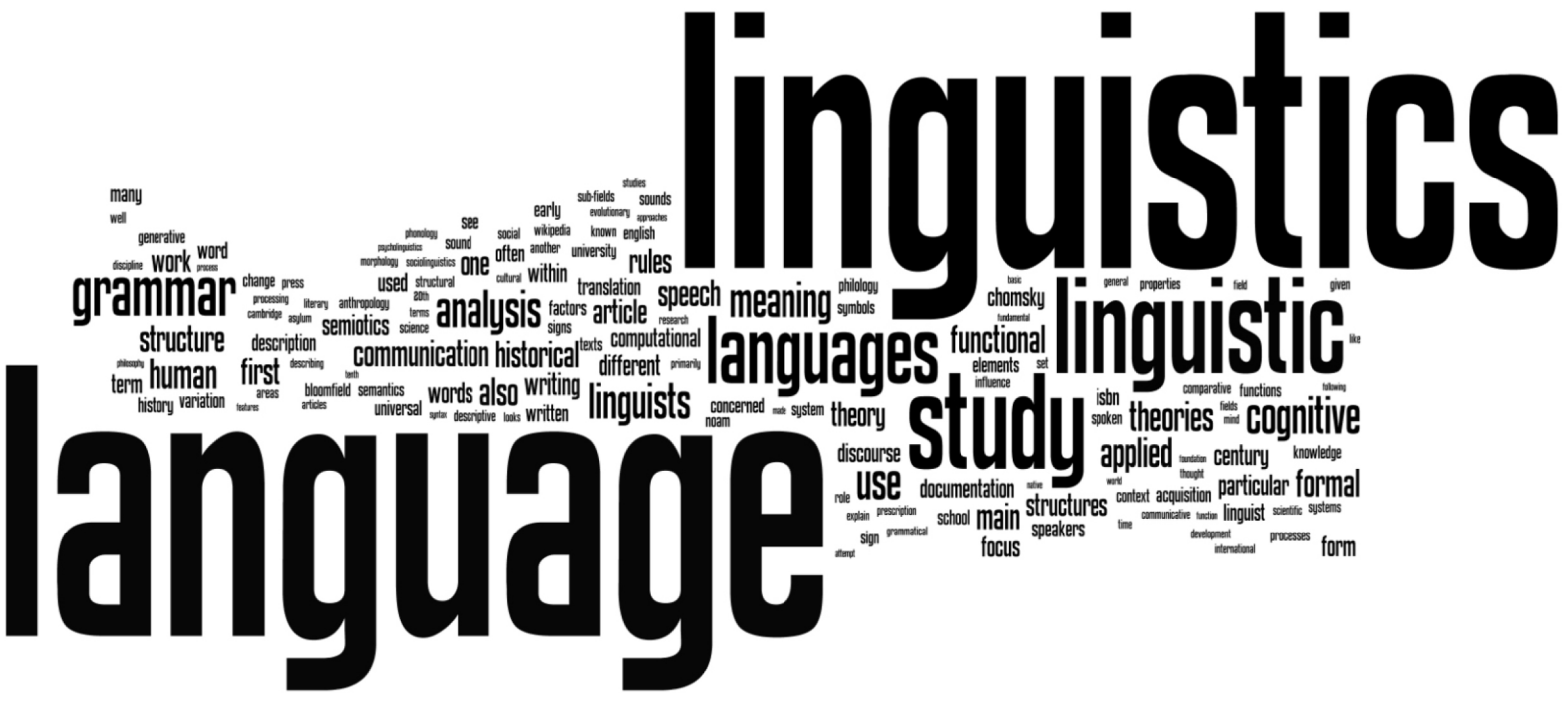
Phonetics and Phonology in Linguistics Anthropology

Eugene Sampson



PHONETICS AND PHONOLOGY IN LINGUISTICS ANTHROPOLOGY

PHONETICS AND PHONOLOGY IN LINGUISTICS ANTHROPOLOGY

Eugene Sampson

BIBLIOTEX
Digital Library

Phonetics and Phonology in Linguistics Anthropology
by Eugene Sampson

www.bibliotex.com

Ebook ISBN: 9781984665249

Published by:
Bibliotex
Canada
Website: www.bibliotex.com

Contents

Chapter 1 Phonetics 1

Chapter 2 Phonology 28

Chapter 3 Modern Phonetics: Main Branches 37

Chapter 4 Language Production 56

Chapter 5 Acoustics 86

Chapter 6 Perception 110

Chapter 7 Phonetic Transcription 160

Chapter 8 Phonologies of the World's Languages 169

Chapter 1

Phonetics

Phonetics is a branch of linguistics that studies how humans produce and perceive sounds, or in the case of sign languages, the equivalent aspects of sign. Phoneticians—linguists who specialize in phonetics—study the physical properties of speech. The field of phonetics is traditionally divided into three sub-disciplines based on the research questions involved such as how humans plan and execute movements to produce speech (articulatory phonetics), how different movements affect the properties of the resulting sound (acoustic phonetics), or how humans convert sound waves to linguistic information (auditory phonetics). Traditionally, the minimal linguistic unit of phonetics is the phone—a speech sound in a language—which differs from the phonological unit of phoneme; the phoneme is an abstract categorization of phones.

Phonetics broadly deals with two aspects of human speech: production—the ways humans make sounds—and perception—the way speech is understood. The communicative modality of a language describes the method by which a language produces and perceives languages. Languages with oral-aural modalities such as English produce speech orally (using the mouth) and perceive speech aurally (using the ears). Sign languages, such as Auslan and ASL, have a manual-visual modality, producing speech manually (using the hands) and perceiving speech visually (using the eyes). ASL and some other sign languages have in addition a manual-manual dialect for use in tactile

signing by deafblind speakers where signs are produced with the hands and perceived with the hands as well.

Language production consists of several interdependent processes which transform a non-linguistic message into a spoken or signed linguistic signal. After identifying a message to be linguistically encoded, a speaker must select the individual words—known as lexical items—to represent that message in a process called lexical selection. During phonological encoding, the mental representation of the words are assigned their phonological content as a sequence of phonemes to be produced. The phonemes are specified for articulatory features which denote particular goals such as closed lips or the tongue in a particular location. These phonemes are then coordinated into a sequence of muscle commands that can be sent to the muscles, and when these commands are executed properly the intended sounds are produced. These movements disrupt and modify an airstream which results in a sound wave. The modification is done by the articulators, with different places and manners of articulation producing different acoustic results. For example, the words *tack* and *sack* both begin with alveolar sounds in English, but differ in how far the tongue is from the alveolar ridge. This difference has large effects on the air stream and thus the sound that is produced. Similarly, the direction and source of the airstream can affect the sound. The most common airstream mechanism is pulmonic—using the lungs—but the glottis and tongue can also be used to produce airstreams.

Language perception is the process by which a linguistic signal is decoded and understood by a listener. In order to perceive speech the continuous acoustic signal must be converted into

discrete linguistic units such as phonemes, morphemes, and words. In order to correctly identify and categorize sounds, listeners prioritize certain aspects of the signal that can reliably distinguish between linguistic categories. While certain cues are prioritized over others, many aspects of the signal can contribute to perception. For example, though oral languages prioritize acoustic information, the McGurk effect shows that visual information is used to distinguish ambiguous information when the acoustic cues are unreliable.

Modern phonetics has three main branches:

- Articulatory phonetics which studies the way sounds are made with the articulators
- Acoustic phonetics which studies the acoustic results of different articulations
- Auditory phonetics which studies the way listeners perceive and understand linguistic signals.

History

Antiquity

The first known phonetic studies were carried out as early as the 6th century BCE by Sanskrit grammarians. The Hindu scholar Pāṇiniis among the most well known of these early investigators, whose four-part grammar, written around 350 BCE, is influential in modern linguistics and still represents "the most complete generative grammar of any language yet written". His grammar formed the basis of modern linguistics and described several important phonetic principles, including voicing. This early account described resonance as being

produced either by tone, when vocal folds are closed, or noise, when vocal folds are open. The phonetic principles in the grammar are considered "primitives" in that they are the basis for his theoretical analysis rather than the objects of theoretical analysis themselves, and the principles can be inferred from his system of phonology.

Modern

Advancements in phonetics after Pāṇini and his contemporaries were limited until the modern era, save some limited investigations by Greek and Roman grammarians. In the millennia between Indic grammarians and modern phonetics, the focus shifted from the difference between spoken and written language, which was the driving force behind Pāṇini's account, and began to focus on the physical properties of speech alone. Sustained interest in phonetics began again around 1800 CE with the term "phonetics" being first used in the present sense in 1841. With new developments in medicine and the development of audio and visual recording devices, phonetic insights were able to use and review new and more detailed data. This early period of modern phonetics included the development of an influential phonetic alphabet based on articulatory positions by Alexander Melville Bell. Known as visible speech, it gained prominence as a tool in the oral education of deaf children.

Before the widespread availability of audio recording equipment, phoneticians relied heavily on a tradition of practical phonetics to ensure that transcriptions and findings were able to be consistent across phoneticians. This training involved both ear training—the recognition of speech sounds—

as well as production training—the ability to produce sounds. Phoneticians were expected to learn to recognize by ear the various sounds on the International Phonetic Alphabet and the IPA still tests and certifies speakers on their ability to accurately produce the phonetic patterns of English (though they have discontinued this practice for other languages). As a revision of his visible speech method, Melville Bell developed a description of vowels by height and backness resulting in 9cardinal vowels. As part of their training in practical phonetics, phoneticians were expected to learn to produce these cardinal vowels in order to anchor their perception and transcription of these phones during fieldwork. This approach was critiqued by Peter Ladefoged in the 1960s based on experimental evidence where he found that cardinal vowels were auditory rather than articulatory targets, challenging the claim that they represented articulatory anchors by which phoneticians could judge other articulations.

Production

Language production consists of several interdependent processes which transform a nonlinguistic message into a spoken or signed linguistic signal. Linguists debate whether the process of language production occurs in a series of stages (serial processing) or whether production processes occur in parallel. After identifying a message to be linguistically encoded, a speaker must select the individual words—known as lexical items—to represent that message in a process called lexical selection. The words are selected based on their meaning, which in linguistics is called semantic information. Lexical selection activates the word's lemma, which contains both semantic and grammatical information about the word.

After an utterance has been planned, it then goes through phonological encoding. In this stage of language production, the mental representation of the words are assigned their phonological content as a sequence of phonemes to be produced. The phonemes are specified for articulatory features which denote particular goals such as closed lips or the tongue in a particular location.

These phonemes are then coordinated into a sequence of muscle commands that can be sent to the muscles, and when these commands are executed properly the intended sounds are produced. Thus the process of production from message to sound can be summarized as the following sequence:

- Message planning
- Lemma selection
- Retrieval and assignment of phonological word forms
- Articulatory specification
- Muscle commands
- Articulation
- Speech sounds

Place of articulation

Sounds which are made by a full or partial constriction of the vocal tract are called consonants. Consonants are pronounced in the vocal tract, usually in the mouth, and the location of this constriction affects the resulting sound. Because of the close connection between the position of the tongue and the resulting sound, the place of articulation is an important concept in many subdisciplines of phonetics.

Sounds are partly categorized by the location of a constriction as well as the part of the body doing the constricting. For example, in English the words *fought* and *thought* are a minimal pair differing only in the organ making the construction rather than the location of the construction. The "f" in *fought* is a labiodental articulation made with the bottom lip against the teeth. The "th" in *thought* is a linguodental articulation made with the tongue against the teeth. Constrictions made by the lips are calledlabials while those made with the tongue are called lingual.

Constrictions made with the tongue can be made in several parts of the vocal tract, broadly classified into coronal, dorsal and radical places of articulation. Coronal articulations are made with the front of the tongue, dorsal articulations are made with the back of the tongue, and radical articulations are made in the pharynx. These divisions are not sufficient for distinguishing and describing all speech sounds. For example, in English the sounds [s] and [ʃ] are both coronal, but they are produced in different places of the mouth. To account for this, more detailed places of articulation are needed based upon the area of the mouth in which the constriction occurs.

Labial

Articulations involving the lips can be made in three different ways: with both lips (bilabial), with one lip and the teeth (labiodental), and with the tongue and the upper lip (linguolabial). Depending on the definition used, some or all of these kinds of articulations may be categorized into the class of labial articulations. Bilabial consonantsare made with both lips. In producing these sounds the lower lip moves farthest to

meet the upper lip, which also moves down slightly, though in some cases the force from air moving through the aperture (opening between the lips) may cause the lips to separate faster than they can come together. Unlike most other articulations, both articulators are made from soft tissue, and so bilabial stops are more likely to be produced with incomplete closures than articulations involving hard surfaces like the teeth or palate. Bilabial stops are also unusual in that an articulator in the upper section of the vocal tract actively moves downwards, as the upper lip shows some active downward movement. Linguolabial consonantsare made with the blade of the tongue approaching or contacting the upper lip. Like in bilabial articulations, the upper lip moves slightly towards the more active articulator. Articulations in this group do not have their own symbols in the International Phonetic Alphabet, rather, they are formed by combining an apical symbol with a diacritic implicitly placing them in the coronal category. They exist in a number of languages indigenous to Vanuatu such as Tangoa.

Labiodental consonants are made by the lower lip rising to the upper teeth. Labiodental consonants are most often fricatives while labiodental nasals are also typologically common. There is debate as to whether true labiodental plosives occur in any natural language, though a number of languages are reported to have labiodental plosives including Zulu, Tonga, and Shubi.

Dorsal

Dorsal consonants are those consonants made using the tongue body rather than the tip or blade and are typically produced at the palate, velum or uvula. Palatal consonantsare

made using the tongue body against the hard palate on the roof of the mouth. They are frequently contrasted with velar or uvular consonants, though it is rare for a language to contrast all three simultaneously, with Jaqaru as a possible example of a three-way contrast. Velar consonantsare made using the tongue body against the velum. They are incredibly common cross-linguistically; almost all languages have a velar stop. Because both velars and vowels are made using the tongue body, they are highly affected by coarticulation with vowels and can be produced as far forward as the hard palate or as far back as the uvula. These variations are typically divided into front, central, and back velars in parallel with the vowel space. They can be hard to distinguish phonetically from palatal consonants, though are produced slightly behind the area of prototypical palatal consonants. Uvular consonants are made by the tongue body contacting or approaching the uvula. They are rare, occurring in an estimated 19 percent of languages, and large regions of the Americas and Africa have no languages with uvular consonants. In languages with uvular consonants, stops are most frequent followed by continuants (including nasals).

Pharyngeal and laryngeal

Consonants made by constrictions of the throat are pharyngeals, and those made by a constriction in the larynx are laryngeal. Laryngeals are made using the vocal folds as the larynx is too far down the throat to reach with the tongue. Pharyngeals however are close enough to the mouth that parts of the tongue can reach them.

Radical consonants either use the root of the tongue or the epiglottis during production and are produced very far back in the vocal tract. Pharyngeal consonants are made by retracting the root of the tongue far enough to almost touch the wall of the pharynx. Due to production difficulties, only fricatives and approximants can produced this way. Epiglottal consonantsare made with the epiglottis and the back wall of the pharynx. Epiglottal stops have been recorded in Dahalo. Voiced epiglottal consonants are not deemed possible due to the cavity between the glottis and epiglottis being too small to permit voicing.

Glottal consonants are those produced using the vocal folds in the larynx. Because the vocal folds are the source of phonation and below the oro-nasal vocal tract, a number of glottal consonants are impossible such as a voiced glottal stop. Three glottal consonants are possible, a voiceless glottal stop and two glottal fricatives, and all are attested in natural languages. Glottal stops, produced by closing the vocal folds, are notably common in the world's languages. While many languages use them to demarcate phrase boundaries, some languages like Huatla Mazatec have them as contrastive phonemes. Additionally, glottal stops can be realized as laryngealization of the following vowel in this language. Glottal stops, especially between vowels, do usually not form a complete closure. True glottal stops normally occur only when they'regeminated.

The larynx

The larynx, commonly known as the "voice box", is a cartilaginous structure in the trachea responsible for phonation. The vocal folds (chords) are held together so that

they vibrate, or held apart so that they do not. The positions of the vocal folds are achieved by movement of the arytenoid cartilages. The intrinsic laryngeal muscles are responsible for moving the arytenoid cartilages as well as modulating the tension of the vocal folds. If the vocal folds are not close or tense enough, they will either vibrate sporadically or not at all. If they vibrate sporadically it will result in either creaky or breathy voice, depending on the degree; if don't vibrate at all, the result will be voicelessness.

In addition to correctly positioning the vocal folds, there must also be air flowing across them or they will not vibrate. The difference in pressure across the glottis required for voicing is estimated at 1 – 2 cm H_2O (98.0665 – 196.133 pascals). The pressure differential can fall below levels required for phonation either because of an increase in pressure above the glottis (superglottal pressure) or a decrease in pressure below the glottis (subglottal pressure). The subglottal pressure is maintained by the respiratory muscles. Supraglottal pressure, with no constrictions or articulations, is equal to about atmospheric pressure. However, because articulations—especially consonants—represent constrictions of the airflow, the pressure in the cavity behind those constrictions can increase resulting in a higher supraglottal pressure.

Lexical access

According to the lexical access model two different stages of cognition are employed; thus, this concept is known as the two-stage theory of lexical access. The first stage, lexical selection provides information about lexical items required to construct the functional level representation. These items are

retrieved according to their specific semantic and syntactic properties, but phonological forms are not yet made available at this stage. The second stage, retrieval of wordforms, provides information required for building the positional level representation.

Articulatory models

When producing speech, the articulators move through and contact particular locations in space resulting in changes to the acoustic signal. Some models of speech production take this as the basis for modeling articulation in a coordinate system that may be internal to the body (intrinsic) or external (extrinsic). Intrinsic coordinate systems model the movement of articulators as positions and angles of joints in the body. Intrinsic coordinate models of the jaw often use two to three degrees of freedom representing translation and rotation. These face issues with modeling the tongue which, unlike joints of the jaw and arms, is a muscular hydrostat—like an elephant trunk—which lacks joints. Because of the different physiological structures, movement paths of the jaw are relatively straight lines during speech and mastication, while movements of the tongue follow curves.

Straight-line movements have been used to argue articulations as planned in extrinsic rather than intrinsic space, though extrinsic coordinate systems also include acoustic coordinate spaces, not just physical coordinate spaces. Models that assume movements are planned in extrinsic space run into an inverse problem of explaining the muscle and joint locations which produce the observed path or acoustic signal. The arm, for example, has seven degrees of freedom and 22 muscles, so

multiple different joint and muscle configurations can lead to the same final position. For models of planning in extrinsic acoustic space, the same one-to-many mapping problem applies as well, with no unique mapping from physical or acoustic targets to the muscle movements required to achieve them. Concerns about the inverse problem may be exaggerated, however, as speech is a highly learned skill using neurological structures which evolved for the purpose.

The equilibrium-point model proposes a resolution to the inverse problem by arguing that movement targets be represented as the position of the muscle pairs acting on a joint. Importantly, muscles are modeled as springs, and the target is the equilibrium point for the modeled spring-mass system. By using springs, the equilibrium point model can easily account for compensation and response when movements are disrupted. They are considered a coordinate model because they assume that these muscle positions are represented as points in space, equilibrium points, where the spring-like action of the muscles converges.

