

Reference of HMM-based Speech Synthesis Engine

“hts_engine API” version 1.10

HTS Working Group
December 25, 2015

1 Engine structures

1.1 Audio

HTS_Audio Audio output wrapper.

size_t	sampling_frequency	- sampling frequency
size_t	max_buff_size	- buffer size of audio output device
short	*buff	- current buffer
size_t	buff_size	- current buffer size
void	*audio_interface	- audio interface specified in compile step

1.2 Model

HTS_Window Window coefficients to calculate dynamic features.

size_t	size	- # of windows (static + deltas)
int	*l_width	- left width of windows
int	*r_width	- right width of windows
double	**coefficient	- window coefficients
size_t	max_width	- maximum width of windows

HTS_Pattern List of patterns in a question and a tree.

char	*string	- pattern string
HTS_Pattern	*next	- pointer to the next pattern

HTS_Question List of questions in a tree.

char	*string	- name of this question
HTS_Pattern	*head	- pointer to the head of pattern list
HTS_Question	*next	- pointer to the next question

HTS_Node List of tree nodes in a tree.

int	index	- index of this node
size_t	pdf	- index of PDF for this node (leaf node only)
HTS_Node	*yes	- pointer to its child node (yes)
HTS_Node	*no	- pointer to its child node (no)
HTS_Node	*next	- pointer to the next node
HTS_Question	*quest	- question applied at this node

HTS_Tree List of decision trees in a model.

HTS_Pattern	*head	- pointer to the head of pattern list for this tree
HTS_Tree	*next	- pointer to the next tree
HTS_Node	*root	- root node of this tree
size_t	state	- state index of this tree

HTS_Model Set of PDFs, decision trees and questions.

size_t	vector_length	- vector length (static features only)
size_t	num_windows	- # of windows for delta
HTS_Boolean	is_msd	- flag for MSD
size_t	ntree	- # of trees
size_t	*npdf	- # of PDFs at each tree
float	***pdf	- PDFs
HTS_Tree	*tree	- pointer to the list of trees
HTS_Question	*question	- pointer to the list of questions

HTS_ModelSet Set of duration models, HMMs and GV models.

char	*hts_voice_version	- version of HTS voice format
size_t	sampling_frequency	- sampling frequency
size_t	frame_period	- frame period
size_t	num_voices	- # of HTS voices
size_t	num_states	- # of HMM states
size_t	num_streams	- # of streams
char	*stream_type	- stream type
char	*fullcontext_format	- fullcontext label format
char	*fullcontext_version	- version of fullcontext label
HTS_Question	*gv_off_context	- GV switch
char	**option	- options for each stream
HTS_Model	*duration	- duration PDFs and trees
HTS_Window	*window	- window coefficients for delta
HTS_Model	**stream	- parameter PDFs and trees
HTS_Model	**gv	- GV PDFs and trees

1.3 Label

HTS_LabelString Individual label string with time information.

HTS_LabelString	*next	- pointer to the next label string
	char *name	- label string
	double start	- start frame specified in the given label
	double end	- end frame specified in the given label

HTS_Label List of label strings.

HTS_LabelString	*head	- pointer to the head of label string
	size_t size	- # of label strings

1.4 State stream

HTS_SStream Individual state stream.

size_t	vector_length	- vector length (static features only)
double	**mean	- mean vector sequence
double	**vari	- variance vector sequence
double	*msd	- MSD parameter sequence
size_t	win_size	- # of windows (static + deltas)
int	*win_l_width	- left width of windows
int	*win_r_width	- right width of windows
double	**win_coefficient	- window coefficients
size_t	win_max_width	- maximum width of windows
double	*gv_mean	- mean vector of GV
double	*gv_vari	- variance vector of GV
HTS_Boolean	*gv_switch	- GV flag sequence

HTS_SStreamSet Set of state stream.

HTS_SStream	*sstream	- state streams
size_t	nstream	- # of streams
size_t	nstate	- # of states
size_t	*duration	- duration sequence
size_t	total_state	- total state
size_t	total_frame	- total frame

1.5 PDF stream

HTS_SMatrices Matrices/Vectors used in the speech parameter generation algorithm.

double	**mean	- mean vector sequence
double	**ivar	- inverse diagonal variance sequence
double	*g	- vector used in the forward substitution
double	**wuw	- $W' U^{-1} W$
double	*wum	- $W' U^{-1} m$

HTS_PStream Individual PDF stream.

size_t	vector_length	- vector length (static features only)
size_t	length	- stream length
size_t	width	- width of dynamic window
double	**par	- output parameter vector
HTS_SMatrices	sm	- matrices for parameter generation
size_t	win_size	- # of windows (static + deltas)
int	*win_l_width	- left width of windows
int	*win_r_width	- right width of windows
double	**win_coefficient	- window coefficients
HTS_Boolean	*msd_flag	- Boolean sequence for MSD
double	*gv_mean	- mean vector of GV
double	*gv_vari	- variance vector of GV
HTS_Boolean	*gv_switch	- GV flag sequence
size_t	gv_length	- frame length for GV calculation

HTS_PStreamSet Set of PDF streams.

