

(외부 Open API) SSML

- 정의 및 정책
- SSML 지원 tag, attribute
 - speak
 - voice
 - prosody
 - break
 - kakao:effect
 - say-as
 - sub
 - audio
- 기타 주의사항
- 사용예 1 : 조사 처리를 하고 싶을 때
- 사용예 2 : 효과음과 같이 음성을 내려보내주고 싶을 때
- 사용예 3 : 새로 나온 영화명, 일반적인 규칙과 다르게 읽어주고 싶을 때
- 사용예 4 : 성경에서는 특이하게 읽는 "10장 10절 → 십장 십절" 쉽게 적용하고 싶을 때

정의 및 정책

- 기본적으로 제공되는 TTS 외에, 원하는대로 **언어 처리를 직접 수정할 수 있도록 제공되는 마크업 랭귀지** 이다.
- 예를 들어, 목소리를 바꾸거나 볼륨을 바꾸는 음성 변경, 상황 별로 특이하게 발음되는 언어 처리, 효과음 합성 등을 할 수 있다.
- w3c에서 제공하는 표준을 지켜 제작되었으며, 표준에 정의되지 않았지만 꼭 필요한 요소는 kakao custom tag, attribute를 통해 제공한다. (<https://www.w3.org/TR/speech-synthesis/#S3.2.3>)

SSML 지원 tag, attribute

- SSML에서 지원할 수 있는 tag와 attribute는 아래와 같다.

*tag명
-attribute명

이탤릭 : possible value

* speak (depth 0 - root)

* voice
- name

* prosody
- rate : *slow, medium, fast*
- volume : *soft, medium, loud*

* break
- time

* kakao:effect
- tone
- friendly

* say-as
- interpret-as
- **spell-out**: 영어를 개별 알파벳으로 읽어줌
- **date**: 날짜 포맷
- **time**: 시간 포맷
- **telephone**: 전화 번호
- **digits**: 숫자를 개별적으로 읽어줌
- **kakao:none**: kakao:effect의 tone 처리 시, 원형 그대로를 보호해주고, 연결된 조사를 자연스럽게 변경해줌.
- **kakao:score**: 스코어를 읽어줌 (ex 3:1 -> 3대1)
- **kakao:number**: 숫자를 특정 방식으로 읽어줌
- **kakao:vocative**: 호격을 지칭함 - ex. <say-as interpret-as="kakao:vocative">영숙</say-as>야 -> "영숙아"

- format
- interpret-as가 date일때
 - d: 일
 - m: 월
 - y: 년
 - 위 값들의 조합 가능 (ex: dmy, my, ymd), ymd의 delimiter는 '-' 또는 '/' 또는 '.'을 사용.
- interpret-as가 time일때
 - h: 시
 - m: 분
 - s: 초
 - z: 시간제 (12 | 24)
 - 위 값들의 조합 가능 (ex: hms12, hm24, ms), hms의 delimiter는 ':'을 사용. z는 delimiter없음
- interpret-as가 kakao:number일때
 - korean: 숫자를 한글로 읽어줌
 - chinese: 숫자를 한자로 읽어줌

- * sub
 - alias

- * audio
 - src : 재생할 오디오 소스 URL
 - clipBegin: 오디오 재생 시작 offset
 - clipEnd : 오디오 재생 끝 offset
 - repeatCount : 몇번 반복할지
 - repeatDur : 재생 시간을 강제로 셋팅. 재생시간보다 clipBegin, clipEnd, repeatCount 값에 의해 정해진 재생보다 값이 큰 경우는 그 시간까지만 재생.
 - soundLevel : 볼륨 조절 (default value : +0dB, maximum range : +/-40dB)
 - speed : 속도 조절 (70~130%까지 사용)

speak

- 기본적으로 모든 음성은 <speak> 태그로 감싸져야 한다
- <speak> 태그 하위로, <speak>를 제외한 모든 tag가 존재할 수 있다.
- 문장, 문단 단위로 적용하는 것을 원칙으로 한다. 한 문장안에서 단어별로 <speak>태그를 감싸지 않는다.

