Chapter 4

Sounds

Human languages may have existed for more than a hundred thousand years, and through most of this time languages have primarily been spoken. Writing was invented only five thousand years ago, and still only a few percentages of the 5,000–7,000 languages in the world are in daily use as written languages. Despite the existence of written languages—and sign languages (the "signed" languages of the deaf)—spoken languages have a privileged status that justifies treating it before written and signed languages.

We shall refer to the study of the sounds in language as **phonology**. The study of the sounds in language is usually divided into two parts, phonetics and phonemics, which can be defined in the following way:

PHONOLOGY The study of the sounds in language.	
PHONETICS The study of the physiological, aerodynamic, and acoustic characteristics of speech sounds.	**PHONEMICS** The study of how speech sounds are organized into systems and utilized in languages.

TABLE 19. *The branches of phonology*

In fact, most linguists these days refer to 'the study of how sounds are organized into systems and utilized in language' as *phonology*, while the term *phonemics* is not used at all, with the unfortunate result that they have no term for 'the study of the sounds in language' in general. We prefer to stick to the terminology introduced in TABLE 19.

4.1   Phones and phonemes

The pronunciation of words are often written in a **phonemic trancription**, which is marked by **slashes**: /…/. The letters of a phonemic transcription represent **phonemes**, which we can define as follows:

Phonemes are sounds that distinguish words (word forms) from each other.

We may compare the English words *pin* /ˈpɪn/ and *pin* /ˈtɪn/, which are distinguished from each other phonologically by the contrast between the /p/ of the first word and the /t/ of the second. Consider also *pin* /ˈpɪn/ and *pan* /ˈpæn/, which are distinguished by the contrast between /ɪ/ versus /æ/, and finally *pin* /ˈpɪn/ versus *pit* /ˈpɪt/, which are distinguished by the contrast between /n/ and /t/. Pairs of words that are distinguished from each other by a contrast between two phonemes are called **minimal pairs**.

As a general rule, a phoneme may vary in pronunciation from one position in the word to another. In English for example, /p/, /t/, and /k/ are **aspirated** (they are followed by a strong breath, like an *h*) in word-initial position, for example in *pin* /ˈpɪn/, *tin* /ˈtɪn/, and *kin* /ˈkɪn/. On the other hand, they are **unaspirated** (they are not followed by a strong breath) when following /s/, like in *spin* /ˈspɪn/, *sting* /ˈstɪŋ/, and *skin* /ˈskɪn/.

The different pronunciation variants of a phoneme are called the **allophones** of that phoneme, and we have a special **phonetic transcription** to write them in. Such a transcription is written in square brackets, […]:

Orthography	Phonemic transcription	Phonetic transciption
pin	/ˈpɪn/	[ˈpʰɪn]
tin	/ˈtɪn/	[ˈtʰɪn]
kin	/ˈkɪn/	[ˈkʰɪn]
spin	/ˈspɪn/	[ˈspɪn]
sting	/ˈstɪŋ/	[ˈstɪŋ]
skin	/ˈskɪn/	[ˈskɪn]

TABLE 25. *Orthography, phonemic and phonetic transcription of six English words*

Sound differences that are **phonetic** (or **allophonic** or **non-distinctive**) in one language may be **phonemic** (or **distinctive**) in another language. For English, the contrast between unaspirated and aspirated stops is a *phonetic* contrast, while there are many languages where the same contrast is *phonemic*, for example in many languages of India—cf. the Hindi words *pāl* /paːl/ 'take care of' and *kān* /kaːn/ 'ear' versus *phāl* /pʰaːl/ 'edge of knife' and *kha\n* /kʰaːn/ 'mine'. We have a tendency not to hear sound contrasts that are phonetic in our own language.

The phonemes may be characterized as *basic level terms* of speech sounds, the highest level at which we can form a mental image of a sound concept. We have no difficulties in forming a mental image of an English /p/-sound, /t/-sound, or /k/-sound, but we are not able to form a mental image of a voiceless stop, which is the hyperonym term of which /p/, /t/, and /k/ are hyponyms. As we have observed earlier, terms *above* the basic level tend to have a more or less technical or scientific «flavor»— which is certainly true of 'voiceless stop'.

Even terms *below* the basic level tend to belong to the world of experts. And there is a level of hyponyms of the basic level phonemes, although the ordinary native speaker is seldom aware of its existence.

