Speech → Magic Voice

User manual

Turn recorded speech into vocabularies for the Commodore 64 Magic Voice speech cartridge (Toshiba T6721A), ready to load with VOC() and play with SAY.

The program is the single file **c64-lpc-encoder.html**. Open it in a normal web browser. It works completely offline: nothing is uploaded anywhere.

Contents

1 · What the program does 3

2 · The screen 4

3 · Importing and selecting a word 4

4 · Converting and listening 5

5 · Settings 6

6 · Adding words and naming 6

7 · The vocabulary gallery 7

8 · Saving and opening projects 8

9 · Export and use on the C64 9

10 · Appearance 9

11 · Tips and troubleshooting 10

1 · What the program does

The Magic Voice cartridge speaks from *vocabularies*: blocks of speech data in memory that you point the cartridge to with **VOC(address)** and then play with **SAY**. This program creates those vocabularies. You import an audio recording, mark the part that contains one word or phrase, convert it to Magic Voice data, listen to how the chip will pronounce it, give it a name and add it to a vocabulary. When all your words are in, you download the result as a disk image (.d64) or as program files (.prg, .bin) and use it on a C64 or an emulator.

The conversion produces exactly the data the T6721A speech chip expects: 50 frames per second, 160 samples per frame at 8 kHz, each frame either voiced, unvoiced or silent. The preview you hear in the browser is a model of that chip, so what you hear is what the chip is told to say.

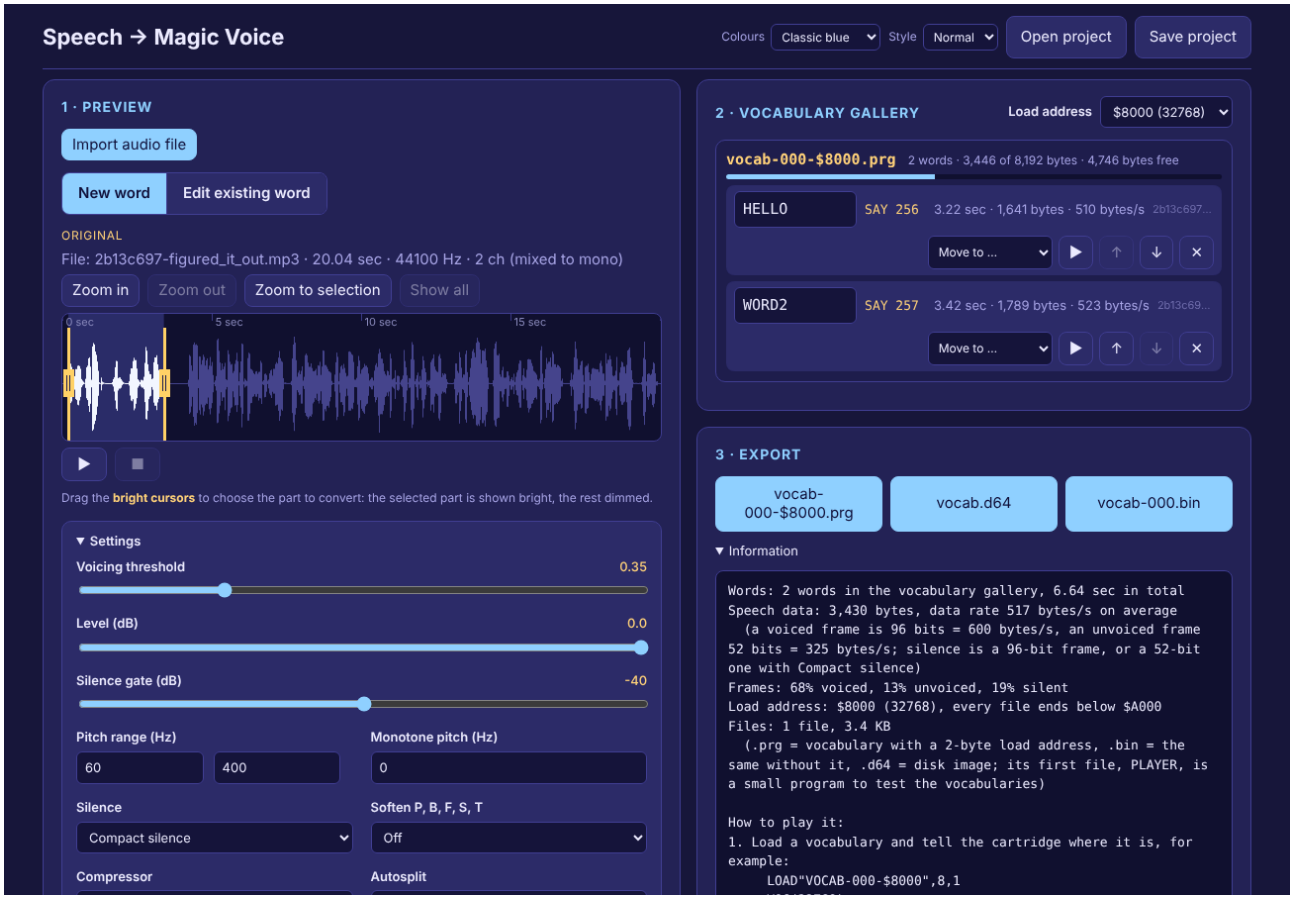
The program has three numbered areas, which you work through in order:

