

Vocal Synthesis

R09921043 電機碩二 林文炤

Outline

Outline

Vocal synthesis

Technologies

Formant-based synthesis

Physical model

Signal model

Experiments

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis

Vocal synthesis technologies

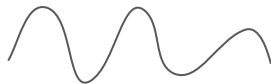
Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



Vocal synthesis technologies

Synthesis from samples

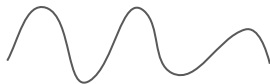
Waveform synthesis

HMM

Deep learning

Formant-based synthesis

現在進站的是



號列車, 往



艾恩葛朗特



8763



Vocal synthesis technologies

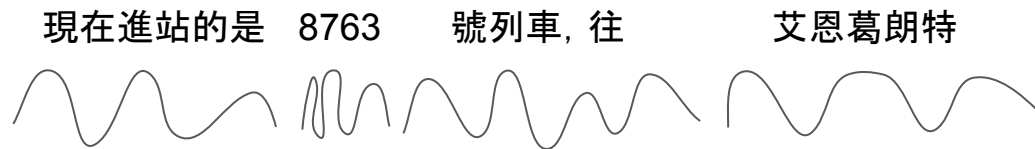
Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



Vocal synthesis technologies

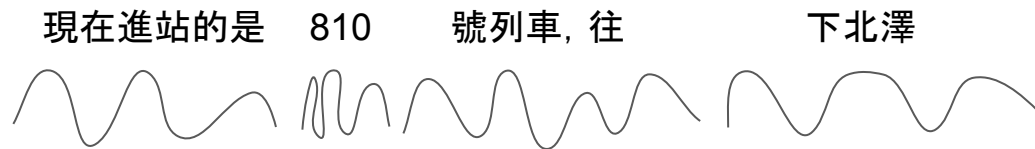
Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



Vocal synthesis technologies

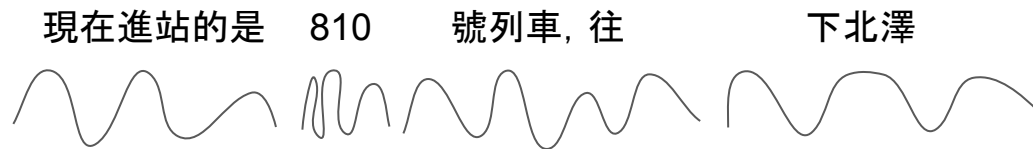
Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



pros: voice sounds natural
easy to make

Vocal synthesis technologies

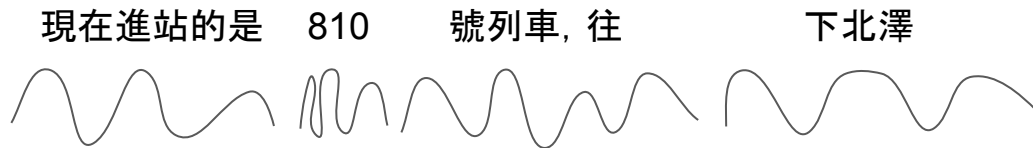
Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



pros: voice sounds natural
easy to make

cons: lack of flexibility
connection between samples are difficult
needs large database

Vocal

Synth

Wave

HMM

Deep learning

Forma

pr



<https://www.youtube.com/watch?v=TTZoWmSvral>

810



https://youtube.com/watch?v=8ahwhnVx_U0



<https://www.youtube.com/watch?v=OWL65w9XEfc>

cons: lack of flexibility
connection between samples are difficult
needs large database

