

# Introduction to Acoustic Phonetics III

Sergey Knyazev

03 June, 2017

# ARTICULATION AND RESONANCE:

## CONSONANTS

earlier: how the movements and positions of the articulators are shaping the **resonating cavities of the vocal tract** to modify the sound source for **vowels**

it is quite similar to that for the **resonant / sonorant** consonants:

- approximants/glides/semivowels (/w/, /j/, /r/?)
- nasals (/m/, /n/, /ɲ/, /ŋ/, /N/ ...),
- liquids /l/ (lateral) and various /r/ (rhotics: **approximants** & vibrants = trill + tap + flap).

# semivowels

the vocal tract is relatively open for the semivowels, as it is for vowels

the **semivowels** are characterized acoustically by **formants**,

**but**

they are consonants, not as vowels:

- behave like consonants (/jadró/ [jidró], *chiabatta*)
- occur on the **periphery of syllables**, and not in the centers or nuclei of syllables

Occasionally, though, liquids and nasals do function as the nuclei “cotton” [katn.]

# w, j

/j/ is a **palatal** glide: the tongue blade must approximate the palate at a position close to that for [i]

/w/ is analogous to /j/: it's starting position is characterized by a high back tongue position and protruded lips, similar to /u/.

/w/ has two places of articulation: the **bilabial** protrusion and the **lingua-palatal** approximation

The articulatory configurations for /j/ and /w/ resemble, respectively, /i/ and /u/ →

the patterns of the **formant frequency change** are usually predictable as the vocal tract configuration changes to that of the vowel following the glide.

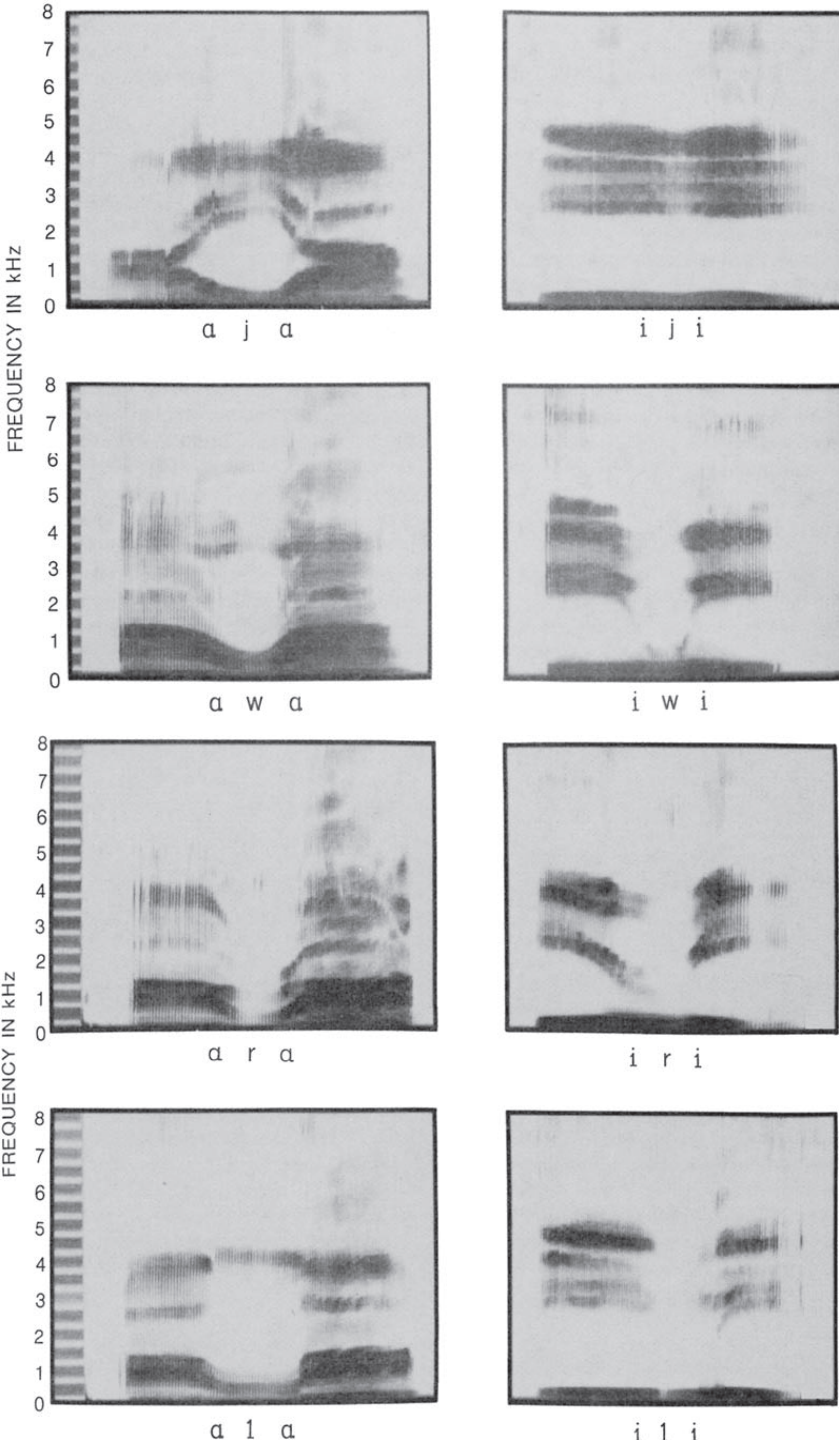
The tongue is very high in the mouth and the **pharyngeal cavity is relatively large** →

**F1 is very low** (as it is for /i/ and /u/) for both glides.

# r

/r/ is produced in syllable-initial position by raising the tongue toward the alveolar ridge, the tip does not touch the alveolar ridge, so some of the airflow and acoustic energy emerges centrally, the lips are usually rounded and the body of the tongue is raised toward the velum  
some speakers produce a retroflex /r/

The **acoustic results** of these tongue tip adjustments are particularly obvious in the third formant: **F<sub>3</sub> falls** below the F<sub>3</sub> frequencies typical of the neighboring vowels.



[w], [j], like vowels, are the result of filtering the complex wave produced at the glottis in the resonating cavities of the vocal tract.

The articulations (only minimum constrictions in the vocal tract), and, as a result the acoustic features (formant structures) of glides resemble those of vowels.

There are, however, some differences

Small constriction in mouth →  
**F<sub>1</sub> lowering, F<sub>2</sub> attenuation**

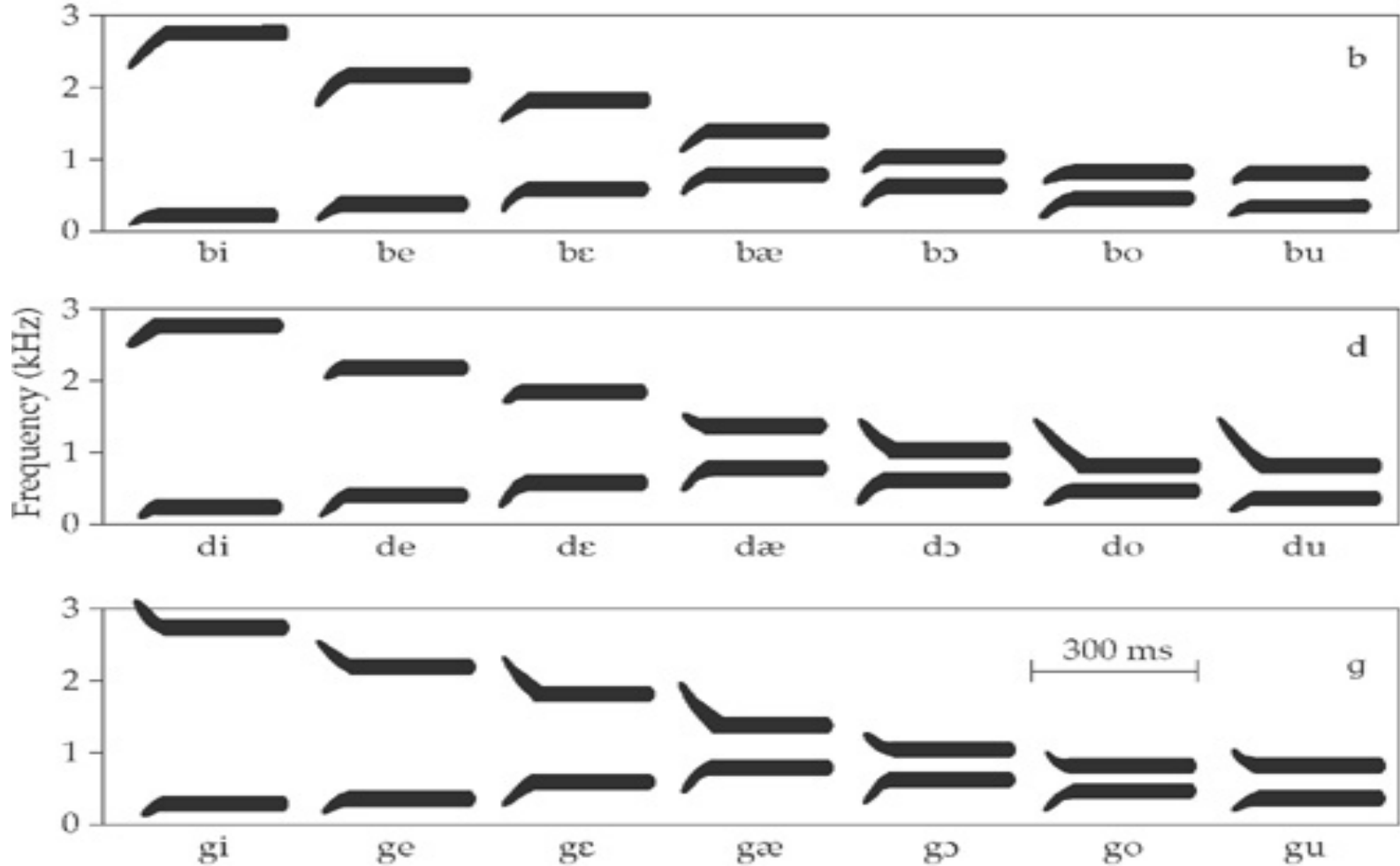
/r/ 3 constrictions (lips, alveolar ridge, velum) →  
**+ low F<sub>3</sub> (300/900/1600)**