- **1 · Preview** – import a recording, choose the part to convert, set the options, convert, listen.
- **2 · Vocabulary gallery** – the words you have added, grouped in vocabularies, with the load address.
- **3 · Export** – download the files and see how to play them on the C64.

1.1 Quick start

1. Click **Import audio file** (or drop a file on the waveform). WAV, MP3, OGG, FLAC and M4A all work.
2. Drag the two bright cursors so that only the word you want is selected. Use ▶ to check.
3. Click **Convert**. Listen to the result with the ▶ under the Converted graph.
4. Type a name in **Word name** and click **Add to vocabulary gallery**.
5. Repeat for every word. Pick a different part of the same recording, or import another file.
6. In area 3, download **vocab.d64** (or the .prg / .bin files).
7. On the C64: load the vocabulary, type `VOC(32768)`, then `SAY 256`.

2 · The screen



The main screen with two words in the gallery. Left: the preview and settings. Right: the vocabulary gallery and the export.

At the top right are the **Colours** and **Style** selectors (section 10) and the **Open project** and **Save project** buttons (section 8). Most buttons show a short explanation when you hold the mouse pointer over them.

3 · Importing and selecting a word

3.1 Import

Click **Import audio file** or drop a file onto the waveform. Stereo files are mixed to mono. The file name, length, sample rate and channel count are shown above the waveform. You can import another file at any time; words already in the gallery keep their own copy of their audio, so they are not affected.

3.2 Choosing the part to convert

The waveform shows the whole recording. The two **bright cursors** mark the start and the end of the part that will be converted: the selected part is drawn bright and the rest is dimmed. Drag the cursors with the mouse. The ▶ button below the waveform plays only the selected part (it pauses and resumes on a second click), and ■ stops.

For precise cutting, zoom in:

- **Zoom in / Zoom out** – step the zoom. Ctrl + mouse wheel zooms around the pointer.
- **Zoom to selection** – fills the view with the selected part for fine adjustment of the cursors.
- **Show all** – back to the whole recording.

- When zoomed, a scroll bar appears. You can also drag the empty background of the waveform, or use Shift + wheel, to scroll.

