Introduction to Acoustic Phonetics IV

Sergey Knyazev

10 June, 2017

nonresonant consonants

Previously:

semivowels and **nasals**,

characterized by a relatively **free flow of air**

and acoustically – by **formant structure**.

Now:

nonresonant consonants:

much more **restricted airflow**

acoustically, little or no of formant structure

nonresonant consonants

The obstructions to the airflow are caused

by the articulators forming constrictions and occlusions within the vocal tract that generate

aperiodicity (**noise**) as the airflow passes through them

This aperiodic source of sound is resonated in the tract in much the same way as the periodic source produced by phonation.

The most **effective resonators for noise** sources are those **immediately anterior to the constrictions and occlusions** that produce them.

Resonant consonants and vowels are **voiced**: **without** a periodic source, many of the resonant sounds would be **inaudible**

Nonresonant consonants:

noise in the speech signal makes the sounds audible whether or not phonation accompanies their articulation →

the possibility of using a single articulation to produce two distinctive speech sounds, one phonated (voiced) and the other unphonated (voiceless):

cognates.

nonresonant consonants = obstruents

- **Stops**
- **Affricates**
- **Fricatives**

Fricatives are continuants, they can be prolonged.

As all the other speech sounds, fricatives are the product of a **sound source** (sometimes two sources) **modified by a resonator** and by the effect of the **sound radiating** at the mouth opening. The fricative noise originates at the articulatory constriction.

Fricatives

The aperiodic source that marks fricatives is created in the vocal tract by sending the breath stream through constrictions formed in the tract.

The airflow must be strong enough and the constriction narrow enough to make the airflow turbulent, creating frication (noisy random vibrations).

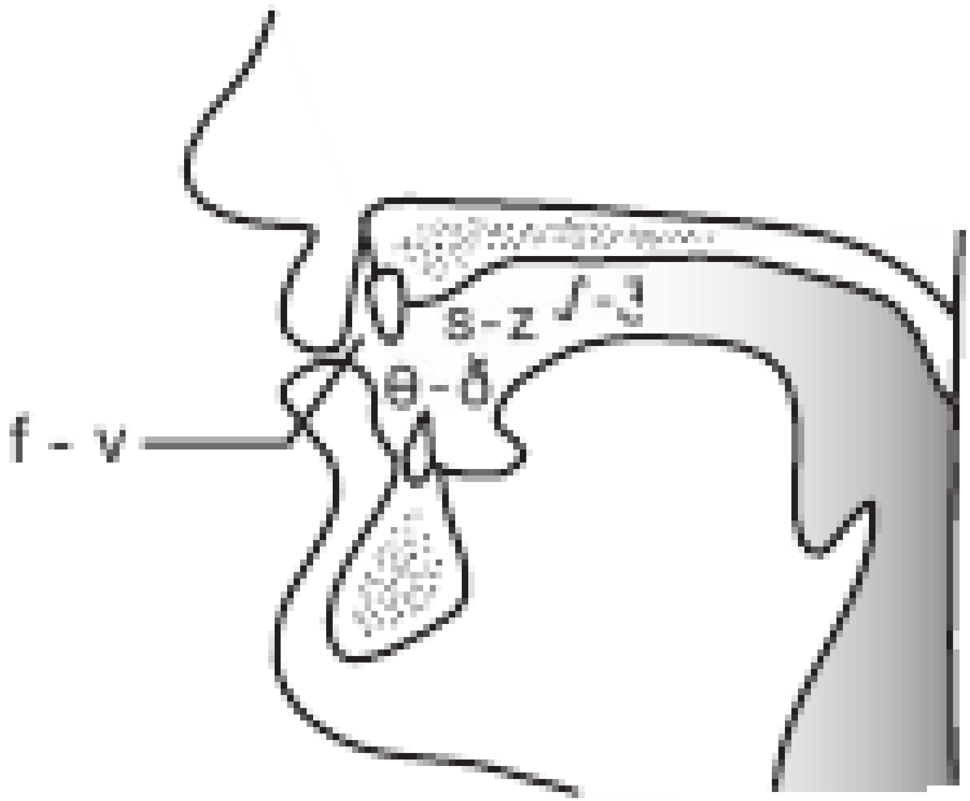
The fricative sounds are thus produced by compressing a continuous flow of air through a constriction formed by closely approximating two articulators.

If the glottis is open, the airstream is made audible only at the point of constriction, but if the glottis is closed with the vocal folds vibrating, the result is two sound sources, the periodic sound of phonation and the aperiodic sound of the airstream passing through the constriction.

To develop enough air pressure in the oral cavity to produce noise, the velopharyngeal port is closed

In English 5 primary places of articulation:

labio-dental, linguadental, alveolar, postalveolar (the front of the hard palate), glottal



Articulation: non-sibilants

labiodental fricatives /f/ and /v/ are formed by bringing the lower lip close to the inferior edges of the upper central incisors.

linguadental fricatives /θ/ and /ð/ are formed by approximation of the tip of the tongue with the upper incisors

labiodental and linguadental fricatives are similar in production, they are also similar in their acoustic properties,

aspirate /h/: the constriction in the larynx at the glottis.

- usually voiceless, but
- can be voiced between phonated sounds (*ahead*),
- the vocal tract takes the shape of whatever vowel is to follow.

The **sibilant**, or strident, fricatives [s] and [ʃ] have louder frication noise than do the others

Articulation: sibilants

for [s] and [z] the constriction is formed between the alveolar ridge and the tongue, but speakers vary in which part of the tongue is elevated

the constriction between the alveolar ridge and the tip of the tongue (a high-point /s/), or the blade of the tongue (low-point)

[s], [z]: a groove along the tongue midline to channel the airstream, accomplished by raising the lateral edges of the tongue to the medial edges of the upper teeth while depressing the center of the tongue: to prevent the airstream from passing over the sides of the tongue,

A second constriction between the upper and lower incisors must be narrow so that the airstream is directed over the edges of the teeth, creating turbulent airflow behind the teeth.

The difficulties in producing /s/ and /z/ for someone who has an open bite or is missing front teeth

The postalveolar /ʃ/, /ʒ/ : the constriction is made a bit farther back, in the postalveolar area, and the midline groove is considerably shallower than for /s/, /z/. In addition, the lips may be somewhat rounded and protruded.

Fricatives: acoustics

The source of noise (acoustic energy) in fricatives is (turbulent) airflow, which is produced as air escapes from a narrow constriction in the vocal tract.

This aperiodic noise is filtered by the vocal tract.

This noise is the acoustic consequence of irregular air molecule motions that are produced when air exits a relatively narrow channel

When fast-moving air in the channel hits inert outside air, the airstream becomes chaotic

Turbulence

The main factors that determine whether airflow is **turbulent:**

- the size of the channel and
- the volume velocity of the airflow (volume of air going past a certain point per unit time).

If 100 cm^3 per second of air flows through a channel, **turbulent airflow** is created if the channel area is less than 10 mm^2, but not if the channel area is 20 mm^2.

It's easier to get turbulent airflow from a narrow straw than a wide one

Fricatives: the aerodynamics of freeways

Particle velocity: the speed of air particles, is like the speed of a car

Volume velocity: the number of air particles passing a point = the number of cars per unit time crossing a particular place.

Laminal flow: all the cars stay in their lanes, and move smoothly

Turbulent flow: cars change lanes a lot; there is some random side-to-side motion in addition to the forward movement.

Channel turbulence: the highway widens from two lanes to six. Cars shoot out of the narrow passage and change lanes when the road widens.

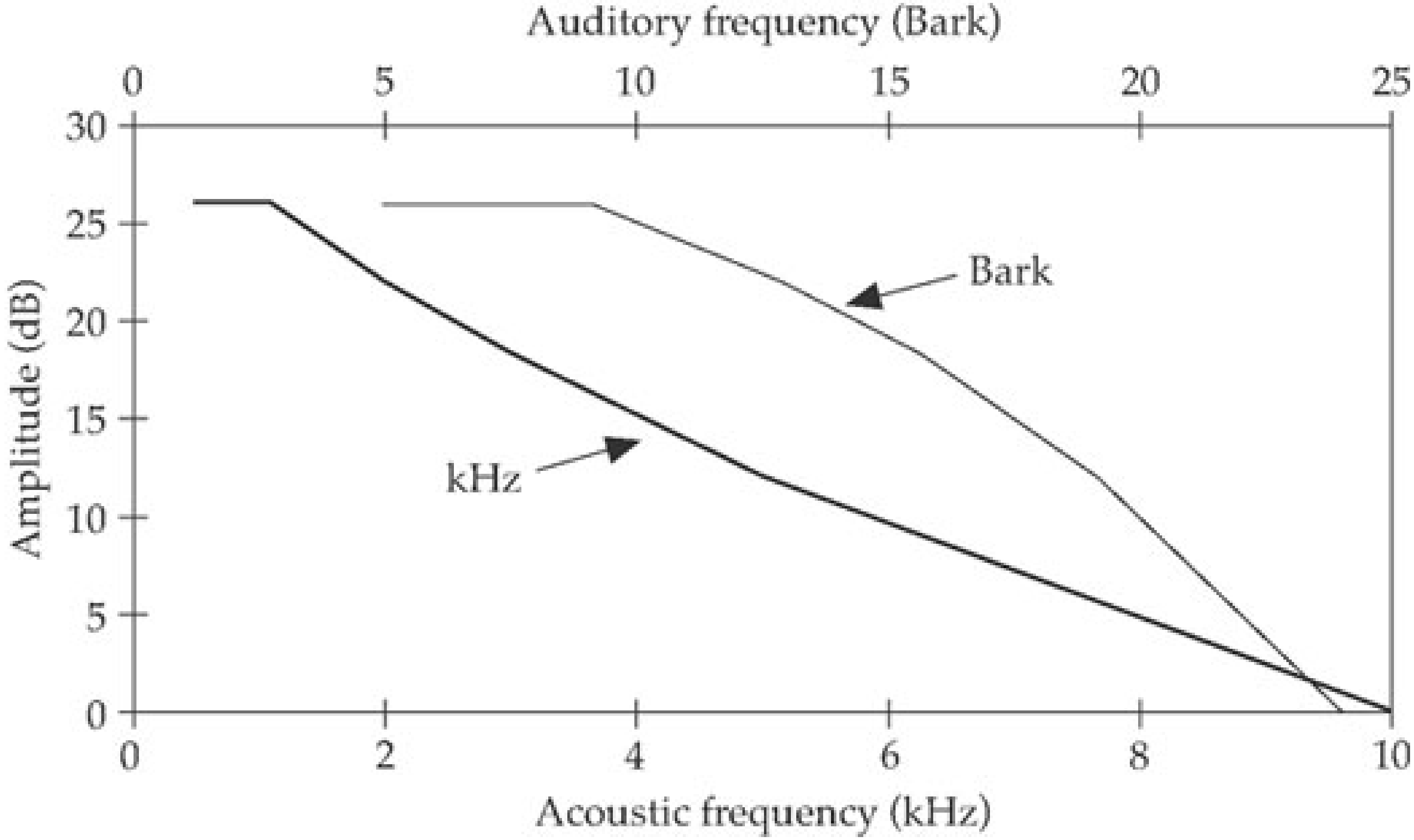
Obstacle turbulence is like the lane-changing that occurs when a car is stalled on the freeway. Cars change lanes to move around the obstacle.

Aerodynamic impedance, the resistance of a channel to airflow, is seen in the difference between two-lane and four-lane roads. Larger channels are generally capable of greater volume velocity than narrow channels. That is, larger channels impede flow less.

If volume velocity remains constant, particle velocity increases as channel width decreases.

When the number of lanes goes from four to two, there is a big slow-down just before the change, and then cars go faster once they get into the narrow part.

In order for the same number of cars to pass through the narrow part as through the wide part (equal volume velocity), the cars have to go faster in the narrow part (higher particle velocity).



Because the motions of air molecules in turbulent airflow are irregular, the sound pressure waves associated with turbulent airflow are random

The spectrum of **turbulent noise** normally seen in fricatives is not completely flat: the spectrum above 1,000 Hz has gradually decreasing amplitude with increasing frequency.