Gestural approaches to speech production propose that articulations are represented as movement patterns rather than particular coordinates to hit. The minimal unit is a gesture that represents a group of "functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech-relevant goal (e.g., a bilabial closure)." These groups represent coordinative structures or "synergies" which view movements not as individual muscle movements but as task-dependent groupings of muscles which work together as a single unit. This reduces the degrees of freedom in articulation planning, a problem especially in

intrinsic coordinate models, which allows for any movement that achieves the speech goal, rather than encoding the particular movements in the abstract representation. Coarticulationis well described by gestural models as the articulations at faster speech rates can be explained as composites of the independent gestures at slower speech rates.

Acoustics

Speech sounds are created by the modification of an airstream which results in a sound wave. The modification is done by the articulators, with different places and manners of articulation producing different acoustic results. Because the posture of the vocal tract, not just the position of the tongue can affect the resulting sound, the manner of articulation is important for describing the speech sound. The words *tack* and *sack* both begin with alveolar sounds in English, but differ in how far the tongue is from the alveolar ridge. This difference has large affects on the air stream and thus the sound that is produced. Similarly, the direction and source of the airstream can affect the sound. The most common airstream mechanism is pulmonic—using the lungs—but the glottis and tongue can also be used to produce airstreams.

Voicing and phonation types

A major distinction between speech sounds is whether they are voiced. Sounds are voiced when the vocal folds begin to vibrate in the process of phonation. Many sounds can be produced with or without phonation, though physical constraints may make phonation difficult or impossible for some articulations. When articulations are voiced, the main source of noise is the

periodic vibration of the vocal folds. Articulations like voiceless plosives have no acoustic source and are noticeable by their silence, but other voiceless sounds like fricatives create their own acoustic source regardless of phonation.

Phonation is controlled by the muscles of the larynx, and languages make use of more acoustic detail than binary voicing. During phonation, the vocal folds vibrate at a certain rate. This vibration results in a periodic acoustic waveform comprising a fundamental frequency and its harmonics. The fundamental frequency of the acoustic wave can be controlled by adjusting the muscles of the larynx, and listeners perceive this fundamental frequency as pitch. Languages use pitch manipulation to convey lexical information in tonal languages, and many languages use pitch to mark prosodic or pragmatic information.

For the vocal folds to vibrate, they must be in the proper position and there must be air flowing through the glottis. Phonation types are modeled on a continuum of glottal states from completely open (voiceless) to completely closed (glottal stop). The optimal position for vibration, and the phonation type most used in speech, modal voice, exists in the middle of these two extremes. If the glottis is slightly wider, breathy voice occurs, while bringing the vocal folds closer together results in creaky voice.

The normal phonation pattern used in typical speech is modal voice, where the vocal folds are held close together with moderate tension. The vocal folds vibrate as a single unit periodically and efficiently with a full glottal closure and no aspiration. If they are pulled farther apart, they do not vibrate

and so produce voiceless phones. If they are held firmly together they produce a glottal stop.

If the vocal folds are held slightly further apart than in modal voicing, they produce phonation types like breathy voice (or murmur) and whispery voice. The tension across the vocal ligaments (vocal cords) is less than in modal voicing allowing for air to flow more freely. Both breathy voice and whispery voice exist on a continuum loosely characterized as going from the more periodic waveform of breathy voice to the more noisy waveform of whispery voice. Acoustically, both tend to dampen the first formant with whispery voice showing more extreme deviations.

Holding the vocal folds more tightly together results in a creaky voice. The tension across the vocal folds is less than in modal voice, but they are heldtightly together resulting in only the ligaments of the vocal folds vibrating. The pulses are highly irregular, with low pitch and frequency amplitude.

Some languages do not maintain a voicing distinction for some consonants, but all languages use voicing to some degree. For example, no language is known to have a phonemic voicing contrast for vowels with all known vowels canonically voiced. Other positions of the glottis, such as breathy and creaky voice, are used in a number of languages, like Jalapa Mazatec, to contrast phonemes while in other languages, like English, they exist allophonically.

There are several ways to determine if a segment is voiced or not, the simplest being to feel the larynx during speech and note when vibrations are felt. More precise measurements can be obtained through acoustic analysis of a spectrogram or

spectral slice. In a spectrographic analysis, voiced segments show a voicing bar, a region of high acoustic energy, in the low frequencies of voiced segments. In examining a spectral splice, the acoustic spectrum at a given point in time a model of the vowel pronounced reverses the filtering of the mouth producing the spectrum of the glottis. A computational model of the unfiltered glottal signal is then fitted to the inverse filtered acoustic signal to determine the characteristics of the glottis. Visual analysis is also available using specialized medical equipment such as ultrasound and endoscopy.

Vowels

Vowels are broadly categorized by the area of the mouth in which they are produced, but because they are produced without a constriction in the vocal tract their precise description relies on measuring acoustic correlates of tongue position.

The location of the tongue during vowel production changes the frequencies at which the cavity resonates, and it is these resonances—known as formants—which are measured and used to characterize vowels.

Vowel height traditionally refers to the highest point of the tongue during articulation. The height parameter is divided into four primary levels: high (close), close-mid, open-mid and low (open). Vowels whose height are in the middle are referred to as mid. Slightly opened close vowels and slightly closed open vowels are referred to as near-close and near-open respectively. The lowest vowels are not just articulated with a lowered tongue, but also by lowering the jaw.

While the IPA implies that there are seven levels of vowel height, it is unlikely that a given language can minimally contrast all seven levels. Chomsky and Halle suggest that there are only three levels, although four levels of vowel height seem to be needed to describe Danish and it's possible that some languages might even need five.

Vowel backness is dividing into three levels: front, central and back. Languages usually do not minimally contrast more than two levels of vowel backness. Some languages claimed to have a three-way backness distinction include Nimboran and Norwegian.

In most languages, the lips during vowel production can be classified as either rounded or unrounded (spread), although other types of lip positions, such as compression and protrusion, have been described. Lip position is correlated with height and backness: front and low vowels tend to be unrounded whereas back and high vowels are usually rounded. Paired vowels on the IPA chart have the spread vowel on the left and the rounded vowel on the right.

Together with the universal vowel features described above, some languages have additional features such as nasality, length and different types of phonation such as voiceless or creaky. Sometimes more specialized tongue gestures such as rhoticity, advanced tongue root, pharyngealization, stridency and frication are required to describe a certain vowel.

Manner of articulation

Knowing the place of articulation is not enough to fully describe a consonant, the way in which the stricture happens

is equally important. Manners of articulation describe how exactly the active articulator modifies, narrows or closes off the vocal tract.

Stops (also referred to as plosives) are consonants where the airstream is completely obstructed. Pressure builds up in the mouth during the stricture, which is then released as a small burst of sound when the articulators move apart. The velum is raised so that air cannot flow through the nasal cavity. If the velum is lowered and allows for air to flow through the nose, the result in a nasal stop. However, phoneticians almost always refer to nasal stops as just "nasals".Affricates are a sequence of stops followed by a fricative in the same place.

Fricatives are consonants where the airstream is made turbulent by partially, but not completely, obstructing part of the vocal tract. Sibilants are a special type of fricative where the turbulent airstream is directed towards the teeth, creating a high-pitched hissing sound.

Nasals (sometimes referred to as nasal stops) are consonants in which there's a closure in the oral cavity and the velum is lowered, allowing air to flow through the nose.

In an approximant, the articulators come close together, but not to such an extent that allows a turbulent airstream.

Laterals are consonants in which the airstream is obstructed along the center of the vocal tract, allowing the airstream to flow freely on one or both sides. Laterals have also been defined as consonants in which the tongue is contracted in such a way that the airstream is greater around the sides than

over the center of the tongue. The first definition does not allow for air to flow over the tongue.

Trills are consonants in which the tongue or lips are set in motion by the airstream. The stricture is formed in such a way that the airstream causes a repeating pattern of opening and closing of the soft articulator(s). Apical trills typically consist of two or three periods of vibration.

Taps and flaps are single, rapid, usually apical gestures where the tongue is thrown against the roof of the mouth, comparable to a very rapid stop. These terms are sometimes used interchangeably, but some phoneticians make a distinction. In a tap, the tongue contacts the roof in a single motion whereas in a flap the tongue moves tangentially to the roof of the mouth, striking it in passing.

During a glottalic airstream mechanism, the glottis is closed, trapping a body of air. This allows for the remaining air in the vocal tract to be moved separately. An upward movement of the closed glottis will move this air out, resulting in it an ejective consonant. Alternatively, the glottis can lower, sucking more air into the mouth, which results in an implosive consonant.

Clicks are stops in which tongue movement causes air to be sucked in the mouth, this is referred to as a velaric airstream. During the click, the air becomes rarefied between two articulatory closures, producing a loud 'click' sound when the anterior closure is released. The release of the anterior closure is referred to as the click influx. The release of the posterior closure, which can be velar or uvular, is the click efflux. Clicks are used in several African language families, such as the Khoisan and Bantu languages.

Pulmonary and subglottal system

The lungs drive nearly all speech production, and their importance in phonetics is due to their creation of pressure for pulmonic sounds. The most common kinds of sound across languages are pulmonic egress, where air is exhaled from the lungs. The opposite is possible, though no language is known to have pulmonic ingressive sounds as phonemes. Many languages such as Swedish use them for paralinguistic articulations such as affirmations in a number of genetically and geographically diverse languages. Both egressive and ingressive sounds rely on holding the vocal folds in a particular posture and using the lungs to draw air across the vocal folds so that they either vibrate (voiced) or do not vibrate (voiceless). Pulmonic articulations are restricted by the volume of air able to be exhaled in a given respiratory cycle, known as the vital capacity.

The lungs are used to maintain two kinds of pressure simultaneously in order to produce and modify phonation. To produce phonation at all, the lungs must maintain a pressure of 3–5 cm H_2O higher than the pressure above the glottis. However small and fast adjustments are made to the subglottal pressure to modify speech for suprasegmental features like stress. A number of thoracic muscles are used to make these adjustments. Because the lungs and thorax stretch during inhalation, the elastic forces of the lungs alone can produce pressure differentials sufficient for phonation at lung volumes above 50 percent of vital capacity. Above 50 percent of vital capacity, the respiratory muscles are used to "check" the elastic forces of the thorax to maintain a stable pressure

differential. Below that volume, they are used to increase the subglottal pressure by actively exhaling air.

During speech, the respiratory cycle is modified to accommodate both linguistic and biological needs. Exhalation, usually about 60 percent of the respiratory cycle at rest, is increased to about 90 percent of the respiratory cycle. Because metabolic needs are relatively stable, the total volume of air moved in most cases of speech remains about the same as quiet tidal breathing. Increases in speech intensity of 18 dB (a loud conversation) has relatively little impact on the volume of air moved. Because their respiratory systems are not as developed as adults, children tend to use a larger proportion of their vital capacity compared to adults, with more deep inhales.

Source–filter theory

The source–filter model of speech is a theory of speech production which explains the link between vocal tract posture and the acoustic consequences. Under this model, the vocal tract can be modeled as a noise source coupled onto an acoustic filter. The noise source in many cases is the larynx during the process of voicing, though other noise sources can be modeled in the same way. The shape of the supraglottal vocal tract acts as the filter, and different configurations of the articulators result in different acoustic patterns. These changes are predictable. The vocal tract can be modeled as a sequence of tubes, closed at one end, with varying diameters, and by using equations for acoustic resonance the acoustic effect of an articulatory posture can be derived. The process of inverse filtering uses this principle to analyze the source

spectrum produced by the vocal folds during voicing. By taking the inverse of a predicted filter, the acoustic effect of the supraglottal vocal tract can be undone giving the acoustic spectrum produced by the vocal folds. This allows quantitative study of the various phonation types.

Subdisciplines

Acoustic phonetics

Acoustic phonetics deals with the acoustic properties of speech sounds. The sensation of sound is caused by pressure fluctuations which cause the eardrum to move. The ear transforms this movement into neural signals that the brain registers as sound. Acoustic waveforms are records that measure these pressure fluctuations.

Articulatory phonetics

Articulatory phonetics deals with the ways in which speech sounds are made.

Auditory phonetics

Auditory phonetics studies how humans perceive speech sounds. Due to the anatomical features of the auditory system distorting the speech signal, humans do not experience speech sounds as perfect acoustic records. For example, the auditory impressions of volume, measured in decibels (dB), does not linearly match the difference in sound pressure.

The mismatch between acoustic analyses and what the listener hears is especially noticeable in speech sounds that have a lot of high-frequency energy, such as certain fricatives. To reconcile this mismatch, functional models of the auditory system have been developed.

Describing sounds

Human languages use many different sounds and in order to compare them linguists must be able to describe sounds in a way that is language independent. Speech sounds can be described in a number of ways. Most commonly speech sounds are referred to by the mouth movements needed to produce them. Consonants and vowels are two gross categories that phoneticians define by the movements in a speech sound. More fine-grained descriptors are parameters such as place of articulation. Place of articulation, manner of articulation, and voicingare used to describe consonants and are the main divisions of the International Phonetic Alphabet consonant chart. Vowels are described by their height, backness, and rounding. Sign language are described using a similar but distinct set of parameters to describe signs: location, movement, hand shape, palm orientation, and non-manual features. In addition to articulatory descriptions, sounds used in oral languages can be described using their acoustics. Because the acoustics are a consequence of the articulation, both methods of description are sufficient to distinguish sounds with the choice between systems dependent on the phonetic feature being investigated.

Consonants are speech sounds that are articulated with a complete or partial closure of the vocal tract. They are

generally produced by the modification of an airstream exhaled from the lungs. The respiratory organs used to create and modify airflow are divided into three regions: the vocal tract (supralaryngeal), the larynx, and the subglottal system. The airstream can be either egressive (out of the vocal tract) or ingressive (into the vocal tract). In pulmonic sounds, the airstream is produced by the lungs in the subglottal system and passes through the larynx and vocal tract. Glottalic sounds use an airstream created by movements of the larynx without airflow from the lungs. Click consonants are articulated through the rarefaction of air using the tongue, followed by releasing the forward closure of the tongue.

Vowels are syllabic speech sounds that are pronounced without any obstruction in the vocal tract. Unlike consonants, which usually have definite places of articulation, vowels are defined in relation to a set of reference vowels called cardinal vowels. Three properties are needed to define vowels: tongue height, tongue backness and lip roundedness. Vowels that are articulated with a stable quality are called monophthongs; a combination of two separate vowels in the same syllable is a diphthong. In the IPA, the vowels are represented on a trapezoid shape representing the human mouth: the vertical axis representing the mouth from floor to roof and the horizontal axis represents the front-back dimension.

Transcription

Phonetic transcription is a system for transcribing phones that occur in a language, whether oral or sign. The most widely known system of phonetic transcription, the International Phonetic Alphabet (IPA), provides a standardized set of symbols

for oral phones. The standardized nature of the IPA enables its users to transcribe accurately and consistently the phones of different languages, dialects, and idiolects. The IPA is a useful tool not only for the study of phonetics, but also for language teaching, professional acting, and speech pathology.

While no sign language has a standardized writing system, linguists have developed their own notation systems that describe the handshape, location and movement. The Hamburg Notation System (HamNoSys) is similar to the IPA in that it allows for varying levels of detail. Some notation systems such as KOMVA and the Stokoe systemwere designed for use in dictionaries; they also make use of alphabetic letters in the local language for handshapes whereas HamNoSys represents the handshape directly. SignWriting aims to be an easy-to-learn writing system for sign languages, although it has not been officially adopted by any deaf community yet.

Sign languages

Unlike spoken languages, words in sign languagesare perceived with the eyes instead of the ears. Signs are articulated with the hands, upper body and head. The main articulators are the hands and arms. Relative parts of the arm are described with the terms proximal and distal. Proximal refers to a part closer to the torso whereas a distal part is further away from it. For example, a wrist movement is distal compared to an elbow movement. Due to requiring less energy, distal movements are generally easier to produce. Various factors – such as muscle flexibility or being consideredtaboo – restrict what can be considered a sign. Native signers do not look at their conversation partner's hands. Instead, their gaze is fixated on

the face. Because peripheral vision is not as focused as the center of the visual field, signs articulated near the face allow for more subtle differences in finger movement and location to be perceived.

Unlike spoken languages, sign languages have two identical articulators: the hands. Signers may use whichever hand they prefer with no disruption in communication. Due to universal neurological limitations, two-handed signs generally have the same kind of articulation in both hands; this is referred to as the Symmetry Condition. The second universal constraint is the Dominance Condition, which holds that when two handshapes are involved, one hand will remain stationary and have a more limited set handshapes compared to the dominant, moving hand. Additionally, it is common for one hand in a two-handed sign to be dropped during informal conversations, a process referred to as weak drop. Just like words in spoken languages, coarticulation may cause signs to influence each other's form. Examples include the handshapes of neighboring signs becoming more similar to each other (assimilation) or weak drop (an instance of deletion).

Chapter 2

Phonology

Phonology is a branch of linguistics that studies how languages or dialects systematically organize their sounds (or signs, in sign languages). The term also refers to the sound system of any particular language variety. At one time, the study of phonology only related to the study of the systems of phonemes in spoken languages. Now it may relate to

- (a) anylinguistic analysis either at a level beneath the word (including syllable, onset and rime, articulatory gestures, articulatory features, mora, etc.), or
- (b) all levels of language where sound or signs are structured to convey linguistic meaning.

Sign languages have a phonological system equivalent to the system of sounds in spoken languages. The building blocks of signs are specifications for movement, location and handshape.

Terminology

The word 'phonology' (as in *the phonology of English*) can also refer to the phonological system (sound system) of a given language. This is one of the fundamental systems which a language is considered to comprise, like its syntax, its morphology and its vocabulary.

Phonology is often distinguished from *phonetics*. While phonetics concerns the physical production, acoustic

transmission and perception of the sounds of speech, phonology describes the way sounds function within a given language or across languages to encode meaning. For many linguists, phonetics belongs to descriptive linguistics, and phonology to theoretical linguistics, although establishing the phonological system of a language is necessarily an application of theoretical principles to analysis of phonetic evidence. Note that this distinction was not always made, particularly before the development of the modern concept of the phoneme in the mid 20th century. Some subfields of modern phonology have a crossover with phonetics in descriptive disciplines such as psycholinguistics and speech perception, resulting in specific areas like articulatory phonology or laboratory phonology.

Derivation and definitions

The word *phonology* comes from Ancient Greekφωνή, *phōnḗ*, "voice, sound," and the suffix *-logy* (which is from Greek λόγος, *lógos*, "word, speech, subject of discussion"). Definitions of the term vary. Nikolai Trubetzkoy in *Grundzüge der Phonologie* (1939) defines phonology as "the study of sound pertaining to the system of language," as opposed to phonetics, which is "the study of sound pertaining to the act of speech" (the distinction between *language* and *speech* being basically Saussure's distinction between *langue* and *parole*). More recently, Lass (1998) writes that phonology refers broadly to the subdiscipline of linguistics concerned with the sounds of language, while in more narrow terms, "phonology proper is concerned with the function, behavior and organization of sounds as linguistic items." According to Clark *et al.* (2007), it means the systematic use of sound to encode meaning in any spoken human language, or the field of linguistics studying this use.

History

Early evidence for a systematic study of the sounds in a language appears in the 4th century BCE *Ashtadhyayi*, a Sanskrit grammar composed by Pāṇini. In particular the *Shiva Sutras*, an auxiliary text to the *Ashtadhyayi*, introduces what may be considered a list of the phonemes of the Sanskrit language, with a notational system for them that is used throughout the main text, which deals with matters of morphology, syntax and semantics.

Ibn Jinni of Mosul, a pioneer in phonology, wrote prolifically in the 10th century on Arabic morphology and phonology of Arabic in works such as *Kitāb Al-Munṣif*, *Kitāb Al-Muḥtasab*, *andKitāb Al-Khaṣāʾiṣ* [ar].