Max. duration of formant transitions: **glides**

**It is the sound of the **acoustic changes** caused by the movements of the articulators that listeners use **to recognize glides****

Limited cases of such syllables as /ji/ (as in “yeast”) and /wu/ (as in “woo”): in both cases, the starting and ending vocal tract configurations are similar.

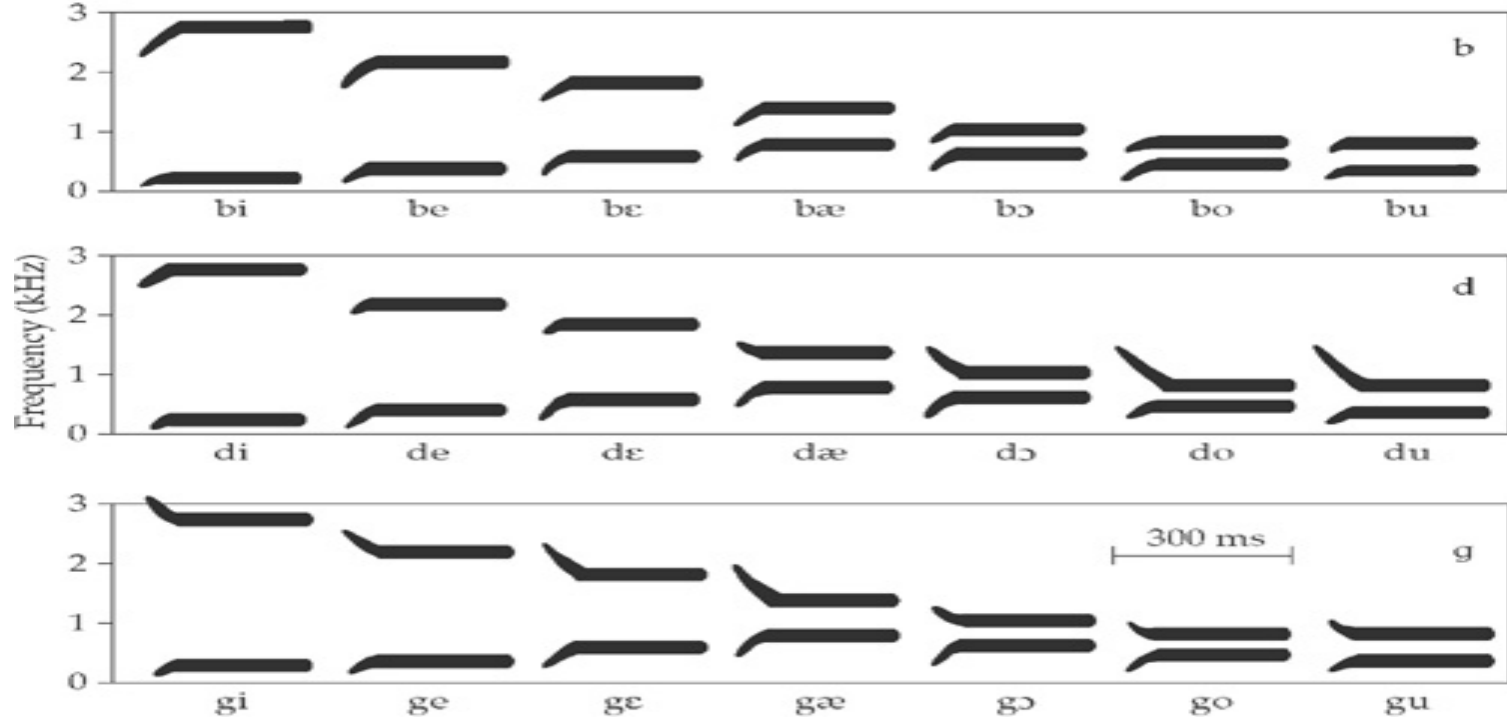
Speakers, however, do not simply produce a long /i/, /u/ Rather, they move the articulators away from the onset configuration and then return them to approximately where they started. This articulatory movement changes the vocal tract resonances and provides the formant frequency changes needed for glide production and perception.



F1 and F2 transition patterns in stop release used to synthesize [b], [d], and [g]  
 followed by various vowels (from Delattre et al. 1955, p. 770),

consonants consist of shutting and release **movements**:  
 the **information** they contain about place of articulation is seen  
 in **formant movements, not values**





Delattre et al. (1955): the perceptually important part of the transition is not actually present in the acoustic signal, but can be derived from a set of syllables such as those shown above.

Superimpose the F2 trajectories of syllables, starting with [d], and extend the F2 trajectories back in time, they seem to intersect at about 1,800 Hz.

This hypothetical starting frequency – the F2 **locus** frequency:

- **Labials:** 700 Hz, a **low**-frequency locus
- **Coronals:** 1800 Hz, a **middle**-level locus
- **Velars:** higher F2 locus.

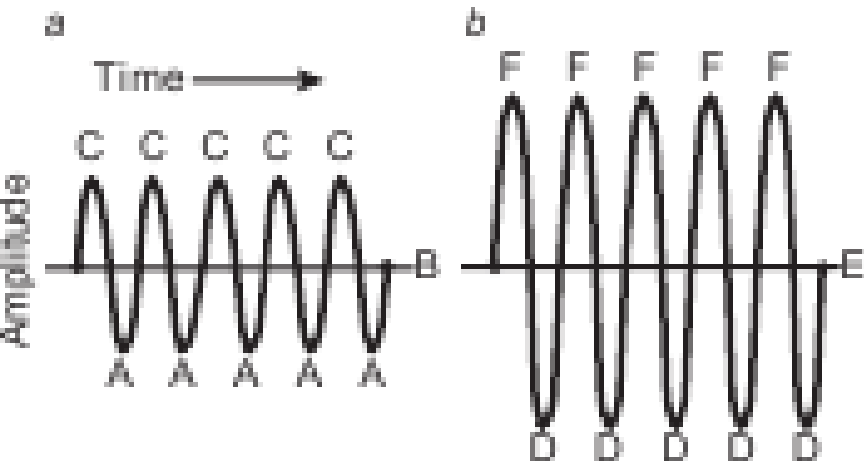
# Nasals and Laterals

are more complicated than vowels:

(it is their vocal tract filtering characteristics that are more complicated).