<speak> 안녕하세요. 반가워요. </speak>

voice

- 음성의 목소리를 변경하기 위해 사용하며, name attribute를 통해 원하는 목소리를 지정한다. 제공되는 목소리는 4가지이다.
- 하위로 <speak>, <voice>를 제외한 모든 tag (kakao:effect, prosody, break, audio, say-as, sub)가 존재할 수 있다.
- 문장, 문단 단위로 적용하는 것을 원칙으로 한다. 한 문장안에서 단어별로 <voice>태그를 감싸지 않는다.

| Attribute | Possible Value |
|-----------|--|
| name | <ul style="list-style-type: none"> • WOMAN_READ_CALM : 여성 차분한 낭독체 (default) • MAN_READ_CALM : 남성 차분한 낭독체 • WOMAN_DIALOG_BRIGHT : 여성 밝은 대화체 • MAN_DIALOG_BRIGHT : 남성 밝은 대화체 |

```
<speak>
  <voice name="WOMAN_READ_CALM"> 지금은 여성 차분한 낭독체입니다.</voice>
  <voice name="MAN_READ_CALM"> 지금은 남성 차분한 낭독체입니다.</voice>
  <voice name="WOMAN_DIALOG_BRIGHT"> 안녕하세요. 여성 밝은 대화체예요.</voice>
  <voice name="MAN_DIALOG_BRIGHT"> 안녕하세요. 남성 밝은 대화체예요.</voice>
</speak>
```

prosody

- 음성의 속도와 볼륨을 변경하기 위해 사용하는 tag 이다. 속도와 볼륨 모두 3가지 value를 제공한다.
- 하위로 <say-as>, <prosody>를 제외한 모든 tag (kakao:effect, break, audio, say-as, sub)가 올 수 있다.
- 문장, 문단 단위로 적용하는 것을 원칙으로 한다. 한 문장 안에서 단어별로 <prosody>태그를 감싸지 않는다.

| Attribute | Possible Value |
|-----------|--|
| rate | <ul style="list-style-type: none"> • slow : 0.9 • medium : 1.0 (default) • fast : 1.1 <p>SSML 표준의 경우 최대값 50% (1.5에 해당), 최소값 -33.3% (0.7에 해당)이므로 표준에 부합</p> |
| volume | <ul style="list-style-type: none"> • soft : 0.7 • medium:1.0 (default) • loud : 1.4 <p>SSML 표준의 경우 최대값 +4.08dB (1.6배에 해당), 최소값 정의 없음이므로 표준에 부합</p> |

```
<say-as>
  <prosody rate="fast" volume="loud">안녕하세요. 반가워요.</prosody>
  <prosody rate="slow" volume="soft">안녕하세요. 반가워요.</prosody>
</say-as>
```

break

- 의도적으로 문장과 문장 사이에 "쉼"을 주기 위해 사용한다.
- <say-as>, <prosody> 아래에 올 수 있으며, <break> 하위에 올 수 있는 tag는 없다.
- 문장과 문장 사이에서만 사용한다. 단어와 단어 사이에 사용할 경우, 억양이 이상할 수 있다.
- <break> tag 앞에 이전 텍스트가 없거나, 특수 문자 등으로 이전 텍스트의 합성이 실패하면 <break>가 적용되지 않는다.
- 단위 ms를 붙여야 한다.

| Attribute | Possible Value |
|-----------|---|
| time | 150(0.15초)~1500(1.5초)까지 지원 default : 150 |

```
<say-as>
  첫 번째 문장입니다.
  <break/> 두 번째 문장입니다 옵션을 안준 경우 기본 셋팅 150입니다.
  세 번째 문장입니다
  <break time="150ms"/> 네 번째 문장입니다. 앞이랑 똑같죠?
  <break time="1500ms"/> 다섯 번째 문장입니다. 길게 쉬죠?
  르ㅁ트ㄷ ㅈ르
  <break time="1500ms"/> 여섯 번째 문장입니다. 앞에 문장이 실패해서 쉬지 않습니다.
</say-as>
```

kakao:effect

- 카카오에서 제공하는 custom tag이다.
- tone attribute를 사용하면, 존댓말을 반말로 바꿀 수 있다. 반말을 존댓말로 변경하는 것은 제공하지 않는다.
- speak를 제외한 모든 tag <kakao:effect>, <prosody>, <break>, <audio>, <say-as>, <sub>가 하위에 올 수 있다.
- 하위에 kakao:effect tag가 다시 여러 번 올 수 있다.

| Attribute | Possible Value |
|-----------|---|
| tone | <ul style="list-style-type: none"> • default: 기본 말투, 변경 없음 • friendly: 친구 같은 말투 |

```

<say-as>
  <say-as format="text" tone="friendly">안녕하세요. 반가워요.
  <say-as format="text" tone="default">잘 지내요? 여긴 존댓말 구역이에요.</say-as>
  <say-as format="text" tone="default">저도 잘 지내요.</say-as>
</say-as>

```

say-as

- 날짜/시간과 같은 축약형 표현이나 전화번호, 스펠링과 같은 자주 사용하는 특수 표현을 해석해서, 상황에 맞게 잘 읽어주기 위해 사용한다.
- format은 interpret-as와 함께 쓰이며, interpret-as에서 date와 같은 일부 value를 지정했을 경우 사용한다.
- <say-as> 하위에 올 수 있는 tag는 없다.