The study of phonology as it exists today is defined by the formative studies of the 19th-century Polish scholar Jan Baudouin de Courtenay, who (together with his students MikołajKruszewski and Lev Shcherba) shaped the modern usage of the term *phoneme* in a series of lectures in 1876–1877. The word *phoneme* had been coined a few years earlier in 1873 by the French linguist A. Dufriche-Desgenettes. In a paper read at 24 May meeting of the Société de Linguistique de Paris, Dufriche-Desgenettes proposed that *phoneme* serve as a one-word equivalent for the German *Sprachlaut*. Baudouin de Courtenay's subsequent work, though often unacknowledged, is considered to be the starting point of modern phonology. He also worked on the theory of phonetic alternations (what is now called allophony and morphophonology), and may have had an influence on the work of Saussure according to E. F. K. Koerner.

An influential school of phonology in the interwar period was the Prague school. One of its leading members was Prince Nikolai Trubetzkoy, whose *Grundzüge der Phonologie* (*Principles of Phonology*), published posthumously in 1939, is among the most important works in the field from this period. Directly influenced by Baudouin de Courtenay, Trubetzkoy is considered the founder of morphophonology, although this concept had also been recognized by de Courtenay. Trubetzkoy also developed the concept of the *archiphoneme*. Another important figure in the Prague school was Roman Jakobson, who was one of the most prominent linguists of the 20th century.

In 1968Noam Chomsky and Morris Halle published *The Sound Pattern of English* (SPE), the basis for generative phonology. In this view, phonological representations are sequences of segments made up of distinctive features. These features were an expansion of earlier work by Roman Jakobson, Gunnar Fant, and Morris Halle. The features describe aspects of articulation and perception, are from a universally fixed set, and have the binary values + or –. There are at least two levels of representation: underlying representation and surface phonetic representation. Ordered phonological rules govern how underlying representationis transformed into the actual pronunciation (the so-called surface form). An important consequence of the influence SPE had on phonological theory was the downplaying of the syllable and the emphasis on segments. Furthermore, the generativists folded morphophonology into phonology, which both solved and created problems.

Natural phonology is a theory based on the publications of its proponent David Stampe in 1969 and (more explicitly) in 1979. In this view, phonology is based on a set of universal phonological processes that interact with one another; which ones are active and which are suppressed is language-specific. Rather than acting on segments, phonological processes act on distinctive features within prosodic groups. Prosodic groups can be as small as a part of a syllable or as large as an entire utterance. Phonological processes are unordered with respect to each other and apply simultaneously (though the output of one process may be the input to another). The second most prominent natural phonologist is Patricia Donegan (Stampe's wife); there are many natural phonologists in Europe, and a few in the U.S., such as Geoffrey Nathan. The principles of natural phonology were extended to morphology by Wolfgang U. Dressler, who founded natural morphology.

In 1976, John Goldsmith introduced autosegmental phonology. Phonological phenomena are no longer seen as operating on *one* linear sequence of segments, called phonemes or feature combinations, but rather as involving *some parallel sequences* of features which reside on multiple tiers. Autosegmental phonology later evolved into feature geometry, which became the standard theory of representation for theories of the organization of phonology as different as lexical phonology and optimality theory.

Government phonology, which originated in the early 1980s as an attempt to unify theoretical notions of syntactic and phonological structures, is based on the notion that all languages necessarily follow a small set of principles and vary according to their selection of certain binary parameters. That

is, all languages' phonological structures are essentially the same, but there is restricted variation that accounts for differences in surface realizations. Principles are held to be inviolable, though parameters may sometimes come into conflict. Prominent figures in this field include Jonathan Kaye, Jean Lowenstamm, Jean-Roger Vergnaud, MonikCharette, and John Harris.

In a course at the LSA summer institute in 1991, Alan Prince and Paul Smolensky developed optimality theory—an overall architecture for phonology according to which languages choose a pronunciation of a word that best satisfies a list of constraints ordered by importance; a lower-ranked constraint can be violated when the violation is necessary in order to obey a higher-ranked constraint. The approach was soon extended to morphology by John McCarthy and Alan Prince, and has become a dominant trend in phonology. The appeal to phonetic grounding of constraints and representational elements (e.g. features) in various approaches has been criticized by proponents of 'substance-free phonology', especially by Mark Hale and Charles Reiss.

An integrated approach to phonological theory that combines synchronic and diachronic accounts to sound patterns was initiated with Evolutionary Phonology in recent years.

Analysis of phonemes

An important part of traditional, pre-generative schools of phonology is studying which sounds can be grouped into distinctive units within a language; these units are known as phonemes. For example, in English, the "p" sound in *pot* is

aspirated (pronounced [pʰ]) while that in *spot* is not aspirated (pronounced [p]). However, English speakers intuitively treat both sounds as variations (allophones) of the same phonological category, that is of the phoneme /p/. (Traditionally, it would be argued that if an aspirated [pʰ] were interchanged with the unaspirated [p] in *spot*, native speakers of English would still hear the same words; that is, the two sounds are perceived as "the same" /p/.) In some other languages, however, these two sounds are perceived as different, and they are consequently assigned to different phonemes. For example, in Thai, Bengali, and Quechua, there are minimal pairs of words for which aspiration is the only contrasting feature (two words can have different meanings but with the only difference in pronunciation being that one has an aspirated sound where the other has an unaspirated one).

Part of the phonological study of a language therefore involves looking at data (phonetic transcriptions of the speech of native speakers) and trying to deduce what the underlying phonemes are and what the sound inventory of the language is. The presence or absence of minimal pairs, as mentioned above, is a frequently used criterion for deciding whether two sounds should be assigned to the same phoneme. However, other considerations often need to be taken into account as well.

The particular contrasts which are phonemic in a language can change over time. At one time, [f] and [v], two sounds that have the same place and manner of articulation and differ in voicing only, were allophones of the same phoneme in English, but later came to belong to separate phonemes. This is one of the main factors of historical change of languages as described in historical linguistics.

The findings and insights of speech perception and articulation research complicate the traditional and somewhat intuitive idea of interchangeable allophones being perceived as the same phoneme. First, interchanged allophones of the same phoneme can result in unrecognizable words. Second, actual speech, even at a word level, is highly co-articulated, so it is problematic to expect to be able to splice words into simple segments without affecting speech perception.

Different linguists therefore take different approaches to the problem of assigning sounds to phonemes. For example, they differ in the extent to which they require allophones to be phonetically similar.

There are also differing ideas as to whether this grouping of sounds is purely a tool for linguistic analysis, or reflects an actual process in the way the human brain processes a language.

Since the early 1960s, theoretical linguists have moved away from the traditional concept of a phoneme, preferring to consider basic units at a more abstract level, as a component of morphemes; these units can be called*morphophonemes*, and analysis using this approach is called morphophonology.

Other topics in phonology

In addition to the minimal units that can serve the purpose of differentiating meaning (the phonemes), phonology studies how sounds alternate, i.e. replace one another in different forms of the same morpheme (allomorphs), as well as, for example, syllable structure, stress, feature geometry, and intonation.

Phonology also includes topics such as phonotactics (the phonological constraints on what sounds can appear in what positions in a given language) and phonological alternation (how the pronunciation of a sound changes through the application of phonological rules, sometimes in a given order which can be feeding or bleeding,) as well as prosody, the study of suprasegmentals and topics such as stress and intonation.

The principles of phonological analysis can be applied independently of modality because they are designed to serve as general analytical tools, not language-specific ones. The same principles have been applied to the analysis of sign languages (see Phonemes in sign languages), even though the sub-lexical units are not instantiated as speech sounds.

Chapter 3

Modern Phonetics: Main Branches

Articulatory phonetics

The field of **articulatory phonetics** is a subfield of phonetics that studies **articulation** and ways that humans produce speech. Articulatory phoneticians explain how humans produce speech sounds via the interaction of different physiological structures. Generally, articulatory phonetics is concerned with the transformation of aerodynamicenergy into acoustic energy. Aerodynamic energy refers to the airflow through the vocal tract. Its potential form is air pressure; its kinetic form is the actual dynamic airflow. Acoustic energy is variation in the air pressure that can be represented as sound waves, which are then perceived by the human auditory system as sound.

Sound is produced simply by expelling air from the lungs. However, to vary the sound quality in a way useful for speaking, two speech organs normally move towards each other to contact each other to create an obstruction that shapes the air in a particular fashion. The point of maximum obstruction is called the *place of articulation*, and the way the obstruction forms and releases is the *manner of articulation*. For example, when making a *p* sound, the lips come together tightly, blocking the air momentarily and causing a buildup of air pressure. The lips then release suddenly, causing a burst of sound. The place of articulation of this sound is therefore called*bilabial*, and the manner is called *stop* (also known as a *plosive*).

Components

The vocal tract can be viewed through an aerodynamic-biomechanic model that includes three main components:

- air cavities
- pistons
- air valves

Air cavities are containers of air molecules of specific volumes and masses. The main air cavities present in the articulatory system are the supraglottal cavity and the subglottal cavity. They are so-named because the glottis, the openable space between the vocal folds internal to the larynx, separates the two cavities. The supraglottal cavity or the orinasal cavity is divided into an oral subcavity (the cavity from the glottis to the lips excluding the nasal cavity) and a nasal subcavity (the cavity from the velopharyngeal port, which can be closed by raising the velum). The subglottal cavity consists of the trachea and the lungs. The atmosphere external to the articulatory stem may also be considered an air cavity whose potential connecting points with respect to the body are the nostrils and the lips.

Pistons are initiators. The term *initiator* refers to the fact that they are used to initiate a change in the volumes of air cavities, and, by Boyle's Law, the corresponding air pressure of the cavity. The term *initiation* refers to the change. Since changes in air pressures between connected cavities lead to airflow between the cavities, initiation is also referred to as an *airstream mechanism*. The three pistons present in the articulatory system are the larynx, the tongue body, and the

physiological structures used to manipulate lung volume (in particular, the floor and the walls of the chest). The lung pistons are used to initiate a pulmonic airstream (found in all human languages). The larynx is used to initiate the glottalic airstream mechanism by changing the volume of the supraglottal and subglottal cavities via vertical movement of the larynx (with a closed glottis). Ejectives and implosivesare made with this airstream mechanism. The tongue body creates a velaric airstream by changing the pressure within the oral cavity: the tongue body changes the mouth subcavity. Click consonants use the velaric airstream mechanism. Pistons are controlled by various muscles.

Valves regulate airflow between cavities. Airflow occurs when an air valve is open and there is a pressure difference between the connecting cavities. When an air valve is closed, there is no airflow. The air valves are the vocal folds (the glottis), which regulate between the supraglottal and subglottal cavities, the velopharyngeal port, which regulates between the oral and nasal cavities, the tongue, which regulates between the oral cavity and the atmosphere, and the lips, which also regulate between the oral cavity and the atmosphere. Like the pistons, the air valves are also controlled by various muscles.

Initiation

To produce any kind of sound, there must be movement of air. To produce sounds that people can interpret as spoken words, the movement of air must pass through the vocal cords, up through the throat and, into the mouth or nose to then leave the body. Different sounds are formed by different positions of

the mouth—or, as linguists call it, "the oral cavity" (to distinguish it from the nasal cavity).

Lack of labials

While most languages make use of purely labial phonemes, a few generally lack them. Examples are Tlingit, Eyak (both Na-Dené), Wichita (Caddoan), and the Iroquoian languages except Cherokee.

Many of these languages are transcribed with /w/ and with labialized consonants. However, it is not always clear to what extent the lips are involved in such sounds. In the Iroquoian languages, for example, /w/ involved little apparent rounding of the lips. See the Tillamook language for an example of a language with "rounded" consonants and vowels that do not have any actual labialization. All of these languages have seen labials introduced under the influence of English.

Coronal consonants

Coronal consonants are made with the tip or blade of the tongue and, because of the agility of the front of the tongue, represent a variety not only in place but in the posture of the tongue. The coronal places of articulation represent the areas of the mouth where the tongue contacts or makes a constriction, and include dental, alveolar, and post-alveolar locations. Tongue postures using the tip of the tongue can be apical if using the top of the tongue tip, laminal if made with the blade of the tongue, or sub-apical if the tongue tip is curled back and the bottom of the tongue is used. Coronals are unique as a group in that every manner of articulation is

attested. Australian languagesare well known for the large number of coronal contrasts exhibited within and across languages in the region.

Dental consonantsare made with the tip or blade of the tongue and the upper teeth. They are divided into two groups based upon the part of the tongue used to produce them: apical dental consonants are produced with the tongue tip touching the teeth; interdental consonants are produced with the blade of the tongue as the tip of the tongue sticks out in front of the teeth. No language is known to use both contrastively though they may exist allophonically.

Alveolar consonantsare made with the tip or blade of the tongue at the alveolar ridge just behind the teeth and can similarly be apical or laminal.

Crosslinguistically, dental consonants and alveolar consonants are frequently contrasted leading to a number of generalizations of crosslinguistic patterns. The different places of articulation tend to also be contrasted in the part of the tongue used to produce them: most languages with dental stops have laminal dentals, while languages with apical stops usually have apical stops. Languages rarely have two consonants in the same place with a contrast in laminality, though Taa (!Xóõ) is a counterexample to this pattern. If a language has only one of a dental stop or an alveolar stop, it will usually be laminal if it is a dental stop, and the stop will usually be apical if it is an alveolar stop, though for example Temne and Bulgarian do not follow this pattern. If a language has both an apical and laminal stop, then the laminal stop is

more likely to be affricated like in Isoko, though Dahalo show the opposite pattern with alveolar stops being more affricated.

Retroflex consonants have several different definitions depending on whether the position of the tongue or the position on the roof of the mouth is given prominence. In general, they represent a group of articulations in which the tip of the tongue is curled upwards to some degree. In this way, retroflex articulations can occur in several different locations on the roof of the mouth including alveolar, post-alveolar, and palatal regions. If the underside of the tongue tip makes contact with the roof of the mouth, it is sub-apical though apical post-alveolar sounds are also described as retroflex. Typical examples of sub-apical retroflex stops are commonly found in Dravidian languages, and in some languages indigenous to the southwest United States the contrastive difference between dental and alveolar stops is a slight retroflexion of the alveolar stop. Acoustically, retroflexion tends to affect the higher formants.

Articulations taking place just behind the alveolar ridge, known as post-alveolar consonants, have been referred to using a number of different terms. Apical post-alveolar consonants are often called retroflex, while laminal articulations are sometimes called palato-alveolar; in the Australianist literature, these laminal stops are often described as 'palatal' though they are produced further forward than the palate region typically described as palatal. Because of individual anatomical variation, the precise articulation of palato-alveolar stops (and coronals in general) can vary widely within a speech community.

Dorsal consonants

Dorsal consonants are those consonants made using the tongue body rather than the tip or blade.

Palatal consonantsare made using the tongue body against the hard palate on the roof of the mouth. They are frequently contrasted with velar or uvular consonants, though it is rare for a language to contrast all three simultaneously, with Jaqaru as a possible example of a three-way contrast.

Velar consonantsare made using the tongue body against the velum. They are incredibly common cross-linguistically; almost all languages have a velar stop. Because both velars and vowels are made using the tongue body, they are highly affected by coarticulation with vowels and can be produced as far forward as the hard palate or as far back as the uvula. These variations are typically divided into front, central, and back velars in parallel with the vowel space. They can be hard to distinguish phonetically from palatal consonants, though are produced slightly behind the area of prototypical palatal consonants.

Uvular consonants are made by the tongue body contacting or approaching the uvula. They are rare, occurring in an estimated 19 percent of languages, and large regions of the Americas and Africa have no languages with uvular consonants. In languages with uvular consonants, stops are most frequent followed by continuants (including nasals).

Radical consonants

Radical consonants either use the root of the tongue or the epiglottis during production.

Pharyngeal consonants are made by retracting the root of the tongue far enough to touch the wall of the pharynx. Due to production difficulties, only fricatives and approximants can produced this way.

Epiglottal consonantsare made with the epiglottis and the back wall of the pharynx. Epiglottal stops have been recorded in Dahalo. Voiced epiglottal consonants are not deemed possible due to the cavity between the glottis and epiglottis being too small to permit voicing.

Glottal consonants

Glottal consonants are those produced using the vocal folds in the larynx. Because the vocal folds are the source of phonation and below the oro-nasal vocal tract, a number of glottal consonants are impossible such as a voiced glottal stop. Three glottal consonants are possible, a voiceless glottal stop and two glottal fricatives, and all are attested in natural languages.

Glottal stops, produced by closing the vocal folds, are notably common in the world's languages. While many languages use them to demarcate phrase boundaries, some languages like Huatla Mazatec have them as contrastive phonemes. Additionally, glottal stops can be realized as laryngealization of the following vowel in this language. Glottal stops, especially

between vowels, do usually not form a complete closure. True glottal stops normally occur only when they'regeminated.

Vowels

Vowels are produced by the passage of air through the larynx and the vocal tract. Most vowels are voiced (i.e. the vocal folds are vibrating). Except in some marginal cases, the vocal tract is open, so that the airstream is able to escape without generating fricative noise.

Variation in vowel quality is produced by means of the following articulatory structures:

Articulators

Glottis

The glottis is the opening between the vocal folds located in the larynx. Its position creates different vibration patterns to distinguish voiced and voiceless sounds. In addition, the pitch of the vowel is changed by altering the frequency of vibration of the vocal folds. In some languages there are contrasts among vowels with different phonation types.

Pharynx

The pharynx is the region of the vocal tract below the velum and above the larynx. Vowels may be made pharyngealized (also *epiglottalized*, *sphincteric* or *strident*) by means of a retraction of the tongue root.

Vowels may also be articulated with advanced tongue root. There is discussion of whether this vowel feature (ATR) is different from the Tense/Lax distinction in vowels.

Velum

The velum—or soft palate—controls airflow through the nasal cavity. Nasals and nasalized sounds are produced by lowering the velum and allowing air to escape through the nose. Vowels are normally produced with the soft palate raised so that no air escapes through the nose. However, vowels may be nasalizedas a result of lowering the soft palate. Many languages use nasalization contrastively.

Tongue

The tongue is a highly flexible organ that is capable of being moved in many different ways. For vowel articulation the principal variations are vowel height and the dimension of backness and frontness. A less common variation in vowel quality can be produced by a change in the shape of the front of the tongue, resulting in a rhotic or rhotacized vowel.

Lips

- The lips play a major role in vowel articulation. It is generally believed that two major variables are in effect: lip-rounding (or labialization) and lip protrusion.

A situation can be considered where (1) the vocal fold valve is closed separating the supraglottal cavity from the subglottal cavity, (2) the mouth is open and, therefore, supraglottal air

pressure is equal to atmospheric pressure, and (3) the lungs are contracted resulting in a subglottal pressure that has increased to a pressure that is greater than atmospheric pressure. If the vocal fold valve is subsequently opened, the previously two separate cavities become one unified cavity although the cavities will still be aerodynamically isolated because the glottic valve between them is relatively small and constrictive. Pascal's Law states that the pressure within a system must be equal throughout the system. When the subglottal pressure is greater than supraglottal pressure, there is a pressure inequality in the unified cavity. Since pressure is a force applied to a surface area by definition and a force is the product of mass and acceleration according to Newton's Second Law of Motion, the pressure inequality will be resolved by having part of the mass in air molecules found in the subglottal cavity move to the supraglottal cavity. This movement of mass is airflow.