One way in which the vocal tract filtering function in **nasals** is different from that in oral vowels is that the width of the resonance peak (the **bandwidth**) of the first formant is larger in nasals as a result of **damping**

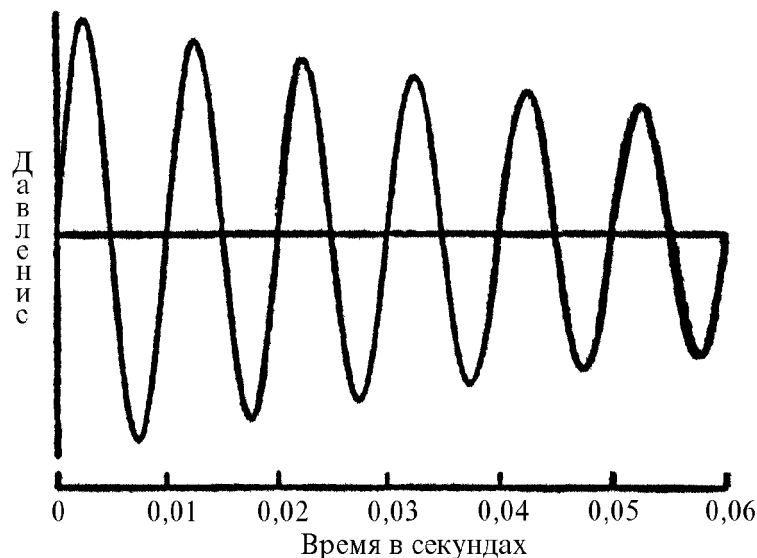
# damping



Simple periodic wave:

**the same amplitude** and  
**the same frequency**  
**all over the time**

**Amplitude** in the **damped**  
**sine waves**  
**decreases over time.**

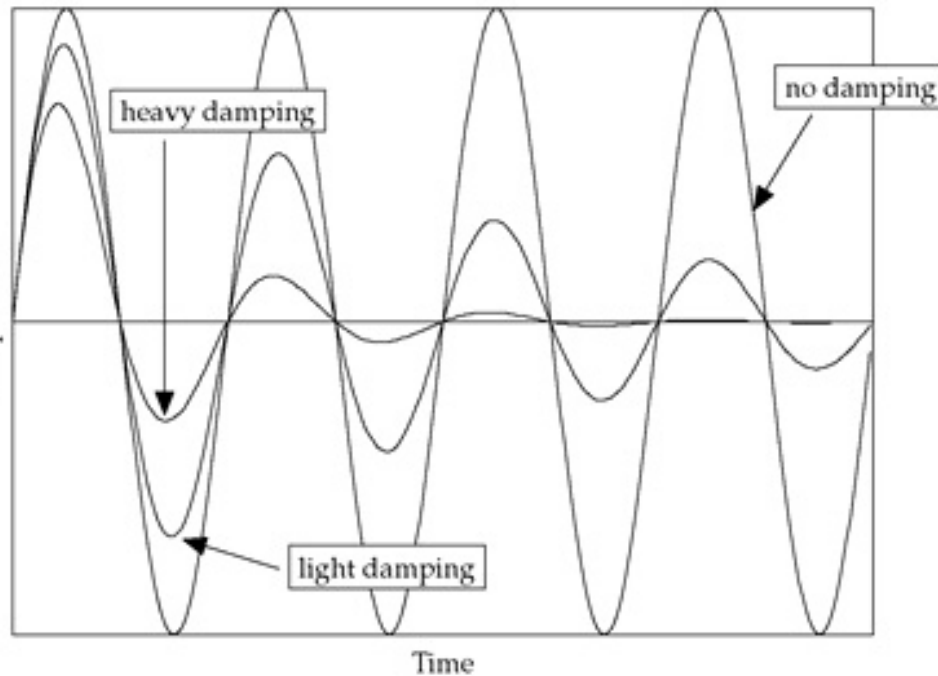


like pushing smb on a swing:

energy you put into the push is  
dissipated by a natural resistance to  
the swinging motion – the pull of  
gravity and the friction as the swing  
moves through the atmosphere.

# damping

An undamped sine wave and two damped sine waves.

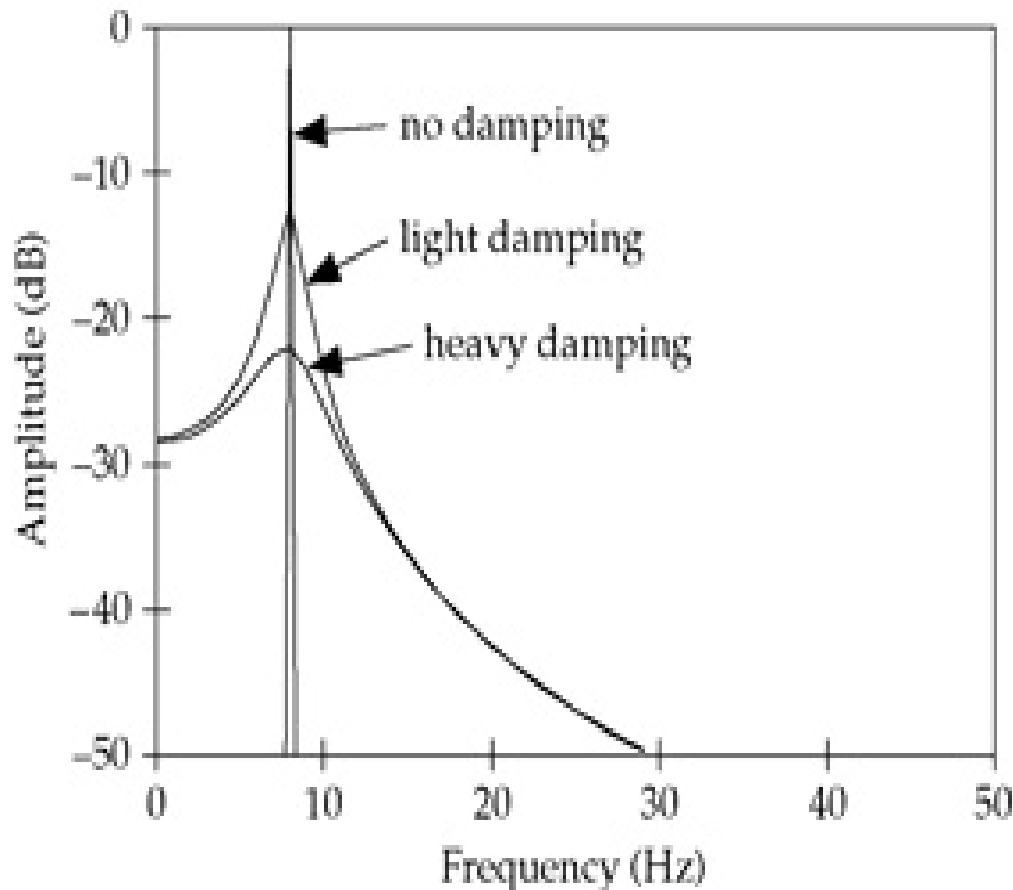


The heavy damping sine wave loses amplitude more quickly than the light damping one

the more heavily damped wave corresponds to swinging on the earth,

the lightly damped wave to swinging on the moon (where gravity is weaker and the atmosphere is less dense).

