 생성 기록/캐시 개선:** TTS 생성 시 요청 파라미터(voice_id, 텍스트, 스타일, 속도, 피치 등)와 결과 파일 경로를 DB에 저장해 이후 믹싱·예약 단계에서 재사용한다[3].
- 음원 믹싱 UI·기능 구현:** 배경음(BGM)과 효과음(SFX) 선택, 볼륨 조절, 미리듣기 제공, 서버에서 FFmpeg 호출로 믹싱을 수행하는 기능을 추가한다.
- 예약 전송 UI/로직 구현:** 전송 대상, 채널, 예약 시간(Asia/Seoul 기준)을 설정하는 UI와 백엔드 스케줄러 연동을 설계한다.
- 사용량·크레딧 대시보드:** GET /usage, GET /voice-usage, GET /credit-balance API를 호출하여 사용량/크레딧 정보를 시각화하고, 크레딧 부족 시 경고 알림을 제공한다[4][5].
- 우선순위·구현 범위 계획:** 단계별로 구현 순서를 제시하여 프로젝트 진행을 지원한다.

⌚ 개선된 사용자 여정 (User Journey)

단계	사용자 행동	시스템 동작	개선 포인트
----	--------	--------	--------

단계	사용자 행동	시스템 동작	개선 포인트
1. 목적 선택	메시지 용도(공지·축사·광고 등)를 선택	선택한 목적을 세션과 DB에 저장하여 이후 단계에서 참고	독립된 UI 단계로 분리, 목적 기반 통계 가능
2. 음성 탐색 및 선택	기본 음성 목록을 언어·성별·스타일로 필터링 후 선택	클라이언트 필터링이 기본이나, 키워드 검색 시 GET /voices/search 호출 가능[1]	대규모 라이브러리 대비 서버 검색 옵션 제공
3. 보이스 클로닝 (선택)	음성 샘플(WAV/MP3, 3 MB 이하)을 업로드	스토리지에 저장 후 POST /custom-voices/create- cloned-voice 호출. 처리 상태 및 결과 voice_id 를 관리한다[2]	업로드/처리 상태 UI, 결과 DB 저장
4. 텍스트 작성 및 길이 예측	전달할 문구(최대 300 자)를 입력 및 수정	POST /text-to- speech/predict- duration 을 호출해 예상 길이를 표시[6]	목적별 템플릿 제공, 길이 초과 경고
5. TTS 생성 및 캐시	텍스트를 선택한 음성으로 생성 요청	캐시 검사 후 POST /text- to-speech 호출; 결과 파일과 메타데이터를 DB 에 저장[3]	동일 요청 재사용, 캐시 히트율 증가
6. 음원 믹싱	BGM 과 SFX 를 선택, 볼륨/타이밍 조절	서버에서 FFmpeg 로 믹싱; 미리듣기 제공; 믹싱 설정과 결과 파일을 DB 에 기록	믹싱 UI 및 프리뷰 제공
7. 검수 및 승인	생성된 음원을 미리듣고 승인/수정	승인 시 다음 단계 진행, 수정 시 이전 단계로 돌아감	승인 상태 관리 및 변경
8. 예약 전송 설정	전송 대상(채널/그룹)과 예약 시간(Asia/Seoul)을 선택	스케줄러에 job 등록; 예약 목록에서 수정/취소 가능	달력·시간 선택기 제공

단계	사용자 행동	시스템 동작	개선 포인트
9. 사용량·크레딧 모니터링	대시보드에서 API 호출량·음성별 사용량·크레딧 잔액 확인	GET /usage, GET /voice-usage, GET /credit-balance 로 데이터를 가져와 시각화[7][5]	경고 알림 및 추세 그래프

■ UI 및 로직 상세 설계

1. 목적 선택 페이지

- 요소:** 라디오 버튼/드롭다운(공지, 축사, 광고, 기타), 선택 설명 도구.
- 데이터 흐름:** 선택된 목적은 세션에 저장 후 DB에 기록. 예약 전송 및 통계 분석시 메타데이터로 활용한다.

2. 음성 탐색/선택

- 클라이언트 필터링:** 첫 로드 시 GET /voices 로 전체 목록을 가져와 언어·성별·스타일·톤 등을 클라이언트에서 필터링한다. 사용자 경험이 빠르다.
- 서버 검색 옵션:** 사용자가 이름이나 키워드 검색을 요청하면 GET /voices/search 를 호출해 이름, 언어, 성별, 연령, 스타일, 감정 등으로 검색한다[1].
- 디자인 판단:** 일반 사용자는 클라이언트 필터로도 충분하나, 음성 라이브러리가 많거나 엔터프라이즈 고객을 위한 옵션으로서 서버 검색 기능을 설정에서 결수 있도록 한다.

3. 보이스 클로닝 UI·스토리지

- 업로드 폼:** 음성 파일 선택(3 MB 이하 WAV/MP3), 클론 음성 이름, 설명 입력. 파일 조건과 보안 정책 안내.
- 업로드·등록 로직:** 프론트엔드가 파일을 서버에 전송 → 백엔드는 스토리지(S3 등)에 저장 후 POST /custom-voices/create-cloned-voice 호출[2].
- 상태 관리:** 응답으로 받은 voice_id 와 등록 상태(대기/처리중/완료/실패)를 DB에 저장하고 UI에서 보여준다. 실패 시 오류 메시지 제공.

