Singing Voice Database

1. The overall database structure.

The purpose of the project is to create a Singing Voice Database (SVDB), which could be used for research of singing expressions in natural singing voice and developing of a Singing Voice Synthesis System.

The SVDB consists of two major parts:

1. The singing musical scale recordings (“Ah-ah-ah” and “La-la-la” versions).
2. The full song singing recordings with different expressions.

Every recording includes two waveform files, which contain the recording of voice of the singer (“vocal recordings”) and the recording of sound near singer glottis (“glottal recordings”). Both parts (the scale and the song) contain the studio recordings of female and male professional singers. The first part includes recordings of Bonnie Lander and Philip Larson. The second part includes recordings of Susan Narucki and Philip Larson.

We assume that the first part will be useful for developing a Singing Voice Synthesis system, while the second part will be used mostly for research.

Both parts contains plain recordings and recordings with special singing expressions described in Table 1 below.

Table 1. Singing expressions

Expression #	**Expression Name**	**Description**
1	***Bounce***	Increased articulation on consonants (slight increase of weight on initial consonants), followed by decrease of weight on adjacent vowels. More rhythmic vitality of a regular sort.
2	***Hollow***	Less articulation of consonants. Modification of vowels to minimize their differences with the addition of "air" in the tone (as opposed to focused tone).
3	***Light***	Minimal initial articulation and weight. Modification of vowels to emphasize "brightness" upper partials.
4	***Soft***	Modification of vowels; some air added, low volume. Consonants are present, but not sharply defined.
5	***Sweet***	Extreme legato. Pure vowels. Consonants present, but without extra articulated weight.
6	***Flat***	Affectless. Consonants and vowels with same weight. Minimizing melodic contour.

7	***Mature***	Emphasis on heavier vibrato in sound, (irregular) emphasis on lower partials of vowels (dark rather than bright).
8	***Sharp***	Emphasis on forward placement of vowel, cutting off lower partials. Aggressive articulation of consonants.
9	***Husky***	Irregular rhythmic inflection in phrasing. Irregular pronunciation of consonants and vowels, additional throat grab noises and air to vowel mix.
10	***Clear***	Purity of vowels and consonants. Emphasis on regularity of pronunciation. Sincere affect.

2. Part 1. Musical Scale Recordings

2.1. Part 1 content description

This part of SVDB contains the following recordings groups:

– The scale (musical notes) performed using “ah” vowel (“ah-ah-ah” recordings).

– The scale (musical notes) performed using “la” syllable (“la-la-la” recordings).

– The transitions between notes performed using “ah” vowel.

– The transitions between notes performed using “la” syllable.

– The scale performed with both “ah” and “la” with singing expressions, listed in Table 1.

Every recording performed with the following musical notes range:

- Female voice (Bonnie Lander): from A3 to A5.
- Male voice (Philip Larson): from A2 to E4.

Vocal and glottal recordings are provided.

For every waveform file the voiced and unvoiced segments are marked in separate annotation file, which has the same name as the waveform file and extension “lab”.

2.2. Directory and files structure

Files of the first part are contained in “Scale recordings” directory.

The recordings are in WAVE PCM format with the following characteristics: 44100 Hz; 16 bit; 1 channel (mono). There is an annotation file (.lab) associated with each wav-file. The .lab-files have the same format, as the TIMIT database files, namely:

<START_POINT > <END_POINT> <SEGMENT_NAME><new-line>

…

<START_POINT > <END_POINT> < SEGMENT_NAME >

where,

START_POINT :== The starting point of the segment, $*10^8$ ms

END_POINT :== The end point of the segment, $*10^8$ ms

SEGMENT_NAME :== The name of the specified segment, **v** for voiced segment and **u** for unvoiced segment.

.wav and .lab files for each recordings have the same name.

The content of the files is described in Table 2.

Table 2. Description of files for scale recordings.