This type of noise occurs in fricatives and in most voiceless sounds.

in [h] turbulence is created when air from the lungs passes through the glottis, because the passage between the vocal folds is relatively narrow, the rate of airflow is relatively high

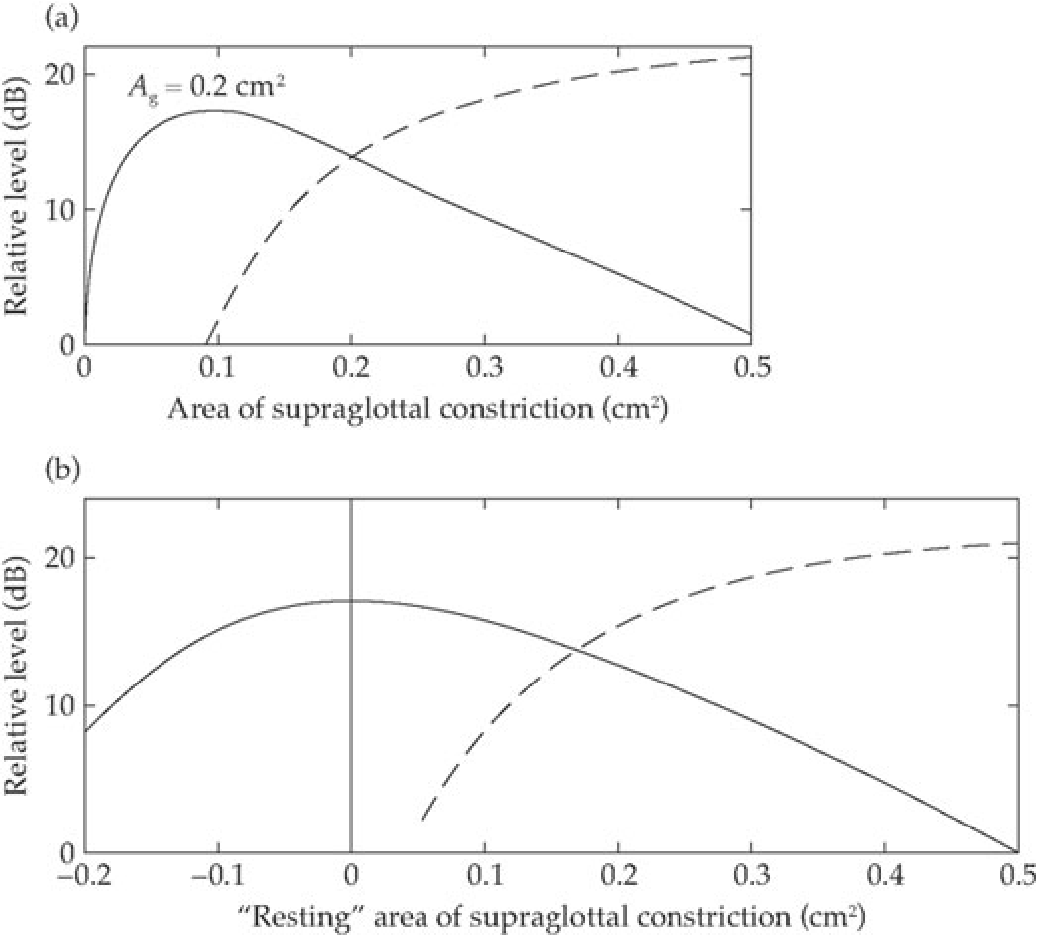
how long one can sustain [h] versus [ɑ]

the main difference between the [h] and the vowel in the word *heed* is that the sound source in [h] is aperiodic turbulent noise produced at the glottis, whereas the sound source in [i] is a complex periodic wave (voicing).

Both the vowel and the [h] have formant values, which are due to the filtering action of the vocal tract

Other fricatives also have acoustic energy in specific vocal tract resonances,

depending the place of articulation



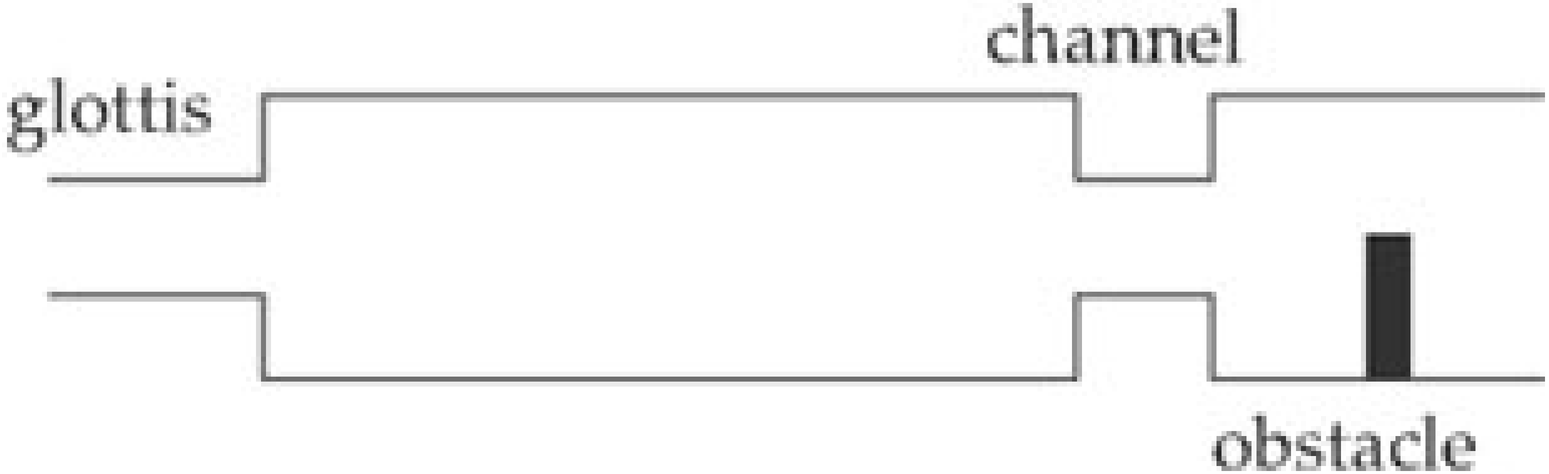
The **amplitude** of turbulent noise is determined by the **velocity of the air** molecules as they pass through a channel.

The faster the air molecules move, the louder the sound.

particle velocity is related to channel area → **for a given rate** of airflow out of the mouth (volume velocity), **the narrower the channel, the louder the turbulent noise**.

1the amplitude of the fricative noise is shown on the vertical axis, and the area of the constriction is shown on the horizontal axis.

2 the same relationship, but with the assumption that air pressure widens the constriction slightly.



in addition to being produced when a jet of air escapes from a narrow channel, turbulent noise is also produced when a jet of air hits a downstream obstacle (**sibilants**).

The presence of an obstacle results in increased amplitude of turbulent noise (a car is blocking a lane),

almost all fricative noises involve turbulence produced by airflow hitting an obstacle.

In [s] and [ʃ] the upper and lower teeth respectively function as turbulence-producing obstacles;

even the amplitude of [f] decreases when you (manually) hold your upper lip up, suggesting that the upper lip functions as an obstacle in the production of the labio-dental fricative.

(labio-dental fricatives are more common in the languages of the world than are bilabial fricatives, because obstacle turbulence tends to be louder than channel turbulence.)

distinction between fricatives

“wall” **source:** a jet of air hits a wall in the vocal tract – [x] (the obstacle is nearly parallel to the airflow)

“obstacle” sources: making a jet of air that hits the front teeth (the obstacle is more nearly at right angles to the airflow)

“lip” source [v] and [f]: the airstream against the upper lip.

The only nonobstacle fricatives are produced at the lips and at the glottis

it is difficult to make bilabial and glottal fricative noises loudly.

/h/ is often produced with oral fricative allophones [x] or [ç].

Voiced fricatives are relatively unusual:

they are surprisingly difficult to produce:

high volume velocity is needed to produce the turbulent noise, and the vibrating vocal cords impede the flow of air through the vocal tract

because during voicing the vocal cords are shut (or nearly so) as much as they are open →

given a comparable amount of air pressure produced by the lungs (the subglottal pressure), the volume velocity during voicing is much lower than it is when the glottis is open.

A certain degree of airflow is necessary in order to produce turbulence → voiced fricatives may lose their frication, and become glides.

This alternation does not necessarily involve a change in the degree of vocal tract constriction:

either a voiceless fricative or a voiced glide for the same degree of vocal tract narrowing.

Acoustic coupling

Constrictions can cause acoustic impedance in much the same way that they can cause aerodynamic impedance, but these two types of impedance are different:

when mouth and nose closed as during a voiced bilabial stop, no air escapes from the vocal tract, but sound does:

sound waves produced at the glottis can vibrate through cheeks

Acoustic coupling refers to acoustic impedance: the degree to which sound is transmitted across some acoustic barrier.

In fricatives, the **front and back cavities are not very well coupled**, and the acoustic impedance at the vocal tract constriction is relatively large.

although air flows through the constriction, relatively little sound is transmitted to or from the back cavity.

Consequently, the **resonant frequencies of the back cavity have relatively little impact** on fricative spectra.

Place of Articulation in Fricatives

a series of fricatives starting from a pharyngeal constriction and moving the constriction forward to the alveolar ridge →

a change in the “pitch” of the fricative

caused by changes in the filtering action of the vocal tract ← **the length of the front cavity of the vocal tract**.

In the tube model: the front cavity from the narrow channel of constriction to the lips, and includes the obstacle.

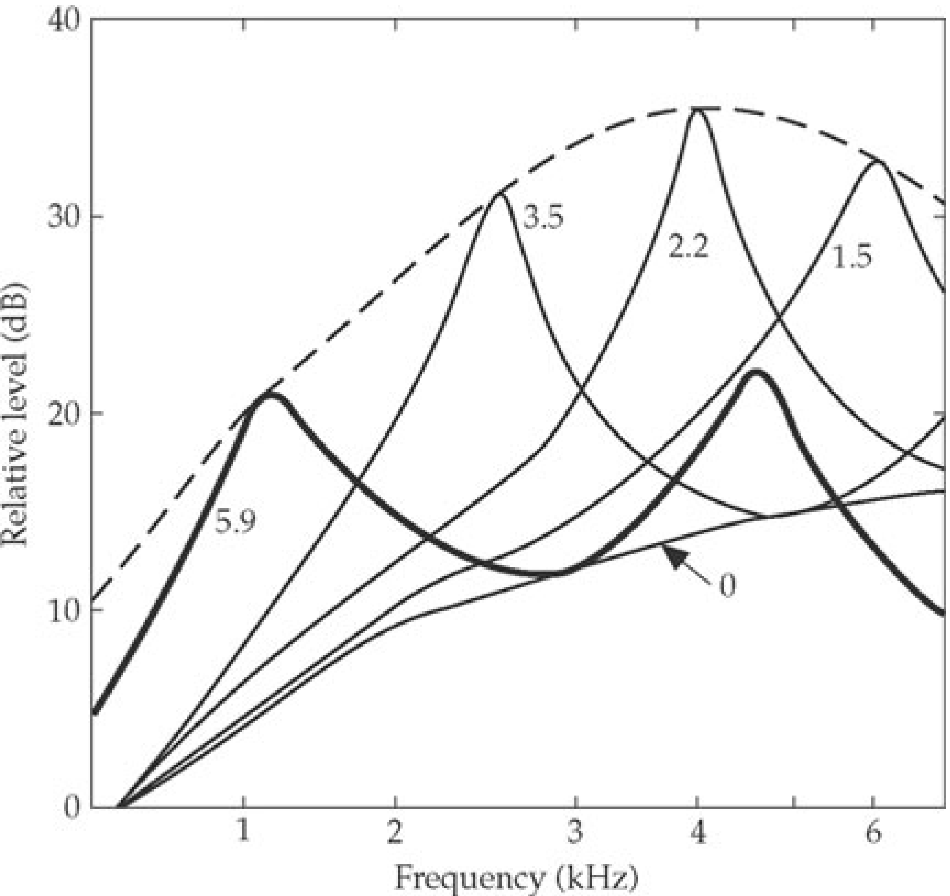
The sound source is located in the front cavity, and the **acoustic coupling** between front and back cavities is weak (the vocal tract is tightly constricted) →

the acoustic filtering action of the vocal tract in fricatives is determined primarily by the resonant frequencies of the front cavity ($F_n = (2n-1)c\,/4\,L$)

When the fricative constriction is located in the pharynx, the front cavity is long, and consequently has lower resonant frequencies than when the fricative constriction is located further forward in the mouth.