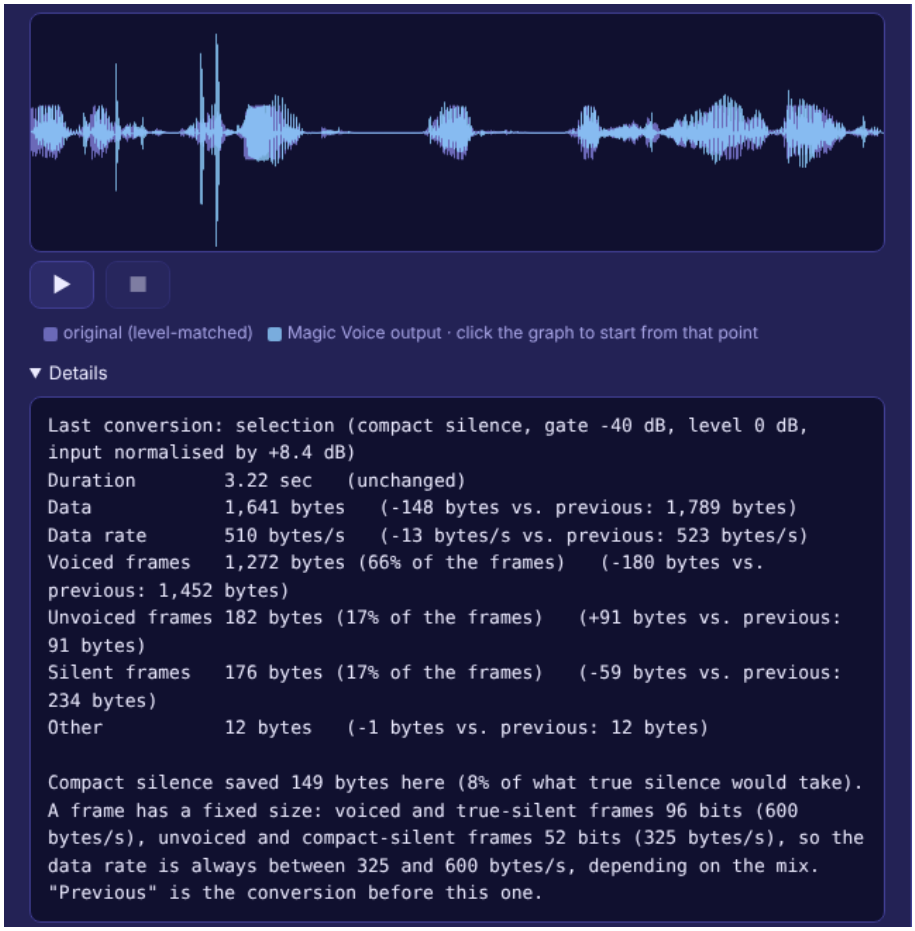
Tip: cut just inside the silence around the word. A little silence at each end is fine; the converter treats pauses efficiently (see *Silence* in section 5).

3.3 New word or existing word

The switch above the waveform has two modes. **New word** (the normal mode) creates a new word from the current selection. **Edit existing word** lets you change a word that is already in the gallery; you can also enter this mode by clicking a word in the gallery. See section 7.

4 · Converting and listening

Click **Convert**. A progress bar shows the work. When it is ready, the **Converted** graph shows your original (dark, level-matched) behind the Magic Voice output (light). Press ▶ under the graph to listen to the chip model; click inside the graph to start playing from that point.



The Converted graph and the Details box after a conversion.

The **Details** box below the graph gives the duration, the number of bytes, the data rate and how many frames are voiced, unvoiced and silent, together with the difference from the previous conversion. This is useful for comparing settings: change one option, convert again and see what it did to the sound and to the size.

Convert also refreshes every word in the gallery whose settings are out of date (marked *out of date* in the list).

4.1 Loudness

The original always plays at the same loudness, before and after converting, so you can compare it fairly with the Magic Voice version. Before conversion the selection is normalised to full scale so the 7-bit energy range of the chip is used completely.