HTS_PStream	*pstream	- PDF streams
size_t	nstream	- # of PDF streams
size_t	total_frame	- total frame

1.6 Generated parameter stream

HTS_GStream Generated parameter stream.

size_t	vector_length	- vector length (static features only)
double	**par	- generated parameter

HTS_GStreamSet Set of generated parameter stream.

size_t	total_nsamp	- total sample
size_t	total_frame	- total frame
size_t	nstream	- # of streams
HTS_GStream	*gstream	- generated parameter streams
double	*gspeech	- generated speech

1.7 Engine

HTS_Condition Synthesis condition.

size_t	sampling_frequency	- sampling frequency
size_t	fperiod	- frame period
size_t	audio_buff_size	- audio buffer size (for audio device)
HTS_Boolean	stop	- stop flag
double	volume	- volume
double	*msd_threshold	- MSD thresholds
double	*gv_weight	- GV weights
HTS_Boolean	phoneme_alignment_flag	- flag for using phoneme alignment in label
double	speed	- speech speed
size_t	stage	- if <i>stage</i> = 0 then <i>gamma</i> = 0 else <i>gamma</i> = -1/ <i>stage</i>
HTS_Boolean	use_log_gain	- log gain flag (for LSP)
double	alpha	- all-pass constant
double	beta	- postfiltering coefficient
double	additional_half_tone	- additional half tone
double	*duration_iw	- weights for duration interpolation
double	**parameter_iw	- weights for parameter interpolation
double	**gv_iw	- weights for GV interpolation

HTS_Engine Engine itself.

HTS_Condition	condition	- synthesis condition
HTS_Audio	audio	- audio output
HTS_ModelSet	ms	- set of duration models, HMMs and GV models
HTS_Label	label	- label
HTS_SStreamSet	sss	- set of state streams
HTS_PStreamSet	pss	- set of PDF streams
HTS_GStreamSet	gss	- set of generated parameter streams

2 Engine functions

2.1 Initialize engine

2.1.1 HTS_Engine_initialize

Type void
Use Initialize engine.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
Attention!! To start engine, first you must call this function.

2.2 Load models

2.2.1 HTS_Engine_load

Type HTS_Boolean
Use Load duration PDFs and trees from files using given file names.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 char **voices - HTS voice file names
 size_t num.voices - # of HTS voices
Attention!! You must initialize engine using HTS_Engine_initialize before calling this function.

2.3 Synthesize speech and set/get synthesis parameters

2.3.1 HTS_Engine_set_sampling_frequency

Type void
Use set sampling frequency.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 size_t i - sampling frequency (Hz), $1 \leq i$

2.3.2 HTS_Engine_get_sampling_frequency

Type size_t
Use get sampling frequency.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.3 HTS_Engine_set_fperiod

Type void
Use set frame shift.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 size_t i - frame shift (point), $1 \leq i$

2.3.4 HTS_Engine_get_fperiod

Type size_t
Use get frame shift.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.5 HTS_Engine_set_audio_buff_size

Type void
Use set buffer size for direct audio output.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t i - buffer size (sample)
Attention!! Default value is 0. If $i = 0$, direct audio play is turned off.

2.3.6 HTS_Engine_get_audio_buff_size

Type size_t
Use get buffer size for direct audio output.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
Attention!! Default value is 0. If $i = 0$, direct audio play is turned off.

2.3.7 HTS_Engine_set_stop_flag

Type void
Use set stop flag.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
HTS_Boolean b - flag
Attention!! Default value is FALSE.

2.3.8 HTS_Engine_get_stop_flag

Type HTS_Boolean
Use get stop flag.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
Attention!! Default value is FALSE.

2.3.9 HTS_Engine_set_volume

Type void
Use set volume in db.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
double f - volume in db
Attention!! Default value is 0.0.

2.3.10 HTS_Engine_get_volume

Type double
Use get volume in db.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.11 HTS_Engine_set_msd_threshold

Type void
Use set MSD threshold.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 size_t stream_index - index of streams
 double f - threshold

2.3.12 HTS_Engine_get_msd_threshold

Type double
Use get MSD threshold.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 size_t stream_index - index of streams

2.3.13 HTS_Engine_set_gv_weight

Type void
Use set GV weight.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 size_t stream_index - index of streams
 double f - GV weight
Attention!! Default value is 1.0.

2.3.14 HTS_Engine_get_gv_weight

Type double
Use get GV weight.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 size_t stream_index - index of streams

2.3.15 HTS_Engine_set_speed

Type void
Use set speech speed.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
double f - speed
Attention!! Default value is 1.0.

2.3.16 HTS_Engine_set_phoneme_alignment_flag

Type void
Use set flag to use phoneme alignment in label.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
HTS_Boolean b - flag
Attention!! Default value is FALSE.

2.3.17 HTS_Engine_set_alpha

Type void
Use set frequency warping parameter alpha.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
double f - alpha, $0.0 \leq f \leq 1.0$

2.3.18 HTS_Engine_get_alpha

Type double
Use get frequency warping parameter alpha.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.19 HTS_Engine_set_beta

Type void
Use set postfiltering coefficient parameter beta.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
double f - beta, $0.0 \leq f \leq 1.0$
Attention!! Default value is 0.0.