Fricative spectra produced by a tube model

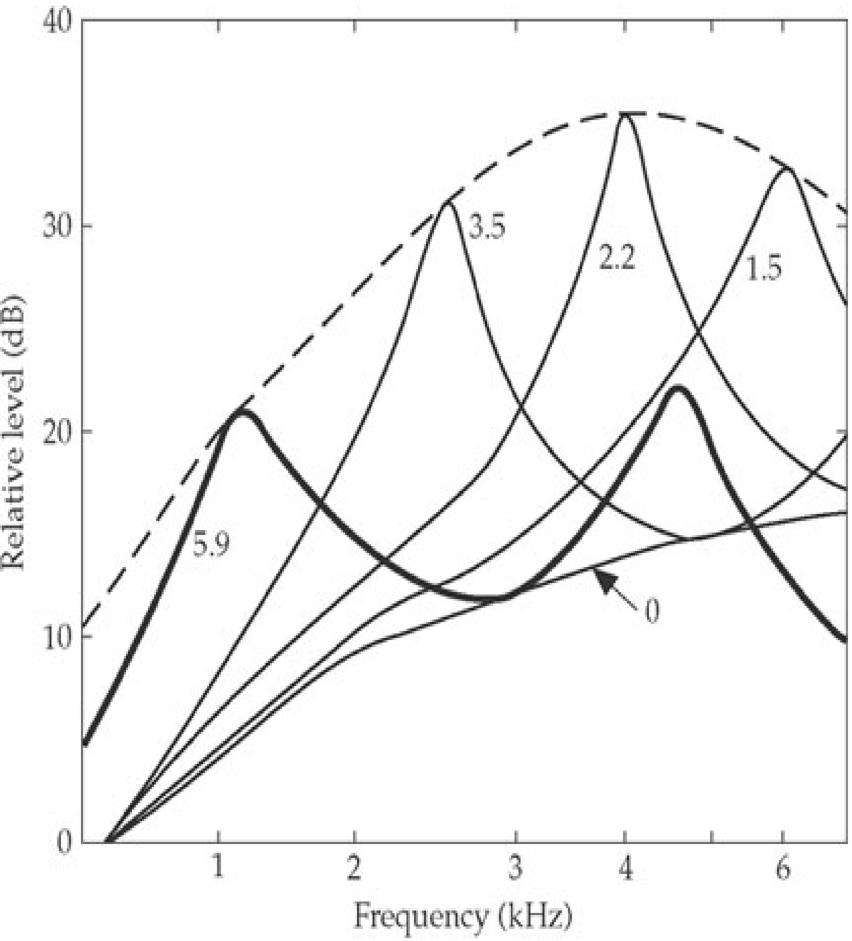
The spectra for five models having different front cavity lengths (indicated in cm).



The shorter the front cavity, the higher the frequency of the lowest spectral peak

In the **pharyngeals** [ħ] and [ʕ] the source of acoustic energy is in the pharynx, and the resonating **vocal tract is relatively long**

Fricative spectra produced by a tube model



Labials [f], [v], produced very far forward in the mouth, may have no vocal tract filtering at all:

made with turbulence at the lips →

there is almost no vocal tract in front of the constriction to filter the sound →

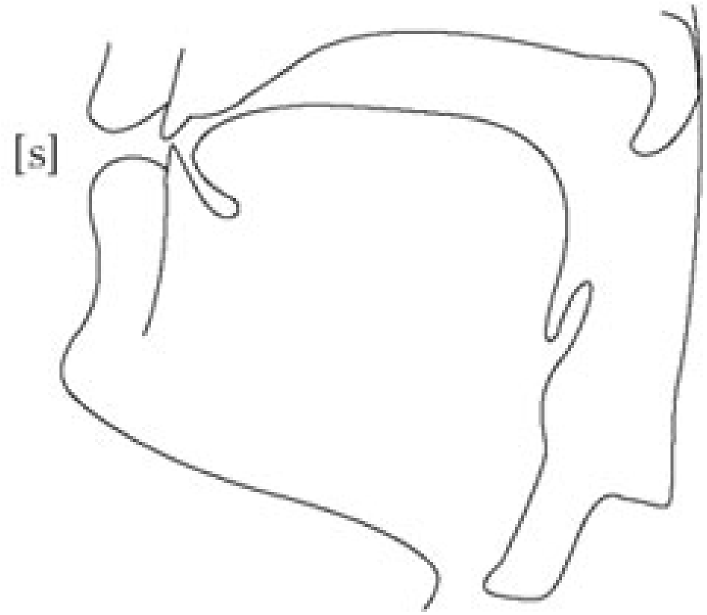
the spectra do not have peaks,

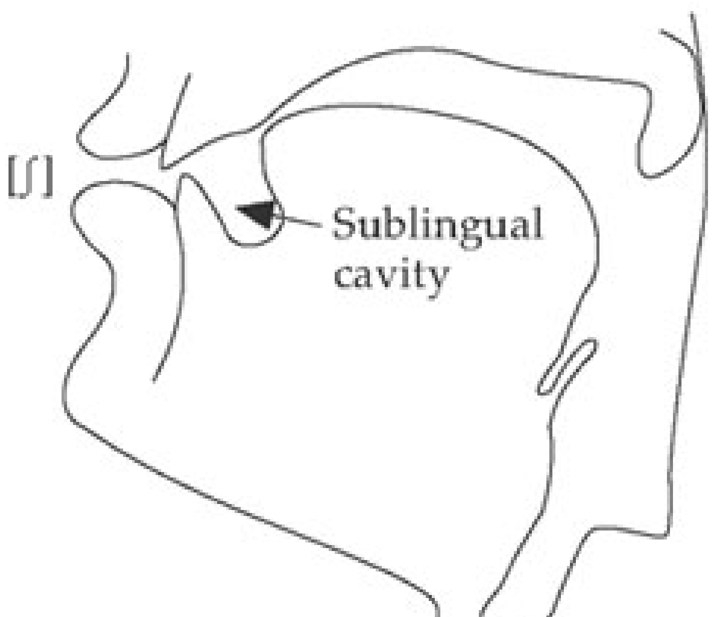
rather have energy spread over a large frequency range

<u>Jakobson et al. (1952):</u>

spectra with a single prominent resonance peak **[compact],** and

with little or no vocal tract filtering or conversely a long front cavity and thus several spectral peaks **[diffuse]**





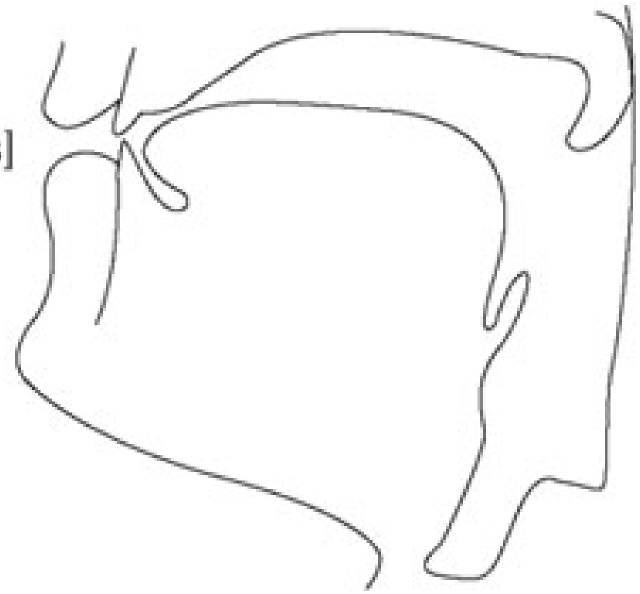
The production of [ʃ] in many languages is affected by the presence of a **sublingual cavity**

(check this: during the pronunciation of a sustained fricative, insert a toothpick between your teeth, and poke your tongue. If you poke the bottom of the tongue, you have evidence that there is a sublingual cavity in the production)

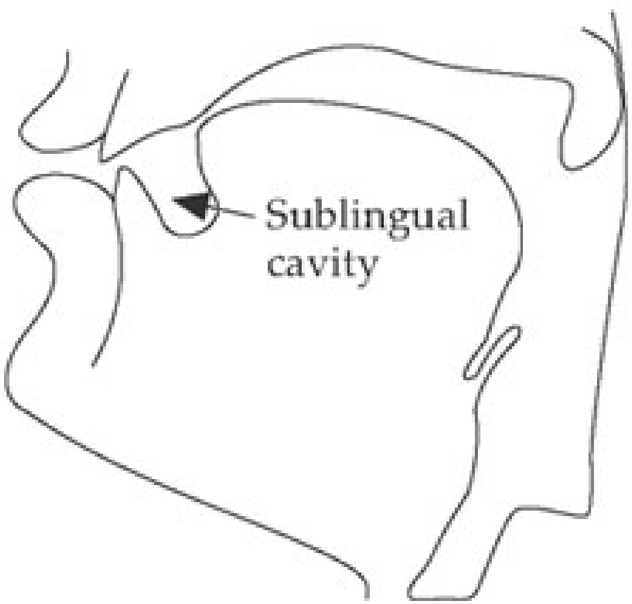
This space below the tongue effectively **adds length to the front cavity** of the vocal tract, and thus **lowers its resonance frequency →**

[s] and [ʃ] have quite similar places of articulation, but their acoustics are markedly different, because [ʃ] has a sublingual cavity and [s] has either a very small one (for tongue-tip-up [s]) or none at all (for tongue-tip-down [s]).

[s]



[ʃ]



Lip rounding has the same acoustic effect in fricatives as in vowels:

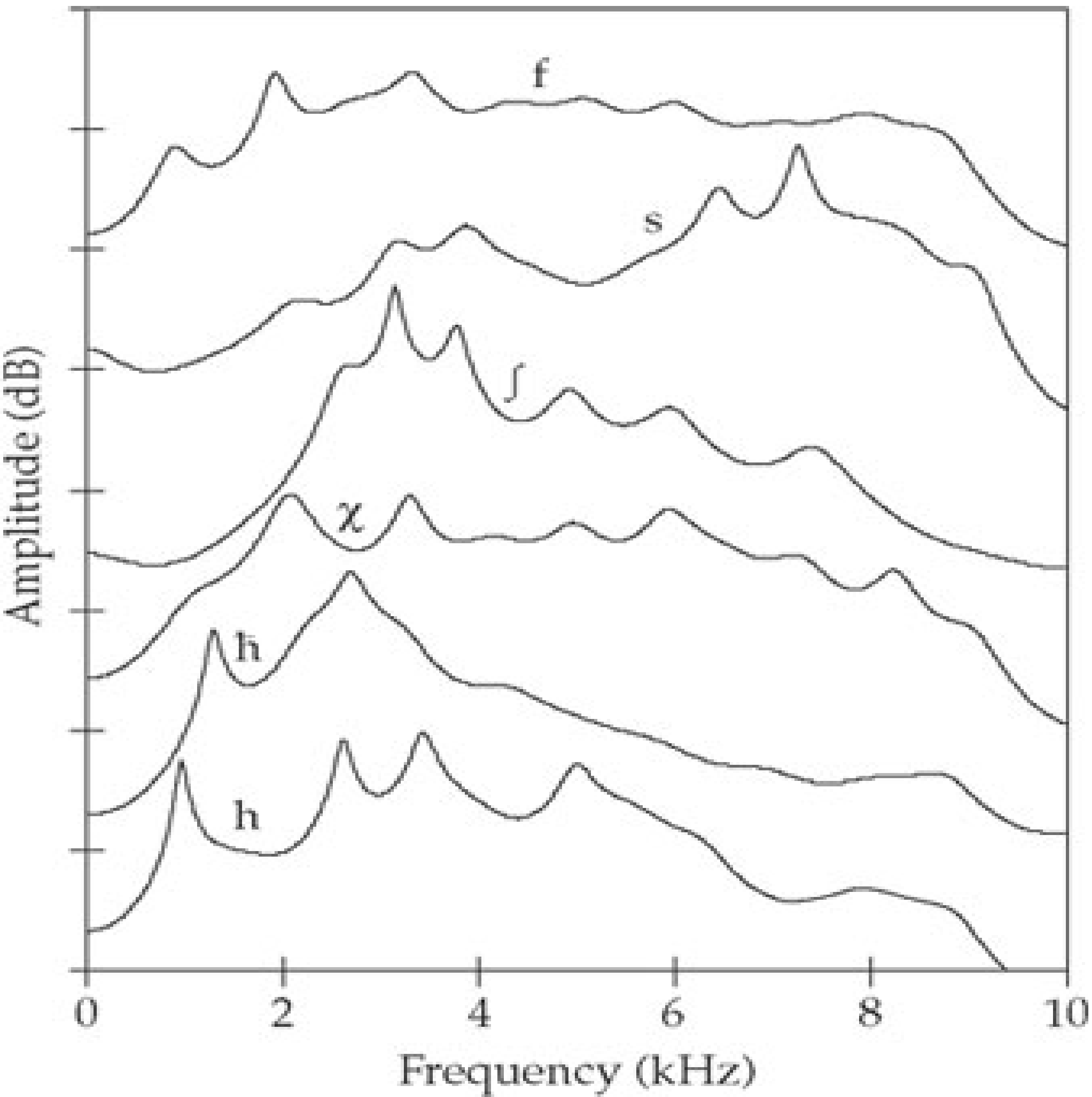
lowers formant frequencies

[ʃ] is usually produced with lip rounding

this *enhances* the acoustic difference between [s] and [ʃ]