The airflow will continue until a pressure equilibrium is reached. Similarly, in an ejective consonant with a glottalicairstream mechanism, the lips or the tongue (i.e., the buccal or lingual valve) are initially closed and the closed glottis (the laryngeal piston) is raised decreasing the oral cavity volume behind the valve closure and increasing the pressure compared to the volume and pressure at a resting state. When the closed valve is opened, airflow will result from the cavity behind the initial closure outward until intraoral pressure is equal to atmospheric pressure. That is, air will flow from a cavity of higher pressure to a cavity of lower pressure until the equilibrium point; the pressure as potential energyis, thus, converted into airflow as kinetic energy.

Sound sources

Sound sources refer to the conversion of aerodynamic energy into acoustic energy. There are two main types of sound sources in the articulatory system: periodic (or more precisely semi-periodic) and aperiodic.

A periodic sound source is vocal fold vibration produced at the glottis found in vowels and voiced consonants. A less common periodic sound source is the vibration of an oral articulator like the tongue found in alveolar trills. Aperiodic sound sources are the turbulent noise of fricative consonants and the short-noise burst of plosive releases produced in the oral cavity.

Voicing is a common period sound source in spoken language and is related to how closely the vocal cords are placed together. In English there are only two possibilities, *voiced* and *unvoiced*. Voicing is caused by the vocal cords held close by each other, so that air passing through them makes them vibrate. All normally spoken vowels are voiced, as are all other sonorants except *h*, as well as some of the remaining sounds (*b*, *d*, *g*, *v*, *z*, *zh*, *j*, and the *th* sound in *this*). All the rest are voiceless sounds, with the vocal cords held far enough apart that there is no vibration; however,

there is still a certain amount of audible friction, as in the sound *h*. Voiceless sounds are not very prominent unless there is some turbulence, as in the stops, fricatives, and affricates; this is why sonorants in general only occur voiced. The exception is during whispering, when all sounds pronounced are voiceless.

Periodic sources

- Non-vocal fold vibration: 20–40 hertz (cycles per second)
- Vocal fold vibration
- Lower limit: 70–80 Hz modal (bass), 30–40 Hz creaky
- Upper limit: 1170 Hz (soprano)

Vocal fold vibration

- larynx:
- cricoid cartilage
- thyroid cartilage
- arytenoid cartilage
- interarytenoid muscles (fold adduction)
- posterior cricoarytenoid muscle (fold abduction)
- lateral cricoarytenoid muscle (fold shortening/stiffening)
- thyroarytenoid muscle (medial compression/fold stiffening, internal to folds)
- cricothyroid muscle (fold lengthening)
- hyoid bone
- sternothyroid muscle (lowers thyroid)
- sternohyoid muscle (lowers hyoid)
- stylohyoid muscle (raises hyoid)
- digastric muscle (raises hyoid)

Experimental techniques

- Plethysmography
- Electromyography
- Photoglottography

- Electrolaryngography
- Magnetic resonance imaging (MRI) / Real-time MRI
- Radiography
- Medical ultrasonography
- Electromagnetic articulography
- Aerometry
- Endoscopy
- Videokymography

Palatography

In order to understand how sounds are made, experimental procedures are often adopted. Palatography is one of the oldest instrumental phonetic techniques used to record data regarding articulators. In traditional, static palatography, a speaker's palate is coated with a dark powder. The speaker then produces a word, usually with a single consonant. The tongue wipes away some of the powder at the place of articulation. The experimenter can then use a mirror to photograph the entire upper surface of the speaker's mouth. This photograph, in which the place of articulation can be seen as the area where the powder has been removed, is called a palatogram.

Technology has since made possible electropalatography (or EPG). In order to collect EPG data, the speaker is fitted with a special prosthetic palate, which contains a number of electrodes. The way in which the electrodes are "contacted" by the tongue during speech provides phoneticians with important information, such as how much of the palate is contacted in different speech sounds, or which regions of the palate are contacted, or what the duration of the contact is.

Acoustic phonetics

Acoustic phonetics is a subfield of phonetics, which deals with acoustic aspects of speech sounds. Acoustic phonetics investigates time domain features such as the mean squared amplitude of a waveform, its duration, its fundamental frequency, or frequency domain features such as the frequency spectrum, or even combined spectrotemporal features and the relationship of these properties to other branches of phonetics (e.g. articulatory or auditory phonetics), and to abstract linguistic concepts such as phonemes, phrases, or utterances.

The study of acoustic phonetics was greatly enhanced in the late 19th century by the invention of the Edisonphonograph. The phonograph allowed the speech signal to be recorded and then later processed and analyzed. By replaying the same speech signal from the phonograph several times, filtering it each time with a different band-pass filter, a spectrogram of the speech utterance could be built up. A series of papers by Ludimar Hermann published in PflügersArchiv in the last two decades of the 19th century investigated the spectral properties of vowels and consonants using the Edison phonograph, and it was in these papers that the term *formant* was first introduced. Hermann also played back vowel recordings made with the Edison phonograph at different speeds to distinguish between Willis' and Wheatstone's theories of vowel production.

Further advances in acoustic phonetics were made possible by the development of the telephone industry. (Incidentally, Alexander Graham Bell's father, Alexander Melville Bell, was a phonetician.) During World War II, work at the Bell Telephone

Laboratories (which invented the spectrograph) greatly facilitated the systematic study of the spectral properties of periodic and aperiodic speech sounds, vocal tract resonances and vowel formants, voice quality, prosody, etc.

Integrated linear prediction residuals (ILPR) was an effective feature proposed by T V Ananthapadmanabha in 1995, which closely approximates the voice source signal. This proved to be very effective in accurate estimation of the epochs or the glottal closure instant. A G Ramakrishnan et al. showed in 2015 that the discrete cosine transform coefficients of the ILPR contains speaker information that supplements the mel frequency cepstral coefficients. Plosion index is another scalar, time-domain feature that was introduced by T V Ananthapadmanabha et al. for characterizing the closure-burst transition of stop consonants.

On a theoretical level, speech acoustics can be modeled in a way analogous to electrical circuits. Lord Rayleigh was among the first to recognize that the new electric theory could be used in acoustics, but it was not until 1941 that the circuit model was effectively used, in a book by Chiba and Kajiyama called "The Vowel: Its Nature and Structure". (This book by Japanese authors working in Japan was published in English at the height of World War II.) In 1952, Roman Jakobson, Gunnar Fant, and Morris Halle wrote "Preliminaries to Speech Analysis", a seminal work tying acoustic phonetics and phonological theory together. This little book was followed in 1960 by Fant "Acoustic Theory of Speech Production", which has remained the major theoretical foundation for speech acoustic research in both the academy and industry. (Fant was himself very involved in the telephone industry.) Other

important framers of the field include Kenneth N. Stevens who wrote "Acoustic Phonetics", Osamu Fujimura, and Peter Ladefoged.

Auditory phonetics

Auditory phonetics is the branch of phonetics concerned with the hearing of speech sounds and with speech perception. It thus entails the study of the relationships between speech stimuli and a listener's responses to such stimuli as mediated by mechanisms of the peripheral and central auditory systems, including certain areas of the brain. It is said to compose one of the three main branches of phonetics along with acoustic and articulatory phonetics, though with overlapping methods and questions.

Physical scales and auditory sensations

- There is no direct connection between auditory sensations and the physical properties of sound that give rise to them. While the physical (acoustic) properties are objectively measurable, auditory sensations are subjective and can only be studied by asking listeners to report on their perceptions. The table below shows some correspondences between physical properties and auditory sensations.

Segmental and suprasegmental

Auditory phonetics is concerned with both segmental (chiefly vowels and consonants) and prosodic (such as stress, tone,

rhythm and intonation) aspects of speech. While it is possible to study the auditory perception of these phenomena without context, in continuous speech all these variables are processed in parallel with significant variability and complex interactions between them. For example, it has been observed that vowels, which are usually described as different from each other in the frequencies of their formants, also have intrinsic values of fundamental frequency (and presumably therefore of pitch) that are different according to the height of the vowel. Thus open vowels typically have lower fundamental frequency than close vowels in a given context, and vowel recognition is likely to interact with the perception of prosody.

In speech research

If there is a distinction to be made between auditory phonetics and speech perception, it is that the former is more closely associated with traditional non-instrumental approaches to phonology and other aspects of linguistics, while the latter is closer to experimental, laboratory-based study. Consequently, the term *auditory phonetics* is often used to refer to the study of speech without the use of instrumental analysis: the researcher may make use of technology such as recording equipment, or even a simple pen and paper (as used by William Labov in his study of the pronunciation of English in New York department stores), but will not use laboratory techniques such as spectrography or speech synthesis, or methods such as EEG and fMRI that allow phoneticians to directly study the brain's response to sound. Most research in sociolinguistics and dialectologyhas been based on auditory analysis of data and almost all pronunciation dictionaries are based on impressionistic, auditory analysis of how words are

pronounced. It is possible to claim an advantage for auditory analysis over instrumental: Kenneth L. Pike stated "Auditory analysis is essential to phonetic study since the ear can register all those features of sound waves, and only those features, which are above the threshold of audibility ... whereas analysis by instruments must always be checked against auditory reaction". Herbert Pilch attempted to define auditory phonetics in such a way as to avoid any reference to acoustic parameters. In the auditory analysis of phonetic data such as recordings of speech, it is clearly an advantage to have been trained in analytical listening. Practical phonetic training has since the 19th century been seen an essential foundation for phonetic analysis and for the teaching of pronunciation; it is still a significant part of modern phonetics. The best-known type of auditory training has been in the system of cardinal vowels; there is disagreement about the relative importance of auditory and articulatory factors underlying the system, but the importance of auditory training for those who are to use it is indisputable. Training in the auditory analysis of prosodic factors such as pitch and rhythm is also important. Not all research on prosody has been based on auditory techniques: some pioneering work on prosodic features using laboratory instruments was carried out in the 20th century (e.g. Elizabeth Uldall's work using synthesized intonation contours, Dennis Fry's work on stress perception or Daniel Jones's early work on analyzing pitch contours by means of manually operating the pickup arm of a gramophone to listen repeatedly to individual syllables, checking where necessary against a tuning fork),. However, the great majority of work on prosody has been based on auditory analysis until the recent arrival of approaches explicitly based on computer analysis of the acoustic signal, such as ToBI, INTSINT or the IPO system.

Chapter 4

Language Production

Language production is the production of spoken or written language. In psycholinguistics, it describes all of the stages between having a concept to express and translating that concept into linguistic forms. These stages have been described in two types of processing models: the lexical access models and the serial models. Through these models, psycholinguists can look into how speeches are produced in different ways, such as when the speaker is bilingual. Psycholinguists learn more about these models and different kinds of speech by using language production research methods that include collecting speech errors and elicited production tasks.

Stages involved

Language production consists of several interdependent processes which transform a nonlinguistic message into a spoken, signed, or written linguistic signal. Though the following steps proceed in this approximate order, there are plenty of interaction and communication between them. The process of message planning is an active area of psycholinguistic research, but researchers have found that it is an ongoing process throughout language production. Research suggests that messages are planned in roughly the same order that they are in an utterance. But, there is also evidence that suggests the verbs that give case may be planned earlier than objects, even when the object is said first. After identifying a message, or part of a message, to be linguistically encoded, a

speaker must selects the individual words—also known as lexical items—to represent that message. This process is called lexical selection. The words are selected based on their meaning, which in linguistics is called semanticinformation. Lexical selection activates the word's lemma, which contains both semantic and grammatical information about the word.

This grammatical information is then used in the next step of language production, grammatical encoding. Critical grammatical information includes characteristics such as the word's syntactic category (noun, verb, etc.), what objects it takes, and grammatical gender if it is presents in the language. Using some of these characteristics as well as information about the thematic roles of each word in the intended message, each word is then assigned the grammatical and thematic role it will have in the sentence.

Function morphemes, like the plural /s/ or the past tense /ɪd/, are added in this stage as well. After an utterance, or part of one, has been formed, it then goes through phonological encoding. In this stage of language production, the mental representation of the words to be spoken is transformed into a sequence of speech sounds to be pronounced. The speech sounds are assembled in the order they are to be produced.

The basic loop occurring in the creation of language consists of the following stages:

- Intended message
- Encode message into linguistic form
- Encode linguistic form into speech motor system

- Sound goes from speaker's mouth to hearer's ear auditory system
- Speech is decoded into linguistic form
- Linguistic form is decoded into meaning

According to the lexical access model (see section below), in terms of lexical access, two different stages of cognition are employed; thus, this concept is known as the two-stage theory of lexical access. The first stage, lexical selection provides information about lexical items required to construct the functional level representation. These items are retrieved according to their specific semantic and syntactic properties, but phonological forms are not yet made available at this stage. The second stage, retrieval of wordforms, provides information required for building the positional level representation.

Models

Serial model

A serial model of language production divides the process into several stages. For example, there may be one stage for determining pronunciation and a stage for determining lexical content. The serial model does not allow overlap of these stages, so they may only be completed one at a time.

Connectionist model

Several researchers have proposed a connectionist model, one notable example being Dell. According to his connectionist model, there are four layers of processing and understanding:

semantic, syntactic, morphological, and phonological. These work in parallel and in series, with activation at each level. Interference and misactivation can occur at any of these stages. Production begins with concepts, and continues down from there. One might start with the concept of a cat: a four-legged, furry, domesticated mammal with whiskers, etc. This conceptual set would attempt to find the corresponding word {cat}.

This selected word would then selects morphological and phonological data /k / at/. The distinction of this model is that, during this process, other elements would also be primed ({rat} might be somewhat primed, for example), as they are physically similar, and so can cause conceptual interference. Errors might also occur at the phoneme level, as many words are phonetically similar, e.g. mat. Substitutions of similar consonant sounds are more likely to occur, e.g. between plosive stop consonants such as d, p and b. Lower primed words are less likely to be chosen, but interference is thought to occur in cases of early selection, where the level of activation of the target and interference words is at the same level.

Lexical access model

This model states that the sentence is made by a sequence of processes generating differing levels of representations. For instance, the functional level representation is made on the a preverbal representation, which is essentially what the speaker seeks to express. This level is responsible for encoding the meanings of lexical items and the way that grammar forms relationships between them. Next, the positional level

representation is built, which functions to encode the phonological forms of words and the order they are found in sentence structures. Lexical access, according to this model, is a process that encompasses two serially ordered and independent stages.

Additional aspects

Fluency

Fluency can be defined in part by prosody, which is shown graphically by a smooth intonation contour, and by a number of other elements: control of speech rate, relative timing of stressed and unstressed syllables, changes in amplitude, changes in fundamental frequency. In other words, fluency can be described as whether someone speaks smoothly and easily. This term is used in speech-language pathology when describing disorders with stuttering or other disfluencies.

Multilingualism

Whether or not a speaker is fluent in one or more languages, the process for producing language remains the same. However, bilinguals speaking two languages within a conversation may have access to both languages at the same time. Three of the most commonly discussed models for multilingual language access are the Bilingual Interactive Activation Plus model, the Revised Hierarchical Model, and the Language Mode model:

- **Bilingual Interactive Activation Plus**, updated from a model made by Dijkstra and Van Heuven, uses solely

bottom-up processing to facilitate bilingual language access. This model suggests that the lexicon for bilingual speakers combines the languages, and access occurs across both languages at the same time.

- **Revised Hierarchical Model**, developed by Kroll and Stewart, is a model suggesting that bilingual brains store meanings in a common place, word-forms are separated by language.
- **Language Mode Model**, made by Grosjean, uses two assumptions to map bilingual language production in a modular way. These assumptions are that a base language is activated in conversation, and that the speaker's other language is activated to relative degrees depending on context. De Bot describes it as overly simple for the complexity of the process and suggests it has room for expansion.

Speakers fluent in multiple languages may inhibit access to one of their languages, but this suppression can only be done once the speaker is at a certain level of proficiency in that language.

A speaker can decide to inhibit a language based on non-linguistic cues in their conversation, such as a speaker of both English and French inhibiting their French when conversing with people who only speak English.

When especially proficient multilingual speakers communicate, they can participate in code-switching. Code-switching has been shown to indicate bilingual proficiency in a speaker, though it had previously been seen as a sign of poor language ability.

Research methods

There are three main types of research into language production: speech error collection, picture-naming, and elicited production. Speech error collection focuses on using the analysis of speech errors made in naturally produced speech. On the other hand, elicited production focuses on elicited speech and is conducted in a lab. Also conducted in a lab, picture-naming focuses on reaction-time data from picture-naming latencies. Although originally disparate, these three methodologies are generally looking at the same underlying processes of speech production.

Speech errors

Speech errors have been found to be common in naturally produced speech. Analysis of speech errors has found that not all are random, but rather systematic and fall into several categories. These speech errors can demonstrate parts of the language processing system, and what happens when that system doesn't work as it should. Language production occurs quickly with speakers saying a little more than 2 words per second; so though errors occur only once out of 1,000 words, they occur relatively often throughout a speaker's day at once every 7 minutes. Some examples of these speech errors that would be collected by psycholinguists are:

- **Anticipation**: The word adds a sound from a word planned for later in the utterance.
- *target*: paddle tennis
- *produced*: taddle tennis

- **Preservation**: The word retains characteristics of a word said previously in an utterance.
- *target*: red wagon
- *produced*: red ragon
- **Blending**: More than one word is being considered in the lexicon and the two intended items "blend" into a single item.
- *target*: shout/yell
- produced: shell
- **Addition**: Additional of linguistics material added to the word.
- *target*: impossible
- *produced*: implossible
- **Substitution**: A whole word of related meaning is replacing another.
- *target*: at low speed it's too heavy
- *produced*: at low speed it's too light
- **Malapropism**: A lay term, in reference to a character Mrs. Malaprop from Sheridan's The Rivals, referring to the incorrect substitution of words.
- Makes no delusions to the past.
- The pineapple of perfection.
- I have interceded another letter from the fellow.
- **Spoonerism**: The switching of the letters from two words in the utterance.
- *target*: slips of the tongue
- *produced*: tips of the slung

Picture-naming

Picture-naming tasks ask participants to look at pictures and name them in a certain way. By looking at the time course for

the responses in these tasks, psycholinguists can learn more about the planning involved in specific phrases. These types of tasks can be helpful for investigating cross-linguistic language production and planning processes.

Elicited Production

Elicited production tasks ask participants to respond to questions or prompts in a particular way. One of the more common types of elicited production tasks is the sentence completion task. These tasks give the participants the beginning of a target sentence, which the participants are then asked to complete. Analyzing these completions can allow psycholinguistics to investigate errors that might be difficult to elicit otherwise.

Place of articulation

In articulatory phonetics, the **place of articulation** (also **point of articulation**) of a consonant is the point of contact where an obstruction occurs in the vocal tract between an **articulatory gesture**, an active articulator (typically some part of the tongue), and a passive location (typically some part of the roof of the mouth). Along with the manner of articulation and the phonation, it gives the consonant its distinctive sound.

The terminology in this article has been developed for precisely describing all the consonants in all the world's spoken languages. No known language distinguishes all of the places described here so less precision is needed to distinguish the sounds of a particular language.

Overview

The human voice produces sounds in the following manner:

- Air pressure from the lungs creates a steady flow of air through the trachea (windpipe), larynx (voice box) and pharynx (back of the throat). Therefore, the air moves out of the lungs through a coordinated action of the diaphragm, abdominal muscles, chest muscles and rib cage.
- The vocal folds in the larynx vibrate, creating fluctuations in air pressure, known as sound waves.
- Resonances in the vocal tract modify these waves according to the position and shape of the lips, jaw, tongue, soft palate, and other speech organs, creating formant regions and so different qualities of sonorant (voiced) sound.
- Mouth radiates the sound waves into the environment.
- Nasal cavity adds resonance to some sounds such as [m] and [n] to give nasal quality of the so-called nasal consonants.