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



Vocal synthesis technologies

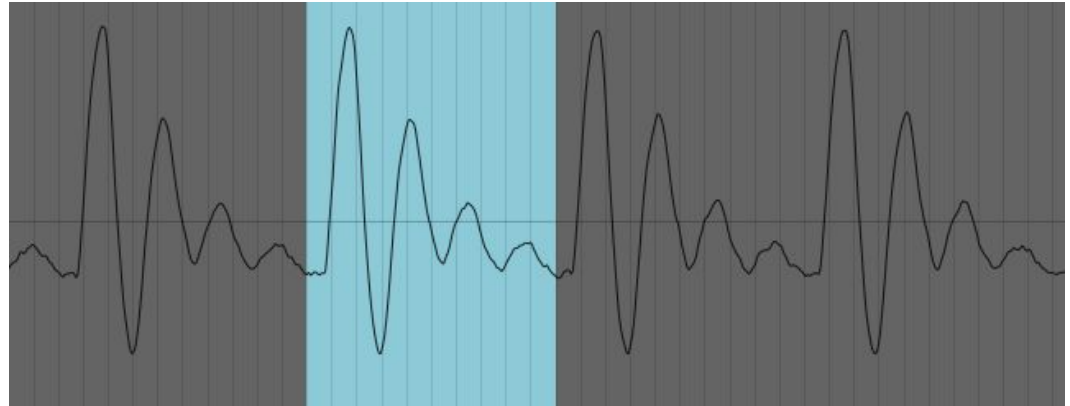
Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



pros: flexible length

easy to adjust compare to sample synthesis

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



pros: flexible length
easy to adjust compare to sample synthesis

cons: hard to adjust pitch without formant-based method
hard to make sound transitions
needs large database