But: [s] in *seen* is unrounded, while in *soon* it has coarticulatory rounding from the following [u]: hear the difference

acoustic spectra of six voiceless fricatives (Egyptian Arabic, female speaker):



[fæːt] *to pass*

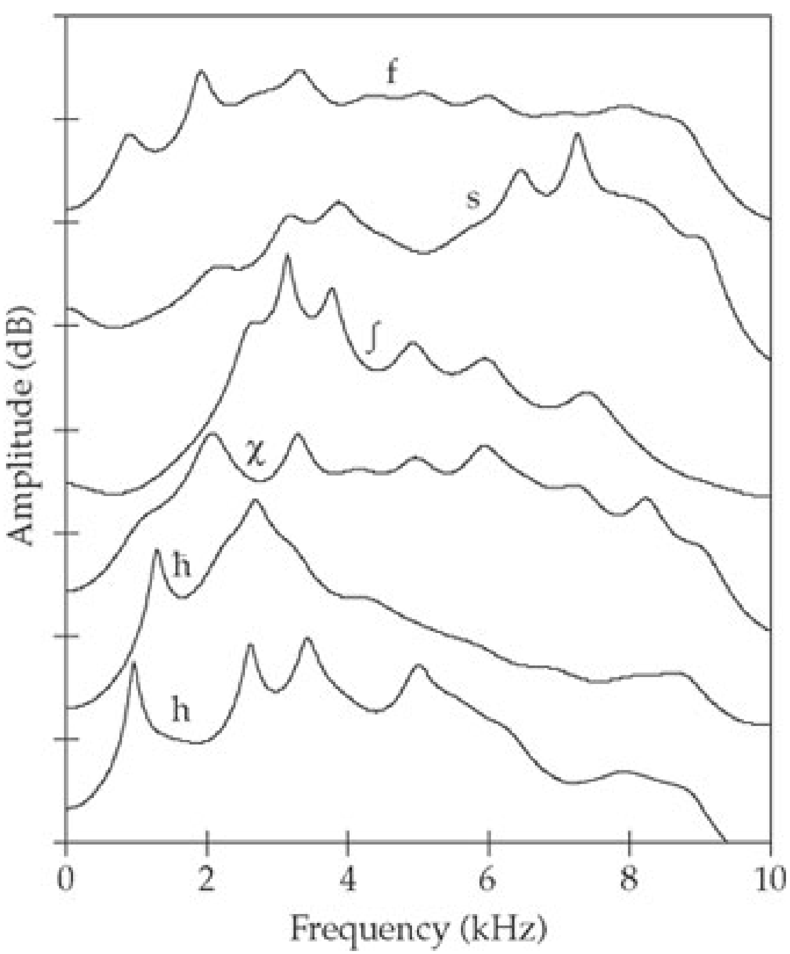
[sæːb] *he left*

[ʃæːb] *to grow old*

[χæːl] *maternal uncle*

[ħæːl] *situation*

[hæːt] *give me*



The spectral peak in [ʃ] occurs at about 3.5 kHz, while the spectral peak in [s] is near 8 kHz,

It is difficult to measure the acoustic characteristics of fricatives (particularly coronals):

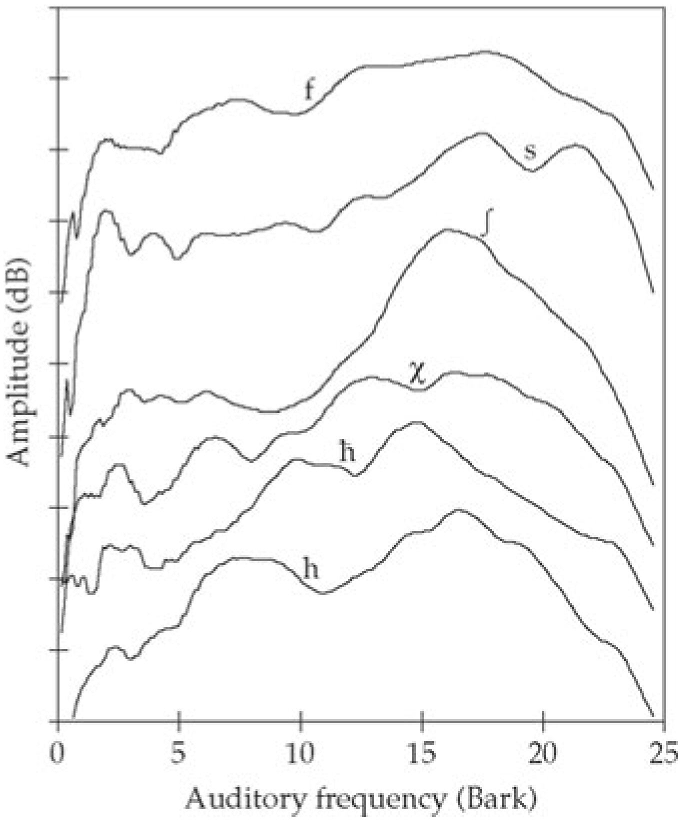
several spectral peaks (and from utterance to utterance one or the other of these may have greatest amplitude) +

inter-speaker variability in the frequencies of the spectral peaks →

center-of-gravity techniques for the characterization of fricative spectra →

the auditory representation of fricatives

the auditory representation of fricatives



the nonlinearity of the auditory frequency scale +

the widths of the auditory critical bands for high frequencies →

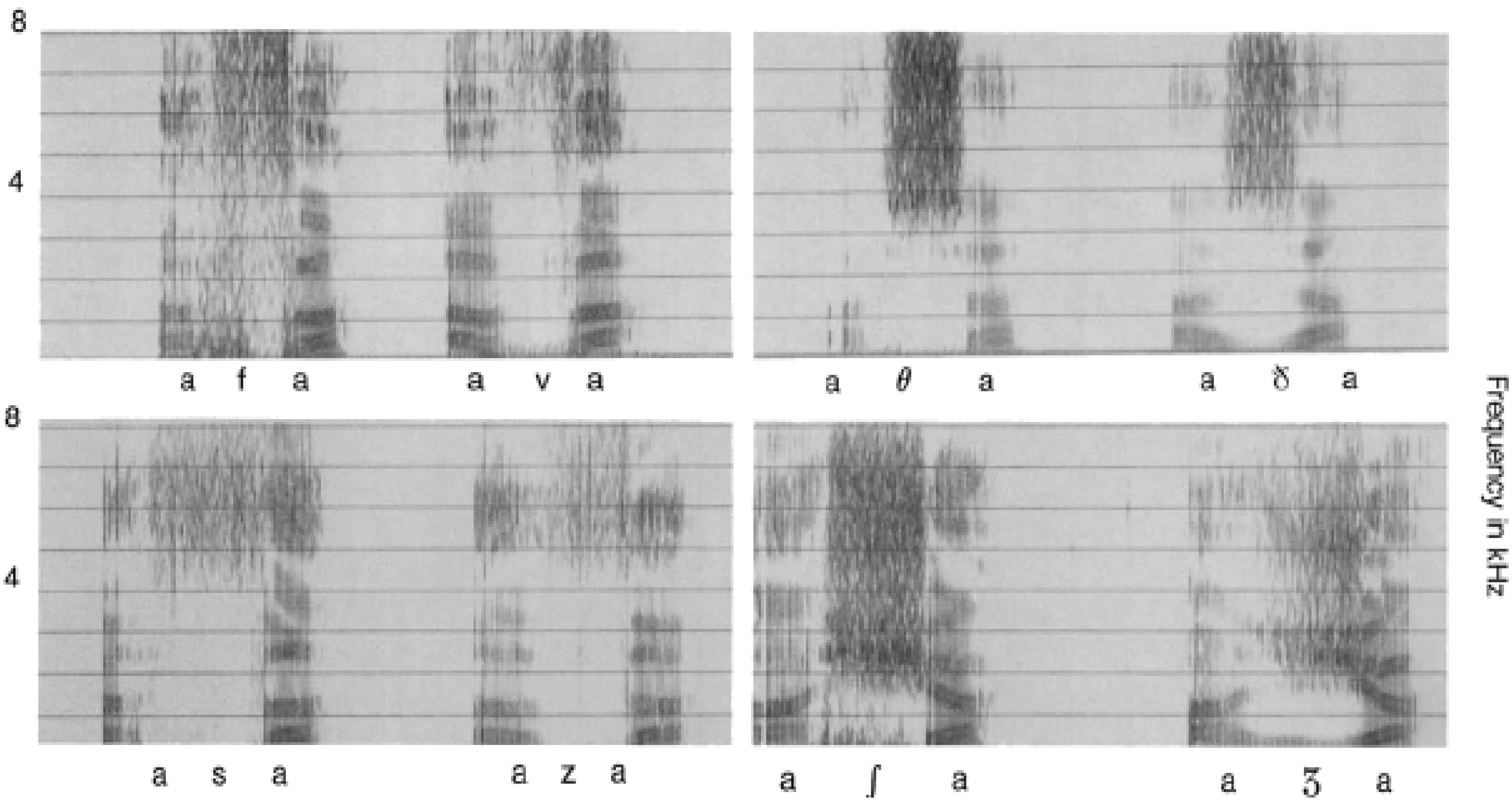
the auditory representation of fricatives differs from their acoustic representation.

The **same**:

- [f] has a flat, diffuse spectrum
- the remaining fricatives show spectral prominences of progressively lower frequency as the constriction moves back

Unlike the acoustic spectra of these sounds, the auditory spectra seem to be more evenly spaced.

the difference between the two broad regions of spectral energy in [s] is reduced.



the spectrum of the fricatives at the lips is determined by the **resonant characteristics of the constriction** and **portion of the vocal tract anterior to the noise source** fricative energy is very low in intensity for /f/, /v/, /θ/, and /ð/, mainly because there is no appreciable resonating cavity anterior to the point of constriction.

Despite the low level of energy, the frequency band is broad.

A narrower band of high-frequency, high-energy noise characterizes the alveolar fricatives /s/ and /z/. Most of the sound energy for /s/ is above 4 kHz, whereas for /ʃ/ it is around 2,000 Hz and above for a male speaker.

Because the point of articulation for /ʃ/ is farther back in the mouth than for /s/, the resonating cavity anterior to it is longer than that for /s/, resulting in its lower frequencies.

The lip rounding and protrusion that are more usually associated with /ʃ/ also lengthen the front cavity and contribute to its lower resonant frequency.

This acoustic effect of lip posture is much the same as we have seen for the back vowels.

source–filter account of consonant production,

The resonances for /s/ are derived from the natural resonant frequency of the constriction and of the cavity in front of the constriction.

The source of the fricative noise is at the anterior edge of the constriction.

The airfilled cavity in front of the noise source can be likened to a tube closed at one end, because the constriction is extremely narrow at the source, and open at the other

These tubes are quarter-wave resonators, the resonance for the anterior cavity:

f = velocity of sound / wavelength = 34,400 cm/s/4 (1 cm) = 8,600Hz

Because of the narrowness of the constriction, the back cavity resonances are severely attenuated and not perceptually salient.

Thus, little energy below 4,000 Hz. The resonances that would have been produced below 4 kHz have been attenuated by the back cavity antiresonances.

most of the significant energy for /s/ above 4,000 Hz

Stops and Affricates

have more complicated articulatory and acoustic properties than vowels or fricatives.

Vowels and fricatives may be treated as static events (acoustic targets, or articulatory postures).