The larynx

The *larynx* or *voice box* is a cylindrical framework of cartilage that serves to anchor the vocal folds. When the muscles of the vocal folds contract, the airflow from the lungs is impeded until the vocal folds are forced apart again by the increasing air pressure from the lungs. The process continues in a periodic cycle that is felt as a vibration (buzzing). In singing, the vibration frequency of the vocal folds determines the pitch

of the sound produced. Voiced phonemes such as the pure vowels are, by definition, distinguished by the buzzing sound of this periodic oscillation of the vocal cords.

The lips of the mouth can be used in a similar way to create a similar sound, as any toddler or trumpeter can demonstrate. A rubber balloon, inflated but not tied off and stretched tightly across the neck produces a squeak or buzz, depending on the tension across the neck and the level of pressure inside the balloon. Similar actions with similar results occur when the vocal cords are contracted or relaxed across the larynx.

Passive places of articulation

The passive place of articulation is the place on the more stationary part of the vocal tract where the articulation occurs and can be anywhere from the lips, upper teeth, gums, or roof of the mouth to the back of the throat. Although it is a continuum, there are several contrastive areas so languages may distinguish consonants by articulating them in different areas, but few languages contrast two sounds within the same area unless there is some other feature which contrasts as well. The following areas are contrastive:

- The upper lip (*labial*)
- The upper teeth, either on the edge of the teeth or inner surface (*dental*)
- The alveolar ridge, the gum line just behind the teeth (*alveolar*)
- The back of the alveolar ridge (*post-alveolar*)
- The hard palate on the roof of the mouth (*palatal*)

- The soft palate further back on the roof of the mouth (*velar*)
- The uvula hanging down at the entrance to the throat (*uvular*)
- The throat itself, a.k.a. the pharynx (*pharyngeal*)
- The epiglottis at the entrance to the windpipe, above the voice box (*epiglottal*)

The regions are not strictly separated. For instance, in some sounds in many languages, the surface of the tongue contacts a relatively large area from the back of the upper teeth to the alveolar ridge, which is common enough to have received its own name, *denti-alveolar*. Likewise, the alveolar and post-alveolar regions merge into each other, as do the hard and soft palate, the soft palate and the uvula, and all adjacent regions. Terms like *pre-velar* (intermediate between palatal and velar), *post-velar* (between velar and uvular), and *upper* vs. *lower* pharyngeal may be used to specify more precisely where an articulation takes place. However, although a language may contrast pre-velar and post-velar sounds, it does not also contrast them with palatal and uvular sounds (of the same type of consonant) so contrasts are limited to the number above, if not always their exact location.

Active places of articulation

The articulatory gesture of the active place of articulation involves the more mobile part of the vocal tract, typically some part of the tongue or lips. The following areas are known to be contrastive:

- The lower lip (*labial*)

- Various parts of the front of the tongue (*coronal*):
- The tip of the tongue (*apical*)
- The upper front surface of the tongue just behind the tip, called the *blade* of the tongue (*laminal*)
- The surface of the tongue *under* the tip (*subapical*)
- The body of the tongue (*dorsal*)
- The base a.k.a. root of the tongue and the throat (*pharyngeal*)
- The aryepiglottic fold inside the throat (*aryepiglottal*)
- The glottis at the very back of the windpipe (*glottal)*

In bilabial consonants, both lips move so the articulatory gesture brings the lips together, but by convention, the lower lip is said to be active and the upper lip passive. Similarly, in linguolabial consonants the tongue contacts the upper lip with the upper lip actively moving down to meet the tongue; nonetheless, the tongue is conventionally said to be active and the lip passive if for no other reason than that the parts of the mouth below the vocal tract are typically active, and those above the vocal tract are typically passive.

In dorsal gestures, different parts of the body of the tongue contact different parts of the roof of the mouth, but it cannot be independently controlled so they are all subsumed under the term *dorsal*. That is unlike coronal gestures involving the front of the tongue, which is more flexible.

The epiglottis may be active, contacting the pharynx, or passive, being contacted by the aryepiglottal folds. Distinctions made in these laryngeal areas are very difficult to observe and are the subject of ongoing investigation, and several still-unidentified combinations are thought possible.

The glottis acts upon itself. There is a sometimes fuzzy line between glottal, aryepiglottal, and epiglottal consonants and phonation, which uses these same areas.

Unlike the passive articulation, which is a continuum, there are five discrete active articulators: the lip (*labial consonants*), the flexible front of the tongue (*coronal consonants*: laminal, apical, and subapical), the middle–back of the tongue (*dorsal consonants*), the root of the tongue together with the epiglottis (*pharyngeal* or *radical consonants*), and the glottis (*glottal consonants*).

The articulators are discrete in that they can act independently of each other, and two or more may work together in what is called*coarticulation* (see below). The distinction, however, between the various coronal articulations, laminal, apical, and subapical is a continuum, without clear boundaries.

Homorganic consonants

Consonants that have the same place of articulation, such as the alveolar sounds /n, t, d, s, z, l/ in English, are said to be *homorganic*. Similarly, labial /p, b, m/ and velar /k, ɡ, ŋ/ are homorganic. A homorganic nasal rule, an instance of assimilation, operates in many languages, where a nasal consonant must be homorganic with a following stop. We see this with English *i**n**tolerable* but *i**m**plausible*; another example is found in Yoruba, where the present tense of *ba* "hide" is *mba* "is hiding", while the present of *sun* "sleep" is *nsun* "is sleeping".

Central and lateral articulation

The tongue contacts the mouth with a surface that has two dimensions: length and width. So far, only points of articulation along its length have been considered. However, articulation varies along its width as well. When the airstream is directed down the center of the tongue, the consonant is said to be *central*. If, however, it is deflected off to one side, escaping between the side of the tongue and the side teeth, it is said to be *lateral*. Nonetheless, for simplicity's sake the place of articulation is assumed to be the point along the length of the tongue, and the consonant may in addition be said to be central or lateral. That is, a consonant may be lateral alveolar, like English /l/ (the tongue contacts the alveolar ridge, but allows air to flow off to the side), or lateral palatal, like Castilian Spanish *ll*/ʎ/. Some Indigenous Australian languages contrast dental, alveolar, retroflex, and palatal laterals, and many Native American languages have lateral fricatives and affricates as well.

Coarticulation

Some languages have consonants with two simultaneous places of articulation, which is calledcoarticulation. When these are doubly articulated, the articulators must be independently movable, and therefore there may be only one each from the major categories *labial, coronal, dorsal* and *pharyngeal.*

The only common doubly articulated consonants are labial–velar stops like [k͡p], [ɡ͡b] and less commonly [ŋ͡m], which are found throughout Western Africa and Central Africa. Other

combinations are rare but include labial–(post)alveolar stops[t͡pd͡ɒn͡m], found as distinct consonants only in a single language in New Guinea, and a uvular–epiglottal stop, [qʡ], found in Somali.

More commonly, coarticulation involves secondary articulation of an approximantic nature. Then, both articulations can be similar such as labialized labial [mʷ] or palatalized velar [kʲ]. That is the case of English [w], which is a velar consonant with secondary labial articulation.

Common coarticulations include these:

- Labialization, rounding the lips while producing the obstruction, as in [kʷ] and English [w].
- Palatalization, raising the body of the tongue toward the hard palate while producing the obstruction, as in Russian[tʲ] and [ɕ].
- Velarization, raising the back of the tongue toward the soft palate (velum), as in the English dark el, [lˠ] (also transcribed [ɫ]).
- Pharyngealization, constriction of the throat (pharynx), such as Arabic "emphatic" [tˤ].

Larynx

The **larynx** (/ˈlærɪŋks/), commonly called the **voice box**, is an organ in the top of the neck involved in breathing, producing sound and protecting the trachea against food aspiration. The opening of larynx into pharynx known as the laryngeal inlet is about 4–5 centimeters in diameter. The larynx houses the vocal cords, and manipulates pitch and volume, which is essential

for phonation. It is situated just below where the tract of the pharynx splits into the trachea and the esophagus. The word larynx (plural larynges) comes from a similar Ancient Greek word (λάρυγξ*lárynx*).

Structure

The triangle-shaped larynx consists largely of cartilages that are attached to one another, and to surrounding structures, by muscles or by fibrous and elastic tissue components. The larynx is lined by a ciliated columnar epithelium. The cavity of the larynx extends from its triangle-shaped inlet, to the epiglottis, and to the circular outlet at the lower border of the cricoid cartilage, where it is continuous with the lumen of the trachea. The mucous membrane lining the larynx forms two pairs of lateral folds that project inward into its cavity. The upper folds are called the vestibular folds. They are also sometimes called the false vocal cords for the rather obvious reason that they play no part in vocalization. The lower pair of folds are known as the vocal cords, which produce sounds needed for speech and other vocalizations. The slit-like space between the left and right vocal cords, called the rimaglottidis, is the narrowest part of the larynx.

The vocal cords and the rimaglottidisare together designated as the glottis. The laryngeal cavity above the vestibular folds is called the vestibule. The very middle portion of the cavity between the vestibular folds and the vocal cords is the ventricle of the larynx, or laryngeal ventricle. The infraglottic cavity is the open space below the glottis.

Location

- In adult humans, the larynx is found in the anterior neck at the level of the cervical vertebrae C3–C6. It connects the inferior part of the pharynx (hypopharynx) with the trachea. The laryngeal skeleton consists of nine cartilages: three single (epiglottic, thyroid and cricoid) and three paired (arytenoid, corniculate, and cuneiform). The hyoid bone is not part of the larynx, though the larynx is suspended from the hyoid. The larynx extends vertically from the tip of the epiglottis to the inferior border of the cricoid cartilage. Its interior can be divided in supraglottis, glottis and subglottis.

Cartilages

There are nine cartilages, three unpaired and three paired (3 pairs=6), that support the mammalian larynx and form its skeleton.

Unpaired cartilages:

- Thyroid cartilage: This forms the Adam's apple (also called the laryngeal prominence). It is usually larger in males than in females. The thyrohyoid membrane is a ligament associated with the thyroid cartilage that connects it with the hyoid bone. It supports the front portion of the larynx.
- Cricoid cartilage: A ring of hyaline cartilage that forms the inferior wall of the larynx. It is attached to the top of the trachea. The median cricothyroid ligament connects the cricoid cartilage to the thyroid cartilage.

- Epiglottis: A large, spoon-shaped piece of elastic cartilage. During swallowing, the pharynx and larynx rise. Elevation of the pharynx widens it to receive food and drink; elevation of the larynx causes the epiglottis to move down and form a lid over the glottis, closing it off.

Paired cartilages:

- Arytenoid cartilages: Of the paired cartilages, the arytenoid cartilages are the most important because they influence the position and tension of the vocal cords. These are triangular pieces of mostly hyaline cartilage located at the posterosuperior border of the cricoid cartilage.
- Corniculate cartilages: Horn-shaped pieces of elastic cartilage located at the apex of each arytenoid cartilage.
- Cuneiform cartilages: Club-shaped pieces of elastic cartilage located anterior to the corniculate cartilages.

Muscles

The muscles of the larynx are divided into *intrinsic* and *extrinsic* muscles. The extrinsic muscles act on the region and pass between the larynx and parts around it but have their origin elsewhere; the intrinsic muscles are confined entirely within the larynx and have their origin and insertion there.

The intrinsic muscles are divided into respiratory and the phonatory muscles (the muscles of phonation). The respiratory muscles move the vocal cords apart and serve breathing. The phonatory muscles move the vocal cords together and serve the

production of voice. The main respiratory muscles are the posterior cricoarytenoid muscles. The phonatory muscles are divided into adductors (lateral cricoarytenoid muscles, arytenoid muscles) and tensors (cricothyroid muscles, thyroarytenoid muscles).

Intrinsic

The intrinsic laryngeal muscles are responsible for controlling sound production.

- Cricothyroid muscle lengthen and tense the vocal cords.
- Posterior cricoarytenoid muscles abduct and externally rotate the arytenoid cartilages, resulting in abducted vocal cords.
- Lateral cricoarytenoid muscles adduct and internally rotate the arytenoid cartilages, increase medial compression.
- Transverse arytenoid muscle adduct the arytenoid cartilages, resulting in adducted vocal cords.
- Oblique arytenoid muscles narrow the laryngeal inlet by constricting the distance between the arytenoid cartilages.
- Thyroarytenoid muscles narrow the laryngeal inlet, shortening the vocal cords, and lowering voice pitch. The internal thyroarytenoid is the portion of the thyroarytenoid that vibrates to produce sound.

Notably the only muscle capable of separating the vocal cords for normal breathing is the posterior cricoarytenoid. If this muscle is incapacitated on both sides, the inability to pull the

vocal cords apart (abduct) will cause difficulty breathing. Bilateral injury to the recurrent laryngeal nerve would cause this condition. It is also worth noting that all muscles are innervated by the recurrent laryngeal branch of the vagus except the cricothyroid muscle, which is innervated by the external laryngeal branch of the superior laryngeal nerve (a branch of the vagus).

Additionally, intrinsic laryngeal muscles present a constitutive Ca-buffering profile that predicts their better ability to handle calcium changes in comparison to other muscles. This profile is in agreement with their function as very fast muscles with a well-developed capacity for prolonged work. Studies suggests that mechanisms involved in the prompt sequestering of Ca (sarcoplasmic reticulum Ca-reuptake proteins, plasma membrane pumps, and cytosolic Ca-buffering proteins) are particularly elevated in laryngeal muscles, indicating their importance for the myofiber function and protection against disease, such as Duchenne muscular dystrophy. Furthermore, different levels of Orai1 in rat intrinsic laryngeal muscles and extraocular muscles over the limb muscle suggests a role for store operated calcium entry channels in those muscles' functional properties and signaling mechanisms.

Extrinsic

- The extrinsic laryngeal muscles support and position the larynx within the mid-cervical region. [trachea.]
- Sternothyroid muscles depress the larynx. (Innervated by ansacervicalis)
- Omohyoid muscles depress the larynx. (Ansa cervicalis)

- Sternohyoid muscles depress the larynx. (Ansa cervicalis)
- Inferior constrictor muscles. (CN X)
- Thyrohyoid muscles elevates the larynx. (C1)
- Digastric elevates the larynx. (CN V_3, CN VII)
- Stylohyoid elevates the larynx. (CN VII)
- Mylohyoid elevates the larynx. (CN V_3)
- Geniohyoid elevates the larynx. (C1)
- Hyoglossus elevates the larynx. (CN XII)
- Genioglossus elevates the larynx. (CN XII)

Nerve supply

The larynx is innervated by branches of the vagus nerve on each side. Sensory innervation to the glottis and laryngeal vestibule is by the internal branch of the superior laryngeal nerve. The external branch of the superior laryngeal nerve innervates the cricothyroid muscle. Motor innervation to all other muscles of the larynx and sensory innervation to the subglottis is by the recurrent laryngeal nerve. While the sensory input described above is (general) visceral sensation (diffuse, poorly localized), the vocal cords also receives general somatic sensory innervation (proprioceptive and touch) by the superior laryngeal nerve.

Injury to the external laryngeal nerve causes weakened phonation because the vocal cords cannot be tightened. Injury to one of the recurrent laryngeal nerves produces hoarseness, if both are damaged the voice may or may not be preserved, but breathing becomes difficult.

Development

In newborn infants, the larynx is initially at the level of the C2–C3 vertebrae, and is further forward and higher relative to its position in the adult body. The larynx descends as the child grows.

Function

Sound generation

Sound is generated in the larynx, and that is where pitch and volume are manipulated. The strength of expiration from the lungs also contributes to loudness.

Manipulation of the larynx is used to generate a source sound with a particular fundamental frequency, or pitch. This source sound is altered as it travels through the vocal tract, configured differently based on the position of the tongue, lips, mouth, and pharynx. The process of altering a source sound as it passes through the filter of the vocal tract creates the many different vowel and consonant sounds of the world's languages as well as tone, certain realizations of stress and other types of linguistic prosody. The larynx also has a similar function to the lungs in creating pressure differences required for sound production; a constricted larynx can be raised or lowered affecting the volume of the oral cavity as necessary in glottalic consonants.

The vocal cords can be held close together (by adducting the arytenoid cartilages) so that they vibrate (see phonation). The muscles attached to the arytenoid cartilages control the degree

of opening. Vocal cord length and tension can be controlled by rocking the thyroid cartilage forward and backward on the cricoid cartilage (either directly by contracting the cricothyroids or indirectly by changing the vertical position of the larynx), by manipulating the tension of the muscles within the vocal cords, and by moving the arytenoids forward or backward. This causes the pitch produced during phonation to rise or fall. In most males the vocal cords are longer and with a greater mass than most females' vocal cords, producing a lower pitch.

The vocal apparatus consists of two pairs of folds, the vestibular folds (false vocal cords) and the true vocal cords. The vestibular folds are covered by respiratory epithelium, while the vocal cords are covered by stratified squamous epithelium. The vestibular folds are not responsible for sound production, but rather for resonance. The exceptions to this are found in Tibetan chanting and Kargyraa, a style of Tuvan throat singing. Both make use of the vestibular folds to create an undertone. These false vocal cords do not contain muscle, while the true vocal cords do have skeletal muscle.

Other

The most important role of the larynx is its protecting function; the prevention of foreign objects from entering the lungs by coughing and other reflexive actions. A cough is initiated by a deep inhalation through the vocal cords, followed by the elevation of the larynx and the tight adduction (closing) of the vocal cords. The forced expiration that follows, assisted by tissue recoil and the muscles of expiration, blows the vocal cords apart, and the high pressure expels the irritating object

out of the throat. Throat clearing is less violent than coughing, but is a similar increased respiratory effort countered by the tightening of the laryngeal musculature. Both coughing and throat clearing are predictable and necessary actions because they clear the respiratory passageway, but both place the vocal cords under significant strain.

Another important role of the larynx is abdominal fixation, a kind of Valsalva maneuver in which the lungs are filled with air in order to stiffen the thorax so that forces applied for lifting can be translated down to the legs. This is achieved by a deep inhalation followed by the adduction of the vocal cords. Grunting while lifting heavy objects is the result of some air escaping through the adducted vocal cords ready for phonation.

Abduction of the vocal cords is important during physical exertion. The vocal cords are separated by about 8 mm (0.31 in) during normal respiration, but this width is doubled during forced respiration.

During swallowing, elevation of the posterior portion of the tongue levers (inverts) the epiglottis over the glottis' opening to prevent swallowed material from entering the larynx which leads to the lungs, and provides a path for a food or liquid bolus to "slide" into the esophagus; the hyo-laryngeal complex is also pulled upwards to assist this process. Stimulation of the larynx by aspirated food or liquid produces a strong coughreflex to protect the lungs.

In addition, intrinsic laryngeal muscles are spared from some muscle wasting disorders, such as Duchenne muscular dystrophy, may facilitate the development of novel strategies

for the prevention and treatment of muscle wasting in a variety of clinical scenarios. ILM have a calcium regulation system profile suggestive of a better ability to handle calcium changes in comparison to other muscles, and this may provide a mechanistic insight for their unique pathophysiological properties

Clinical significance

Disorders

There are several things that can cause a larynx to not function properly. Some symptoms are hoarseness, loss of voice, pain in the throat or ears, and breathing difficulties.

