



BEST PRACTICES > Prompting

Controls

Learn how to control delivery, pronunciation & emotion of text to speech.

 We are actively working on *Director's Mode* to give you even greater control over outputs.

This guide provides techniques to enhance text-to-speech outputs using ElevenLabs models. Experiment with these methods to discover what works best for your needs. These techniques provide a practical way to achieve nuanced results until advanced features like *Director's Mode* are rolled out.

Pauses

Use `<break time="x.xs" />` for natural pauses up to 3 seconds.

 Using too many break tags in a single generation can cause instability. The AI might speed up, or introduce additional noises or audio artifacts. We are working on resolving this.

Example



"Hold on, let me think." `<break time="1.5s" />` "Alright, I've got it."

- **Consistency:** Use `<break>` tags consistently to maintain natural speech flow. Excessive use can lead to instability.
- **Voice-Specific Behavior:** Different voices may handle pauses differently, especially those trained with filler sounds like "uh" or "um".



Need help?

Voice chat

Powered by ElevenLabs [Conversational AI](#)

ElevenLabs

Example



"It... well, it might work." "Wait – what's that noise?"



Pronunciation

Phoneme Tags

Specify pronunciation using [SSML phoneme tags](#). Supported alphabets include [CMU](#), Arpabet and the [International Phonetic Alphabet \(IPA\)](#).



Phoneme tags are only compatible with "Eleven Flash v2", "Eleven Turbo v2" and "Eleven English v1" [models](#).

CMU Arpabet Example

IPA Example



```
1 <phoneme alphabet="cmu-arpabet" ph="M AE1 D IH0 S AH0 N">
2   Madison
3 </phoneme>
```

We recommend using CMU Arpabet for consistent and predictable results with current AI models. While IPA can be effective, CMU Arpabet generally offers more reliable performance.

Phoneme tags only work for individual words. If for example you have a name with a first and last name that you want to be pronounced a certain way, you will need to create a phoneme tag for each word.

Ensure correct stress marking for multi-syllable

Need help?

For example:

Correct usage

Incorrect usage

Powered by ElevenLabs [Conversational AI](#)



ElevenLabs

3 </phoneme>



Alias Tags

For models that don't support phoneme tags, you can try writing words more phonetically. You can also employ various tricks such as capital letters, dashes, apostrophes, or even single quotation marks around a single letter or letters.

As an example, a word like "trapezii" could be spelt "trapezli" to put more emphasis on the "ii" of the word.

You can either replace the word directly in your text, or if you want to specify pronunciation using other words or phrases when using a pronunciation dictionary, you can use alias tags for this. This can be useful if you're generating using Multilingual v2 or Turbo v2.5, which don't support phoneme tags. You can use pronunciation dictionaries with Studio, Dubbing Studio and Speech Synthesis via the API.

For example, if your text includes a name that has an unusual pronunciation that the AI might struggle with, you could use an alias tag to specify how you would like it to be pronounced:

```
<lexeme>
  <grapheme>Claughton</grapheme>
  <alias>Cloffton</alias>
</lexeme>
```

If you want to make sure that an acronym is always delivered in a certain way whenever it is encountered in your text, you can use an alias tag to specify this:

```
<lexeme>
  <grapheme>UNK</grapheme>
  <alias>United Nations</alias>
</lexeme>
```

Need help?

ElevenLabs

Some of our tools, such as Studio and Dubbing Studio, allow you to create and upload a pronunciation dictionary. These allow you to specify the pronunciation of certain words, such as character or brand names, or to specify how acronyms should be read.



Pronunciation dictionaries allow this functionality by enabling you to upload a lexicon or dictionary file that specifies pairs of words and how they should be pronounced, either using a phonetic alphabet or word substitutions.

Whenever one of these words is encountered in a project, the AI model will pronounce the word using the specified replacement.