This is not so with stops and affricates:

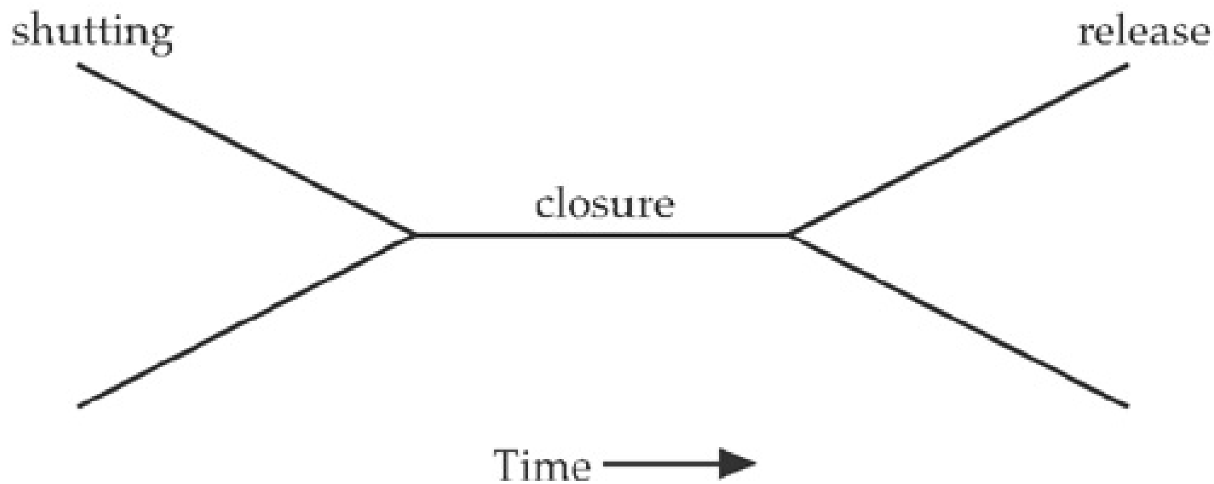
- The main articulatory posture during a stop is complete closure of the vocal tract, acoustically silence.
- However, languages use a great variety of stops, utilizing different places of articulation, stop release sounds, and accompanying noises

→ it is useful to identify in the articulation of stops and affricates

three stages,

corresponding to three intervals of time, each of which can be characterized in terms of the acoustic theory of speech production as a combination of a sound **source** and a vocal tract **filtering** function:

1) the movement of an articulator toward a stop closure (the shutting movement);
2) the closure itself;
3) the release of the closure.



Source: phonation types

Three main types of voicing occur cross-linguistically: modal, creaky, and breathy voicing:

the amount of time that the vocal folds are open during each glottal cycle.

Aerodynamics in fricatives: oral volume velocity is related to glottal impedance; voiceless sounds have greater oral airflow than voiced sounds →

glottal impedance in voiced sounds varies, depending on the amount of time that the vocal folds are open during each glottal cycle.

Phonation types

Modal voicing occurs in all languages, and some distinguish it from one or both of the other types of voicing.

Modal: the vocal folds are closed during half of each glottal cycle and open during the other half (approximately). Thus, the proportion of time that the glottis is open (the **open quotient**) during each cycle is 0.5.

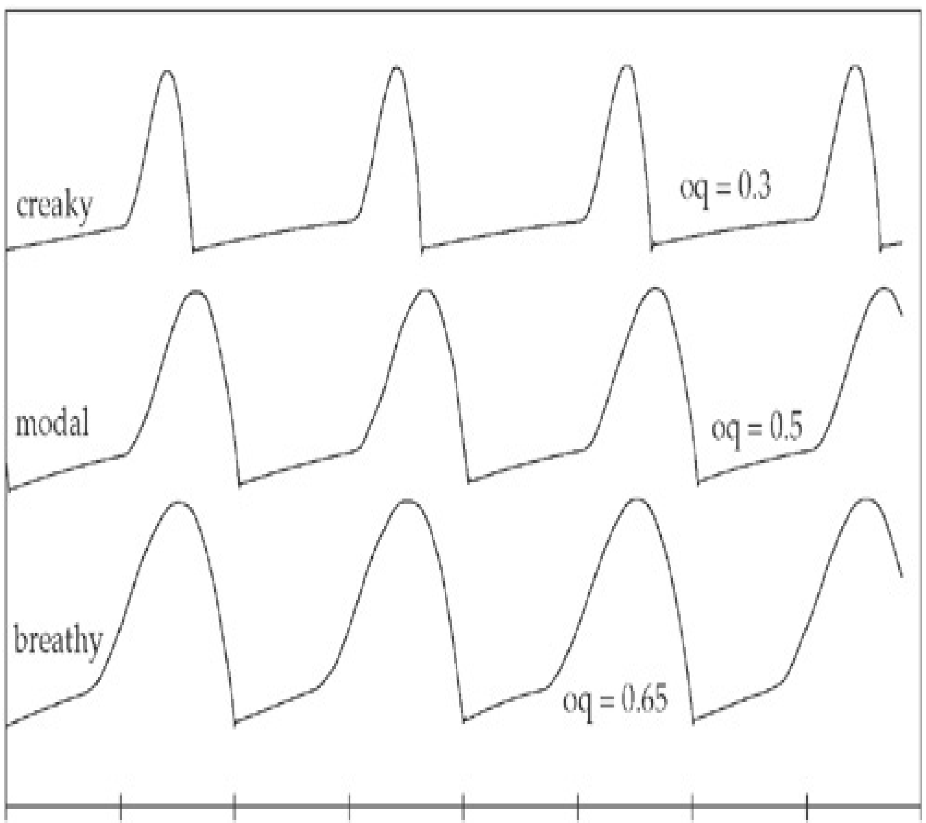
Creaky: the vocal folds are held together loosely, and air bubbles up through them →

longer closed phase of the glottal period and a comparably shorter open phase (and thus a smaller open quotient).

Breathy: the vocal folds vibrate, but without much contact (for some people the vocal folds do not completely close during breathy voicing) →

the glottis is open for a relatively long portion of each glottal cycle.

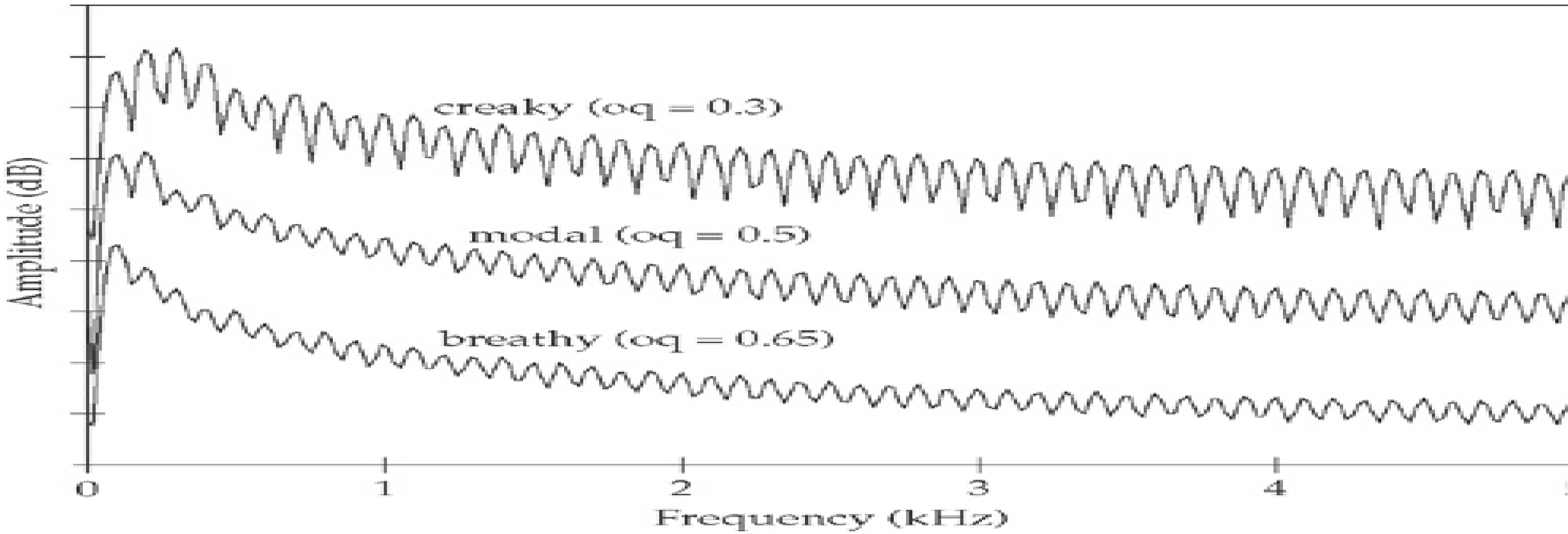
Glottal waveforms produced by varying the open quotient (oq) of the glottis



top: **creaky voice**, the glottis closed over 70 % of each glottal cycle, with a quick glottal opening (30 %)

In the second and third waveforms the open quotient is increased to 50 and 65 % of the glottal cycle, respectively:

modal and breathy voice



Fourier's theorem → any change in the shape of the voicing waveform (which is a complex periodic wave) results in a spectral change as well.

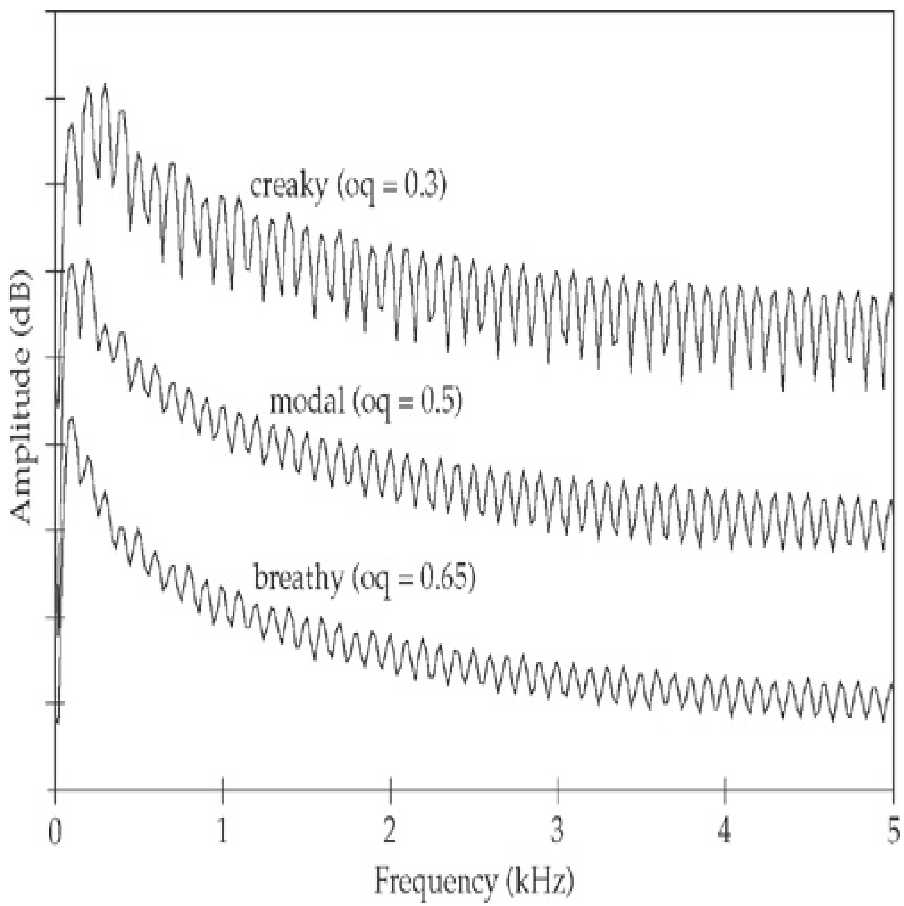
The BV waveform is more like a sine wave than either creaky or modal → the dominance of the first harmonic. All the other harmonics have lower amplitude

The CV waveform is least like a sine wave: 2, 3, 4 harmonics have higher amplitude than the first.

The difference between the amplitude of the first the second harmonics is a reliable way to measure the type of phonation

But: changes in F0, loudness and vowel quality can alter the relative amplitudes of the first and second harmonics.

