

Sega ASR1600/C V2 Memory usage

1. Sample memory usage for a speech application

The memory usage is divided into a read-only and a read/write section to indicate that the memory of the first section does not have write accesses and therefor could be put in fast ROM. However on the Dreamcast platform all memory resides in the main RAM.

Read-only memory

- Code
 - ✓ Program code 107kb ¹
 - ✓ Constant data 27kb
- Speech data
 - ✓ Language model (4bits american english) 533kb
 - ✓ Context data (100 isolated words) 4kb

Read/write memory

- Dynamic heap memory

Heap usage varies according to engine setting. For the default parameters and a context of 100 words we get: 152kb

Total:	823kb
--------	-------

¹ Number is given for an average application, linker will only include the modules which are used.

2. Detailed memory information

2.1 Language models

American English

V1.0.1

	Fixed	Float
4bits	asr16v2_aev101_fixed_4b.lng 533 kb	asr16v2_aev101_float_4b.lng 643 kb
8bits	asr16v2_aev101_fixed_8b.lng 838 kb	asr16v2_aev101_float_8b.lng 947 kb

Japanese

V1.1.0 Sega model

	Fixed	Float
4bits	asr16v2_jpv110_fixed_4b.lng 494 kb	asr16v2_jpv110_float_4b.lng 604 kb
8bits	asr16v2_jpv110_fixed_8b.lng 799 kb	asr16v2_jpv110_float_4b.lng 908 kb

2.2 Contexts

The context filesize is approximately **270+36.N Words**

It depends ofcourse on how big the transcription² of that word is.

Example context sizes:

- 20 isolated words 1k ctx file
- 100 isolated words 4k ctx file

2.3 Heap memory

For the floating point version we typically have around 150k of maximum heap memory usage³. In a total of about 550 allocations (average allocation size is 273)

Note: This does not count the loss of heap memory due to overhead or fragmentation.

The memory usage can rise quickly if the context branching factor is big and a high accuracy is set.

The fixed point typically has 190k of maximum heap memory usage or 40k more than the floating point version.

2.4 Userwords

To store a userword (without the recorded speech) it only takes the size of that word in speech units.

This is typically around 40 bytes for an isolated word. In general, the longer the utterance the bigger the userword.

² L&H phonetic representation of a word

³ This is with default asr parameters.