4.2 Speech organs

Strictly speaking, human beings do not have «speech organs» the way we have «visual organs» (eyes) or «auditory organs» (ears). The speech organs, the totality of which we have also referred to as the vocal tract, is simply those parts of the mouth and the respiratory tract that are used to produce speech sounds. While seeing is the primary function of the eyes and hearing the primary function of the ears, the primary functions of the speech organs are breathing, eating, and drinking. The vocal tract may be construed as a "quartet" with the these members:

«The Vocal Tract Quartet»			
The lungs	The larynx	The nasal cavity	The pharynx

& the windpipe			& the mouth

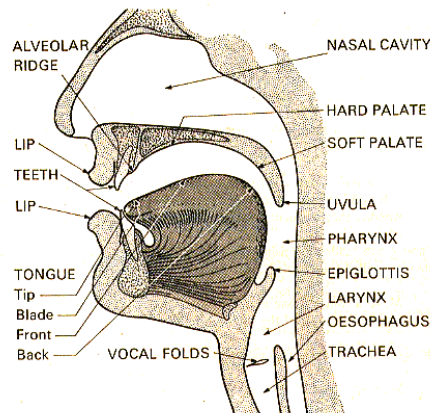The members of the "quartet"—except the lungs—are depicted in FIG. 16.

FIGURE 16. *The speech organs (except the lungs)*

4.2.1 The lungs and the windpipe

By increasing and decreasing the volume of the **lungs**, air is caused to flow into and out of the lungs. In all languages some—often most or all—sounds are produced on the basis of air flowing out of the lungs and the **windpipe** (trachea), through the throat and out of the mouth and/or the nasal cavity. We can also use the tongue and parts of the throat to cause air to flow through shorter parts of the vocal tract.

4.2.2 The larynx

The **larynx** is a box-like structure sitting on the top of the windpipe—cf. FIG. 17, where the larynx is seen from the front, with the structure known as the Adam's apple in the middle.

Inside the larynx are the vocal cords, two flaps of muscle across the windpipe. The vocal cords can be held in different positions, affecting the airflow from the lungs in different ways. The opening between the vocal cords is the **glottis**, seen from above in PICT. 3. The vocal cords are the light folds surrounding most of the glottis.
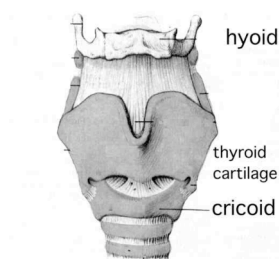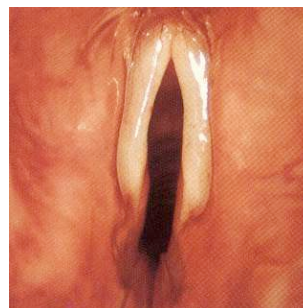

FIGURE 17. *The larynx*

PICTURE 3. *The glottis and the vocal cords*

4.2.3 The nasal cavity

3

The interior of the nose, the **nasal cavity**—cf. FIG. 16—is directly connected to the pharynx, so that air can flow from the pharynx, through the nasal cavity, and out. The passage between the nasal cavity and the pharynx—called the **nasal port**—can be opened and closed by lowering and raising the **velum** or **soft palate**, a muscular flap at the back of the roof of the mouth—cf. FIG. 16.

4.2.4 The oro-pharyngeal tract

The **mouth cavity** and the **pharynx** (cf. FIG. 16), which is the tube starting right above the **larynx** (where the air-passage and the food-passage (*oesophagus*) divide on their way down) and going up to the back of the nose, constitute the **oro-pharyngeal tract**, which plays an important part in speech sound production.

Without comparison, the oro-pharyngeal tract is the quartet member with the heaviest and most challenging tasks. It is no coincidence that in many languages all over the world, language is referred to metonymically as «tongue» or «mouth». From a functional point of view, the mouth is the most important part of the oro-pharyngeal tract, and the tongue is the most important part of the mouth: without a tongue, no speech.

4.3   Speech production

Speech production has two basic components: **initiation** and **modulation**. Without these components there is simply no speech.

4.3.1 Initiation

Initiation is a movement of a speech organ to create an outgoing or ingoing flow of air through the vocal tract. The main *initiation organs* are the *lungs*, which work like a pair of bellows. When muscles are used to decrease the lung volume, the air pressure inside the lungs becomes higher than the air pressure outside the body, causing an outgoing **airstream** through the vocal tract. When muscles are used to increase the lung volume, the air pressure inside the lungs becomes lower than outside the body, causing an ingoing airstream through the vocal tract.