- Acute laryngitis is the sudden inflammation and swelling of the larynx. It is caused by the common cold or by excessive shouting. It is not serious. Chronic laryngitis is caused by smoking, dust, frequent yelling, or prolonged exposure to polluted air. It is much more serious than acute laryngitis.
- Presbylarynx is a condition in which age-related atrophy of the soft tissues of the larynx results in weak voice and restricted vocal range and stamina. Bowing of the anterior portion of the vocal colds is found on laryngoscopy.
- Ulcersmay be caused by the prolonged presence of an endotracheal tube.
- Polyps and vocal cord nodules are small bumps caused by prolonged exposure to tobacco smoke and vocal misuse, respectively.

- Two related types of cancer of the larynx, namely squamous cell carcinoma and verrucous carcinoma, are strongly associated with repeated exposure to cigarette smoke and alcohol.
- Vocal cord paresis is weakness of one or both vocal cords that can greatly impact daily life.
- Idiopathic laryngeal spasm.
- Laryngopharyngeal reflux is a condition in which acid from the stomach irritates and burns the larynx. Similar damage can occur with gastroesophageal reflux disease (GERD).
- Laryngomalacia is a very common condition of infancy, in which the soft, immature cartilage of the upper larynx collapses inward during inhalation, causing airway obstruction.
- Laryngeal perichondritis, the inflammation of the perichondrium of laryngeal cartilages, causing airway obstruction.
- Laryngeal paralysis is a condition seen in some mammals (including dogs) in which the larynx no longer opens as wide as required for the passage of air, and impedes respiration. In mild cases it can lead to exaggerated or "raspy" breathing or panting, and in serious cases can pose a considerable need for treatment.
- Duchenne muscular dystrophy, intrinsic laryngeal muscles (ILM) are spared from the lack of dystrophin and may serve as a useful model to study the mechanisms of muscle sparing in neuromuscular diseases. Dystrophic ILM presented a significant increase in the expression of calcium-binding proteins. The increase of calcium-binding proteins in dystrophic

> ILM may permit better maintenance of calcium homeostasis, with the consequent absence of myonecrosis. The results further support the concept that abnormal calcium buffering is involved in these neuromuscular diseases.

Treatments

Patients who have lost the use of their larynx are typically prescribed the use of an electrolarynx device. Larynx transplants are a rare procedure. The world's first successful operation took place in 1998 at the Cleveland Clinic, and the second took place in October 2010 at the University of California Davis Medical Center in Sacramento.

Other animals

Pioneering work on the structure and evolution of the larynx was carried out in the 1920s by the British comparative anatomist Victor Negus, culminating in his monumental work *The Mechanism of the Larynx* (1929). Negus, however, pointed out that the descent of the larynx reflected the reshaping and descent of the human tongue into the pharynx. This process is not complete until age six to eight years. Some researchers, such as Philip Lieberman, Dennis Klatt, Bart de Boer and Kenneth Stevens using computer-modeling techniques have suggested that the species-specific human tongue allows the vocal tract (the airway above the larynx) to assume the shapes necessary to produce speech sounds that enhance the robustness of human speech. Sounds such as the vowels of the words see and do, [i] and [u], (in phonetic notation) have been

shown to be less subject to confusion in classic studies such as the 1950 Peterson and Barney investigation of the possibilities for computerized speech recognition.

In contrast, though other species have low larynges, their tongues remain anchored in their mouths and their vocal tracts cannot produce the range of speech sounds of humans. The ability to lower the larynx transiently in some species extends the length of their vocal tract, which as Fitch showed creates the acoustic illusion that they are larger. Research at Haskins Laboratories in the 1960s showed that speech allows humans to achieve a vocal communication rate that exceeds the fusion frequency of the auditory system by fusing sounds together into syllables and words. The additional speech sounds that the human tongue enables us to produce, particularly [i], allow humans to unconsciously infer the length of the vocal tract of the person who is talking, a critical element in recovering the phonemes that make up a word.

Non-mammals

Most tetrapod species possess a larynx, but its structure is typically simpler than that found in mammals. The cartilages surrounding the larynx are apparently a remnant of the original gill arches in fish, and are a common feature, but not all are always present. For example, the thyroid cartilage is found only in mammals. Similarly, only mammals possess a true epiglottis, although a flap of non-cartilagenousmucosais found in a similar position in many other groups. In modern amphibians, the laryngeal skeleton is considerably reduced; frogs have only the cricoid and arytenoid cartilages, while salamanders possess only the arytenoids.

Vocal folds are found only in mammals, and a few lizards. As a result, many reptiles and amphibians are essentially voiceless; frogs use ridges in the trachea to modulate sound, while birds have a separate sound-producing organ, the syrinx.

History

The ancient Greek physician Galen first described the larynx, describing it as the "first and supremely most important instrument of the voice"

Chapter 5

Acoustics

Manner of articulation

In articulatory phonetics, the **manner of articulation** is the configuration and interaction of the articulators (speech organs such as the tongue, lips, and palate) when making a speech sound. One parameter of manner is *stricture,* that is, how closely the speech organs approach one another. Others include those involved in the r-like sounds (taps and trills), and the sibilancy of fricatives. The concept of manner is mainly used in the discussion of consonants, although the movement of the articulators will also greatly alter the resonant properties of the vocal tract, thereby changing the formant structure of speech sounds that is crucial for the identification of vowels. For consonants, the place of articulation and the degree of phonation of voicing are considered separately from manner, as being independent parameters. Homorganic consonants, which have the same place of articulation, may have different manners of articulation. Often nasality and laterality are included in manner, but some phoneticians, such as Peter Ladefoged, consider them to be independent.

Broad Classifications

Manners of articulation with substantial obstruction of the airflow (stops, fricatives, affricates) are called**obstruents**. These are prototypically voiceless, but voiced obstruents are

extremely common as well. Manners without such obstruction (nasals, liquids, approximants, and alsovowels) are called **sonorants** because they are nearly always voiced. Voiceless sonorants are uncommon, but are found in Welsh and Classical Greek (the spelling "rh"), in Standard Tibetan (the "lh" of Lhasa), and the "wh" in those dialects of English that distinguish "which" from "witch".

Sonorants may also be called**resonants**, and some linguists prefer that term, restricting the word 'sonorant' to non-vocoidresonants (that is, nasals and liquids, but not vowels or semi-vowels). Another common distinction is between **occlusives** (stops, nasals and affricates) and **continuants** (all else).

Stricture

From greatest to least stricture, speech sounds may be classified along a cline as stop consonants (with *occlusion*, or blocked airflow), fricative consonants (with partially blocked and therefore strongly turbulent airflow), approximants (with only slight turbulence), and vowels (with full unimpeded airflow). Affricates often behave as if they were intermediate between stops and fricatives, but phonetically they are sequences of a stop and fricative.

Over time, sounds in a language may move along the cline toward less stricture in a process called lenition or towards more stricture in a process called fortition.

Other parameters

Sibilants are distinguished from other fricatives by the shape of the tongue and how the airflow is directed over the teeth. Fricatives at coronal places of articulation may be sibilant or non-sibilant, sibilants being the more common.

Flaps (also called taps) are similar to very brief stops. However, their articulation and behavior are distinct enough to be considered a separate manner, rather than just length. The main articulatory difference between flaps and stops is that, due to the greater length of stops compared to flaps, a build-up of air pressure occurs behind a stop which does not occur behind a flap. This means that when the stop is released, there is a burst of air as the pressure is relieved, while for flaps there is no such burst.

Trills involve the vibration of one of the speech organs. Since trilling is a separate parameter from stricture, the two may be combined. Increasing the stricture of a typical trill results in a trilled fricative. Trilled affricates are also known.

Nasal airflow may be added as an independent parameter to any speech sound. It is most commonly found in nasal occlusives and nasal vowels, but nasalized fricatives, taps, and approximants are also found. When a sound is not nasal, it is called*oral.* Laterality is the release of airflow at the side of the tongue. This can be combined with other manners, resulting in lateral approximants (such as the pronunciation of the letter L in the English word "let"), lateral flaps, and lateral fricatives and affricates.

Individual manners

- **Stop**, often called a plosive, is an oral occlusive, where there is *occlusion* (blocking) of the oral vocal tract, and no nasal air flow, so the air flow stops completely. Examples include English/p t k/ (voiceless) and /b d ɡ/ (voiced). If the consonant is voiced, the voicing is the only sound made during occlusion; if it is voiceless, a stop is completely silent. What we hear as a /p/ or /k/ is the effect that the *onset* of the occlusion has on the preceding vowel, as well as the release burst and its effect on the following vowel. The shape and position of the tongue (the *place* of articulation) determine the resonant cavity that gives different stops their characteristic sounds. All languages have stops.
- **Nasal**, a nasal occlusive, where there is occlusion of the oral tract, but air passes through the nose. The shape and position of the tongue determine the resonant cavity that gives different nasals their characteristic sounds. Examples include English /m, n/. Nearly all languages have nasals, the only exceptions being in the area of Puget Sound and a single language on Bougainville Island.
- **Fricative**, sometimes called **spirant**, where there is continuous *frication* (turbulent and noisy airflow) at the place of articulation. Examples include English /f, s/ (voiceless), /v, z/ (voiced), etc. Most languages have fricatives, though many have only an/s/. However, the Indigenous Australian languages are almost completely devoid of fricatives of any kind.

- **Sibilants** are a type of fricative where the airflow is guided by a groove in the tongue toward the teeth, creating a high-pitched and very distinctive sound. These are by far the most common fricatives. Fricatives at coronal (front of tongue) places of articulation are usually, though not always, sibilants. English sibilants include /s/ and /z/.
- **Lateral fricatives** are a rare type of fricative, where the frication occurs on one or both sides of the edge of the tongue. The "ll" of Welsh and the "hl" of Zulu are lateral fricatives.
- **Affricate**, which begins like a stop, but this releases into a fricative rather than having a separate release of its own. The English letters "ch" [tʃ] and "j" [d͡ʒ] represent affricates. Affricates are quite common around the world, though less common than fricatives.
- **Flap**, often called a **tap**, is a momentary closure of the oral cavity. The "tt" of "utter" and the "dd" of "udder" are pronounced as a flap [ɾ] in North American and Australian English. Many linguists distinguish *taps* from *flaps*, but there is no consensus on what the difference might be. No language relies on such a difference. There are also **lateral flaps**.
- **Trill**, in which the articulator (usually the tip of the tongue) is held in place, and the airstream causes it to vibrate. The double "r" of Spanish "perro" is a trill. Trills and flaps, where there are one or more brief occlusions, constitute a class of consonant called **rhotics**.
- **Approximant**, where there is very little obstruction. Examples include English /w/ and /r/. In some

languages, such as Spanish, there are sounds that seem to fall between *fricative* and *approximant*.

- One use of the word **semivowel**, sometimes called a **glide**, is a type of approximant, pronounced like a vowel but with the tongue closer to the roof of the mouth, so that there is slight turbulence. In English, /w/ is the semivowel equivalent of the vowel /u/, and /j/ (spelled "y") is the semivowel equivalent of the vowel /i/ in this usage. Other descriptions use *semivowel* for vowel-like sounds that are not syllabic, but do not have the increased stricture of approximants. These are found as elements in diphthongs. The word may also be used to cover both concepts. The term **glide** is newer than **semivowel**, being used to indicate an essential quality of sounds such as /w/and /j/, which is the movement (or **glide**) from their initial position (/u/ and /i/, respectively) to a following vowel.
- **Lateral approximants**, usually shortened to **lateral**, are a type of approximant pronounced with the side of the tongue. English /l/ is a lateral. Together with the *rhotics*, which have similar behavior in many languages, these form a class of consonant called **liquids**.

Other airstream initiations

All of these manners of articulation are pronounced with an airstream mechanism called pulmonic egressive, meaning that the air flows outward, and is powered by the lungs (actually the ribs and diaphragm). Other airstream mechanisms are possible. Sounds that rely on some of these include:

- **Ejectives**, which are *glottalic egressive*. That is, the airstream is powered by an upward movement of the glottis rather than by the lungs or diaphragm. Stops, affricates, and occasionally fricatives may occur as ejectives. All ejectives are voiceless, or at least transition from voiced to voiceless.
- **Implosives**, which are *glottalic ingressive*. Here the glottis moves downward, but the lungs may be used simultaneously (to provide voicing), and in some languages no air may actually flow into the mouth. Implosive stops are not uncommon, but implosive affricates and fricatives are rare. Voiceless implosives are also rare.
- **Clicks**, which are *lingual ingressive*. Here the back of the tongue is used to create a vacuum in the mouth, causing air to rush in when the forward occlusion (tongue or lips) is released. Clicks may be oral or nasal, stop or affricate, central or lateral, voiced or voiceless. They are extremely rare in normal words outside Southern Africa. However, English has a click in its "tsk tsk" (or "tut tut") sound, and another is often used to say "giddy up" to a horse.
- Combinations of these, in some analyses, in a single consonant: *linguo-pulmonic* and *linguo-glottalic (ejective)* consonants, which are clicks released into either a pulmonic or ejective stop/fricative.

Breathing

Breathing (or **ventilation**) is the process of moving air out and in the lungs to facilitate gas exchange with the internal environment, mostly to flush out carbon dioxide and bring in

oxygen. All aerobic creatures need oxygen for cellular respiration, which uses the oxygen to break down foods for energy and produces carbon dioxide as a waste product. Breathing, or "external respiration", brings air into the lungs where gas exchange takes place in the alveoli through diffusion. The body's circulatory system transports these gases to and from the cells, where "cellular respiration" takes place.

The breathing of all vertebrates with lungs consists of repetitive cycles of inhalation and exhalation through a highly branched system of tubes or airways which lead from the nose to the alveoli. The number of respiratory cycles per minute is the breathing or respiratory rate, and is one of the four primary vital signs of life. Under normal conditions the breathing depth and rate is automatically, and unconsciously, controlled by several homeostatic mechanisms which keep the partial pressures of carbon dioxide and oxygen in the arterial blood constant. Keeping the partial pressure of carbon dioxide in the arterial blood unchanged under a wide variety of physiological circumstances, contributes significantly to tight control of the pH of the extracellular fluids (ECF). Over-breathing (hyperventilation) and under-breathing (hypoventilation), which decrease and increase the arterial partial pressure of carbon dioxide respectively, cause a rise in the pH of ECF in the first case, and a lowering of the pH in the second. Both cause distressing symptoms.

Breathing has other important functions. It provides a mechanism for speech, laughter and similar expressions of the emotions. It is also used for reflexes such as yawning, coughing and sneezing. Animals that cannot thermoregulate by

perspiration, because they lack sufficient sweat glands, may lose heat by evaporation through panting.

Mechanics

The lungs are not capable of inflating themselves, and will expand only when there is an increase in the volume of the thoracic cavity. In humans, as in the other mammals, this is achieved primarily through the contraction of the diaphragm, but also by the contraction of the intercostal muscles which pull the rib cage upwards and outwards as shown in the diagrams on the right. During forceful inhalation (Figure on the right) the accessory muscles of inhalation, which connect the ribs and sternum to the cervical vertebrae and base of the skull, in many cases through an intermediary attachment to the clavicles, exaggerate the pump handle and bucket handle movements (see illustrations on the left), bringing about a greater change in the volume of the chest cavity. During exhalation (breathing out), at rest, all the muscles of inhalation relax, returning the chest and abdomen to a position called the “resting position”, which is determined by their anatomical elasticity. At this point the lungs contain the functional residual capacity of air, which, in the adult human, has a volume of about 2.5–3.0 liters.

During heavy breathing (hyperpnea) as, for instance, during exercise, exhalation is brought about by relaxation of all the muscles of inhalation, (in the same way as at rest), but, in addition, the abdominal muscles, instead of being passive, now contract strongly causing the rib cage to be pulled downwards (front and sides). This not only decreases the size of the rib cage but also pushes the abdominal organs upwards against

the diaphragm which consequently bulges deeply into the thorax. The end-exhalatory lung volume is now less air than the resting "functional residual capacity". However, in a normal mammal, the lungs cannot be emptied completely. In an adult human, there is always still at least one liter of residual air left in the lungs after maximum exhalation.

Diaphragmatic breathing causes the abdomen to rhythmically bulge out and fall back. It is, therefore, often referred to as "abdominal breathing". These terms are often used interchangeably because they describe the same action.

When the accessory muscles of inhalation are activated, especially during labored breathing, the clavicles are pulled upwards, as explained above. This external manifestation of the use of the accessory muscles of inhalation is sometimes referred to as clavicular breathing, seen especially during asthma attacks and in people with chronic obstructive pulmonary disease.

Passage of air

Upper airways

Ideally, air is breathed first out and secondly in through the nose. The nasal cavities (between the nostrils and the pharynx) are quite narrow, firstly by being divided in two by the nasal septum, and secondly by lateralwalls that have several longitudinal folds, or shelves, called nasal conchae, thus exposing a large area of nasal mucous membrane to the air as it is inhaled (and exhaled). This causes the inhaled air to take up moisture from the wet mucus, and warmth from the

underlying blood vessels, so that the air is very nearly saturated with water vapor and is at almost body temperature by the time it reaches the larynx. Part of this moisture and heat is recaptured as the exhaled air moves out over the partially dried-out, cooled mucus in the nasal passages, during exhalation. The sticky mucus also traps much of the particulate matter that is breathed in, preventing it from reaching the lungs.

Lower airways

The anatomy of a typical mammalian respiratory system, below the structures normally listed among the "upper airways" (the nasal cavities, the pharynx, and larynx), is often described as a **respiratory tree** or **tracheobronchial tree** (figure on the left). Larger airways give rise to branches that are slightly narrower, but more numerous than the "trunk" airway that gives rise to the branches. The human respiratory tree may consist of, on average, 23 such branchings into progressively smaller airways, while the respiratory tree of the mouse has up to 13 such branchings. Proximal divisions (those closest to the top of the tree, such as the trachea and bronchi) function mainly to transmit air to the lower airways. Later divisions such as the respiratory bronchioles, alveolar ducts and alveoli are specialized for gas exchange.

The trachea and the first portions of the main bronchi are outside the lungs. The rest of the "tree" branches within the lungs, and ultimately extends to every part of the lungs.

The alveoli are the blind-ended terminals of the "tree", meaning that any air that enters them has to exit the same way it came.

A system such as this creates dead space, a term for the volume of air that fills the airways at the end of inhalation, and is breathed out, unchanged, during the next exhalation, never having reached the alveoli. Similarly, the dead space is filled with alveolar air at the end of exhalation, which is the first air to breathed back into the alveoli during inhalation, before any fresh air which follows after it. The dead space volume of a typical adult human is about 150 ml.

Gas exchange

The primary purpose of breathing is to refresh air in the alveoli so that gas exchange can take place in the blood. The equilibration of the partial pressures of the gases in the alveolar blood and the alveolar air occurs by diffusion. After exhaling, adult human lungs still contain 2.5–3 L of air, their functional residual capacity or FRC. On inhalation, only about 350 mL of new, warm, moistened atmospheric air is brought in and is well mixed with the FRC. Consequently, the gas composition of the FRC changes very little during the breathing cycle. This means that the pulmonary, capillary blood always equilibrates with a relatively constant air composition in the lungs and the diffusion rate with arterial blood gases remains equally constant with each breath. Body tissues are therefore not exposed to large swings in oxygen and carbon dioxide tensions in the blood caused by the breathing cycle, and the peripheral and central chemoreceptors measure only gradual changes in dissolved gases. Thus the homeostatic control of the breathing rate depends only on the partial pressures of oxygen and carbon dioxide in the arterial blood, which then also maintains a constant pH of the blood.