5 · Settings

The Settings box belongs to each word: a word keeps the settings it was converted with. Changing a setting only affects the next word you convert, or the word you are editing (**Autosplit** is the one exception). **Reset to defaults** puts every setting back to its default.

Voicing threshold	Default 0.35. Decides whether a stretch of sound is voiced (buzzy, like vowels) or unvoiced (hissy, like S). Lower gives more buzzy frames, higher more hissy frames. Change it only if vowels sound whispery or hisses sound buzzy.
Level (dB)	Default 0. Lowers the output level from full scale (down to −12 dB). The selection is always normalised before conversion, so this is the way to make a word quieter.
Silence gate (dB)	Default −40. Frames this far below the loud parts of the speech become silent frames. Below about −42 dB the chip's lowest energy step is silence anyway, so lower values change little.
Pitch range (Hz)	Default 60 to 400. The lowest and highest voice pitch the analysis looks for. Narrow it for a steady voice (for example 80–250 for a typical male voice) if the pitch jumps.
Monotone pitch (Hz)	Default 0 (natural pitch). Any other value fixes the pitch to that frequency and gives the classic robot voice.
Silence	**Compact silence** (default) stores pauses as the cheapest frame (52 bits, 325 bytes/s) at the lowest energy: a hiss about 50 dB below full scale, practically inaudible, and about 45% less data in the pauses. **True silence** uses real silent frames (96 bits, 600 bytes/s): no hiss but more data.
Soften P, B, F, S, T	Off, Mild or Strong. Bursts and hissing sounds can come out of the synthesiser louder and sharper than in the recording. This option checks every such frame against the recording, turns it down where it is too loud, drops frames that are only smear and caps the hiss relative to the vowel level. Mild corrects only what is clearly too much; Strong does more.
Compressor	Off, Mild, Medium (default) or Strong. Makes soft passages louder and loud passages softer, then scales the whole recording to the maximum level, so the chip's 7-bit energy range is used as well as possible (the best signal to noise ratio). Pauses and background noise are not amplified. Ratios are about 1.6:1, 2.5:1 and 4:1.
Autosplit	Disabled (default) or Enabled. See section 6.

Suggested approach: start with the defaults. If a word sounds muffled or uneven, try the Compressor one step stronger. If P, T, S or F sound too harsh or smeared, try Soften. Compare each version with the ▶ buttons and the size in Details.

6 · Adding words and naming

After a conversion you add the word to the gallery:

- Type a name in **Word name**: letters A–Z and digits 0–9, up to 16 characters, unique in the gallery. The name is what you will type in `SAY"NAME"` on the C64. **Suggest name** fills in the next free name (WORD1, WORD2, …).

- Click **Add to vocabulary gallery**. The word appears in area 2 with its word number, for example SAY 256.