4. TTS 생성 및 캐시 로직

- **요청 구성:** 텍스트(300 자 제한), voice_id, 언어, 스타일, 속도(speed), 피치(pitch_shift), 톤(similarity) 등을 JSON 으로 구성하여 POST /text-to-speech 에 전송한다[3].
- **캐시 확인:** DB 에서 동일한 요청(문구+voice_id+설정)의 생성 기록이 있는지 확인. 있으면 저장된 오디오 파일 URL 을 반환한다.
- **새 요청 처리:** 캐시가 없을 경우 API 호출 후 결과 오디오를 스토리지에 저장하고 메타데이터(요청 매개변수, 파일 경로, 생성 시각, 예측 길이)를 DB 에 기록한다.
- **길이 예측:** 합성 전에 POST /predict-duration 을 호출해 예상 시간을 표시하여 사용량과 막싱 길이를 사용자에게 안내한다[6].

5. 음원 믹싱 기능

- **UI:**
- BGM 라이브러리와 SFX 라이브러리 목록.
- 각 트랙별 볼륨 슬라이더 및 시작 시간 조절.
- 미리듣기/미리보기 버튼.
- **서버 처리:** 사용자가 선택한 음성(TTS 결과), BGM, SFX 를 오디오 믹싱 엔진(FFmpeg 등)에 전송해 볼륨과 길이를 조정한 후 믹싱한다.
- **프리뷰:** 혼합 결과를 임시 스트림으로 제공해 사용자가 확인할 수 있도록 한다.
- **메타데이터 저장:** 믹싱에 사용한 재료(음성 파일, BGM ID, SFX ID, 볼륨 값, 타이밍)를 DB 에 저장해 향후 재편집이나 반복 사용을 지원한다.

6. 예약 전송 UI·로직

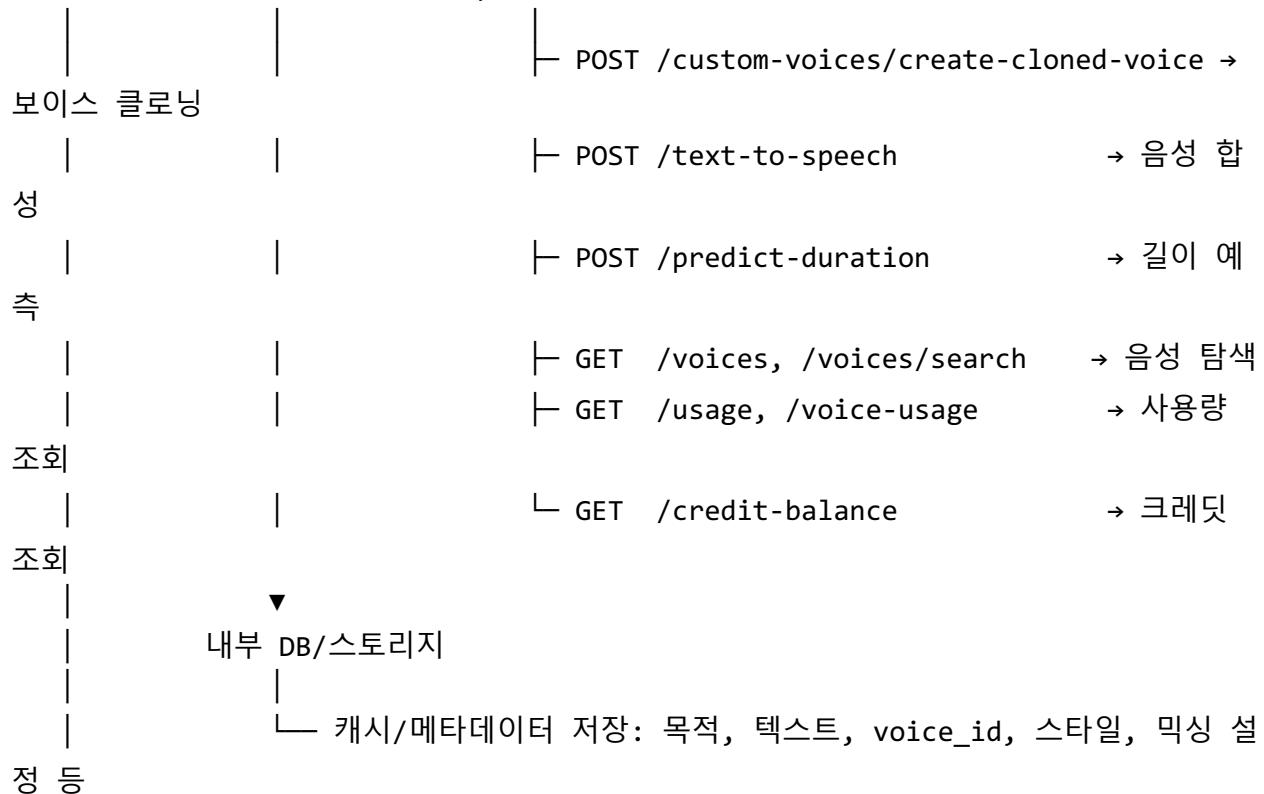
- **UI 요소:** 전송 채널(문자, 전화, 앱 푸시 등), 대상 그룹 선택, 예약 날짜·시간(Asia/Seoul), 반복 옵션, 메시지 미리보기.
- **스케줄링:** 입력 데이터를 DB 에 저장하고, 백엔드 스케줄러(예: Celery Beat, Cron)로 job 을 등록한다. job 에는 목적, 음원 파일 경로, 대상 리스트, 재시도 정책을 포함한다.
- **관리:** 예약 목록 페이지에서 예약 상태(예약됨/진행중/완료/실패)를 확인하고 수정 또는 취소를 할 수 있도록 한다.