No	File Name	Male\Female Voice	Material	Number of instances	Expression
1	1-1.wav	Male	Notes: “ah-ah”	2	*No expression*
2	1-1.lab				
3	1-1 glottal.wav				
4	1-2.wav	Male	Notes: “la-la”	2	*No expression*
5	1-2.lab				
6	1-2 glottal.wav				
7	1-3.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*No expression*
8	1-3.lab				
9	1-3 glottal.wav				
10	1-4.wav	Male	Transitions: “ah-ah”	1	*No expression*
11	1-4.lab				
12	1-4 glottal.wav				
13	1-5.wav	Male	Transitions: “la-la”	1	*No expression*
14	1-5.lab				
15	1-5 glottal.wav				
16	1-6.wav	Male	Transitions: “la-la”, “ah-ah”	1	*No expression*
17	1-6.lab				
18	1-6 glottal.wav				
19	1-7.wav	Male	Transitions: “la-la”, “ah-ah”	1	*No expression*
20	1-7.lab				
21	1-7 glottal.wav				
22	1-8.wav	Male	Notes: “ah-ah”	2	*Bounce*
23	1-8.lab				
24	1-8 glottal.wav				
25	1-9.wav	Male	Notes: “la-la”	2	*Bounce*
26	1-9.lab				
27	1-9 glottal.wav				
28	1-10.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Hallow*
29	1-10.lab				
30	1-10 glottal.wav				
31	1-11.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Light*
32	1-11.lab				
33	1-11 glottal.wav				
34	1-12.wav	Male	Notes: “ah-ah”, “la-	2+2	*Soft*

35	1-12.lab		la”		
36	1-12 glottal.wav				
37	1-13.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Sweet*
38	1-13.lab				
39	1-13 glottal.wav				
40	1-14.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Flat*
41	1-14.lab				
42	1-14 glottal.wav				
43	1-15.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Mature*
44	1-15.lab				
45	1-15 glottal.wav				
46	1-16.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Sharp*
47	1-16.lab				
48	1-16 glottal.wav				
49	1-17.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Clear*
50	1-17.lab				
51	1-17 glottal.wav				
52	1-18.wav	Male	Notes: “ah-ah”, “la-la”	2+2	*Husky*
53	1-18.lab				
54	1-18 glottal.wav				
55	1-19.wav	Female	Notes: “ah-ah”, “la-la”	2+2	No expression
56	1-19.lab				
57	1-19 glottal.wav				
58	1-20.wav	Female	Transitions: “ah-ah”	1	No expression
59	1-20.lab				
60	1-20 glottal.wav				
61	1-21.wav	Female	Notes: “ah-ah”	2	*Bounce*
62	1-21.lab				
63	1-21 glottal.wav				
64	1-22.wav	Female	Notes: “la-la”	2	*Bounce*
65	1-22.lab				
66	1-22 glottal.wav				
67	1-23.wav	Female	Notes: “ah-ah”	2	*Hallow*
68	1-23.lab				
69	1-23 glottal.wav				
70	1-24.wav	Female	Notes: “la-la”	2	*Hallow*
71	1-24.lab				
72	1-24 glottal.wav				

73	1-25.wav	Female	Notes: “ah-ah”, “la-la”	2	*Light*
74	1-25.lab				
75	1-25 glottal.wav				
76	1-26.wav	Female	Notes: “ah-ah”, “la-la”	2+1	*Soft*
77	1-26.lab				
78	1-26 glottal.wav				
79	1-27.wav	Female	Notes: “la-la”	1	*Soft*
80	1-27.lab				
81	1-27 glottal.wav				
82	1-28.wav	Female	Notes: “ah-ah”, “la-la”	2+2	*Sweet*
83	1-28.lab				
84	1-28 glottal.wav				
85	1-29.wav	Female	Notes: “ah-ah”, “la-la”	2+2	*Mature*
86	1-29.lab				
87	1-29 glottal.wav				
88	1-30.wav	Female	Notes: “ah-ah”, “la-la”	2+2	*Sharp*
89	1-30.lab				
90	1-30 glottal.wav				
91	1-31.wav	Female	Notes: “ah-ah”, “la-la”	2+2	*Husky*
92	1-31.lab				
93	1-31 glottal.wav				
94	1-32.wav	Female	Notes: “ah-ah”, “la-la”	2+2	*Clear*
95	1-32.lab				
96	1-32 glottal.wav				
97	1-33.wav	Female	Transitions: “ah-ah”	1	*Vibrato*
98	1-33.lab				
99	1-33 glottal.wav				

