Speech production

Gray's Anatomy



- ► Find the bit where the noise comes out.
- Take it to bits.
- See how it works.

http://en.wikipedia.org/wiki/Vocal_tract

Vocal apparatus



Apparently from Borden et al, 1994.

Physically quite large

- Lungs, throat.
- Oral cavity, nasal cavity.
- ► Tongue, lips, glottis, teeth.

Vocal folds



Vocal "chords" are perhaps the most obvious part.

Used even in whispered speech.

Vowels and consonants

This is a rough separation at best

Vowels are resonant sounds that can be held as if singing. Consonants are everything else Including some things that are difficult to distinguish from vowels.

Vowels are vocalised

- The vocal chords vibrate, imparting pitch.
- The oral cavity resonates, imparting character or quality. Consonants involve other articulators.
 - ▶ Tongue, lips.
 - Stops, fricatives, may or may not be vocalised.



©The Wire Magazine

Donald Ayler, 1942–2007 (with Albert Ayler, 1936–1970)



©The Wire Magazine

Donald Ayler, 1942–2007 (with Albert Ayler, 1936–1970)

Modeling the vocal tract

Vocal tract models have two elements:

- 1. Excitation from the vocal chords.
- 2. System response from the vocal tract shape.

I will gloss over this. For a full description, see

Thomas F. Quatieri.

Discrete-Time Speech Signal Processing: principles and practice.

Signal Processing Series. Prentice Hall PTR, Upper Saddle River, New Jersey, 07458, 2002.

Excitation

Excitation is important, it has two forms:

- 1. Voiced sounds.
 - Pulse train (not a pure tone)
 - Corresponds to vocal chords opening.
 - Has a distinct pitch.
 - Imposes a distinct fundamental period in the spectra.
- 2. Unvoiced sounds.
 - White (-ish, pink maybe) noise.
 - ▶ No pitch, but perception of pitch can come from formants.
 - Spectra are smooth.

These two forms get you quite a long way in understanding speech.

System response 1

- Blah blah lossless tubes.
- Blah blah partial differential equations.
- Blah blah piston in infinite wall.

System response 2



The vocal tract is a pretty complicated tube. However, it can be summarised:

- There are multiple resonances. These correspond to poles.
- Some sounds, notably nasals, introduce zeros.
- There is a zero corresponding to the radiation impedance.

Vocalisation

Simplistic frequency domain view



Data driven

Why so blasé?

- In speech and language, hand-crafting a solution tends to only get you so far.
- It's better to learn what you can from data.
 - Provide a model that is capable of representing what you want.
 - Provide an algorithm capable of training the model.
 - Provide enough data to allow the algorithm to train the model.

Vowels

Vowel shapes







i: "eee"



ae "ah"



a "aaa"



u: "ooo"



The vowel quadrilateral



And formalised

VOWELS

Front Central Back Close 1 H uu • u ΙΥ Ω Close-mid е 9 0 Y e Open-mid ε Э G (X Λ æ D Open а Œ n Where symbols appear in pairs, the one to the right represents a rounded vowel.

The engineer's approach



Figure 9.1 Upper part: spectrogram of the word *August*. Lower part: three out of a possible 75 graphs representing the energy in a one-hundredth of a second section of this word.

Petersen-Barney data



- Formant frequencies of 76 American English speakers (around 1952).
- First vs. second formant frequencies are plotted.
- American vowels are separable in two dimensions.

Petersen-Barney rotated





Italian



- Italian vowels are very close to cardinal vowels.
- The mid ones are interchangeable to an extent with dialect So in this sense there are only 5.

Spanish



- Simplest of all European languages
 Phonetically speaking!
- Just 5 cardinal vowels.
- No need for a quadrilateral "The Spanish vowel triangle"

English (British RP)



- English is not like Mediterranean languages.
- Classed as Germanic.
- Central vowels.
- Un-stressed vowel "schwa" [ə].

French



- Actually quite close to Italian.
- A single central vowel Like English
- Bonus: extra dimension

Rounding

Generally: Back vowels are rounded. Front vowels are un-rounded.

- French and German (for example)
 Front vowels can be rounded or unrounded
- N.B.: Rounding is not the only other dimension. Length, stress, rhoticity.

Vowels are optional sometimes



Said Sjlbvdnzv resident Grg Hmphrs, 67: With just a few key letters, I could be George Humphries. That is my dream.

– The Onion

Difficulties for non-natives

Non-native speakers find some vowels very tricky.