Control

The rate and depth of breathing is automatically controlled by the respiratory centers that receive information from the peripheral and central chemoreceptors. These chemoreceptors continuously monitor the partial pressures of carbon dioxide and oxygen in the arterial blood. The first of these sensors are the central chemoreceptors on the surface of the medulla oblongata of the brain stem which are particularly sensitive to pH as well as the partial pressure of carbon dioxide in the blood and cerebrospinal fluid. The second group of sensors measure the partial pressure of oxygen in the arterial blood. Together the latter are known as the peripheral chemoreceptors, and are situated in the aortic and carotid bodies.

Information from all of these chemoreceptors is conveyed to the respiratory centers in the pons and medulla oblongata, which responds to fluctuations in the partial pressures of carbon dioxide and oxygen in the arterial blood by adjusting the rate and depth of breathing, in such a way as to restore the partial pressure of carbon dioxide to 5.3 kPa (40 mm Hg), the pH to 7.4 and, to a lesser extent, the partial pressure of oxygen to 13 kPa (100 mm Hg). For example, exercise increases the production of carbon dioxide by the active muscles. This carbon dioxide diffuses into the venous blood and ultimately raises the partial pressure of carbon dioxide in the arterial blood. This is immediately sensed by the carbon dioxide chemoreceptors on the brain stem. The respiratory centers respond to this information by causing the rate and depth of breathing to increase to such an extent that the partial pressures of carbon dioxide and oxygen in the arterial blood

return almost immediately to the same levels as at rest. The respiratory centers communicate with the muscles of breathing via motor nerves, of which the phrenic nerves, which innervate the diaphragm, are probably the most important.

Automatic breathing can be overridden to a limited extent by simple choice, or to facilitate swimming, speech, singing or other vocal training. It is impossible to suppress the urge to breathe to the point of hypoxia but training can increase the ability to hold one's breath.

Conscious breathing practices have been shown to promote relaxation and stress relief but have not been proven to have any other health benefits.

Other automatic breathing control reflexes also exist. Submersion, particularly of the face, in cold water, triggers a response called the diving reflex. This has the initial result of shutting down the airways against the influx of water. The metabolic rate slows right down. This is coupled with intense vasoconstriction of the arteries to the limbs and abdominal viscera, reserving the oxygen that is in blood and lungs at the beginning of the dive almost exclusively for the heart and the brain.

The diving reflex is an often-used response in animals that routinely need to dive, such as penguins, seals and whales. It is also more effective in very young infants and children than in adults.

Composition

Inhaled air is by volume 78% nitrogen, 20.95% oxygen and small amounts of other gases including argon, carbon dioxide, neon, helium, and hydrogen.

The gas exhaled is 4% to 5% by volume of carbon dioxide, about a 100 fold increase over the inhaled amount. The volume of oxygen is reduced by a small amount, 4% to 5%, compared to the oxygen inhaled. The typical composition is:

- 5.0–6.3% water vapor
- 79% nitrogen
- 13.6–16.0% oxygen
- 4.0–5.3% carbon dioxide
- 1% argon
- parts per million (ppm) of hydrogen, from the metabolic activity of microorganisms in the large intestine.
- ppm of carbon monoxide from degradation of heme proteins.
- 1 ppm of ammonia.
- Trace many hundreds of volatile organic compounds especially isoprene and acetone. The presence of certain organic compounds indicate disease.

In addition to air, underwater divers practicing technical diving may breathe oxygen-rich, oxygen-depleted or helium-rich breathing gas mixtures. Oxygen and analgesic gases are sometimes given to patients under medical care. The atmosphere in space suits is pure oxygen. However, this is kept at around 20% of Earthbound atmospheric pressure to regulate the rate of inspiration.

Effects of ambient air pressure

Breathing at altitude

Atmospheric pressure decreases with the height above sea level (altitude) and since the alveoli are open to the outside air through the open airways, the pressure in the lungs also decreases at the same rate with altitude. At altitude, a pressure differential is still required to drive air into and out of the lungs as it is at sea level. The mechanism for breathing at altitude is essentially identical to breathing at sea level but with the following differences:

The atmospheric pressure decreases exponentially with altitude, roughly halving with every 5,500 metres (18,000 ft) rise in altitude. The composition of atmospheric air is, however, almost constant below 80 km, as a result of the continuous mixing effect of the weather. The concentration of oxygen in the air (mmols O_2 per liter of air) therefore decreases at the same rate as the atmospheric pressure. At sea level, where the ambient pressure is about 100 kPa, oxygen contributes 21% of the atmosphere and the partial pressure of oxygen (P_{O2}) is 21 kPa (i.e. 21% of 100 kPa). At the summit of Mount Everest, 8,848 metres (29,029 ft), where the total atmospheric pressure is 33.7 kPa, oxygen still contributes 21% of the atmosphere but its partial pressure is only 7.1 kPa (i.e. 21% of 33.7 kPa = 7.1 kPa). Therefore, a greater volume of air must be inhaled at altitude than at sea level in order to breathe in the same amount of oxygen in a given period.

During inhalation, air is warmed and saturated with water vapor as it passes through the nose and pharynx before it

enters the alveoli. The *saturated* vapor pressure of water is dependent only on temperature; at a body core temperature of 37 °C it is 6.3 kPa (47.0 mmHg), regardless of any other influences, including altitude. Consequently, at sea level, the *tracheal* air (immediately before the inhaled air enters the alveoli) consists of: water vapor (P_{H2O} = 6.3 kPa), nitrogen (P_{N2} = 74.0 kPa), oxygen (P_{O2} = 19.7 kPa) and trace amounts of carbon dioxide and other gases, a total of 100 kPa. In dry air, the P_{O2} at sea level is 21.0 kPa, compared to a P_{O2} of 19.7 kPa in the tracheal air (21% of [100 – 6.3] = 19.7 kPa). At the summit of Mount Everest tracheal air has a total pressure of 33.7 kPa, of which 6.3 kPa is water vapor, reducing the P_{O2} in the tracheal air to 5.8 kPa (21% of [33.7 – 6.3] = 5.8 kPa), beyond what is accounted for by a reduction of atmospheric pressure alone (7.1 kPa).

The pressure gradient forcing air into the lungs during inhalation is also reduced by altitude. Doubling the volume of the lungs halves the pressure in the lungs at any altitude. Having the sea level air pressure (100 kPa) results in a pressure gradient of 50 kPa but doing the same at 5500 m, where the atmospheric pressure is 50 kPa, a doubling of the volume of the lungs results in a pressure gradient of the only 25 kPa.

In practice, because we breathe in a gentle, cyclical manner that generates pressure gradients of only 2–3 kPa, this has little effect on the actual rate of inflow into the lungs and is easily compensated for by breathing slightly deeper. The lower viscosity of air at altitude allows air to flow more easily and this also helps compensate for any loss of pressure gradient.

All of the above effects of low atmospheric pressure on breathing are normally accommodated by increasing the respiratory minute volume (the volume of air breathed in — *or* out — per minute), and the mechanism for doing this is automatic. The exact increase required is determined by the respiratory gases homeostatic mechanism, which regulates the arterial P_{O2} and P_{CO2}. This homeostatic mechanism prioritizes the regulation of the arterial P_{CO2} over that of oxygen at sea level. That is to say, at sea level the arterial $P_{CO2}$is maintained at very close to 5.3 kPa (or 40 mmHg) under a wide range of circumstances, at the expense of the arterial P_{O2}, which is allowed to vary within a very wide range of values, before eliciting a corrective ventilatory response. However, when the atmospheric pressure (and therefore the atmospheric P_{O2}) falls to below 75% of its value at sea level, oxygen homeostasisis given priority over carbon dioxide homeostasis. This switch-over occurs at an elevation of about 2,500 metres (8,200 ft). If this switch occurs relatively abruptly, the hyperventilation at high altitude will cause a severe fall in the arterial P_{CO2} with a consequent rise in the pH of the arterial plasma leading to respiratory alkalosis. This is one contributor to high altitude sickness. On the other hand, if the switch to oxygen homeostasis is incomplete, then hypoxia may complicate the clinical picture with potentially fatal results.

Breathing at depth

Pressure increases with the depth of water at the rate of about one atmosphere — slightly more than 100 kPa, or onebar, for every 10 meters. Air breathed underwater by divers is at the ambient pressure of the surrounding water and this has a complex range of physiological and biochemical implications. If

not properly managed, breathing compressed gasses underwater may lead to several diving disorders which include pulmonary barotrauma, decompression sickness, nitrogen narcosis, and oxygen toxicity. The effects of breathing gasses under pressure are further complicated by the use of one or more special gas mixtures.

Air is provided by a diving regulator, which reduces the high pressure in a diving cylinder to the ambient pressure. The breathing performance of regulators is a factor when choosing a suitable regulator for the type of diving to be undertaken. It is desirable that breathing from a regulator requires low effort even when supplying large amounts of air. It is also recommended that it supplies air smoothly without any sudden changes in resistance while inhaling or exhaling. In the graph, right, note the initial spike in pressure on exhaling to open the exhaust valve and that the initial drop in pressure on inhaling is soon overcome as the Venturi effect designed into the regulator to allow an easy draw of air. Many regulators have an adjustment to change the ease of inhaling so that breathing is effortless.

Respiratory disorders

Abnormal breathing patterns include Kussmaul breathing, Biot's respiration and Cheyne–Stokes respiration.

Other breathing disorders include shortness of breath (dyspnea), stridor, apnea, sleep apnea (most commonly obstructive sleep apnea), mouth breathing, and snoring. Many conditions are associated with obstructed airways. Chronic mouth breathing may be associated with illness. Hypopnea

refers to overly shallow breathing; hyperpnea refers to fast and deep breathing brought on by a demand for more oxygen, as for example by exercise. The terms hypoventilation and hyperventilation also refer to shallow breathing and fast and deep breathing respectively, but under inappropriate circumstances or disease. However, this distinction (between, for instance, hyperpnea and hyperventilation) is not always adhered to, so that these terms are frequently used interchangeably.

A range of breath testscan be used to diagnose diseases such as dietary intolerances. A rhinomanometer uses acoustic technology to examine the air flow through the nasal passages.

Society and culture

The word "spirit" comes from the Latin*spiritus*, meaning breath. Historically, breath has often been considered in terms of the concept of life force. The Hebrew Bible refers to God breathing the breath of life into clay to make Adam a living soul (nephesh). It also refers to the breath as returning to God when a mortal dies. The terms spirit, prana, the Polynesian mana, the Hebrew ruach and the psyche in psychology are related to the concept of breath.

In T'ai chi, aerobic exerciseis combined with breathing exercises to strengthen the diaphragm muscles, improve posture and make better use of the body's qi. Different forms of meditation, and yoga advocate various breathing methods. A form of Buddhist meditation called anapanasati meaning mindfulness of breath was first introduced by Buddha. Breathing disciplines are incorporated into meditation, certain

forms of yoga such as pranayama, and the Buteyko method as a treatment for asthma and other conditions.

In music, some wind instrument players use a technique called circular breathing. Singers also rely on breath control.

Common cultural expressions related to breathing include: "to catch my breath", "took my breath away", "inspiration", "to expire", "get my breath back".

Breathing and mood

Certain breathing patterns have a tendency to occur with certain moods. Due to this relationship, practitioners of various disciplines consider that they can encourage the occurrence of a particular mood by adopting the breathing pattern that it most commonly occurs in conjunction with. For instance, and perhaps the most common recommendation is that deeper breathing which utilizes the diaphragm and abdomen more can encourage relaxation. Practitioners of different disciplines often interpret the importance of breathing regulation and its perceived influence on mood in different ways. Buddhists may consider that it helps precipitate a sense of inner-peace, holistic healers that it encourages an overall state of health and business advisers that it provides relief from work-based stress.

Breathing and physical exercise

During physical exercise, a deeper breathing pattern is adapted to facilitate greater oxygen absorption. An additional reason for the adoption of a deeper breathing pattern is to strengthen the body's core. During the process of deep

breathing, the thoracic diaphragm adopts a lower position in the core and this helps to generate intra-abdominal pressure which strengthens the lumbar spine. Typically, this allows for more powerful physical movements to be performed. As such, it is frequently recommended when lifting heavy weights to take a deep breath or adopt a deeper breathing pattern.

Source–filter model

The **source–filter model** represents speech as a combination of a sound source, such as the vocal cords, and a linear acoustic filter, the vocal tract. While only an approximation, the model is widely used in a number of applications such as speech synthesis and speech analysis because of its relative simplicity. It is also related to linear prediction. The development of the model is due, in large part, to the early work of Gunnar Fant, although others, notably Ken Stevens, have also contributed substantially to the models underlying acoustic analysis of speech and speech synthesis. Fant built off the work of Tsutomu Chiba and Masato Kajiyama, who first showed the relationship between a vowel's acoustic properties and the shape of the vocal tract.

An important assumption that is often made in the use of the source–filter model is the independence of source and filter. In such cases, the model should more accurately be referred to as the "independent source–filter model".

History

In 1942, Chiba and Kajiyama published their research on vowel acoustics and the vocal tract in their book, *The Vowel: Its*

nature and structure. By creating models of the vocal tract using X-ray photography, they were able to predict the formant frequencies of different vowels, establishing a relationship between the two. Gunnar Fant, a pioneering speech scientist, used Chiba and Kajiyama's research involving X-ray photography of the vocal tract to interpret his own data of Russian speech sounds in *Acoustic Theory of Speech Production*, which established the source–filter model.

Applications

To varying degrees, different phonemescan be distinguished by the properties of their source(s) and their spectral shape. Voiced sounds (e.g., vowels) have at least one source due to mostly periodic glottal excitation, which can be approximated by an impulse train in the time domain and by harmonics in the frequency domain, and a filter that depends on, for example, tongue position and lip protrusion. On the other hand, fricatives, such as [s] and [f], have at least one source due to turbulent noise produced at a constriction in the oral cavity or pharynx. So-called *voiced fricatives*, such as [z] and [v], have two sources - one at the glottis and one at the supra-glottal constriction.

Speech synthesis

In implementation of the source–filter model of speech production, the sound source, or excitation signal, is often modelled as a periodic impulse train, for voiced speech, or white noise for unvoiced speech. The vocal tract filter is, in the simplest case, approximated by an all-pole filter, where the coefficients are obtained by performing linear prediction to

minimize the mean-squared error in the speech signal to be reproduced. Convolution of the excitation signal with the filter response then produces the synthesised speech.

Modeling human speech production

In human speech production, the sound source is the vocal folds, which can produce a periodic sound when constricted or an aperiodic (white noise) sound when relaxed. The filter is the rest of the vocal tract, which can change shape through manipulation of the pharynx, mouth, and nasal cavity. Fant roughly compares the source and filter to phonation and articulation, respectively. The source produces a number of harmonics of varying amplitudes, which travel through the vocal tract and are either amplified or attenuated to produce a speech sound.

Chapter 6

Perception

Speech perception

Categorical perception is involved in processes of perceptual differentiation. People perceive speech sounds categorically, that is to say, they are more likely to notice the differences *between* categories (phonemes) than *within* categories. The perceptual space between categories is therefore warped, the centers of categories (or "prototypes") working like a sieve or like magnets for incoming speech sounds.

In an artificial continuum between a voiceless and a voiced bilabial plosive, each new step differs from the preceding one in the amount ofVOT. The first sound is a pre-voiced[b], i.e. it has a negative VOT. Then, increasing the VOT, it reaches zero, i.e. the plosive is a plain unaspirated voiceless [p]. Gradually, adding the same amount of VOT at a time, the plosive is eventually a strongly aspirated voiceless bilabial [p^h]. (Such a continuum was used in an experiment by Lisker and Abramson in 1970.

The sounds they used are available online.) In this continuum of, for example, seven sounds, native English listeners will identify the first three sounds as /b/ and the last three sounds as /p/ with a clear boundary between the two categories. A two-alternative identification (or categorization) test will yield a discontinuous categorization function (see red curve in Figure 4).

In tests of the ability to discriminate between two sounds with varying VOT values but having a constant VOT distance from each other (20 ms for instance), listeners are likely to perform at chance level if both sounds fall within the same category and at nearly 100% level if each sound falls in a different category (see the blue discrimination curve in Figure 4).

The conclusion to make from both the identification and the discrimination test is that listeners will have different sensitivity to the same relative increase in VOT depending on whether or not the boundary between categories was crossed. Similar perceptual adjustment is attested for other acoustic cues as well.

Top-down influences

In a classic experiment, Richard M. Warren (1970) replaced one phoneme of a word with a cough-like sound. Perceptually, his subjects restored the missing speech sound without any difficulty and could not accurately identify which phoneme had been disturbed, a phenomenon known as the phonemic restoration effect. Therefore, the process of speech perception is not necessarily uni-directional.

Another basic experiment compared recognition of naturally spoken words within a phrase versus the same words in isolation, finding that perception accuracy usually drops in the latter condition. To probe the influence of semantic knowledge on perception, Garnes and Bond (1976) similarly used carrier sentences where target words only differed in a single phoneme (bay/day/gay, for example) whose quality changed along a continuum. When put into different sentences that each

naturally led to one interpretation, listeners tended to judge ambiguous words according to the meaning of the whole sentence . That is, higher-level language processes connected with morphology, syntax, or semantics may interact with basic speech perception processes to aid in recognition of speech sounds.

It may be the case that it is not necessary and maybe even not possible for a listener to recognize phonemes before recognizing higher units, like words for example. After obtaining at least a fundamental piece of information about phonemic structure of the perceived entity from the acoustic signal, listeners can compensate for missing or noise-masked phonemes using their knowledge of the spoken language. Compensatory mechanisms might even operate at the sentence level such as in learned songs, phrases and verses, an effect backed-up by neural coding patterns consistent with the missed continuous speech fragments, despite the lack of all relevant bottom-up sensory input.

Acquired language impairment

The first ever hypothesis of speech perception was used with patients who acquired an auditory comprehension deficit, also known as receptive aphasia. Since then there have been many disabilities that have been classified, which resulted in a true definition of "speech perception". The term 'speech perception' describes the process of interest that employs sub lexical contexts to the probe process. It consists of many different language and grammatical functions, such as: features, segments (phonemes), syllabic structure (unit of pronunciation), phonological word forms (how sounds are

grouped together), grammatical features, morphemic (prefixes and suffixes), and semantic information (the meaning of the words). In the early years, they were more interested in the acoustics of speech. For instance, they were looking at the differences between /ba/ or /da/, but now research has been directed to the response in the brain from the stimuli. In recent years, there has been a model developed to create a sense of how speech perception works; this model is known as the dual stream model.

This model has drastically changed from how psychologists look at perception. The first section of the dual stream model is the ventral pathway. This pathway incorporates middle temporal gyrus, inferior temporal sulcus and perhaps the inferior temporal gyrus. The ventral pathway shows phonological representations to the lexical or conceptual representations, which is the meaning of the words. The second section of the dual stream model is the dorsal pathway. This pathway includes the sylvianparietotemporal, inferior frontal gyrus, anterior insula, and premotor cortex. Its primary function is to take the sensory or phonological stimuli and transfer it into an articulatory-motor representation (formation of speech).

Aphasia

Aphasia is an impairment of language processing caused by damage to the brain. Different parts of language processing are impacted depending on the area of the brain that is damaged, and aphasia is further classified based on the location of injury or constellation of symptoms. Damage to Broca's area of the brain often results in expressive aphasia which manifests

as impairment in speech production. Damage to Wernicke's area often results in receptive aphasia where speech processing is impaired.