To provide a pronunciation dictionary file, open the settings for a project and upload a file in either TXT or the [.PLS format](#). When a dictionary is added to a project it will automatically recalculate which pieces of the project will need to be re-converted using the new dictionary file and mark these as unconverted.

Currently we only support pronunciation dictionaries that specify replacements using phoneme or alias tags.

Both phonemes and aliases are sets of rules that specify a word or phrase they are looking for, referred to as a grapheme, and what it will be replaced with. Please note that searches are case sensitive. When checking for a replacement word in a pronunciation dictionary, the dictionary is checked from start to end and only the very first replacement is used.

Pronunciation Dictionary examples

Here are examples of pronunciation dictionaries in both CMU Arpabet and IPA, including a phoneme to specify the pronunciation of "Apple" and an alias to replace "UN" with "United Nations":

CMU Arpabet Example

```
1  <?xml version="1.0" encoding="UTF-8"?
2  <lexicon version="1.0"
3      xmlns="http://www.w3.org/2005/01/xml-namespace-instance"
4      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

IPA Example

Need help?

ElevenLabs

```
7      alphabet="cmu-arpabet" xml:lang="en-GB">
8      <lexeme>
9          <grapheme>apple</grapheme>
10         <phoneme>AE P AH L</phoneme>
11     </lexeme>
12     <lexeme>
13         <grapheme>UN</grapheme>
14         <alias>United Nations</alias>
15     </lexeme>
16 </lexicon>
```



To generate a pronunciation dictionary `.pls` file, there are a few open source tools available:

- [Sequitur G2P](#) - Open-source tool that learns pronunciation rules from data and can generate phonetic transcriptions.
- [Phonetisaurus](#) - Open-source G2P system trained on existing dictionaries like CMUdict.
- [eSpeak](#) - Speech synthesizer that can generate phoneme transcriptions from text.
- [CMU Pronouncing Dictionary](#) - A pre-built English dictionary with phonetic transcriptions.

Emotion

Convey emotions through narrative context or explicit dialogue tags. This approach helps the AI understand the tone and emotion to emulate.

Example



You're leaving?" she asked, her voice trembling with sadness. "That's it!" he exc]

Need help?

Explicit dialogue tags yield more predictable results, but they can also be removed in post-production using an audio editor if unwanted.

Generated by ElevenLabs [Conversational AI](#)

ElevenLabs

The pacing of the audio is highly influenced by the audio used to create the voice. When creating your voice, we recommend using longer, continuous samples to avoid pacing issues like unnaturally fast speech.



For control over the speed of the generated audio, you can use the speed setting. This allows you to either speed up or slow down the speed of the generated speech. The speed setting is available in Text to Speech via the website and API, as well as in Studio and Conversational AI. It can be found in the voice settings.

The default value is 1.0, which means that the speed is not adjusted. Values below 1.0 will slow the voice down, to a minimum of 0.7. Values above 1.0 will speed up the voice, to a maximum of 1.2. Extreme values may affect the quality of the generated speech.

Pacing can also be controlled by writing in a natural, narrative style.

Example



"I... I thought you'd understand," he said, his voice slowing with disappointment.

Tips

- > Common Issues
- > Tips for Improving Output

Creative control

While we are actively developing a "Director's over outputs, here are some interim techniques

Need help?



Narrative styling

Powered by ElevenLabs [Conversational AI](#)

ElevenLabs

2 Layered outputs

Generate sound effects or speech in segments and layer them together using audio editing software for more complex compositions.



3 Phonetic experimentation

If pronunciation isn't perfect, experiment with alternate spellings or phonetic approximations to achieve desired results.

4 Manual adjustments

Combine individual sound effects manually in post-production for sequences that require precise timing.

5 Feedback iteration

Iterate on results by tweaking descriptions, tags, or emotional cues.

Was this page helpful?

Yes

No

Edit this page

< Previous

Normalization

Learn how to normalize text for Text to Speech.

Next >

Built with fern

Need help?