Another spectral differences:

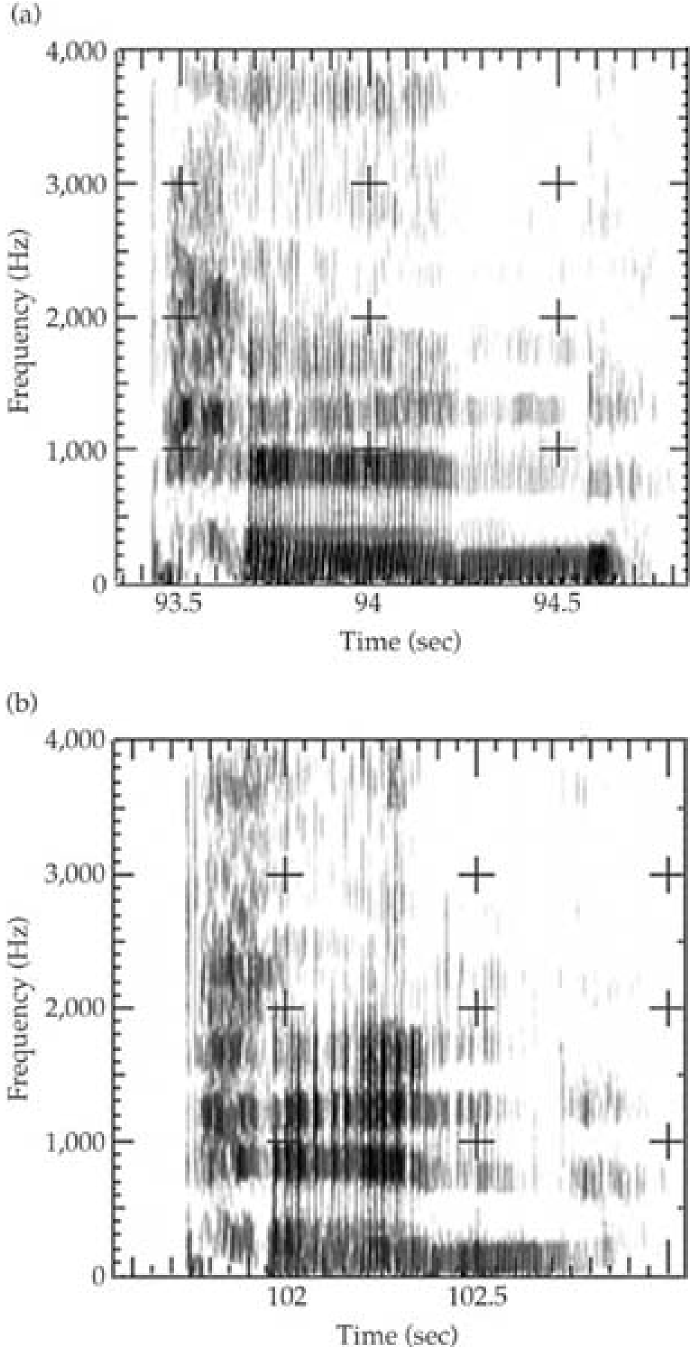


1 the slope of the spectrum (harder to measure in practical terms because of the influence of vocal tract resonances):

the amplitudes of the harmonics in the **breathy voiced spectrum**

drop off more quickly as frequency increases than in the modal or creaky spectra.

2 The muscular tension involved in creaky voice often results in slower vocal fold vibration, so **creaky voice may have a lower fundamental frequency** than modal or breathy phonation.



Spectrograms of the English word “pin” contrasting (a) modal phonation with (b) creaky voice.

On a spectrogram the voicing pulses during creaky voice

are further apart than they are in modal voicing →

a reliable acoustic correlate of creaky voice in English.

Similarly the long open period in breathy voice results in a bit of time in each glottal cycle during which the glottis is open enough to allow a high rate of airflow, and thus some glottal frication (**aspiration noise**) is produced during each glottal cycle in breathy voicing

Phonation types

Stops and affricates may differ distinctively in different languages,

depending on the phonation type with which they are produced.

These voicing distinctions can occur in any of the three stages of stop and affricate production.

Sound sources in stops and affricates

The most common source of acoustic energy during the *shutting stage* of stop consonant production is voicing.

Pre-glottalized stops are produced with creaky voicing in the closing stage, and

Pre-aspirated stops have a bit of aspiration noise (turbulence generated at the glottis) during the closing stage.

During *closure,* voicing is the only possible sound source:

it is aero-dynamically impeded, though not prevented.

Air flows through the vocal folds into the occluded vocal tract, but soon

the vocal tract fills up, so the air from the lungs has nowhere to go, and voicing can no longer be maintained →

Either voicing persists through only part of the closure, or the speaker performs a maneuver to expand the oral cavity:

- larynx lowering (→ glottalic ingressive sounds),
- tongue-root advancement
- reduction of tension in cheek or pharynx muscle (to allow for passive expansion during the closure).

How we can hear, and see on spectrograms, voicing during closures?

it is like the music when a neighbor plays too loud:

as with stop voicing, there may be no direct air passage for the sound (if windows are closed);

but one can hear the stereo because the floor transmits the music at the floor's resonant frequencies.

this wall-rattling music does not sound the same as in the room with the loudspeakers

All one can hear is the bass: the beat, but not the melody.

This is because floors and walls have low resonant frequencies (they are big, massive things), so they resonate best to low notes.

Similarly, when the vocal cords vibrate during a stop closure, the walls of the vocal tract transmit the sound, but only the low-frequency components.

Thus, in spectrograms, voicing during stop closures appears as a low-frequency band, called a **voice bar**, at the bottom of the spectrogram.

Pre-nasalized stops such as [mb] sometimes arise historically from voiced stops in which voicing during the closure is facilitated by opening the nose

So this variant realization of the closure stage may also be motivated by the aerodynamic constraints on voicing during stop closure.

During the ***release*** portion of stops, several different types of sound source are found:

- stop **release burst**, when increased air pressure in the vocal tract is released: air rushes out of the mouth at high speed, producing a **pressure impulse** that lasts only 2-3 ms.
- release can also be voiced (with modal, creaky, or breathy voicing)
- or aspirated.

It is also possible to release a stop by lowering the velum (nasal release) when a stop is followed by a homorganic nasal: [hidn.].

stops in absolute final position are often **unreleased** in English, although some recent research has suggested that release, even if not audible, is more likely to occur than not.

two sound sources in aspirated stops

Burst noise is produced **at the place of articulation** of a consonant (like fricative noise), whereas

aspiration noise is produced **at the glottis →**

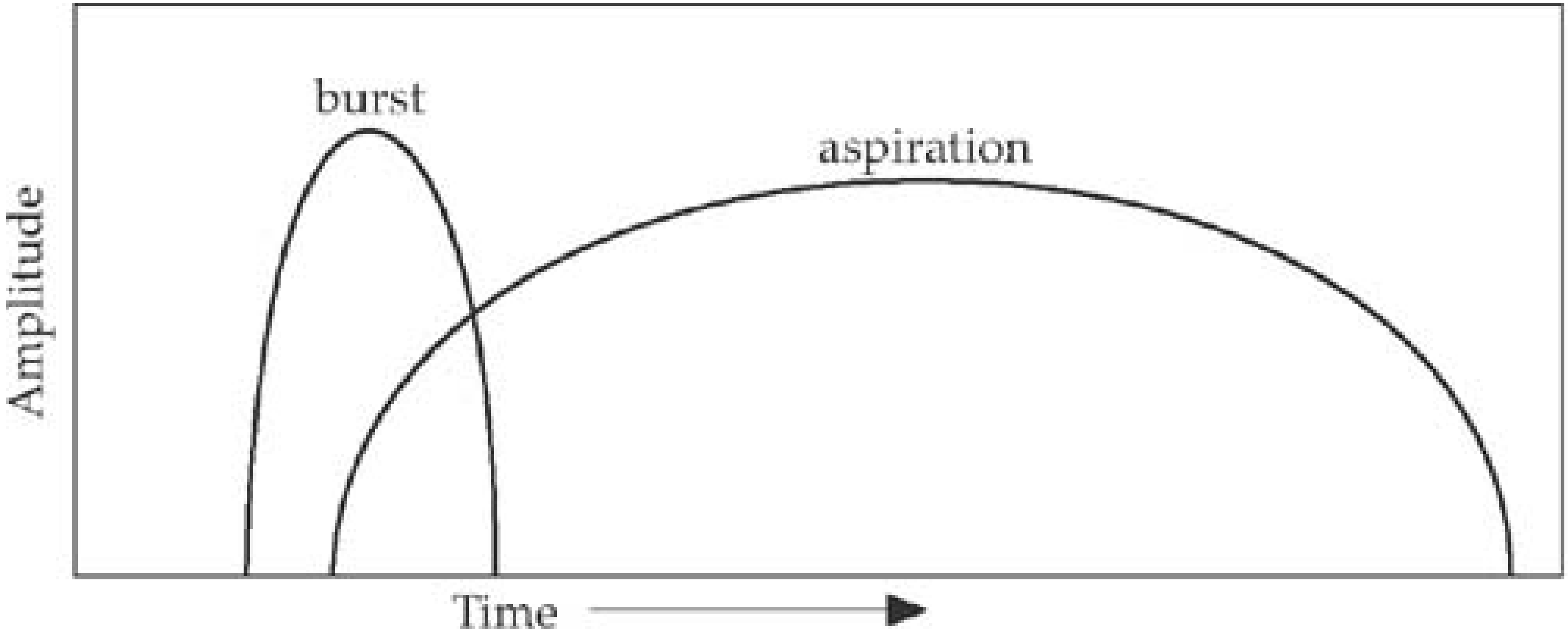
the vocal tract filter for the two types of sources is different +

the burst is very short in duration, whereas aspiration noise may be quite long.

For a few milliseconds after a stop closure is released, the constriction is too narrow to allow the amount of airflow needed to produce aspiration at the glottis.

Thus, immediately after the release, conditions are right for the burst (high-pressure buildup, very narrow opening) but not for the aspiration;

while later, as the constriction is opened further, conditions are right for aspiration (when the constriction is open enough for a high volume of airflow)



The relative dominance of these two sound sources:

the burst and

aspiration noises

in the release of an aspirated stop

how to make **stop release bursts**:

pulmonic stops,

ejectives (glottalic egressive sounds):

the air pressure behind the stop closure,

which is generated by closing the glottis and raising the larynx during the stop closure,

is usually greater than that in pulmonic stops, and the resulting burst has greater amplitude.

the stop release bursts of **implosives** (glottalic ingressive sounds), on the other hand, are weaker than those of pulmonic stops:

the intraoral air pressure buildup during the closure state is usually not very large.

how to make **stop release bursts**:

Click release bursts are typically much louder than any other type of release burst:

the difference between air pressure in the mouth and atmospheric air pressure is quite large.

Since clicks are produced with a very small air cavity between the tongue and the roof of the mouth,

a relatively small movement of the tongue produces a very large change in intraoral air pressure,

and hence makes it possible to produce a loud ingressive release burst.

Vocal Tract Filter Functions in Stops

Two types of vocal tract filtering affect the acoustic properties of stop consonants:

1. the vocal tract configuration immediately following closure shapes the spectral qualities of the release burst
2. the vocal tract configurations and movements just before and after stop closure shape formant movements in the shutting and release stages of stop production.

The noise source of burst is shaped by the resonances of the portion of the vocal tract in front of the stop closure,

because the closure at the time of the release is still quite narrow, and thus

the front and back cavities of the vocal tract are not acoustically coupled →

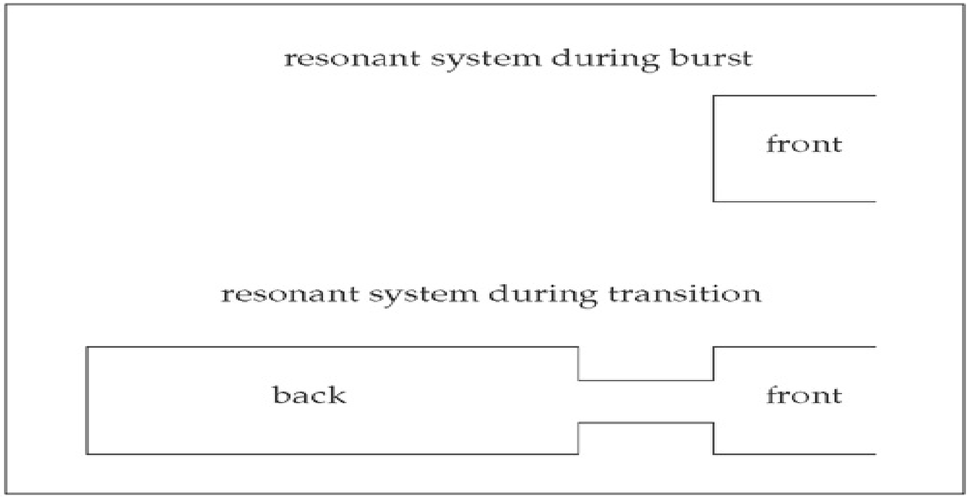
- bursts are acoustically similar to fricatives.

bursts are acoustically similar to fricatives:

Labial stops do not have a front cavity, the release bursts have no formant peaks, and energy is spread diffusely at all frequencies.