Airstreams can also be initiated by a piston-like movement of the larynx or the tongue: (1) Close the opening between the vocal cords and move the larynx up or down like a piston, or (2) close the opening between the back of the tongue and the back of the roof of the mouth and move the tongue backwards or forwards. The former type of initiation creates implosives and ejectives, while the latter type creates clicks. Implosives are found in many Asian and African languages, while clicks are only found in languages in Southern Africa.

4.3.2 Modulation

Modulation is an organic posture or movement of a speech organ that changes the airstream in such a way that a sound is created. There are three types of modulation: **phonation**, **nasality**, and **articulation**.

PHONATION

Phonation is modulation in the glottis, which—as mentioned earlier—is the opening between the vocal cords. Two postures of the glottis are found in all languages: **voiced** posture and **voiceless** posture.

In the **voiced** posture, the glottis is quite narrow, so that the air-stream from the lungs causes the vocal cords to vibrate—cf. Fig. 18. Sounds produced with this type of phonation are called **voiced sounds**, like the sounds in the English words *wall* /ˈwɒːl/, *Monday* /ˈmʌndeɪ/, and *busybody* /ˈbɪzibɑːdi/.

In the **voiceless** posture, the glottis is wide open, and the vocal cords do not vibrate—cf. Fig. 19. Sounds produced with this phonation type are called **voiceless sounds**, like the consonants of the English words *fat* /ˈfæt/, *post* /ˈpoʊst/, and *appetite* /ˈæpɪtaɪt/. The voiceless posture is also the posture of normal breathing.

FIGURE 18. *The glottis in the voiced posture.*   FIGURE 19. *The glottis in the voiceless posture.*

Several other postures of the glottis are possible, including whisper, breathy voice, creaky voice, and *glottal closure*, which are treated in more advanced books.

NASALITY
As mentioned earlier, the nasal port may be opened and closed by lowering and raising the velum, the velum or soft palate, a muscular flap at the back of the roof of the mouth—cf. Fig. 16. Sounds produced with a lowered velum are called **nasal** or **nasalized sounds**, while those produced with a raised velum are called **oral sounds**. The term *nasal* is used when there is a complete closure somewhere in the mouth; if not, the term *nasalized* is used. Some linguists use *nasal* in both meanings.
The consonants in the English word *meaning* /ˈmiːnɪŋ/ are nasal, while those in *batik* /bəˈtiːk/ are oral. Many languages all over the world distinguish between oral and nasalized vowels. In the French word *chanter* /ʃɑ̃te/ 'sing', the first vowel is nasal (this is marked with the /˜/ above the vowel sign) and the second is oral. Another example is Hindi /yahã̃/ 'here'.

ARTICULATION
Articulation is a modulation with a certain manner and place in the oro-pharyngeal tract.

Manner of articulation is type of stricture, while **place of articulation** is the location of the stricture in the oro-pharyngeal tract. For example, the initial consonant of the English word *pen* /ˈpen/ is produced with a complete closure (a manner) between the lips (a place), while the initial consonant of the English word *sun* /ˈsʌn/ is produced with a narrow passage causing friction in the air-stream (a manner) between the tongue blade and the gums (a place).

As already mentioned, the articulation is the most complicated part of speech production. Manner of articulation is treated in more detail in section 4.4 and place of articulation in section 4.5.

SUMMARY

The division of labor between the members of "The Vocal Tract Quartet" is summed up in TABLE 20. Dark grey shows the most important function of a quartet member, lighter grey shows less important functions.

	initiation	phonation	nasality	articulation
lungs	■■■			
larynx	▨	■■■		
nasal cavity			■■■	
oro-pharyngeal cavity	▫			■■■

TABLE 20. *The functions of the "Vocal Tract Quartet" members.*

4.4 Manner of articulation

Speech sounds can be classified into several types on the basis of the **manner of articulation**. In a few cases, information about other modulations—nasality and place of articulation—is also included in the classification. The following five types are most important: stops, nasals, fricatives, liquids, and vowels.

4.4.1 Stops

Stops are defined by a complete closure in oro-pharyngeal tract and by a raised velum, preventing air from escaping through the nose. Stops are found in all languages. Voiced stop are a little less common than voiceless stops, and the former type is only found in languages with the latter type.