Vocal synthesis technologie

Synthesis from samples

Waveform synthesis

HMM

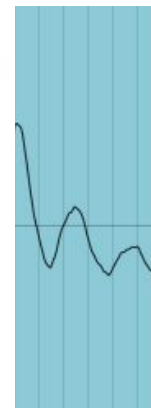
Deep learning

Formant-based synthesis

p **VOCALOID™**

<https://www.vocaloid.com/>

sample synthesis



t-based method

Illustration by iXima © Crypton Future Media, INC. www.piapro.net piapro

<https://blog.piapro.net/2016/07/z1607081-1.html>

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



https://www.ntu.edu.tw/spotlight/2016/887_20160818.html

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis



https://www.ntu.edu.tw/spotlight/2016/887_20160818.html



<https://case.ntu.edu.tw/blog/?p=33164>

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

based on physical model

Formant-based synthesis

Vocal synthesis technologies

Synthesis from samples

Waveform synthesis

HMM

Deep learning

Formant-based synthesis

based on physical model

simplified to signal process

Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowels

(Almost) independant from notes

Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowels

(Almost) independant from notes



Formant

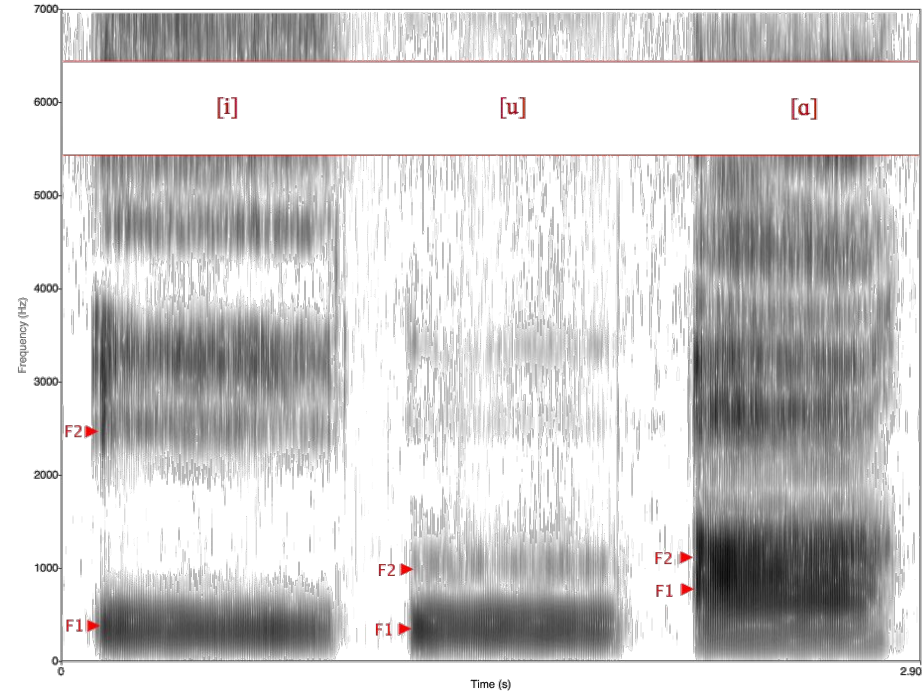
Resonant peaks from human's mouth and

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowel

(Almost) independant from notes



Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowels

(Almost) independant from notes

Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowels

(Almost) independant from notes

Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

different shape and size

2 formants can express informations of vowels

(Almost) independant from notes

Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowels

(Almost) independant from notes

Formant

Resonant peaks from human's mouth and throat

Determines vowel

Personal characteristics ex: gender, age

2 formants can express informations of vowels

(Almost) independant from notes

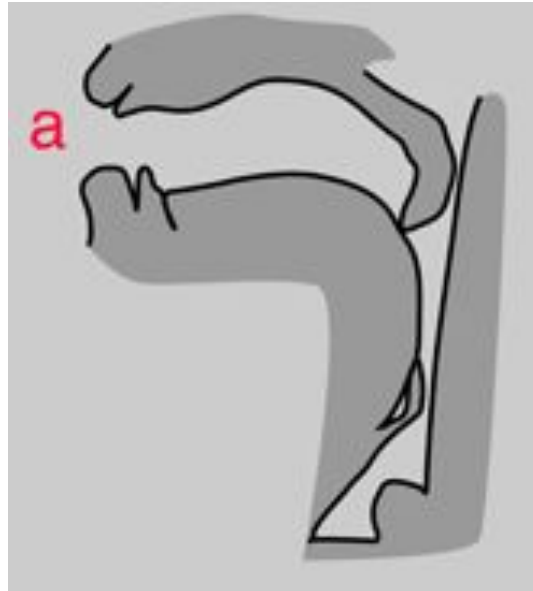
Physical model of human voice

vocal cord

laryngeal cavity

oral cavity

nasal cavity



<http://hyperphysics.phy-astr.gsu.edu/hbase/Music/vowel.html>

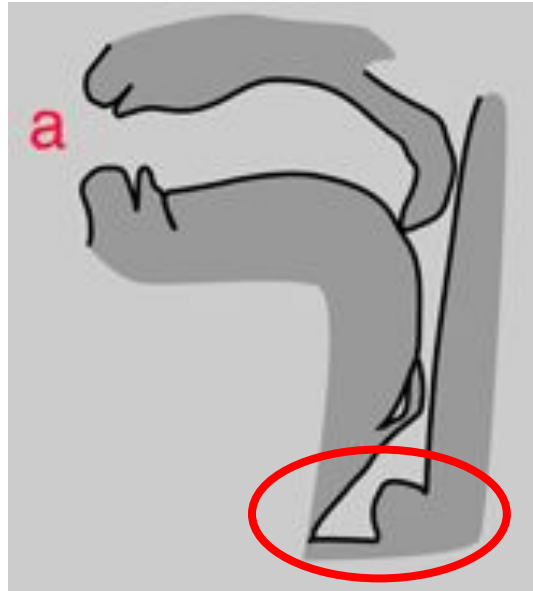
Physical model of human voice

vocal cord

laryngeal cavity

oral cavity

nasal cavity



<http://hyperphysics.phy-astr.gsu.edu/hbase/Music/vowel.html>

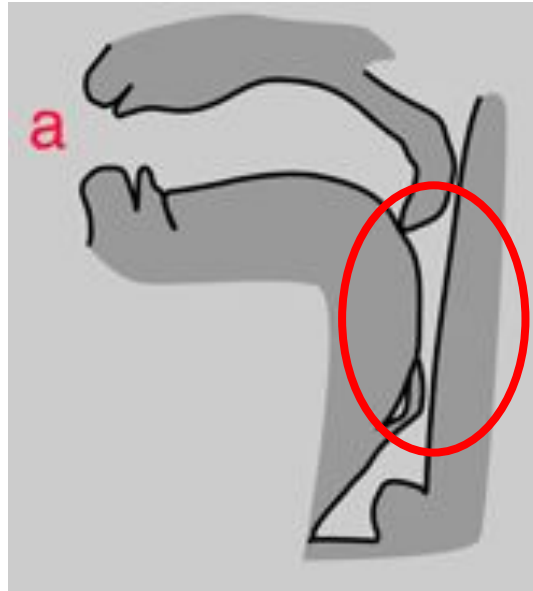
Physical model of human voice

vocal cord

laryngeal cavity

oral cavity

nasal cavity



<http://hyperphysics.phy-astr.gsu.edu/hbase/Music/vowel.html>