Dental and **alveolar** stops have a short front cavity, hence high-frequency peaks of spectral energy.

Palatal and **velar** stops have a longer front cavity, thus lower-frequency peaks in the spectrum and generally more formant structure than other stop bursts.



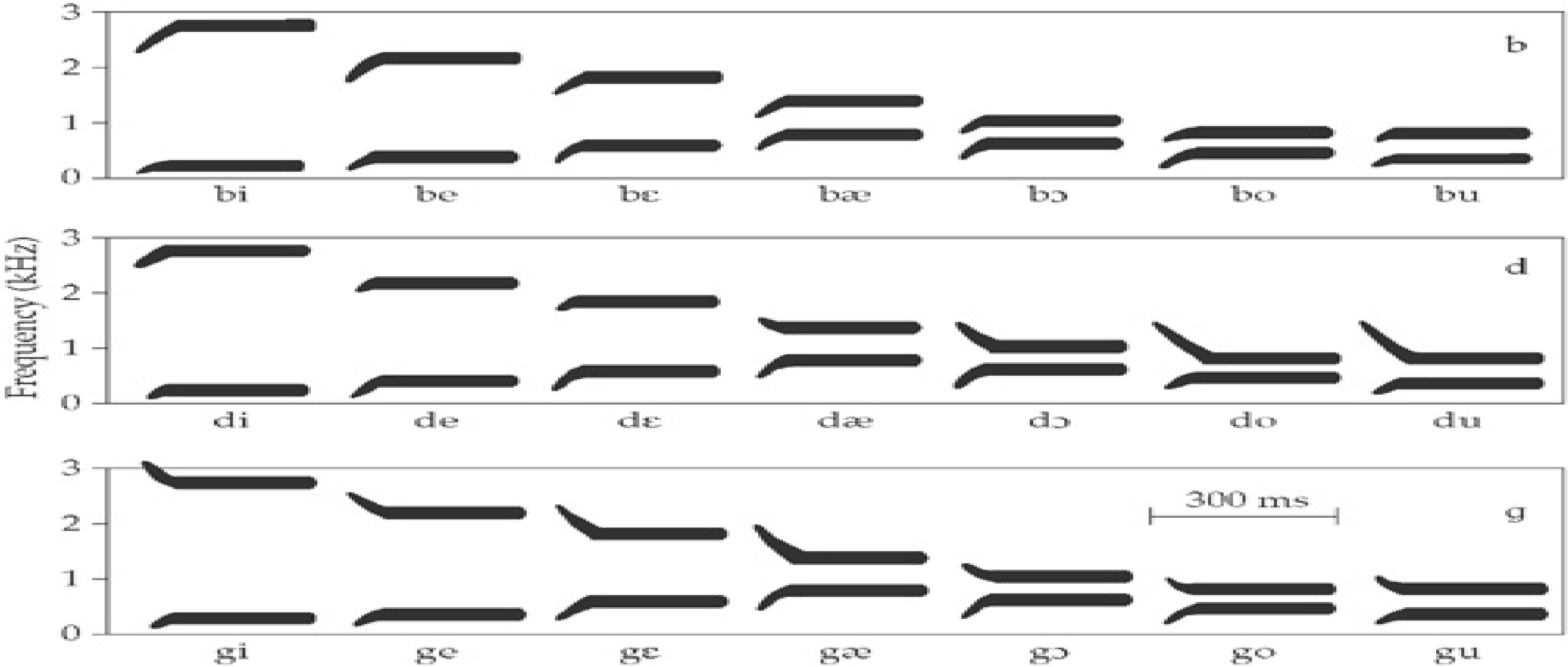
In the portion of the stop release **after the burst and in the shutting stage**,

the front and back cavities of the vocal tract are acoustically coupled, the source is located at the glottis, both for voicing and for aspiration →

the acoustic properties of these stages are better understood in terms of the models we used to discuss vowel formant values (perturbation theory and two-tube models):

shutting and release are movements, not postures; so, the information they contain about place of articulation is seen in formant movements, rather than in particular formant values.

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



the hypothetical starting frequency of the F2: **locus frequency**

Labials have a low-frequency locus, alveolars a middle-level locus, and velars a somewhat higher one.

The locus is not actually reached: it is due to changes in the acoustic coupling of the front and back cavities during stop

F1

the the F1 frequency is positively correlated with the size of mouth opening:

should be minimal during stop articulation, as the oral cavity is completely occluded.

opening the tract raises F_1 in the case of initial stops, and closing the tract will force the f F_1 down in the case of final stops.

The degree of rise or fall varies, depending on the F_1 of the neighboring sounds.

the increase or decrease in F1 is much greater when a stop precedes or follows a vowel such as /ɑ/, which has a high F_1, than when it precedes or follows a vowel with a low F_1, such as /i/.

The vocal tract's acoustic properties change suddenly when the front and back cavities are coupled.

At the instant of stop release, the front and back cavities are uncoupled, so the resonant frequencies of the back cavity do not have an impact on the speech spectrum;

as the closure is opened further, the front and back cavities become coupled →

the release burst of [d] has higher-frequency resonances than does the release burst of [g], while [g] has a higher F2 locus than [d].

The release burst of [d] is shaped by a front cavity resonance (which has a high frequency), but the F2 locus is associated with a back cavity resonance.

Both the burst and the F2 locus of [g] are associated with a front cavity resonance.

- The rapid change in the coupling of front and back cavities also explains why all stop consonants have a low-frequency F1 near the stop closure.
- During closure the two cavities are uncoupled, thus there is no first formant.
- The Helmholtz resonator that gives us F1 starts with a very narrow neck (and thus a low F1).
- As the constriction widens, the neck of the Helmholtz resonator increases, and thus the frequency of F1 increases.
- Also, as the closure is released and the constriction becomes wider, the front and back cavities become acoustically coupled.
- The opposite happens during the shutting stage. The first formant starts relatively high, and falls as the constriction is formed.

Formant transitions

All formants are lower near the stop closure in ***labial*** (bilabial) stops as a result of having a constriction located at a velocity node (at the lips) for all the vocal tract resonances.

The F2 locus for ***coronal*** (dental and alveolar) stops is a bit higher than the second resonant frequency of a uniform tube. The F3 near the closure in coronal stops is also higher than in the uniform tube ← a two-tube model with a constriction near the front of the vocal tract

In ***dorsal*** (palatal and velar) stops there is a convergence of F2 and F3 near the stop closure, because the location of the closure is near the intersection of the front and back cavity resonances.

the voicing classes of stop

the stops of English can be grouped into voiceless (/p/, /t/, /k/) and voiced (/b/, /d/, or /g/).

In many contexts, such as the intervocalic context, this simply means that phonation continues throughout the period of articulatory closure for /b/, /d/, and /g/ and ceases during the closure for /p/, /t/, and /k/:

a classic example of the voiced–voiceless difference.

In other contexts, however, the difference between the members of the cognate pairs is less straightforward.

VOT

in syllable-initial position before a vowel or resonant consonant,

speakers of English do not normally phonate during the closures of either /p/, /t/, /k/ or /b/, /d/, /g/

the opposition is maintained by a difference in the **timing of the onset of phonation** relative to the release burst of the stop:

phonation for /b/, /d/, /g/ begins at or very shortly after stop release,

whereas there is a delay of at least 50 ms before phonation begins after the release of /p/, /t/, /k/.

This relative timing of stop release and the initiation of phonation: **voice onset time (VOT)** (Lisker & Abramson 1964).

VOT is measured in milliseconds as the duration between the burst of stop release and the first vocal pulse that can be observed.

If the onset of phonation follows stop release, VOT values are positive; if voicing onset precedes stop release, VOT values are negative.

Thus, if phonation begins 75 ms after stop release, the VOT value is +75;

if phonation onset occurs 85 ms before stop release, the VOT is −85.

VOT is the result of a coordinated articulatory strategy labial and glottal adjustments used to produce initial prevocalic /p/ and /b/ in English

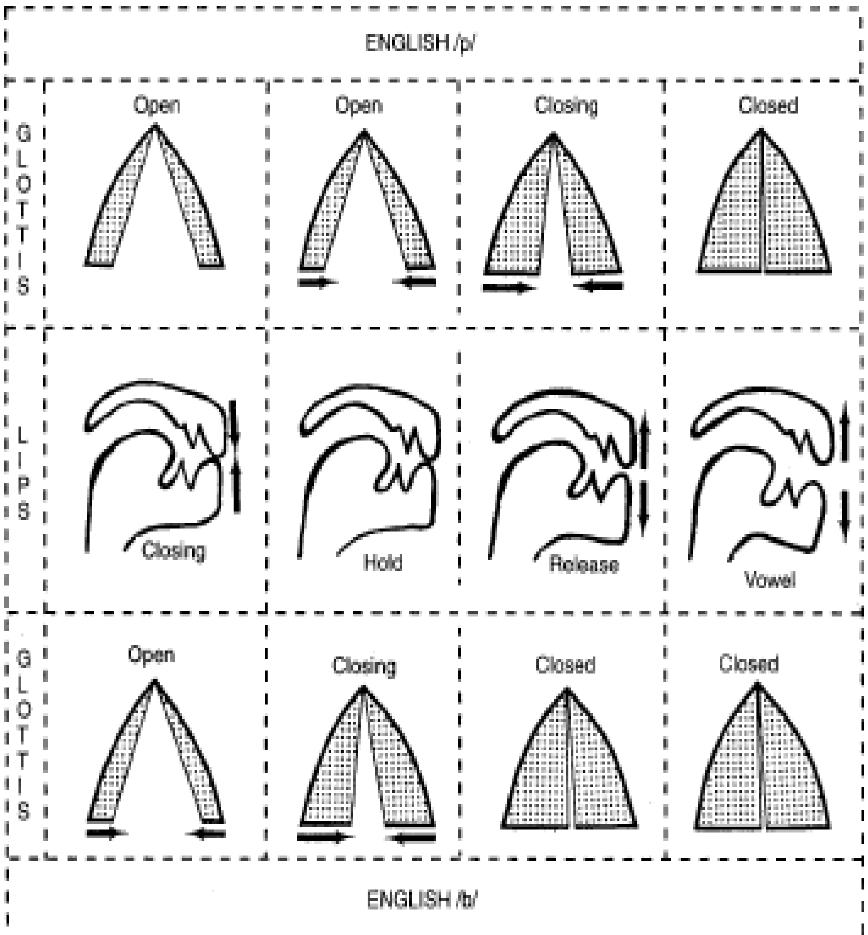


FIGURE 6.14 The relationships between glottal states and labial articulation for prevocalic /p/ and /b/ in English.

at the moment of the occlusion, the vocal folds have begun to adduct for /b/, adduction continues during the hold period by the time the /b/ is released, the folds are in phonatory position, ready to vibrate.

This accounts for the low, positive VOT values (between 0 and 10 ms) that characterize /b/.

In contrast, the folds do not begin to adduct for /p/ at the moment of stop release the glottis is still partially open.

Vocal fold adduction is not complete until sometime after the release of the stop, during the articulation of the following vowel.

This accounts for the long positive VOT values (+60 to +70 ms or more) that characterize /p/.

It also accounts for the presence of another acoustic feature that differentiates the voiced from the voiceless stops: **aspiration**.