In English we find the *voiceless stops* /p t k/ and the *voiced stops* /b d g/, as illustrated by the words *pet* /ˈpet/, *cat* /ˈkæt/, *bed* /ˈbed/ and *dog* /ˈdɑːg/.

4.4.2 Nasals

Nasals are defined by a complete closure in the mouth and by a lowered velum, allowing air to escape through the nose. Nasals are found in the overwhelming majority of the languages of the world. Voiceless nasals are rare, and are only found in addition to the voiced ones.

In English we find three nasals, /m n ŋ/, which are all voiced, as illustrated by the word *meaning* /ˈmiːnɪŋ/.

4.4.3 Fricatives

Fricatives are defined by a narrow stricture in the oro-pharyngeal tract, causing audible friction as air passes by. Fricatives are common, but languages often have fewer fricatives than stops or nasals. Voiced fricatives are less common than voiceless ones, and are only found in addition to voiceless fricatives.

English has an exceptionally high number of fricatives, the four *voiceless* /f θ s ʃ/ and the four *voiced* /v ð z ʒ/, as illustrated by the words *thief* /ˈθiːf/, *fish* /ˈfɪʃ/, *these* /ˈðiːz/, *vase* /ˈveɪs/, *vision* /ˈvɪʒən/.

6

4.4.4 Fricatives

Fricatives

4.4.5 Liquids

Liquids are a less homogeneous group than stops, nasals, and fricatives, but the term is a very useful common denominator for "l-sounds" and "r-sounds". The typical liquid is produced with a constriction between the front part of the tongue (i.e., the tip or the blade; cf. 4.5.1) and the roof of the mouth; it is oral and voiced. The constriction is «lighter» than the one found in stops, nasals, and fricatives: either shorter than in stops/nasals or more open than in fricatives.

Most languages have at least one liquid. Korean and many Bantu languages have a single liquid that varies in pronunciation between [l] and [ɾ]. [l] is a **lateral**, which is characterized by an air-stream passing the tongue on one or two sides through openings that are wide enough to avoid friction. [ɾ] is **tap**, which is like a very short [d], that is, a *momentary voiced stop*: the tip of the tongue touches swiftly the roof of the mouth. Another common liquid is the **trill** [r], produced by letting the tip of the tongue vibrate against the roof of the mouth two or more times.

4.5   Place of articulation

The initial consonants of the English words *pan* /ˈpæn/, *tan* /ˈtæn/, and *can* /ˈkæn/ are all *voiceless stops*, but they differ in **place of articulation**:

> The **place of articulation** of a speech sound is the place in the oro-pharyngeal tract with the strongest constriction during the production of the sound.

To produce /p/, we make a *complete closure* between the lower lip and the upper lip; to produce /t/, we make a complete closure between the tip or the blade of the tongue and the alveolar ridge; to produce /k/, we make a complete closure between the back of the tongue and the velum.

At each *place of articulation*, there is a constriction between an **active articulator** and a **passive articulator**. The *active articulators* are the lower lip and the tongue, while the *passive articulators* are the upper lip, the upper teeth, the roof of the mouth, and the rear wall. Different languages utilize a different number of articulators, and the most important ones are presented in TABLE 21.

ACTIVE ARTICULATORS	PASSIVE ARTICULATORS
labium	**labium**
apex	**dentes**
lamina	**alveoli**
dorsum	**postalveoli**
	palatum
	velum
	uvular

TABLE 21. *The active and passive articulators*

The *active articulator* **labium** ('lip' in Latin) is *the lower lip*; the **apex** ('tip' in Latin) is *the tip of the tongue*; the **lamina** ('blade' in Latin) is *the blade of the tongue*; while the **dorsum** ('back' in Latin) is *the back of the tongue*. Sometimes it is practical to use **corona** as a common denominator for *the apex plus the lamina*.