# damping

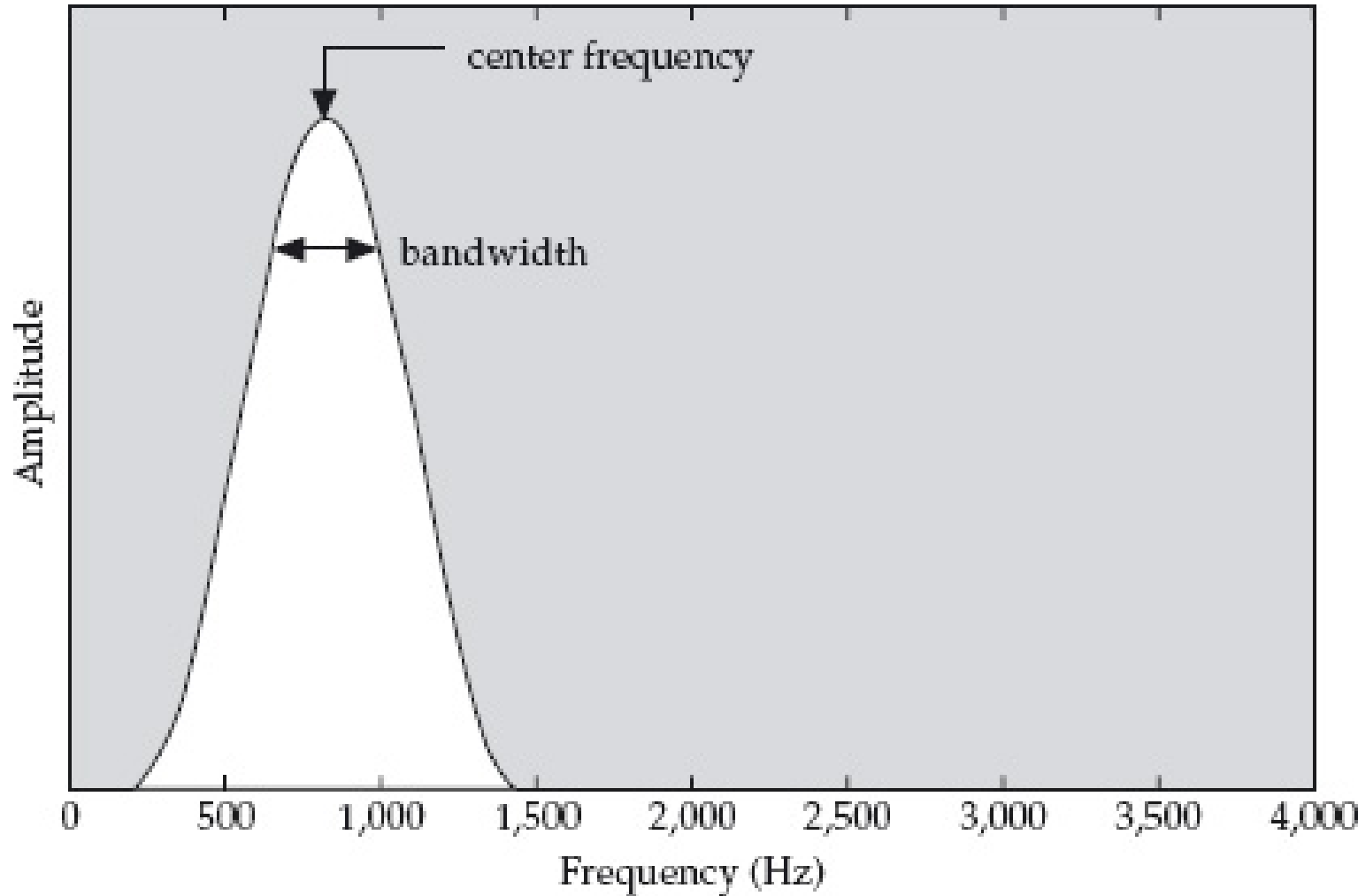


An undamped sine wave, two damped sine waves identical in frequency and phase

- the spectrum of a sine wave: a line
- a damped sine wave does not have an exactly sinusoidal form → more complex spectra
- energy spread over other frequencies near the peak.

The “light damping” waveform is more similar to a sine wave than the “heavy damping” waveform: it has a narrower peak.

# bandwidth



- The spectral result of **damping** the sine wave is **to broaden the peak around the sine wave's frequency**: with more damping the peak gets wider.

# formant bandwidths

The sound waves that resonate in the vocal tract might go on vibrating infinitely, but

- the **sound energy is absorbed by the soft walls and the inertia of air**

Thus, formants have certain **bandwidths**, because the resonant **frequencies** of the vocal tract are **damped**

In **nasals**

the **greater** surface **area** of the vocal tract means that the walls of the **vocal tract absorb more energy** than in non-nasal sounds, and

the **greater volume of air** means that the **inertia of air** within the vocal tract **absorbs more sound** as well.

# Nasal Stops

**Uvular** nasal [N] is the **simplest** nasal consonant to describe.

When the uvula is lowered and the dorsum of the tongue raised to produce an uvular nasal, the vocal tract can be described as a uniform tube that is closed at the glottis and open at the nostrils.

If we know the length of the tube, we can calculate its resonant frequencies, because this is a quarter-wave resonator (like the vocal tract configuration for schwa)



# Nasal Stops → low $F_1$

Fant (1960):

the distance from the uvula to the nares is 12.5 cm,  
the distance from the uvula to the glottis is 9 cm →  
a total length from glottis to nares  $\approx$  21.5 cm.

The four lowest resonant frequencies of the tube  
( $c$  is the speed of sound in air in cm/s):

$$F_1 = c/4L = 35,000/4 \times 21.5 = 35,000/86 = 407 \text{ Hz}$$

$$F_2 = 3c/4L = 1,221 \text{ Hz}$$

$$F_3 = 5c/4L = 2,035 \text{ Hz}$$

$$F_4 = 7c/4L = 2,849 \text{ Hz}$$

The assumption that the vocal tract can be modeled by a uniform tube is wrong: there is a **constriction** of the nasal tract **at the nostrils** (the nose has permanent “lip rounding” 😊).

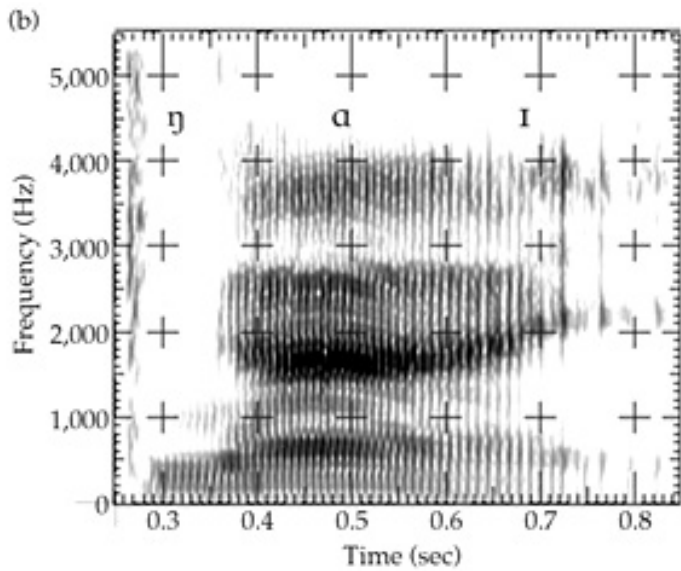
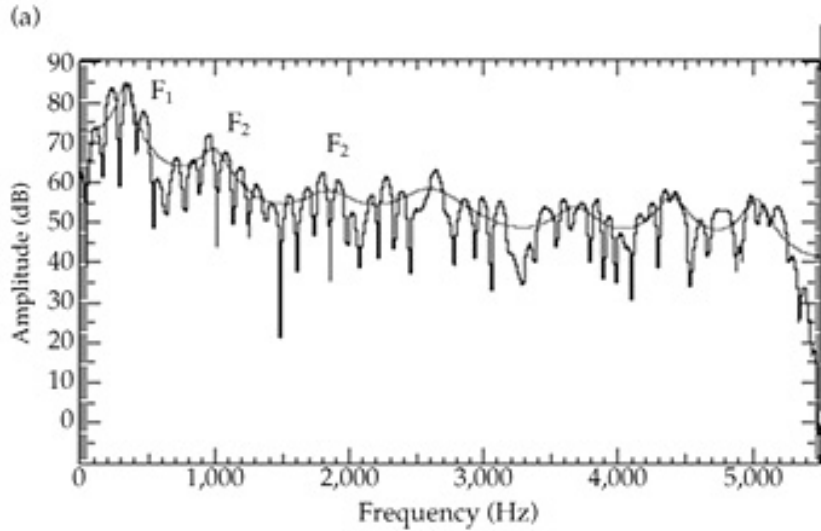
Perturbation theory:

there is a velocity maximum at the nostrils →  
**each of the resonant frequencies will be lower**  
than the calculations above suggest.