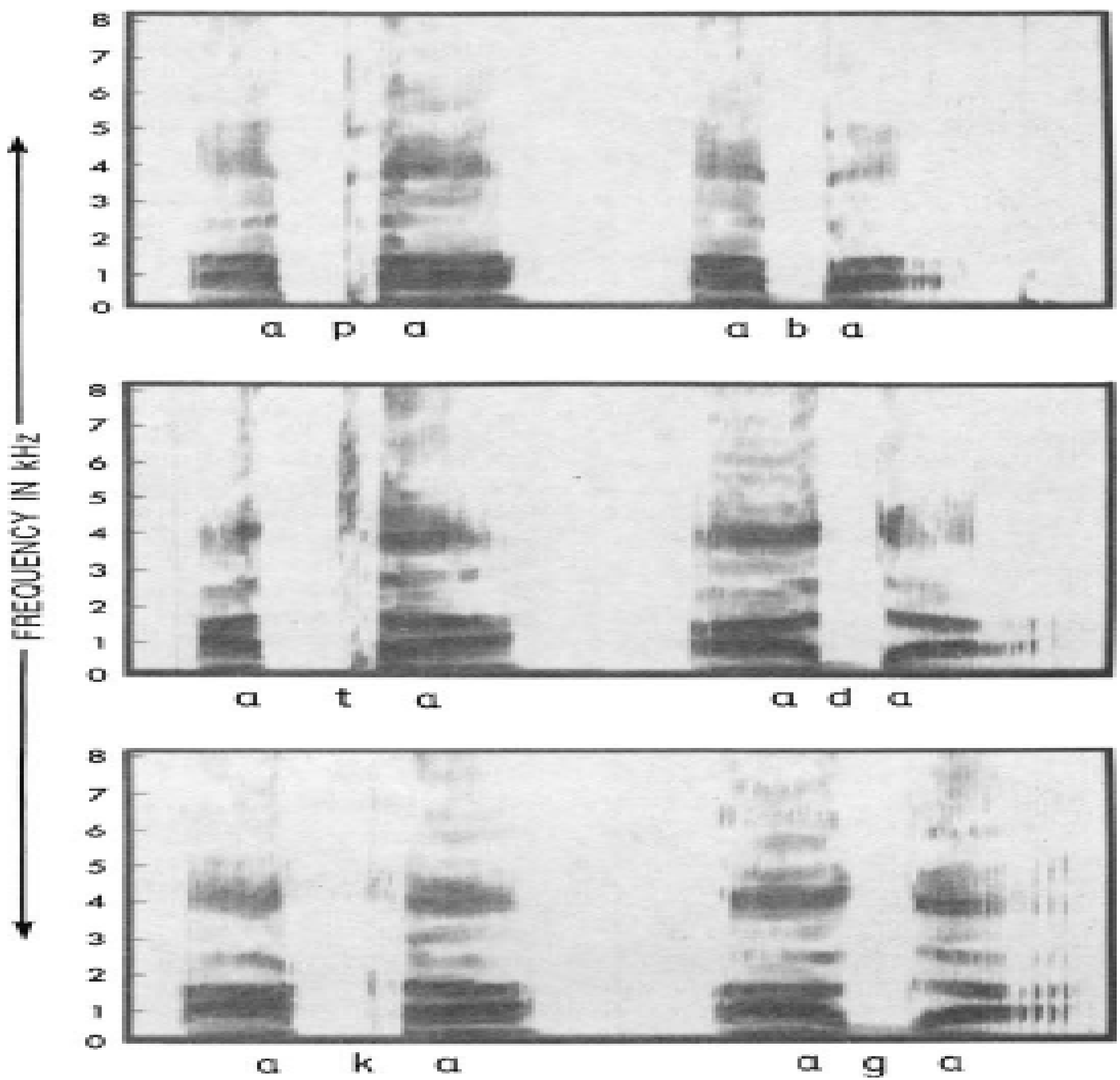
aspiration

Aspiration, a period of voicelessness after stop release, is associated only with /p/, /t/, /k/ in English:

the open glottis at the moment of stop release allows the breath stream to flow freely into the upper vocal tract without generating phonation.

The reverse is true for /b/, /d/, /g/: the closed glottis at the moment of stop release forces the breath stream to set the vocal folds into vibration and to carry the periodic phonated vibrations into the upper vocal tract.

Phonation and aspiration are therefore complementary phenomena in English. The VOT differences between /p/, /t/, /k/ and /b/, /d/, /g/ and the presence versus absence of aspiration in those stops are clearly visible


FREQUENCY IN kHz
8
7
6
5
4
3
2
1
0
a p a
a b a
a t a
a d a
a k a
a g a

Other languages may use different timing contrasts to distinguish between voicing classes of stops

- Spanish, Italian, Russian, French /b/, /d/, /g/, for example, display negative values of VOT: phonation commences during the hold period for the stops because vocal fold adduction begins during the closing phase of stop articulation.
- In contrast, the VOT values for /p/, /t/, /k/ in these languages are much like the values for /b/, /d/, /g/ in English.
- speakers of Russian often appear to English-speaking listeners to produce only voiced stops: none of the stops is significantly aspirated
- differentiating voiced from voiceless stops requires complex and highly coordinated motor activity

perception

- **although F2 transitions may appear to be fairly indistinct in spectrograms (especially when a large frequency range is displayed), they are enhanced in the auditory representation**
- release bursts are more prominent: after a brief period of silence, the auditory system responds more strongly than it does in a period of continuing sound

Affricates

The release phase of a stop may have frication rather than voicing or aspiration → affricates.

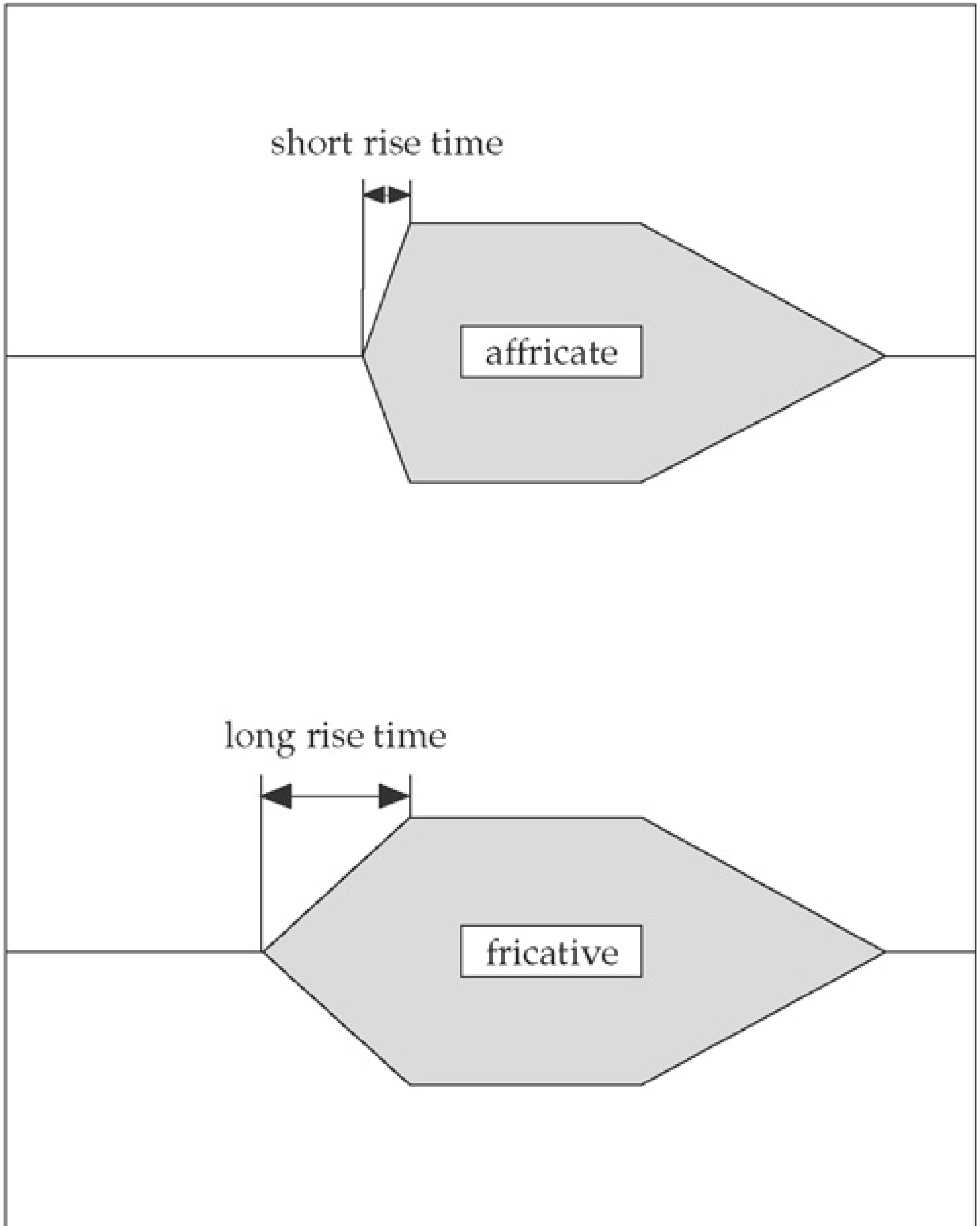
The frication noise in affricates is usually produced at the same place of articulation as the stop

(heterorganic affricates do occur: [tx] in Navajo)

The main acoustic distinction between an affricate and a sequence of a stop + fricative is that

the amplitude of frication noise rises quickly to full amplitude in affricates, and more slowly in fricatives.

This property has been called **rise time**,



rise time

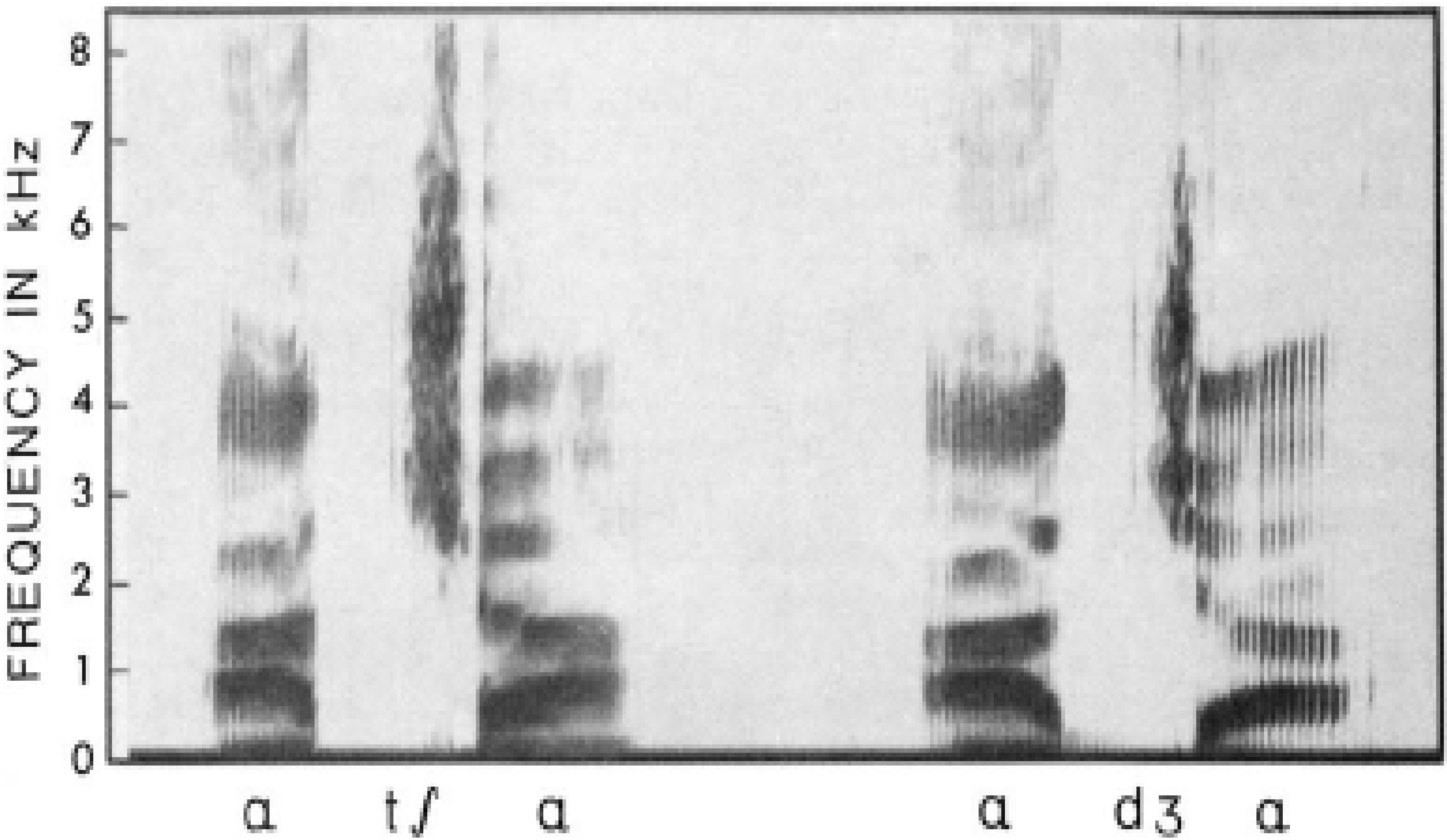


FIGURE 6.15 Spectrograms of [ɑtʃɑ] and [ɑdʒɑ].

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.