2.3.20 HTS_Engine_get_beta

Type double
Use get postfiltering coefficient parameter beta.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
Attention!! Default value is 0.0.

2.3.21 HTS_Engine.add_half_tone

Type void
Use add half tone.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
double f - half tone

2.3.22 HTS_Engine.set_duration_interpolation_weight

Type void
Use set weight for duration interpolation.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t voice_index - index of duration models
double f - interpolation weight

2.3.23 HTS_Engine.get_duration_interpolation_weight

Type double
Use get weight for duration interpolation.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t voice_index - index of duration models

2.3.24 HTS_Engine.set_parameter_interpolation_weight

Type void
Use set weight for parameter interpolation.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t voice_index - index of parameter models
size_t stream_index - index of streams
double f - interpolation weight

2.3.25 HTS_Engine.get_parameter_interpolation_weight

Type double
Use get weight for parameter interpolation.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t voice_index - index of parameter models
size_t stream_index - index of streams

2.3.26 HTS_Engine_set_gv_interpolation_weight

Type void
Use set weight for GV interpolation.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t voice_index - index of GV models
size_t stream_index - index of streams
double f - interpolation weight

2.3.27 HTS_Engine_get_gv_interpolation_weight

Type double
Use get weight for GV interpolation.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t voice_index - index of GV models
size_t stream_index - index of streams

2.3.28 HTS_Engine_get_total_state

Type size_t
Use get total # of state.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.29 HTS_Engine_set_state_mean

Type void
Use set mean value of state.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t stream_index - index of streams
size_t state_index - index of states
size_t vector_index - index of vector
double f - mean value

2.3.30 HTS_Engine_get_state_mean

Type double
Use get mean value of state.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t stream_index - index of streams
size_t state_index - index of states
size_t vector_index - index of vector

2.3.31 HTS_Engine_get_state_duration

Type size_t
Use get state duration.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t state_index - index of states

2.3.32 HTS_Engine_get_nvoices

Type size_t
Use get # of HTS voices.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.33 HTS_Engine_get_nstream

Type size_t
Use get # of stream.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.34 HTS_Engine_get_nstate

Type size_t
Use get # of state.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.35 HTS_Engine_get_fullcontext_label_format

Type const char *
Use get fullcontext label format defined in HTS voice.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.36 HTS_Engine_get_fullcontext_label_version

Type const char *
Use get fullcontext label version defined in HTS voice.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.37 HTS_Engine_get_total_frame

Type size_t
Use get total # of frame.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.38 HTS_Engine_get_nsamples

Type size_t
Use get # of samples.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.39 HTS_Engine_get_generated_parameter

Type size_t
Use get generated parameter.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t stream_index - index of streams
size_t frame_index - index of frames
size_t vector_index - index of vector

2.3.40 HTS_Engine_get_generated_speech

Type size_t
Use get generated speech.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
size_t index - index of samples

2.3.41 HTS_Engine_synthetize_from_fn

Type HTS_Boolean
Use synthesize speech from file name.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
char *fn - label file name

2.3.42 HTS_Engine_synthetize_from_strings

Type HTS_Boolean
Use synthesize speech from string list.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
char **lines - label string list
size_t num_lines - # of lines

2.3.43 HTS_Engine_generate_from_fn

Type HTS_Boolean
Use generate state sequence from file name (1/3 synthesis step)
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
char *fn - label file name

2.3.44 HTS_Engine_generate_from_strings

Type HTS_Boolean
Use generate state sequence from string list (1/3 synthesis step)
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 char **lines - label string list
 size_t num_lines - # of lines

2.3.45 HTS_Engine_generate_parameter_sequence

Type HTS_Boolean
Use generate parameter sequence (2/3 synthesis step)
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.46 HTS_Engine_generate_sample_sequence

Type HTS_Boolean
Use generate sample sequence (3/3 synthesis step)
Arguments HTS_Engine *engine - pointer to HTS_Engine structure

2.3.47 HTS_Engine_save_information

Type void
Use output trace information.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 FILE *fp - output file pointer

Attention!!

2.3.48 HTS_Engine_save_label

Type void
Use output label with time.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 FILE *fp - output file pointer

Attention!!

2.3.49 HTS_Engine_save_generated_parameter

Type void
Use output generated parameter.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
 FILE *fp - output file pointer

Attention!!

2.3.50 HTS_Engine_save_generated_speech

Type void
Use output generated speech.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
FILE *fp - output file pointer

Attention!!

2.3.51 HTS_Engine_save_riff

Type void
Use output riff format file.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
FILE *fp - output file pointer

Attention!!

2.3.52 HTS_Engine_refresh

Type void
Use free label, state streams, PDF streams and generated parameter streams per one time synthesis
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
Attention!!

2.4 Free engine

2.4.1 HTS_Engine_clear

Type void
Use free engine.
Arguments HTS_Engine *engine - pointer to HTS_Engine structure
Attention!!