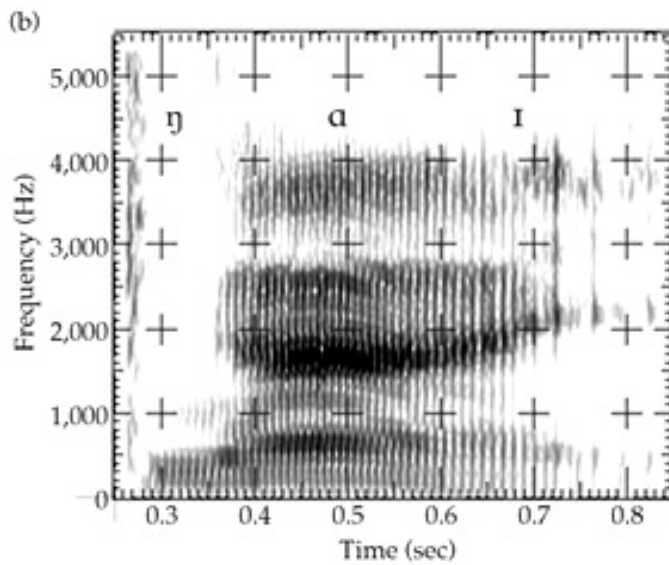
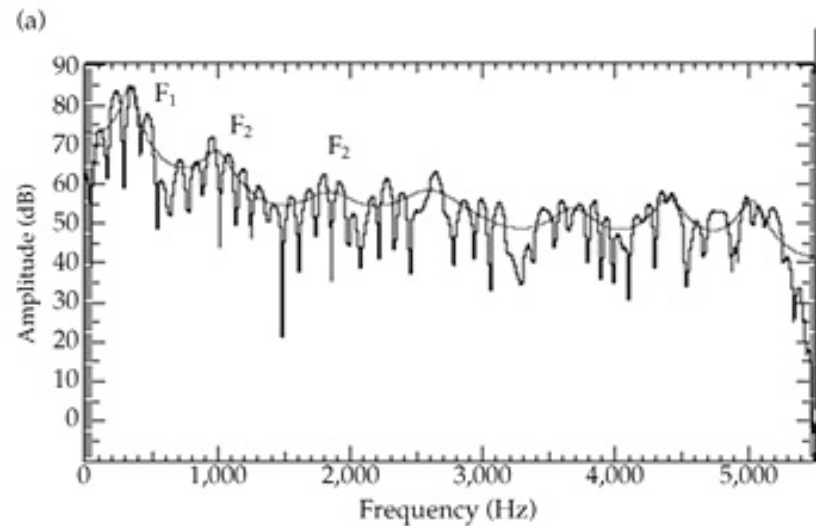
It is hard to make quantitative predictions about the formant frequencies of [N] (the shape of the nasal passage varies), but  
**the formant values will be spaced more closely** in the uvular nasal than they are in [ə];

for a male vocal tract the interval between formants in [ə] is  $\approx 1,000$  Hz, whereas in [N] it is  $\approx 800$  Hz.

***'easy'***



# Thai [ŋâːɪ] ‘easy’



Nasal consonants have **lower amplitude** (than vowels):

larger resonant passageway → damping of the formants

the **vocal tract is more constricted** at the opening to the nasal passages

the vocal tract has **side cavities** like the sinuses (not only those)

# [m]

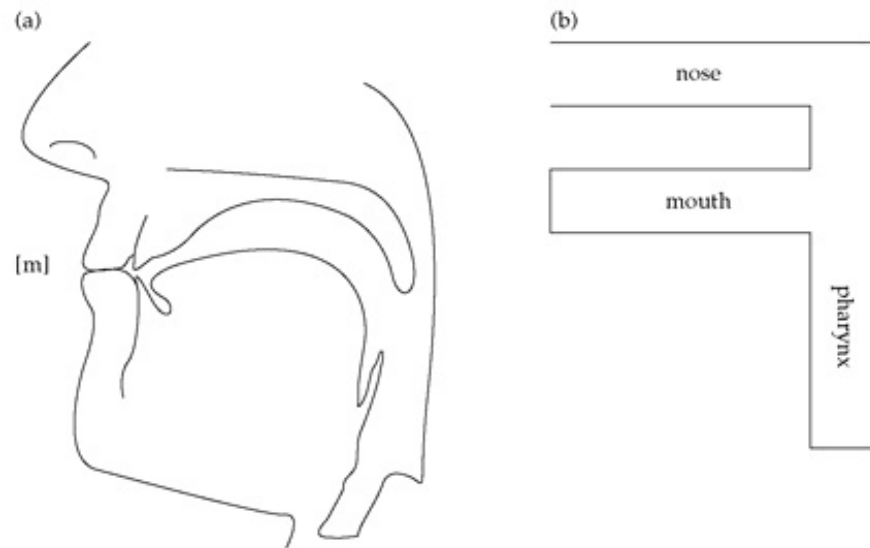
The main difference between [N] and [m] is that **the mouth cavity forms a side branch** in the resonant tube.

The mouth cavity can be modeled as a tube closed at one end (the lips) and open at the other (the uvula), with a length of about 9 cm.

We can therefore calculate the resonances of the mouth cavity as we did for [ə], [N]:

$$c/4L = 35,000 / (4 \times 9) = \mathbf{972}$$

$$\mathbf{3c/4L = 2,917 \text{ Hz}}$$



# anti-resonance

The resonant frequencies of the mouth cavity in nasals are not like that we've seen before:

the mouth cavity **is a side branch** of a larger resonator: it doesn't open directly to the atmosphere.

So frequency components in the sound source that are near the **resonant frequencies of a side cavity** resonate in the side branch without making an appearance in the acoustic output of the acoustic tube system

They are “**absorbed**” in the side branch →

# anti-resonance

the **frequency components** in [m] that are **near the resonant frequencies of the mouth cavity are canceled**, and become **anti-resonances** (anti-formants) in the acoustic output.

**Formants** show up in the spectrum as **peaks** of sound energy, and **anti-formants** show up as pronounced spectral **valleys**.

There are valleys in the spectrum of schwa, but these are the result of a lack of resonance; some frequency components are simply not enhanced as much as others.

In contrast to this, there are some frequency components **in [m]** that are **actively subtracted** from the spectrum.

# Another effect of an anti-formant in the spectrum:

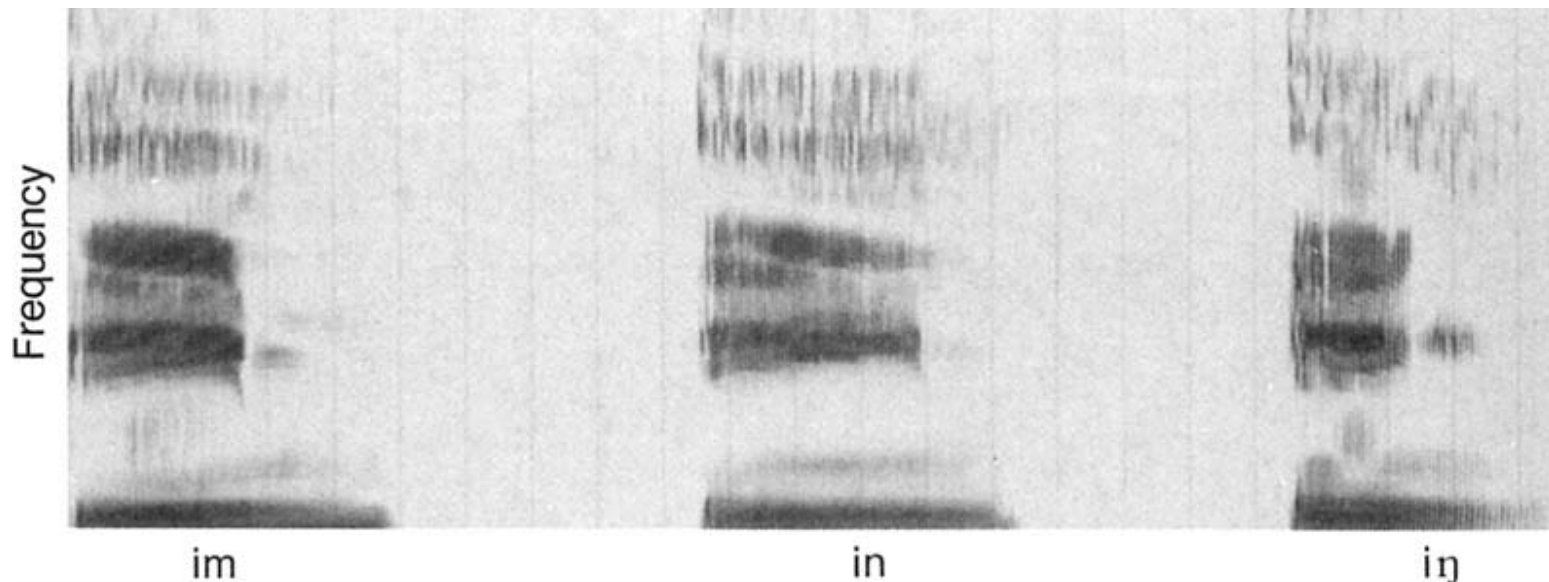
amplitudes of all the formants **above** it are reduced:

nasals have **more energy at the low end** of the spectrum

(+ increased bandwidths also cause the formant amplitudes to be reduced) →

in spectrograms nasals are **lighter than the nearby oral vowels**

The **frequencies of anti-formants** in the spectrum of a nasal stop  
**depend on the length of the mouth cavity**





# [n], [m]

- the mouth cavity in [n] is about 5.5 cm long.