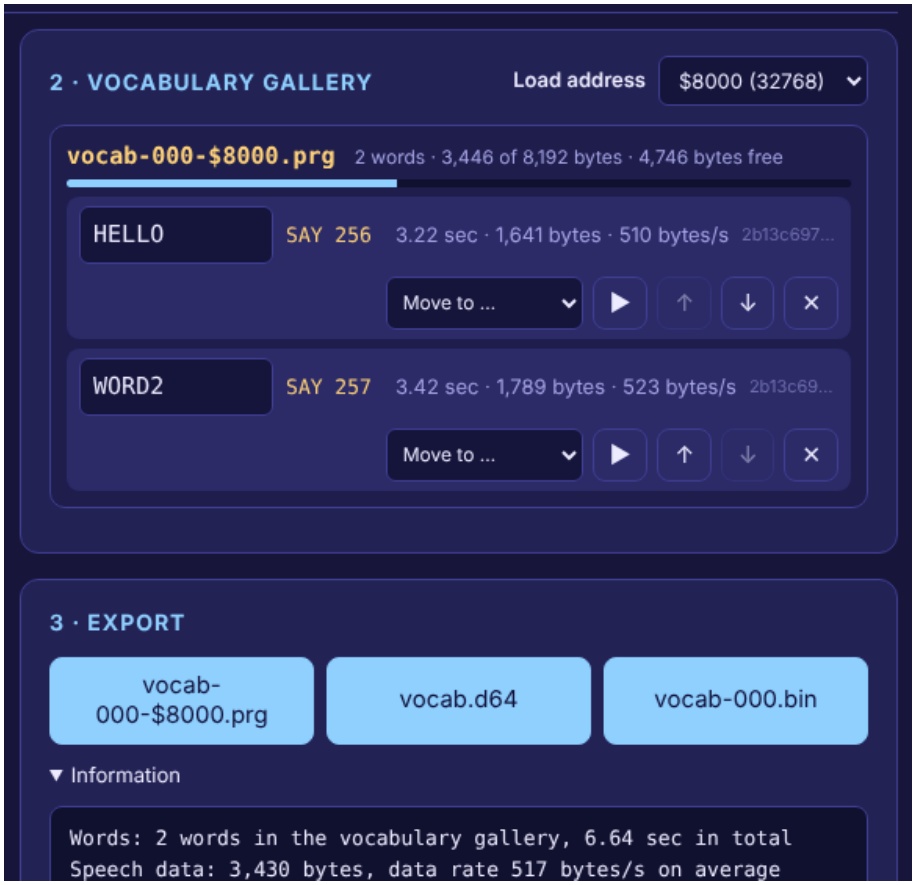
6.1 Long selections and Autosplit

One vocabulary has limited memory (see section 7). If the selection is too long for it, there are two outcomes:

- **Autosplit disabled:** the word is flagged as too long, the part that does not fit is shown in red, and the word is left out of the export. Shorten the selection or lower the load address.
- **Autosplit enabled:** the selection is cut at pauses into several words and spread over as many vocabularies as needed. The pieces are named with a dash and a number: a name such as WORD1 becomes WORD1-001, WORD1-002, and so on. The cuts are shown in the converted graph.

The dash makes the pieces easy to recognise and avoids a clash with a word that you named WORD1 yourself.

7 · The vocabulary gallery



The vocabulary gallery with two words, the load address selector and the memory bar.

Each box in the gallery is one vocabulary and becomes one file. Its header shows the file name (for example `vocab-000-$8000.prg`), the number of words, and how many bytes of the available memory are used. A bar shows the fill level and turns red when the vocabulary is too full. Word numbers start at **256** in every vocabulary; numbers below 256 are the cartridge's own built-in words.

7.1 Load address

The **Load address** selector (top right of the gallery, in steps of $100) is the memory address where the vocabulary will be loaded and the number you give to `VOC()`. The default is $8000 (32768). Every file is cut so that it ends below $A000, where the cartridge ROM starts, so a lower address gives more room per vocabulary and a higher address gives less.

7.2 Working with words

Name field	Click in the name to rename the word (letters and digits only).
SAY number	The word number for SAY, valid once the vocabulary is loaded and VOC() has been called. Moving or reordering words changes the numbers.
Information	Length in seconds, size in bytes, data rate and the source file with the converted part. *Out of date* means the settings have changed since the word was converted.
Move to …	Moves the word to another vocabulary, or to a new one. Words can be grouped in any way you like.
▶	Plays this word through the Magic Voice model.
↑ ↓	Moves the word up or down within its vocabulary (changes the word numbers).
×	Removes the word from the gallery.
Click a word	Selects it for editing (see 7.3).

7.3 Editing a word

Click a word in the gallery, or choose **Edit existing word** and pick one. Its recording, cursors and settings are loaded. Change the cursors or settings, press **Convert** to hear the result, and then use **Save changes** to store them in that word. **Add to vocabulary gallery** stores a *copy* instead, which must have a different name.