3. Part 2. The Song Recordings

3.1. Part 2 content description

This part contains the recording of the song “Twinkle, twinkle, little star” performed by female and male singers, Susan Narucki and Philip Larson, with/without

different singing expressions. The types of expressions are described in Table 1 above. The vocal recordings are accompanied by glottal recordings.

For several recordings (namely, male recordings with expressions “flat”, “bounce” and “hollow”) there are annotation file with all the vowel phonemes marked.

The ARPABET code is used for phonetic transcription. The phonetic transcription of the whole song as well as the vowel phonemes are presented in Table 3.

Table 3. The text and phonetic transcription of “Twinkle, twinkle, little star” song

No	Word	Transcription	Vowel phonemes
1.	Twinkle	ˈT,W,IX,NG,K,L,	IX
2.	Twinkle	ˈT,W,IX,NG,K,L,	IX
3.	Little	ˈL,IX,T,L,	IX
4.	Star	S,T,AA,	AA
5.	How	H,AW,	AW
6.	I	AY,	AY
7.	Wonder	ˈW,AH,N,D,AX,	AH, AX
8.	What	W,AH,T,	AH
9.	You	J,UX,	UX
10.	Are	AA,	AA
11.	Up	AH,P,	AH
12.	Above	AX,ˈB,AH,V,	AX , AH
13.	The	DH,AX,	AX
14.	World	W,ER,L,D,	ER
15.	So	S,OW,	OW
16.	High	H,AY,	AY
17.	Like	L,AY,K,	AY
18.	A	AX,	AX
19.	Diamond	ˈD,AY,M,AX,N,D,	AY, AX
20.	In	IX,N,	IX
21.	The	DH,AX,	AX
22.	Sky	S,K,AY,	AY
23.	Twinkle	ˈT,W,IX,NG,K,L,	IX
24.	Twinkle	ˈT,W,IX,NG,K,L,	IX
25.	Little	ˈL,IX,T,L,	IX
26.	Star	S,T,AA,	AA
27.	How	H,AW,	AW
28.	I	AY,	AY
29.	Wonder	ˈW,AH,N,D,AX,	AH, AX
30.	What	W,AH,T,	AH
31.	You	J,UX,	UX
32.	Are	AA,	AA

33.	When	W,EH,N,	EH
34.	The	DH,AX,	AX
35.	Blazing	'B,L,EY,Z,IX,NG,	EY , IX
36.	Sun	S,AH,N,	AH
37.	Is	IX,Z,	IX
38.	Gone	G,AH,N,	AH
39.	When	W,EH,N,	EH
40.	There's	DH,AXR,Z,	AXR
41.	Nothing	'N,AH,TH,IX,NG,	AH , IX
42.	He	H,EE,	EE
43.	Shines	SH,AY,N,Z,	AY
44.	Upon	AX,'P,AH,N,	AX , AH
45.	Then	DH,EH,N,	EH
46.	You	J,UX,	UX
47.	Show	SH,OW,	OW
48.	Your	J,AO,	AO
49.	Little	'L,IX,T,L,	IX
50.	Light	L,AY,T,	AY
51.	Twinkle	'T,W,IX,NG,K,L,	IX
52.	Twinkle	'T,W,IX,NG,K,L,	IX
53.	Through	TH,R,UX,	UX
54.	The	DH,AX,	AX
55.	Night	N,AY,T,	AY
56.	Twinkle	'T,W,IX,NG,K,L,	IX
57.	Twinkle	'T,W,IX,NG,K,L,	IX
58.	Little	'L,IX,T,L,	IX
59.	Star	S,T,AA,	AA
60.	How	H,AW,	AW
61.	I	AY,	AY
62.	Wonder	'W,AH,N,D,AX,	AH, AX
63.	What	W,AH,T,	AH
64.	You	J,UX,	UX
65.	Are	AA,	AA
66.	In	IX,N,	IX
67.	The	DH,AX,	AX
68.	Dark	D,AA,K,	AA
69.	Blue	B,L,UX,	UX
70.	Sky	S,K,AY,	AY
71.	So	S,OW,	OW
72.	Deep	D,EE,P,	EE
73.	Through	TH,R,UX,	UX
74.	My	M,AY,	AY