| Attribute | Possible Value |
|--------------|---|
| interpret-as | <p>단일 사용 가능</p> <ul style="list-style-type: none"> • spell-out: 영어 단어를 개별 알파벳으로 읽어줌 • digits: 숫자를 개별적으로 읽어줌 (일, 이, 삼..) • telephone: 전화번호로 읽어줌 • kakao:none: kakao:effect의 tone 처리 시, 원형 그대로를 보호해주고, 연결된 조사를 자연스럽게 변경해줌. • kakao:score: 스코어를 읽어줌 (예 3:1 → 3대1) • kakao:vocative: 호격 조사 처리를 하도록 함 <p>format attribute를 반드시 지정해주어야 함</p> <ul style="list-style-type: none"> • date: 날짜로 읽어줌 • time: 시간으로 읽어줌 • kakao:number: 숫자 표현. 한글식(한/하나, 두/둘, 세/셋), 한자식(일, 이, 삼)으로 읽어줌 |
| format | <p>date의 다양한 포맷</p> <ul style="list-style-type: none"> • d: 일 • m: 월 • y: 년 <p>d(일), m(월), y(년)의 다양한 조합 가능 (예 dmy, my, ymd, d, m 등)</p> <p>날짜 표기는 '-', '/', '.' 중 한가지를 사용</p> <p>time의 다양한 포맷</p> <ul style="list-style-type: none"> • h: 시 • m: 분 • s: 초 • 시간제: 12 or 24 <p>hms 시간제의 다양한 조합 가능 (예 hms12, hm24, ms, hm, s 등)</p> <p>시간 표기는 ':' 만 사용 가능</p> <p>시간제를 표시 안할 경우, 오전/오후를 말하지 않고 숫자를 12시간제로 변화하여 읽어줌</p> <p>kakao:number 숫자 포맷, 숫자 뒤의 수량 단위명사를 태그 안에 같이 넣어줘야 함</p> <ul style="list-style-type: none"> • korean: 한/하나, 두/둘, 세/셋 • chinese: 일, 이, 삼 |

단일 사용 가능 value 예시

```

<say-as interpret-as="kakao:effect tone="friendly">
  <say-as interpret-as="spell-out">kakao</say-as>를 스펠링으로 읽어줍니다.
  <say-as interpret-as="digits">1987</say-as> 숫자 낱개로 읽어줍니다.
  <say-as interpret-as="telephone">82-010-3421-6063</say-as> 휴대폰 번호를 잘 읽어줍니다.
  <say-as interpret-as="kakao:none">부처님 오신날</say-as>가 조사 처리 잘 해줍니다.
  <say-as interpret-as="kakao:score">3:1</say-as>로 이겼습니다. 스코어로 인식하고 읽어줍니다.
  <say-as interpret-as="kakao:vocative">영숙</say-as>야.
  <say-as interpret-as="kakao:vocative">톤</say-as>야.
</kakao:effect>
</say-as>

```

format 사용 예시

```

<say-as interpret-as="date" format="dmy">10.6.85</say-as> 입니다.
<say-as interpret-as="date" format="my">10.1985</say-as> 입니다.
<say-as interpret-as="date" format="md">10-6</say-as> 입니다.
<say-as interpret-as="date" format="d">10</say-as> 입니다.
<say-as interpret-as="time" format="hms12">13:16:45</say-as> 입니다.
<say-as interpret-as="time" format="hm24">13:50</say-as> 입니다.
<say-as interpret-as="time" format="ms">10:59</say-as> 입니다.
<say-as interpret-as="time" format="h">13</say-as> 입니다.
<say-as interpret-as="kakao:number" format="chinese">13장</say-as> 입니다.
<say-as interpret-as="kakao:number" format="korean">13장</say-as> 입니다.
</say-as>

```

sub

- 일반적인 발음이 아니라 상황에 맞도록 특별하게 발음을 지정해야 할 경우, 사용한다.
- value 입력시, 가능한한 한글로 작성해야 해석이 용이하다.
- 하위에 올 수 있는 tag는 없다.

| Attribute | Possible Value |
|-----------|------------------------|
| alias | 읽어주기를 원하는 한글 독음을 직접 입력 |

```

<say-as><sub alias="알루미늄">Al</sub>은 단단하다.</say-as>

```

audio

- 외부 음원을 speak 채널을 통해 제공하기 위하여 사용한다.
- 효과음 등을 음성과 같이 제공하기에 용이하다.
- 음원의 시작/종료지점, 반복회수 등을 지정할 수 있다.
- <speak>, <prosody> 하위에 올 수 있다.