A uniform 5.5 cm tube that is closed at one end and open at the other has resonant frequencies at 1,591 and 4,773 Hz;

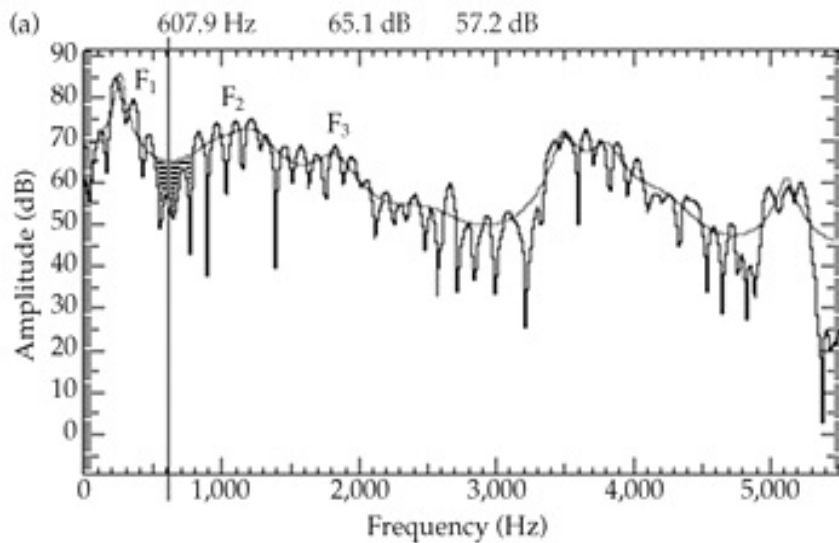
thus we expect

the spectrum of [n] to have an **anti-formants** at **≈1,600 Hz** and **≈4,800 Hz**.

The mouth cavity in [m] is about 9 cm long, → anti-formants at less than **1,000** and **≈3,000 Hz**.

The frequencies of the anti-formants are cues to the place of articulation in these (isolated) nasals.

# [mâɪ] “no”



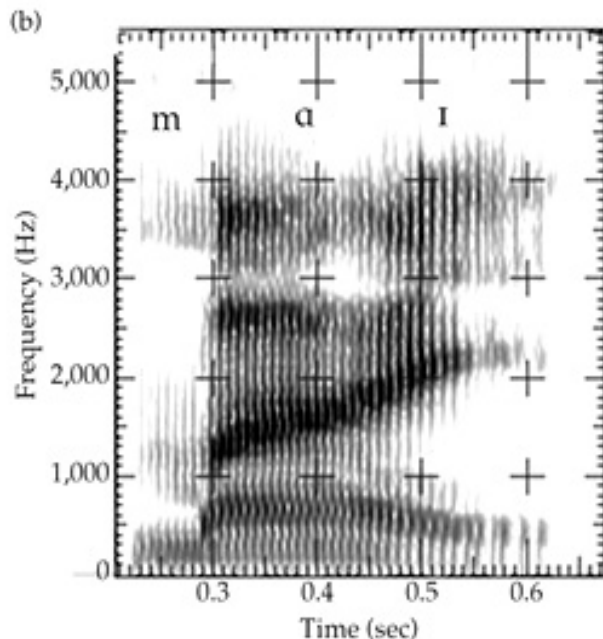
the first anti-formant is indicated by shading and a vertical cursor

an **anti-formant** may or may not appear as a white band in the spectrogram

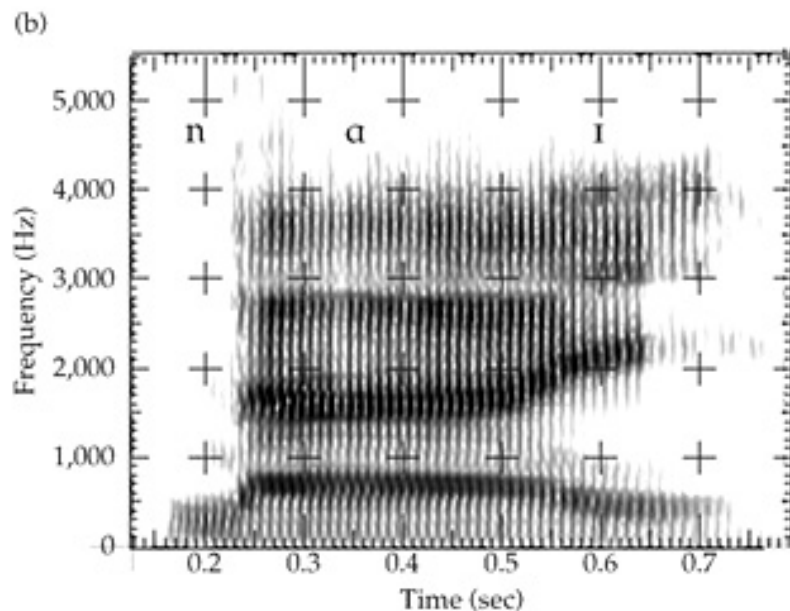
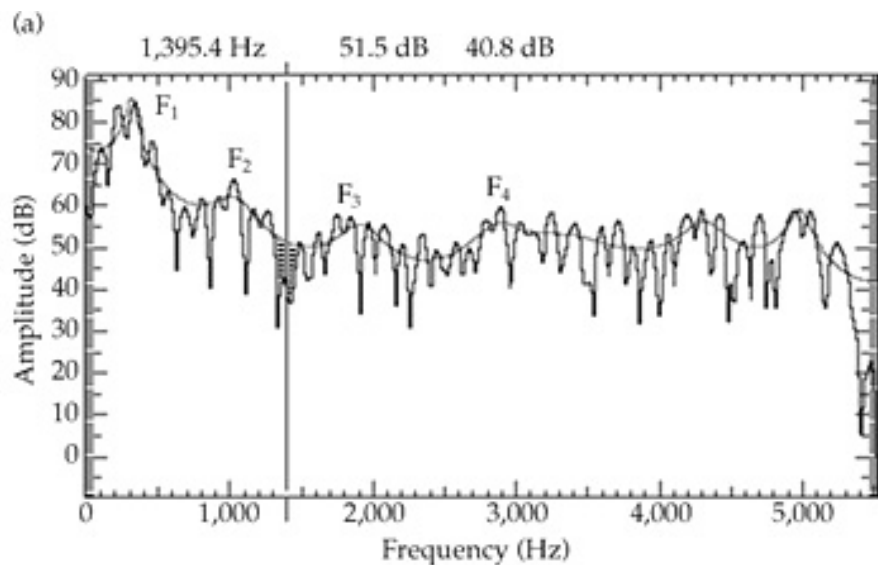
if there are no formants near the anti-formant, it appears **white** in the spectrogram (here);

but if the frequency of the anti-formant is about the same as the frequency of a formant, the net result is to

**weaken the formant peak**



# [na:I] “master”



the spectrum of [n] should  
have anti-formants at  
higher frequencies  
(1,600 Hz and 4,800 Hz)

# the main properties of nasals

- **low F1** (sometimes called the “nasal formant”),
- **close spacing** between formants, and
- the presence of **anti-formants**, whose frequencies are determined by the place of articulation

# Perceiving anti-formants

Repp (1986):

listeners could identify nasal murmurs extracted from syllables starting with [n] and [m] correctly 72% of the time.

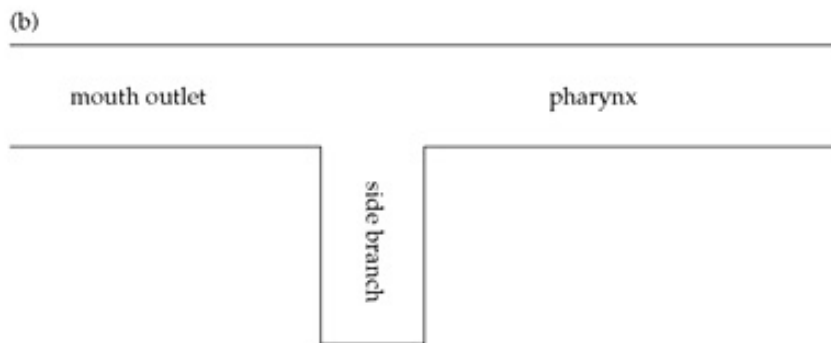
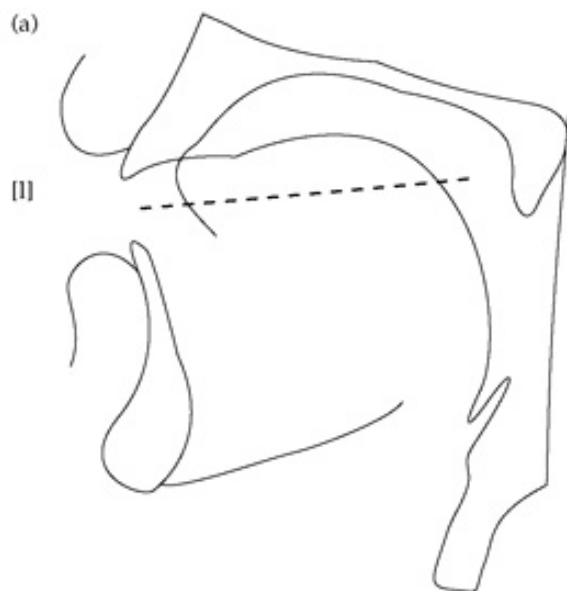
when the vowel formant transitions were included, performance was 94% correct (with only 10 ms of the vowel)

→ the frequencies of the anti-formants in [m] and [n] are not really very important perceptual cues for these consonants.