8 · Saving and opening projects

Save project writes a file called **S2MV_project.json** containing the recordings, cut points and settings of all words, so you can continue later. **Open project** loads such a file again and rebuilds the gallery.

- Save your project regularly, so you can always continue from the file.
- Only project files saved by the current version are accepted. A word whose settings are incomplete is skipped and listed by name.
- The project is separate from the exported vocabulary files: export the .d64/.prg/.bin files to use the words on a C64.

9 · Export and use on the C64

9.1 Files

Area 3 offers the downloads. With a single vocabulary file you get three buttons; with several you get zip files and the disk images.

.prg	The vocabulary with a 2-byte load address in front. Named like `vocab-000-$8000.prg`, where the end of the name is the load address. Load it with LOAD"NAME",8,1.
.bin	The same data without the load address: the raw bytes as they sit at the VOC() address.
.d64	A disk image containing the vocabulary files and, as the first file, the small test program **PLAYER**. If the vocabularies do not fit on one disk, several images are made.
.zip	With more than one vocabulary: all .prg, .bin or .d64 files in one download.

Below the buttons, the **Information** box lists the number of words and files, total sizes, the data rate and the share of voiced, unvoiced and silent frames. It also shows the exact commands for your current vocabulary.

9.2 Playing a word

Load the vocabulary into memory, tell the cartridge where it is with **VOC**, then play a word with **SAY**, either by number or by name:

```
LOAD"VOCAB-000-$8000",8,1
VOC(32768)
SAY 256
SAY"HELLO"
```

- Do **not** RUN a vocabulary file and do not autostart it. It is data, not a program.
- The number in `VOC()` is the load address you chose, in decimal ($8000 = 32768).
- For another vocabulary, load that file and call `VOC()` again. The numbering starts at 256 in each file.
- SAY, VOC and RATE are Magic Voice keywords. In a BASIC program on disk they are stored as the cartridge's own tokens, which is how the PLAYER program uses them.

9.3 The PLAYER test program

To try a vocabulary without typing, attach the .d64 and enter:

```
LOAD"PLAYER",8
RUN
```

The menu lists the vocabularies on the disk and offers two options:

- **1** – play all words of one vocabulary (you are asked for its number when there are several).
- **2** – play the first word of every vocabulary, loading each in turn.

After a vocabulary has been loaded, the program prints PRESS SPACE and waits for the space bar before it goes on. The words are played in order and the program prints which word is playing.

10 · Appearance

The two selectors at the top right change the look of the program. Both choices are remembered in your browser.

- **Colours:** Classic blue (default), CRT green, CRT amber and Black & white.

- **Style:** Normal, or CRT, which adds a monospace font, a soft glow on the text, square corners and faint scanlines for an 80s look. Any colour theme can be combined with either style.

The waveform graphs follow the chosen colours.

11 · Tips and troubleshooting

The word is "too long for one vocabulary"	One vocabulary holds roughly 14 seconds of speech at the worst case ($8000 start). Shorten the selection, lower the load address, spread the words over more vocabularies, or enable Autosplit.
Word sounds too quiet or uneven	Use the Compressor (Medium or Strong) and keep Level at 0 dB.
P, T, S, F sound harsh or smeared	Set Soften P, B, F, S, T to Mild or Strong.
Pauses hiss	Switch Silence to True silence (costs more data): it uses real silent frames, so there is no hiss.
Voice warbles or jumps in pitch	Narrow the pitch range, or set a Monotone pitch for a robot voice.
SAY"NAME" says nothing	Check that the vocabulary is loaded at the address you gave to VOC(), that VOC() was called after loading, and that the name matches the one in the gallery.
A word in an imported project is missing	Its settings were incomplete; the program lists such words when you open the project. Convert the word again from the original recording.

Sound quality note: the Magic Voice chip is a very low-bitrate synthesiser, so speech will always sound clearly synthetic. Short, clear, slowly spoken words recorded without background noise give the best results.