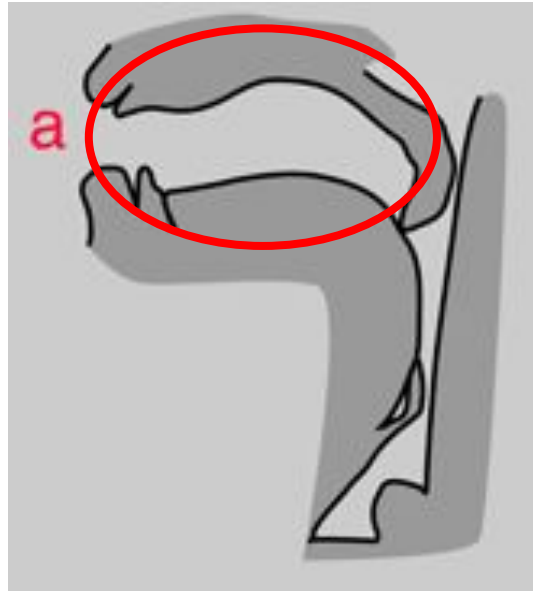
Physical model of human voice

vocal cord

laryngeal cavity

oral cavity

nasal cavity



<http://hyperphysics.phy-astr.gsu.edu/hbase/Music/vowel.html>

Signal model

Use saw wave oscillator to model vocal cord

Model cavities with linear filters

Signal model

Use saw wave oscillator to model vocal cord

Model cavities with linear filters

Signal model

Use saw wave oscillator to model vocal cord

Model cavities with linear filters

saw waves have all harmonics

Signal model

Use saw wave oscillator to model vocal cord

Model cavities with linear filters

Pros and Cons

Pros:

space friendly

easy to adjust

Cons:

hard to make voice natural

hard to synthesize consonants

Voice imitation experiment

Use vocaloid-初音ミクV4X as reference

Analyze formants of sound samples

Use saw wave synthesizer with LPF as vocal cord model

Use parametric EQ to get ideal frequency response

Voice imitation experiment

Use vocaloid-初音ミクV4X as reference

Analyze formants of sound samples

Use saw wave synthesizer with LPF as vocal cord model

Use parametric EQ to get ideal frequency response

Properties of Japanese

Voice imitation experiment

Use vocaloid-初音ミクV4X as reference

Analyze formants of sound samples

Use saw wave synthesizer with LPF as vocal cord model

Use parametric EQ to get ideal frequency response

Properties of Japanese

あ い う え お ん

A I U E O N

Voice imitation experiment

Use vocaloid-初音ミクV4X as reference

Analyze formants of sound samples

Use saw wave synthesizer with LPF as vocal cord model

Use parametric EQ to get ideal frequency response

Properties of Japanese

あ い う え お ん

no transitions

A I U E O N

Future development

Consonant synthesis

Different tones ex: modal voice/falsetto voice, different emotions

Formant will change with pitch in real case

Saw wave is not a good model of vocal cord

Harmonic analysis

Future development

Consonant synthesis

Different tones ex: modal voice/falsetto voice, different emotions

Formant will change with pitch in real case

Saw wave is not a good model of vocal cord

Harmonic analysis

愛を言い合う王老い、居合をおいおい終え、
以往、良い葵を上へ鋭意植え、青い魚を和え、
アイアイを得、硫黄覆う庵へ、御家を負う甥を追う。

<https://id.fnshr.info/2014/10/12/vowel-sentence-ja/>

Future development

Consonant synthesis

Different tones ex: modal voice/falsetto voice, different emotions

Formant will change with pitch in real case

Saw wave is not a good model of vocal cord

Harmonic analysis