# Laterals:

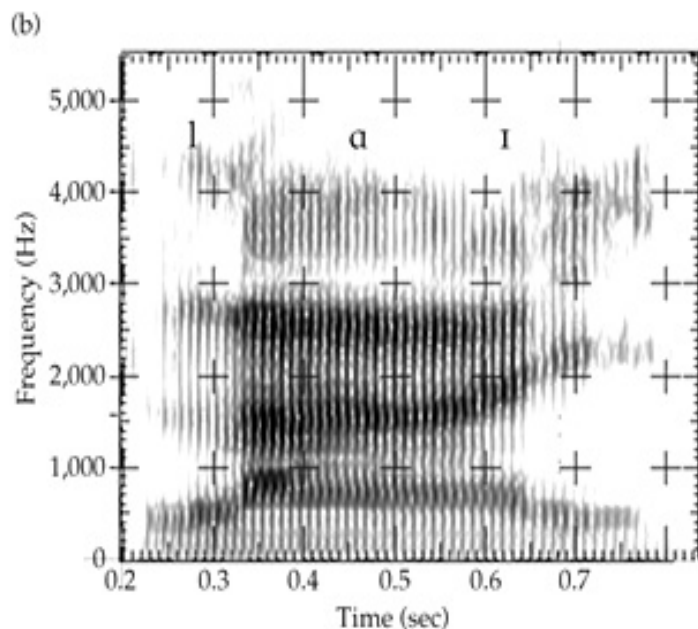
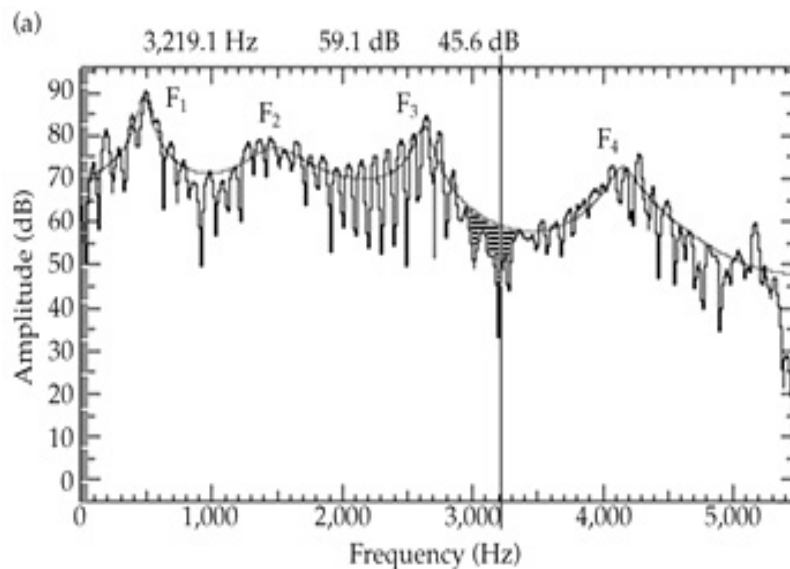
similar to the analysis of nasals:

A small **pocket of air on top of the tongue** acts as a **side branch** (to the main acoustic channel which curves around one or both sides of the tongue) produce **anti-formants** in the spectrum, just as the mouth cavity does in nasals



There may be some asymmetry in the sizes of lateral openings on the left and right sides of the tongue: the opening around one side is more open than the other.

## [laɪ] “stripes”



vocal tract can be modeled as a  
uniform tube that has  
**a short side branch**

$$F = c/\lambda$$

Fant: pocket 4 cm, vocal tract 16 cm  
(10 cm from the glottis to the branch and  
a lateral outlet cavity of 6 cm)

## The resonances

of the vocal tract: 531 / 1,594 /  
2,656 Hz, ( $\approx$  [ə])

of the pocket:  $c/4L = \mathbf{2,125\ Hz}$

Since this is a **side cavity**, this  
resonance becomes an **anti-**  
**resonance** in the output.

oversimplified, of course:

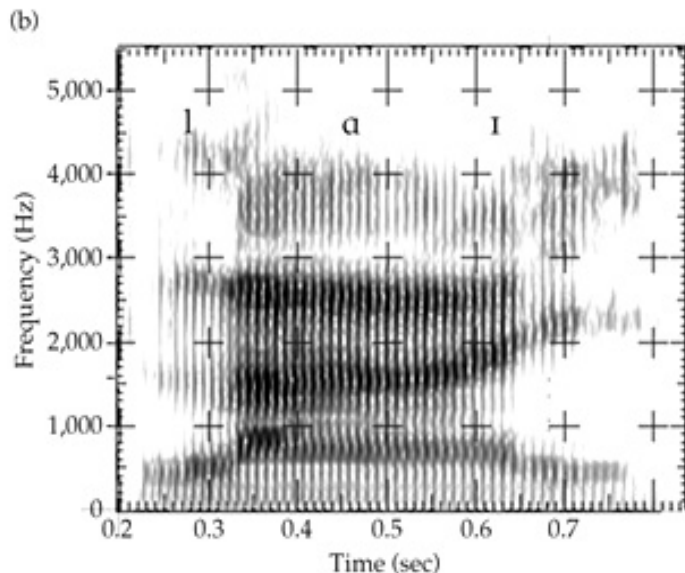
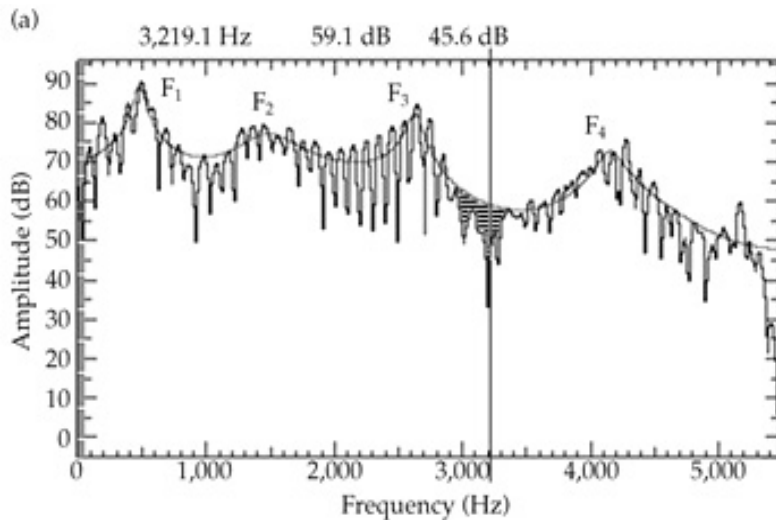
the outlet cavity has a smaller diameter than does the tube from the glottis to the lateral constriction →

the frequency of the F1 is lower than it would be in a uniform tube,

syllable-final [l] is also produced with a dorsal constriction that lowers the frequency of the second formant.

Still, the tube model's basic prediction is that

the spectral “signature” of laterality – **an anti-formant near the frequency of F3.**





# Laterals & Nasals

As with nasal sounds, the presence of an anti-formant causes the amplitudes of all higher formants to be reduced.