7. 사용량·크레딧 대시보드

- **데이터 수집:**
- GET /usage 로 전체 API 사용량을 시간 베kt별로 조회한다[7].
- GET /voice-usage 로 기간별 특정 보이스 사용량을 조회해 인기 음성을 분석한다[5].
- GET /credit-balance 로 잔여 크레딧을 확인하고, 남은 양이 임계값 이하일 때 경고한다[4].
- **시각화:** 월별 또는 주별 사용량 그래프, 음성별 사용량 순위, 크레딧 잔액 및 예측 소진일을 표시한다.
- **알림:** 크레딧이 일정 수준 아래로 떨어지거나 일일 사용량이 급증하면 관리 페이지나 이메일로 경고 메시지를 발송한다.

④ 데이터 흐름 및 시스템 구성

사용자 UI — API 레이어 —> Supertone API



|

↳ 스케줄러(예약 전송) & 오디오 믹싱 엔진

- **UI:** React/Vue 기반 SPA로 구현하며, 각 단계 컴포넌트는 상태 관리 라이브러리로 연결한다.
- **API 레이어:** 백엔드(Node.js/Express 등)가 Supertone API 호출과 DB 저장, 스케줄러 등록, 믹싱 엔진 인터페이스를 담당한다.
- **DB/스토리지:** 텍스트, voice_id, 생성 요청, 결과 파일 경로, 믹싱 파라미터, 목적 정보 등을 저장한다. 캐싱을 위해 적절한 인덱스 설계가 필요하다.
- **스케줄러 및 믹싱 엔진:** 예약 전송을 담당하는 worker 와 FFmpeg 를 사용한 오디오 믹싱 서비스.

17

우선순위·구현 계획

단계	(예시)	기간	
		주요 작업	비고
1. 기반 로직 정비	1~2 주	목적 선택 UI, 세션/DB 저장, TTS 요청 캐시 및 메타데이터 저장 로직 구현; 초기 크레딧 조회 화면 (GET /credit- balance)	기본 UI 와 캐싱 로직 구축
2. 보이스 클로닝 기능	2~3 주	업로드 UI, 스토리지 연동, POST /custom-voices/create-cloned-voice 호출, 처리 상태 표시	기능별 상태 관리 필요
3. 음원 믹싱 및 예약 전송	3~4 주	믹싱 UI(BGM, SFX, 볼륨), 서버 믹싱 엔진 연동; 예약 설정 UI 및 스케줄러 로직 개발	믹싱 후 파일과 메타데이터 저장
4. 사용량·대시보드 고도화	4~5 주	GET /usage, GET /voice-usage 호출로 통계 그래프 및 인기 음성 리스트 구현; 크레딧 경고 시스템 개발; 서버 검색 옵션 UI 추가	운영을 위한 모니터링 강화
5. 추가 개선 (선택)	추후	AI 감정 조정, 다중 음성 대화 합성,	확장 기능

단계	기간 (예시)	주요 작업	비고
		Webhook 기반 전송 결과 콜백, CMS 통합 등	

✓ 결론

이번 설계 개선을 통해 다음과 같은 효과를 기대할 수 있다:

- **목적 선택과 캐싱 로직**을 도입하여 향후 분석 및 예약 전송에서 일관성 있는 데이터를 제공.
- **보이스 클로닝·음원 믹싱·예약 전송 기능** 추가로 사용자 맞춤형 음성 콘텐츠를 보다 쉽게 생성·전달할 수 있으며, **프리뷰와 상태 관리**로 작업 신뢰도를 높인다.
- **사용량 및 크레딧 대시보드**를 구축해 비용 관리와 시스템 운영에 도움이 된다.
- **단계별 구현 우선순위**를 제시하여 개발 자원 계획에 유용하다.

이 설계서를 기반으로 차례대로 개발을 진행하면, 고품질의 통합 음성 합성 서비스를 구현할 수 있다.

[1] Search voices - Supertone API Documentation

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

[2] Create cloned voice - Supertone API Documentation

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

[3] [6] Create speech - Supertone API Documentation

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

[4] Get credit balance - Supertone API Documentation

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

[5] Get voice usage - Supertone API Documentation

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

[7] Get usage - Supertone API Documentation

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