The *passive articulator* **labium** is *the upper lip*; the **dentes** ('teeth' in Latin) is *the upper teeth*; the **alveoli** ('basins' in Latin) is *the alveolar ridge or the gum*s, minus the back half, which is called the **postalveoli**; the **palatum** and the **velum** are *the hard palate* and in FIG. 16, respectively. The lower part of the velum is the **uvula**. There are adjectives corresponding to the nouns in Table 21: **labial, apical, laminal, coronal, dorsal; labial, dental, alveolar, postalveolar, palatal, velar, uvular.**

Terms for active and passive articulators are combined to define **places of articulation**. For example, when the active articulator is apical and the passive articulator is alveolar, the place of articulation is **apical-alveolar**. Two terms are irregular; we use **bilabial** instead of «labial-labial» and **labiodental** instead of «labial-dental». Quite often it is practical to mention only the active or only the passive articulator. For example, **alveolar** is a common denominator for *apical-alveolar* and *laminal-alveolar*. For **palatal**, **velar**, and **uvular** sounds, it is superfluous to mention the active articulator, which in these cases is always the *dorsum*.

In TABLE 22 we present some common sounds in the languages of the world, defined with parts of the terminology introduced so far.

		bilabial	labio-dental	dental	alveolar	post-alveolar	retro-flex	palatal	velar	uvular
stop	vc	p			t		ʈ	c	k	q
	vcl	b			d		ɖ	ɟ	g	ɢ
nasal	vc	m			n		ɳ	ɲ	ŋ	ɴ
fricative	vc	ɸ	f	θ	s	ʃ		ç	x	χ
	vcl	β	v	ð	z	ʒ			ɣ	ʁ
approx.	vc	w	ʋ		l	ɹ	ɭ	j		
tap	vc				ɾ					
flap	vc						ɽ			
trill	vc				r					ʀ

Abbreviations: vc = voiced; vcl = voiceless; approx. = approximant.

TABLE 22. Some common speech sounds.

In TABLE 22, we have included a column with **retroflex** consonants, which strictly are **sublaminal-prepalatal**, that is, the active articulator is the area behind the apex, under the tongue, while the passive articulator is the front part of the palatum. However, the symbols for the retroflex consonants are also commonly used for **apical-postalveolar**sounds. Retroflex consonants are common in India, cf. Hindi *ḍyūṭī* /ɖjuːʈiː/ 'duty'.

4.6 Vowels

While stops, nasals, fricatives, and liquids are **consonants**, we are now going to take a look at the **vowels**. The *prototypical consonant* has a relatively strong constriction in the oro-pharyngeal tract, while such a constriction is absent in vowels. Furthermore,

the *prototypical consonant* belongs to the *margin of a syllable* (cf. the next paragraph), while the *prototypical vowel* belongs to the *nucleus of a syllable*.

Practically all languages distinguish at least three vowels, while the most common vowel inventory in the world is /i e a o u/, which is found for example in Swahili and Fula. We shall come back to a more detailed description of differences in vowel quality later.

Vowels are traditionally analyzed with a terminology that differs from that used for consonants. Three main dimensions are used in this classification:

• **Horizontal tongue position**:   front / central / back
• **Vertical tongue position**:   close / close-mid / open-mid / open
• **Lip position**:   unrounded / rounded

Before explaining how this terminology is used, we present TABLE 23, where some vowel symbols are placed in a chart in accordance with there classification. All cells could have been filled, but we have left out some less commonly used symbols.

	front		central		back	
	unrounded	rounded	unrounded	rounded	unrounded	rounded
close	i	y			ɯ	u
close-mid	e	ø	ɘ		ɤ	o
open-mid	ɛ					ɔ
open	a				ɑ	

TABLE 23. *Some important vowel symbols*

Let us start with **i**, which is **front**, **close**, and **unrounded**. *Front & close* means that the body of the tongue is pushed forward and upward, so that dorsum is as close to palatum as possible without creating audible friction. *Unrounded* means that the lips are spread, like in a smile. This is the vowel in English *feel* /ˈfiːl/, where /ː/ tells that the vowel is long.

From **i** we move diagonally down to **ɑ**, which is **back, open,** and **unrounded**. *Back & open* means that the body of the tongue is pulled backwards and downwords, so that the root of the tongue is as close to the back wall of the pharynx as possible without creating audible friction. This is the vowel in English *father* /ˈfɑːðər/.

From **ɑ** we move upwards to **u**, which is **back, close,** and **rounded**. *Back & close* means that the body of the tongue is pulled backwards and upwards, so that the back of the tongue is as close to the velum as possible without creating audible friction. *Rounded* means that the lips are rounded, with a posture resembling the letter O when you look at them in a mirror. This is the vowel in English *fool* /ˈfuːl/.