## Differencies:

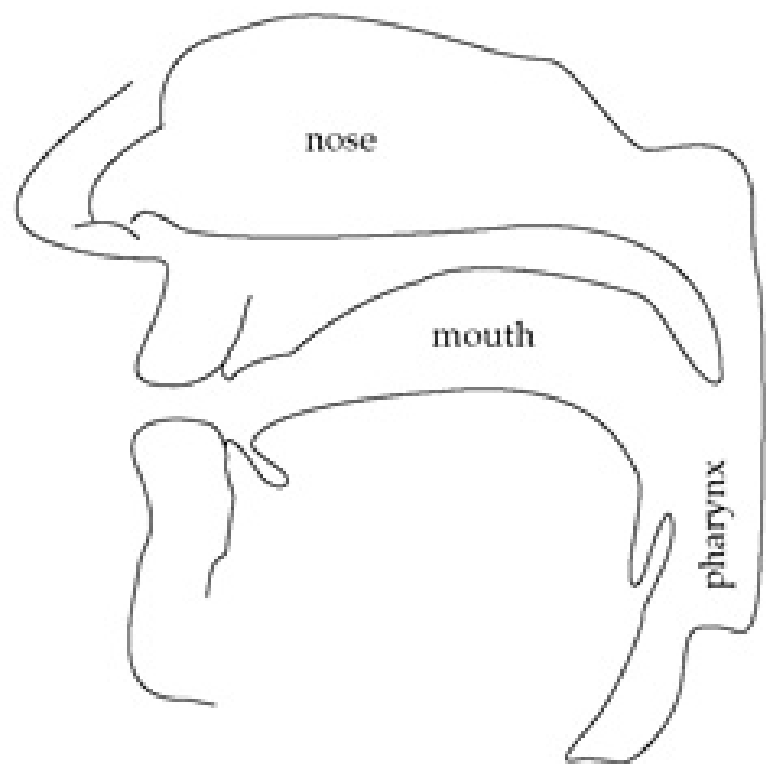
the **average spacing of the formants is wider in laterals** than it is in nasals ← the primary resonant tube in nasals is longer than it is in laterals.

the average spacing between formants in nasals is about 800 Hz, while in laterals it is 1,000 Hz.

# Nasal vowels:

the most complicated configuration of the vocal tract: **2 resonant systems at once:**

the pharynx cavity + the mouth cavity,  
the pharynx cavity + the nasal cavity.



- the pharynx + mouth system – the **oral tract** – has resonances at about 500 Hz, 1,500 Hz, and 2,500 Hz;
- the pharynx + nose system – the **nasal tract** – 400, 1,200, 2,000 Hz.

**All these formants are present in the spectrum of nasalized vowels.**

Of course, the resonant frequencies of the oral tract can be modified by movements of the tongue and lips, constriction of the nose at the nares.

Still, the tube model predicts that **the spectrum of a nasalized vowel will have a lot of formants.**

	<i>Nasal formants</i>	<i>Oral formants</i>	<i>Anti-formants</i>
	( $L = 21.5$ cm)	( $L = 17.5$ cm)	( $L = 12.5$ cm)
$F_1$	407	500	680
$F_2$	1,221	1,500	2,040
$F_3$	2,035	2,500	–

Above with nasal consonants: closed mouth has resonances that become **anti-resonances** (anti-formants) in the spectrum

In nasalized vowels:

the **resonances of the nasal cavity become anti-formants**, analogous to the anti-formants in nasal consonants, but now the **mouth is more open than the nose**, so the acoustic coupling between the mouth and the atmosphere is greater than between the nasal passage and the atmosphere.

**below 1,000 Hz** non-nasalized vowels have one resonance (F1), nasalized vowels have three: an oral formant (F1o), a nasal formant (F1n), and an anti-formant (A1).

- F1o depends on the positions of the lips and tongue, and
- **A1 depends on the degree of nasalization:** increases as the degree of nasal coupling increases → may cancel F1n when there is slight nasalization and cancel F1o at a higher degree.

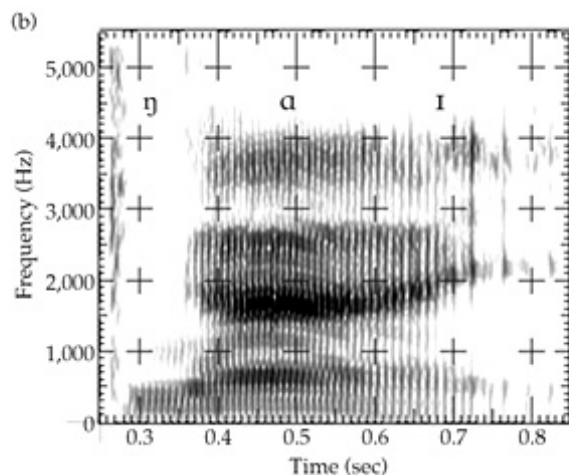
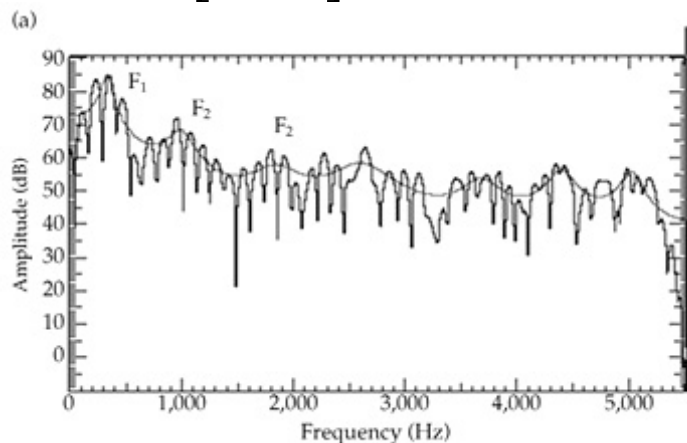
[a] with heavy nasalization can appear on a spectrogram to have a much lower F1 than normal because F1n is low (and will always be low, regardless of the vowel) and A1 is high enough to cancel some energy of F1o.

This type of complicated interplay between nasal formants, oral formants, and nasal anti-formants that occurs in the region of F1 also happens at other Fs.

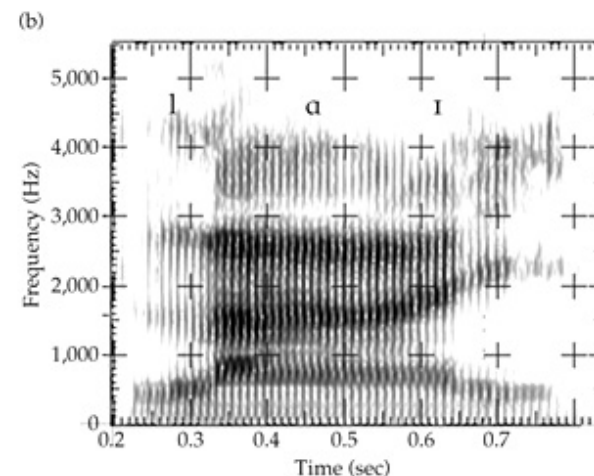
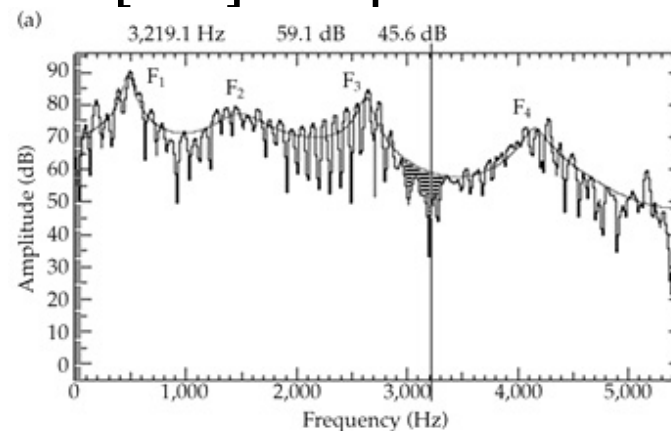
no surprise that **the number of distinctive nasal vowels in languages is usually smaller (and never larger) than the number of oral vowels.**

There is **no articulatory problem** in producing nasalized vowels, but they are **acoustically and perceptually more complex** than oral vowels.

## [na:I] “master”



## [la:I] “stripes”



interplay of formants and anti-formants in the F1 region of nasalized vowels:

- bandwidth of F1 is increased: usually two formants in the region, not one  
[na:i] vowel is very nasalized, closely spaced formants are like the F1 and F2 of [u], seem to merge into one
- the F1 of [la:I] is easier to identify than the F1 in the nasalized vowel

# Based on:

Кодзасов С.В., Кривнова О.Ф. *Общая фонетика*.  
М., 2001.

Johnson, Keith. *Acoustic and Auditory Phonetics*.  
*3rd edition*. Wiley-Blackwell (2011).

Ladefoged Peter. *Elements of Acoustic Phonetics*.  
University Of Chicago Press, 1995.

*Speech science primer : physiology, acoustics, and  
perception of speech* / Lawrence J. Raphael,  
Gloria J. Borden, Katherine S. Harris. – 6th ed.  
Lippincott Williams & Wilkins. 2007.