75.	Curtains	ˈK,ER,T,N,Z,	ER
76.	Often	ˈAH,F,N,	AH
77.	Peep	P,EE,P,	EE
78.	For	F,AO,	AO
79.	You	J,UX,	UX
80.	Never	ˈN,EH,V,AX,	EH , AX
81.	Close	K,L,OW,Z,	OW
82.	Your	J,AO,	AO
83.	Eyes	AY,Z,	AY
84.	Till	T,IX,L,	IX
85.	The	DH,AX,	AX
86.	Morning	ˈM,AO,N,IX,NG,	IX
87.	Sun	S,AH,N,	AH
88.	Does	D,AH,Z,	AH
89.	Rise	R,AY,Z,	AY
90.	Twinkle	ˈT,W,IX,NG,K,L,	IX
91.	Twinkle	ˈT,W,IX,NG,K,L,	IX
92.	Little	ˈL,IX,T,L,	IX
93.	Star	S,T,AA,	AA
94.	How	H,AW,	AW
95.	I	AY,	AY
96.	Wonder	ˈW,AH,N,D,AX,	AH, AX
97.	What	W,AH,T,	AH
98.	You	J,UX,	UX
99.	Are	AA,	AA

3.2. Directory and files structure

Files of the first part are contained in “Song recordings” directory.

The recordings are in WAVE PCM format with the following characteristics: 44100 Hz; 16 bit; 1 channel (mono). The annotation files, if present, have the same name, that the .wav-files, and the .lab extension. The format of .lab files is identical to those described in section 2.2. The SEGMENT_NAME is the name of the vowel for parts of recordings, representing vowels, and is “!” for the parts of recordings between two vowels. The content of files is described in Table 4.

Table 4. Description of files for song recordings.

No	Male/Female voice	File Name	Expression
1	Male	2-1.wav	*No expression*
2		2-1.lab	
3		2-1 glottal.wav	
4		2-2.wav	*Bouncy*
5		2-2.lab	
6		2-2 glottal.wav	
7		2-3.wav	*Hollow*
8		2-3.lab	
9		2-3 glottal.wav	
10		2-4.wav	*Bright*
11		2-4 glottal.wav	
12		2-5.wav	*Soft*
13		2-5 glottal.wav	
14		2-6.wav	*Operatic*
15		2-6 glottal.wav	
16		2-7.wav	*Flat*
17		2-7 glottal.wav	
18		2-8.wav	*Sweet*
19		2-8 glottal.wav	
20		2-9.wav	*Light*
21		2-9 glottal.wav	
22	Female	2-10.wav	*Bounce*
23		2-10 glottal.wav	
24		2-11.wav	*Clear*
25		2-11 glottal.wav	
26		2-12.wav	*Flat*
27		2-12 glottal.wav	
28		2-13.wav	*Hollow*
29		2-13 glottal.wav	
30		2-14.wav	*Husky*
31		2-14 glottal.wav	
32		2-15.wav	*Light*

33		2-15 glottal.wav	
34		2-16.wav	*Mature*
35		2-16 glottal.wav	
36		2-17.wav	*Sharp*
37		2-17 glottal.wav	
38		2-18.wav	*Soft*
39		2-18 glottal.wav	
40		2-19.wav	*Sweet*
41		2-19 glottal.wav	
42		2-20.wav	*Jazz with hollow and husky*
43		2-20 glottal.wav	