Let us move to the fourth corner, to **a**, which is **front, open,** and **unrounded**. *Front & open* means that the body of the tongue is pushed forward and downward as much as possible. This vowel is found as the first element in the vowel in English *price* /ˈpraɪs/. This vowel is also found in French *chat* /ʃa/ 'cat' and German *da* /ˈdaː/ 'there'.

Between **i** and **a** you can move the dorsum up and down and stop at whatever height you like, but in this system two vertical tongue positions between close and open have been given a special status: **close-mid** and **open-mid**. These positions are obtained by moving the tongue in three equally long steps between the close and

open: i – e – ɛ – a. Here, **e** is the vowel of French *bébé* /bebe/ 'baby' and **ɛ** is the vowel of French *bête* /bɛːt/ 'animal, beast'.

In whatever position you have your tongue, the lips can vary between **unrounded** and **rounded**. For example, you can add rounding to **i** – **e** and get **y** – **ø**, which are the vowels of French *nu* /ny/ 'naked' and *nœud* /nø/ 'knot'; **y** is also found in Turkish *üzüm* /yzym/ 'grapes'.

Let us then go the vowels **u** – **o** – **ɔ**, which are all *back & rounded*, but which differ by being **close**, **close-mid**, and **open-mid**, respectively. These are the vowels of French *bout* /bu/ 'end', *beau* /bo/ 'beautiful', and *coteau* /kɔto/ 'hillside'. If we remove the rounding of **u** – **o**, we get **ɯ** – **ɤ**, which are the vowels of Vietnamese *hư* /hɯ/ 'spoiled' and *cơm* /kɤm/ 'rice'; **ɯ** is also the vowel of Turkish *kıptı* /kɯptɯ/ 'Gipsy'.

The **central**, **close-mid**, **unrounded** vowel **ə** is the vowel of the first syllable of English *again* /əˈgeɪn/. Central is the horizontal tongue position halfway between front and back.

4.7 Syllables

Speakers of most languages are able to tell how many **syllables** there are in a certain linguistic unit. Speakers of English are generally able to tell that the words *ape* /ˈeɪp/, *baboon* /bæˈbuːn/, and *gorilla* /gəˈrɪlə/ have one, two, and three syllables, respectively. Still, it is difficult to define the syllable in a simple way, and we shall simply present some characteristic features of the syllable.

The English word *book* /ˈbʊk/ consists of one syllable, and may be said to have the structure in FIG. 20 — that is, it has three parts, **onset**, **nucleus**, and **coda**. The onset and the coda constitute the margin of the syllable, that is, its non-nucleus parts.

syllable		
onset	**nucleus**	**coda**
/b	ʊ	k/

FIGURE 20. The syllable /ˈbʊk/

The vowel is the only sound type that is allowed to occur in the nucleus in all languages, while only consonants are allowed to occur in the margin in all languages. Some languages accept more than one consonant in the onset and in the coda, but such congestions of sound are rather rare. European languages (and a few others) are exceptional here — cf. English words like *strength* /ˈstreŋθ/, where the onset is /str/ and the coda is /ŋθ/ — but the languages of Europe constitute only about 3% of the languages of the world.

The only syllable type that is found in all languages is onset+nucleus, with one consonant in the onset and one vowel in the nucleus. There is a strong — but not exceptionless — tendency in the languages of the world for onsets to have a «richer» inventory of sounds than the coda.

4.8 Phonotactics

As mentioned a few lines above, *strength* /ˈstreŋθ/ is an acceptable word in English, but in most languages of the world it is an impossible word, for two reasons.

First, because English has got sounds, for example /θ/, that are quite rare; /θ/ is found in Arabic and Swahili, but not in Hausa, Fula, Turkish, Modern Hebrew, Persian, Hindi, Indonesian, Chinese, Korean, and Japanese.

Secondly, the sound combinations in English are more complicated than in most languages outside Europe. The sound combination principles in a language are called the **phonotactic principles** of that language. They can be formulated on the basis of the word or the syllable. Consider the Japanese sentence in (49), where /. / marks syllable boundaries within words, that is, where one syllable ends and the next starts. Japanese onsets and codas never contain more than one consonant.

(49) *A Japanese sentence*

Zidoosya o motte Amerikazin wa arimasen ka?
/dʒi.doː.ʃa o mot.te a.me.ɾi.ka.dʒin wa a.ɾi.ma.sen ka/
'Aren't there any Americans who don't own cars?'