| Attribute | Possible Value |
|-----------|----------------|
|-----------|----------------|

| | |
|-------------|---|
| clipBegin | 음원의 시작 시간 세팅, 초 단위로 3분부터 시작하고 싶을 경우 180s로 설정 소수점 첫 번째 자리까지 가능 음원의 길이 보다 큰 값을 주면 무시됨 clipBegin 이 clipEnd 보다 크면 음원이 만들어지지 않음 단위로 s를 붙힐 수도 있고, 떨 수도 있음 clipBegin과 clipEnd를 쌍으로 쓸 필요 없음 |
| clipEnd | 음원의 끝 시간, 초 단위로 세팅 소수점 첫 번째 자리까지 가능 음원의 길이 보다 큰 값을 주면 무시됨 단위로 s를 붙힐 수도 있고, 떨 수도 있음 clipBegin과 clipEnd를 쌍으로 쓸 필요 없음 |
| repeatCount | 반복 횟수 (자연수, 단위 없음) clipBegin과 clipEnd가 적용된 후 반복됨 |
| repeatDur | 반복 시간, 초 단위로 세팅 소수점 첫 번째 자리까지 가능 음원의 시작과 끝을 기준으로 반복하며, clipBegin과 clipEnd가 설정되어 있을 경우, 구간 반복함 재생시간보다 clipBegin, clipEnd, repeatCount 값에 의해 정해진 재생보다 값이 큰 경우는 그 시간까지만 재생 repeatDur와 repeatCount가 둘 다 설정되어 있을 경우, repeatCount를 무시하고 repeatDur를 기준으로 동작함 단위로 s를 붙힐 수도 있고, 떨 수도 있음 |
| soundLevel | 볼륨 조절 (+- n dB) 소수점 첫째자리까지 가능 값을 너무 높게 주거나 낮게 주면, 클리핑이 발생할 수 있음 단위로 dB를 붙힐 수도 있고, 떨 수도 있음 |
| src | url |
| speed | 속도 조절 (%) 소수점 첫째자리까지 가능 단위 없어도 가능 70~130%까지만 사용 권장 |

```
<speaK>
  <audio src="http://www.noiseaddicts.com/samples_1w72b820/274.mp3" clipBegin="1.1"
clipEnd="3s" repeatDur="10s" soundLevel="3dB" speed="70%"/>
</speaK>
```

기타 주의할점

- escape 처리가 필요하다.

사용예 1 : 조사 처리를 하고 싶을 때

- 문장 조사처리가 필요할경우, <say-as interpret-as="kakao:none">을 사용한다.

- 조사나 어미 처리가 필요한 부분을 <say-as interpret-as="kakao:none">으로 감싼다.

```
< speak >
  오늘은 < say-as interpret-as="kakao:none">부처님</ say-as >가 태어난 날로, < say-as interpret-
as="kakao:none">부처님 오신날</ say-as >예요.
</ speak >
```

사용예 2 : 효과음과 같이 음성을 내려보내주고 싶을 때

- 효과음과 음성을 같이 내려줄 경우, <audio>를 사용한다.

```
< speak >
  < audio src="http://www.noiseaddicts.com/samples_1w72b820/274.mp3" clipBegin="1.1"
clipEnd="3s"/> 안녕하세요?
</ speak >
```

사용예 3 : 새로 나온 영화명, 일반적인 규칙과 다르게 읽어주고 싶을 때

- 영화 1987의 경우, 기존의 "천구백팔십칠"이 아니라 "일구팔칠"로 읽어야 한다. 이처럼, 기존의 읽는 방식과 다르게 읽어줘야 할때 <say-as>나 <sub> 태그를 사용한다
- <say-as>에서 정해진 규칙으로 변경이 불가능할 경우, <sub>를 활용한다.

say-as를 활용할 경우

```
< speak >
  영화 < say-as interpret-as="digits">1987</ say-as >은, 흥행 1위입니다.
</ speak >
```

sub를 활용할 경우

```
< speak >
  영화 < sub alias="일구팔칠">1987</ sub >은, 흥행 1위입니다.
</ speak >
```

사용예 4 : 성경에서는 특이하게 읽는 "10장 10절 → 십장 십절" 쉽게 적용하고 싶을때

- 기본 룰은 "10장"을 "열장"으로 읽는 것이다. 그러나 성경에서는 "십장"으로 읽어야 한다. 이 경우, <say-as>를 활용하여 숫자 읽는 규칙을 정의할 수 있다.

<speaK>

 마태복음 <say-as interpret-as="kakao:number" format="chinese">10</say-as>장 <say-as
interpret-as="kakao:number" format="chinese">10</say-as>절 말씀입니다.

</speaK>