- Particularly close together ones.
- Particularly rounded vs. back.

Pitch vs. Peach



Dessus vs. Dessous



(I am native English and can't hear the difference)

Diphthongs



"eight five zero" [ert farv ziro]

A diphthong is a vowel that mutates into a different vowel.

- ► Cow.
- ► Buy.

Often represented by lines on the vowel quadrilateral.

Consonants

Resonance

Vowels are about resonances.

- ▶ The vocal tract is "open".
- The sounds can be held for some time.

By contrast, consonants are about stopping resonance.

- The tongue touches part of the vocal tract.
- Air-flow is restricted somehow.

Sagittal sections





There are lots of named parts.

Example vocabulary



These are adjectives: Labial The lips. **Dental** The teeth. Alveolar The ridge behind the upper teeth. Alveoli are the holes your teeth sit in. Velar The roof of the mouth, or soft palate. Uvular The dangly thing at the back of your mouth.

IPA chart

The IPA is characteristically terse:

| | Bilabial | | Labiodental | | Dental Alveolar | | Postalveolar | | Retroflex | | Palatal | | Velar | | Uvular | | Pharyngeal | | Glottal | | | |
|------------------------|----------|---|-------------|---|-----------------|---|--------------|---|-----------|---|---------|----|-------|---|--------|---|------------|---|---------|---|---|---|
| Plosive | р | b | | | | | t | d | | | t | d | с | f | k | g | q | G | | | ? | |
| Nasal | | m | | ŋ | n | | | | | | | η | | ր | | ŋ | | N | | | | |
| Trill | | В | | | r | | | | | | | | | | | | | R | | | | |
| Tap or Flap | | | | | ſ | | | | | | | r | | | | | | | | | | |
| Fricative | ф | β | f | v | θ | ð | s | Z | S | 3 | ş | Z, | ç | j | х | y | χ | R | ħ | ٢ | h | ĥ |
| Lateral fricative | | | | | | | ł | ß | | | | | | | | | | | | | | |
| Approximant | | | | υ | | | | r | | | | ન | | j | | щ | | | | | | |
| Lateral approximant | | | | | | | | 1 | | | | l | | λ | | L | | | | | | |

CONSONANTS (PULMONIC)

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Labio-dental



Upper teeth and lower lip.[f], [v].

Bilabial



Both lips.[p], [b].

Alveolar



- Tongue, and the area just behind the teeth.
- In English, there are three quite close areas:
 - Dental, [θ] [δ].
 - ▶ Alveolar, [s], [z], [t], [d].
 - Post alveolar, $[\int], [_3]$.

Velar



 The tongue and the roof of the mouth, the hard palette.

▶ [k], [g].

Voicing



"zero seven"

Many consonants can be voiced or unvoiced. Consider the difference between these:

- ▶ [p] and [b].
- ▶ [k] and [g].
- ▶ [s] and [z].

Nasals



"oh nine one"

Nasals are actually a lot like vowels.

- In normal vowels, the nasal cavity is closed at the uvula.
- In nasals, the oral cavity is closed at
 - ▶ the lips: [m].
 - the alveolar ridge or teeth: [n].
 - ▶ the velum: [ŋ].
- In both cases, an open path exists between larynx and air.

Swedish nasal

Engwall & Badin: Collecting and analysing two- and three-dimensional MRI data for Swedish



Figure 3. Midsagittal images of log/, lumu/ and log/. In all pictures, an artefact due to minor movements of the velum can be seen. Note the absence of the teeth in the pictures. The position of the incisors in the midsagittal plane could however be inferred from lumu/, where the subject protrudes his tongue tip between the teeth, leaving clear marks in the tongue.

Fricatives



"zero seven" [zirou sevən]

Fricatives involve friction

- In English, it's (labio-)dental or alveolar
- Unvoiced: [f], [θ], [s], [∫].
- ▶ Voiced: [v], [ð], [z], [ʒ].



"six two" [siks tu:]

Stops are momententary complete obstructions

- ▶ Unvoiced: [p], [t], [k].
- ▶ Voiced: [b], [d], [g].
- Same articulation points as the nasals.

Also

Consonants also include:

- Semi-vowels
 - ▶ [w], [j].
 - Very like vowels, but a little more obstructed.
- Liquids
 - ► [J], [l].
 - Again, somewhat closer to vowels, but with some obstruction.

Notice that the English "r" is [1]. [r] is a trill.