Aphasia with impaired speech perception typically shows lesions or damage located in the left temporal or parietal lobes. Lexical and semantic difficulties are common, and comprehension may be affected.

Agnosia

Agnosia is "the loss or diminution of the ability to recognize familiar objects or stimuli usually as a result of brain damage". There are several different kinds of agnosia that affect every one of our senses, but the two most common related to speech are speech agnosia and phonagnosia.

Speech agnosia: Pure word deafness, or speech agnosia, is an impairment in which a person maintains the ability to hear, produce speech, and even read speech, yet they are unable to understand or properly perceive speech. These patients seem to have all of the skills necessary in order to properly process speech, yet they appear to have no experience associated with speech stimuli. Patients have reported, "I can hear you talking, but I can't translate it". Even though they are physically receiving and processing the stimuli of speech, without the ability to determine the meaning of the speech, they essentially are unable to perceive the speech at all. There are no known treatments that have been found, but from case studies and experiments it is known that speech agnosia is related to lesions in the left hemisphere or both, specifically right temporoparietal dysfunctions.

Phonagnosia: Phonagnosia is associated with the inability to recognize any familiar voices. In these cases, speech stimuli can be heard and even understood but the association of the speech to a certain voice is lost. This can be due to "abnormal processing of complex vocal properties (timbre, articulation, and prosody—elements that distinguish an individual voice". There is no known treatment; however, there is a case report of an epileptic woman who began to experience phonagnosia along with other impairments. Her EEG and MRI results showed "a right cortical parietal T2-hyperintense lesion without gadolinium enhancement and with discrete impairment of water molecule diffusion". So although no treatment has been discovered, phonagnosia can be correlated to postictal parietal cortical dysfunction.

Infant speech perception

Infants begin the process of language acquisition by being able to detect very small differences between speech sounds. They can discriminate all possible speech contrasts (phonemes). Gradually, as they are exposed to their native language, their perception becomes language-specific, i.e. they learn how to ignore the differences within phonemic categories of the language (differences that may well be contrastive in other languages – for example, English distinguishes two voicing categories of plosives, whereas Thai has three categories; infants must learn which differences are distinctive in their native language uses, and which are not). As infants learn how to sort incoming speech sounds into categories, ignoring irrelevant differences and reinforcing the contrastive ones, their perception becomes categorical. Infants learn to contrast different vowel phonemes of their native language by

approximately 6 months of age. The native consonantal contrasts are acquired by 11 or 12 months of age. Some researchers have proposed that infants may be able to learn the sound categories of their native language through passive listening, using a process called statistical learning. Others even claim that certain sound categories are innate, that is, they are genetically specified (see discussion about innate vs. acquired categorical distinctiveness).

If day-old babies are presented with their mother's voice speaking normally, abnormally (in monotone), and a stranger's voice, they react only to their mother's voice speaking normally. When a human and a non-human sound is played, babies turn their head only to the source of human sound. It has been suggested that auditory learning begins already in the pre-natal period.

One of the techniques used to examine how infants perceive speech, besides the head-turn procedure mentioned above, is measuring their sucking rate. In such an experiment, a baby is sucking a special nipple while presented with sounds. First, the baby's normal sucking rate is established. Then a stimulus is played repeatedly. When the baby hears the stimulus for the first time the sucking rate increases but as the baby becomes habituated to the stimulation the sucking rate decreases and levels off. Then, a new stimulus is played to the baby. If the baby perceives the newly introduced stimulus as different from the background stimulus the sucking rate will show an increase. The sucking-rate and the head-turn method are some of the more traditional, behavioral methods for studying speech perception. Among the new methods (see Research

methods below) that help us to study speech perception,near-infrared spectroscopy is widely used in infants.

It has also been discovered that even though infants' ability to distinguish between the different phonetic properties of various languages begins to decline around the age of nine months, it is possible to reverse this process by exposing them to a new language in a sufficient way. In a research study by Patricia K. Kuhl, Feng-Ming Tsao, and Huei-Mei Liu, it was discovered that if infants are spoken to and interacted with by a native speaker of Mandarin Chinese, they can actually be conditioned to retain their ability to distinguish different speech sounds within Mandarin that are very different from speech sounds found within the English language. Thus proving that given the right conditions, it is possible to prevent infants' loss of the ability to distinguish speech sounds in languages other than those found in the native language.

Cross-language and second-language

A large amount of research has studied how users of a language perceive foreign speech (referred to as cross-language speech perception) or second-language speech (second-language speech perception). The latter falls within the domain of second language acquisition.

Languages differ in their phonemic inventories. Naturally, this creates difficulties when a foreign language is encountered. For example, if two foreign-language sounds are assimilated to a single mother-tongue category the difference between them will be very difficult to discern. A classic example of this situation is the observation that Japanese learners of English will have

problems with identifying or distinguishing English liquid consonants/l/ and /r/ (see Perception of English /r/ and /l/ by Japanese speakers).

Best (1995) proposed a Perceptual Assimilation Model which describes possible cross-language category assimilation patterns and predicts their consequences. Flege (1995) formulated a Speech Learning Model which combines several hypotheses about second-language (L2) speech acquisition and which predicts, in simple words, that an L2 sound that is not too similar to a native-language (L1) sound will be easier to acquire than an L2 sound that is relatively similar to an L1 sound (because it will be perceived as more obviously "different" by the learner).

In language or hearing impairment

Research in how people with language or hearing impairment perceive speech is not only intended to discover possible treatments. It can provide insight into the principles underlying non-impaired speech perception. Two areas of research can serve as an example:

Listeners with aphasia

Aphasia affects both the expression and reception of language. Both two most common types, expressive aphasia and receptive aphasia, affect speech perception to some extent. Expressive aphasia causes moderate difficulties for language understanding. The effect of receptive aphasia on understanding is much more severe. It is agreed upon, that aphasics suffer from perceptual deficits. They usually cannot

fully distinguish place of articulation and voicing. As for other features, the difficulties vary. It has not yet been proven whether low-level speech-perception skills are affected in aphasia sufferers or whether their difficulties are caused by higher-level impairment alone.

Listeners with cochlear implants

- Cochlear implantation restores access to the acoustic signal in individuals with sensorineural hearing loss. The acoustic information conveyed by an implant is usually sufficient for implant users to properly recognize speech of people they know even without visual clues. For cochlear implant users, it is more difficult to understand unknown speakers and sounds. The perceptual abilities of children that received an implant after the age of two are significantly better than of those who were implanted in adulthood. A number of factors have been shown to influence perceptual performance, specifically: duration of deafness prior to implantation, age of onset of deafness, age at implantation (such age effects may be related to the Critical period hypothesis) and the duration of using an implant. There are differences between children with congenital and acquired deafness. Postlingually deaf children have better results than the prelingually deaf and adapt to a cochlear implant faster. In both children with cochlear implants and normal hearing, vowels and voice onset time becomes prevalent in development before the ability to discriminate the place of articulation. Several

months following implantation, children with cochlear implants can normalize speech perception.

Noise

One of the fundamental problems in the study of speech is how to deal with noise. This is shown by the difficulty in recognizing human speech that computer recognition systems have. While they can do well at recognizing speech if trained on a specific speaker's voice and under quiet conditions, these systems often do poorly in more realistic listening situations where humans would understand speech without relative difficulty. To emulate processing patterns that would be held in the brain under normal conditions, prior knowledge is a key neural factor, since a robust learning history may to an extent override the extreme masking effects involved in the complete absence of continuous speech signals.

Music-language connection

Research into the relationship between music and cognition is an emerging field related to the study of speech perception. Originally it was theorized that the neural signals for music were processed in a specialized "module" in the right hemisphere of the brain. Conversely, the neural signals for language were to be processed by a similar "module" in the left hemisphere. However, utilizing technologies such as fMRI machines, research has shown that two regions of the brain traditionally considered exclusively to process speech, Broca's and Wernicke's areas, also become active during musical activities such as listening to a sequence of musical chords.

Other studies, such as one performed by Marques et al. in 2006 showed that 8-year-olds who were given six months of musical training showed an increase in both their pitch detection performance and their electrophysiological measures when made to listen to an unknown foreign language.

Conversely, some research has revealed that, rather than music affecting our perception of speech, our native speech can affect our perception of music. One example is the tritone paradox. The tritone paradox is where a listener is presented with two computer-generated tones (such as C and F-Sharp) that are half an octave (or a tritone) apart and are then asked to determine whether the pitch of the sequence is descending or ascending. One such study, performed by Ms. Diana Deutsch, found that the listener's interpretation of ascending or descending pitch was influenced by the listener's language or dialect, showing variation between those raised in the south of England and those in California or from those in Vietnam and those in California whose native language was English. A second study, performed in 2006 on a group of English speakers and 3 groups of East Asian students at University of Southern California, discovered that English speakers who had begun musical training at or before age 5 had an 8% chance of having perfect pitch.

Speech phenomenology

The experience of speech

Casey O'Callaghan, in his article *Experiencing Speech*, analyzes whether "the perceptual experience of listening to speech differs in phenomenal character" with regards to

understanding the language being heard. He argues that an individual's experience when hearing a language they comprehend, as opposed to their experience when hearing a language they have no knowledge of, displays a difference in *phenomenal features* which he defines as "aspects of what an experience is like" for an individual.

If a subject who is a monolingual native English speaker is presented with a stimulus of speech in German, the string of phonemes will appear as mere sounds and will produce a very different experience than if exactly the same stimulus was presented to a subject who speaks German.

He also examines how speech perception changes when one learning a language. If a subject with no knowledge of the Japanese language was presented with a stimulus of Japanese speech, and then was given the exact *same* stimuli after being taught Japanese, this *same* individual would have an extremely *different* experience.

Research methods

The methods used in speech perception research can be roughly divided into three groups: behavioral, computational, and, more recently, neurophysiological methods.

Behavioral methods

Behavioral experiments are based on an active role of a participant, i.e. subjects are presented with stimuli and asked to make conscious decisions about them. This can take the form of an identification test, a discrimination test, similarity

rating, etc. These types of experiments help to provide a basic description of how listeners perceive and categorize speech sounds.

Sinewave Speech

Speech perception has also been analyzed through sinewave speech, a form of synthetic speech where the human voice is replaced by sine waves that mimic the frequencies and amplitudes present in the original speech. When subjects are first presented with this speech, the sinewave speech is interpreted as random noises.

But when the subjects are informed that the stimuli actually is speech and are told what is being said, "a distinctive, nearly immediate shift occurs" to how the sinewave speech is perceived.

Computational methods

Computational modeling has also been used to simulate how speech may be processed by the brain to produce behaviors that are observed. Computer models have been used to address several questions in speech perception, including how the sound signal itself is processed to extract the acoustic cues used in speech, and how speech information is used for higher-level processes, such as word recognition.

Neurophysiological methods

Neurophysiological methods rely on utilizing information stemming from more direct and not necessarily conscious (pre-

attentative) processes. Subjects are presented with speech stimuli in different types of tasks and the responses of the brain are measured. The brain itself can be more sensitive than it appears to be through behavioral responses. For example, the subject may not show sensitivity to the difference between two speech sounds in a discrimination test, but brain responses may reveal sensitivity to these differences. Methods used to measure neural responses to speech include event-related potentials, magnetoencephalography, and near infrared spectroscopy. One important response used with event-related potentials is the mismatch negativity, which occurs when speech stimuli are acoustically different from a stimulus that the subject heard previously.

Neurophysiological methods were introduced into speech perception research for several reasons:

Behavioral responses may reflect late, conscious processes and be affected by other systems such as orthography, and thus they may mask speaker's ability to recognize sounds based on lower-level acoustic distributions.

Without the necessity of taking an active part in the test, even infants can be tested; this feature is crucial in research into acquisition processes. The possibility to observe low-level auditory processes independently from the higher-level ones makes it possible to address long-standing theoretical issues such as whether or not humans possess a specialized module for perceiving speech or whether or not some complex acoustic invariance (see lack of invariance above) underlies the recognition of a speech sound.

Theories

Motor theory

Some of the earliest work in the study of how humans perceive speech sounds was conducted by Alvin Liberman and his colleagues at Haskins Laboratories. Using a speech synthesizer, they constructed speech sounds that varied in place of articulation along a continuum from /bɑ/ to /dɑ/ to /gɑ/. Listeners were asked to identify which sound they heard and to discriminate between two different sounds. The results of the experiment showed that listeners grouped sounds into discrete categories, even though the sounds they were hearing were varying continuously. Based on these results, they proposed the notion of categorical perception as a mechanism by which humans can identify speech sounds.

More recent research using different tasks and methods suggests that listeners are highly sensitive to acoustic differences within a single phonetic category, contrary to a strict categorical account of speech perception.

To provide a theoretical account of the categorical perception data, Liberman and colleagues worked out the motor theory of speech perception, where "the complicated articulatory encoding was assumed to be decoded in the perception of speech by the same processes that are involved in production" (this is referred to as analysis-by-synthesis). For instance, the English consonant /d/ may vary in its acoustic details across different phonetic contexts (see above), yet all /d/'s as perceived by a listener fall within one category (voiced alveolar plosive) and that is because "linguistic representations are

abstract, canonical, phonetic segments or the gestures that underlie these segments". When describing units of perception, Liberman later abandoned articulatory movements and proceeded to the neural commands to the articulators and even later to intended articulatory gestures, thus "the neural representation of the utterance that determines the speaker's production is the distal object the listener perceives". The theory is closely related to the modularity hypothesis, which proposes the existence of a special-purpose module, which is supposed to be innate and probably human-specific.

The theory has been criticized in terms of not being able to "provide an account of just how acoustic signals are translated into intended gestures" by listeners. Furthermore, it is unclear how indexical information (e.g. talker-identity) is encoded/decoded along with linguistically relevant information.

Exemplar theory

Exemplar models of speech perception differ from the four theories mentioned above which suppose that there is no connection between word- and talker-recognition and that the variation across talkers is "noise" to be filtered out.

The exemplar-based approaches claim listeners store information for both word- and talker-recognition. According to this theory, particular instances of speech sounds are stored in the memory of a listener. In the process of speech perception, the remembered instances of e.g. a syllable stored in the listener's memory are compared with the incoming stimulus so that the stimulus can be categorized. Similarly,

when recognizing a talker, all the memory traces of utterances produced by that talker are activated and the talker's identity is determined. Supporting this theory are several experiments reported by Johnson that suggest that our signal identification is more accurate when we are familiar with the talker or when we have visual representation of the talker's gender. When the talker is unpredictable or the sex misidentified, the error rate in word-identification is much higher.

The exemplar models have to face several objections, two of which are (1) insufficient memory capacity to store every utterance ever heard and, concerning the ability to produce what was heard, (2) whether also the talker's own articulatory gestures are stored or computed when producing utterances that would sound as the auditory memories.

Acoustic landmarks and distinctive features

Kenneth N. Stevens proposed acoustic landmarks and distinctive features as a relation between phonological features and auditory properties. According to this view, listeners are inspecting the incoming signal for the so-called acoustic landmarks which are particular events in the spectrum carrying information about gestures which produced them. Since these gestures are limited by the capacities of humans' articulators and listeners are sensitive to their auditory correlates, the lack of invariance simply does not exist in this model. The acoustic properties of the landmarks constitute the basis for establishing the distinctive features. Bundles of them uniquely specify phonetic segments (phonemes, syllables, words).

In this model, the incoming acoustic signal is believed to be first processed to determine the so-called landmarks which are special spectral events in the signal; for example, vowels are typically marked by higher frequency of the first formant, consonants can be specified as discontinuities in the signal and have lower amplitudes in lower and middle regions of the spectrum. These acoustic features result from articulation. In fact, secondary articulatory movements may be used when enhancement of the landmarks is needed due to external conditions such as noise. Stevens claims that coarticulation causes only limited and moreover systematic and thus predictable variation in the signal which the listener is able to deal with. Within this model therefore, what is called the lack of invariance is simply claimed not to exist.

Landmarks are analyzed to determine certain articulatory events (gestures) which are connected with them. In the next stage, acoustic cues are extracted from the signal in the vicinity of the landmarks by means of mental measuring of certain parameters such as frequencies of spectral peaks, amplitudes in low-frequency region, or timing.

The next processing stage comprises acoustic-cues consolidation and derivation of distinctive features. These are binary categories related to articulation (for example [+/- high], [+/- back], [+/- round lips] for vowels; [+/- sonorant], [+/- lateral], or [+/- nasal] for consonants.

Bundles of these features uniquely identify speech segments (phonemes, syllables, words). These segments are part of the lexicon stored in the listener's memory. Its units are activated in the process of lexical access and mapped on the original

signal to find out whether they match. If not, another attempt with a different candidate pattern is made. In this iterative fashion, listeners thus reconstruct the articulatory events which were necessary to produce the perceived speech signal. This can be therefore described as analysis-by-synthesis.

This theory thus posits that the distal object of speech perception are the articulatory gestures underlying speech. Listeners make sense of the speech signal by referring to them. The model belongs to those referred to as analysis-by-synthesis.

Fuzzy-logical model

The fuzzy logical theory of speech perception developed by Dominic Massaro proposes that people remember speech sounds in a probabilistic, or graded, way. It suggests that people remember descriptions of the perceptual units of language, called prototypes. Within each prototype various features may combine. However, features are not just binary (true or false), there is a fuzzy value corresponding to how likely it is that a sound belongs to a particular speech category. Thus, when perceiving a speech signal our decision about what we actually hear is based on the relative goodness of the match between the stimulus information and values of particular prototypes. The final decision is based on multiple features or sources of information, even visual information (this explains the McGurk effect). Computer models of the fuzzy logical theory have been used to demonstrate that the theory's predictions of how speech sounds are categorized correspond to the behavior of human listeners.

Speech mode hypothesis

Speech mode hypothesis is the idea that the perception of speech requires the use of specialized mental processing. The speech mode hypothesis is a branch off of Fodor's modularity theory (see modularity of mind). It utilizes a vertical processing mechanism where limited stimuli are processed by special-purpose areas of the brain that are stimuli specific.

Two versions of speech mode hypothesis:

- Weak version – listening to speech engages previous knowledge of language.
- Strong version – listening to speech engages specialized speech mechanisms for perceiving speech.

Three important experimental paradigms have evolved in the search to find evidence for the speech mode hypothesis. These are dichotic listening, categorical perception, and duplex perception. Through the research in these categories it has been found that there may not be a specific speech mode but instead one for auditory codes that require complicated auditory processing. Also it seems that modularity is learned in perceptual systems. Despite this the evidence and counter-evidence for the speech mode hypothesis is still unclear and needs further research.

Direct realist theory

The direct realist theory of speech perception (mostly associated with Carol Fowler) is a part of the more general theory of direct realism, which postulates that perception allows us to have direct awareness of the world because it

involves direct recovery of the distal source of the event that is perceived. For speech perception, the theory asserts that the objects of perception are actual vocal tract movements, or gestures, and not abstract phonemes or (as in the Motor Theory) events that are causally antecedent to these movements, i.e. intended gestures. Listeners perceive gestures not by means of a specialized decoder (as in the Motor Theory) but because information in the acoustic signal specifies the gestures that form it. By claiming that the actual articulatory gestures that produce different speech sounds are themselves the units of speech perception, the theory bypasses the problem of lack of invariance.

Hearing

Hearing, or **auditory perception**, is the ability to perceive sounds through an organ, such as an ear, by detecting vibrations as periodic changes in the pressure of a surrounding medium. The academic field concerned with hearing is auditory science.

Sound may be heard through solid, liquid, or gaseous matter. It is one of the traditional five senses. Partial or total inability to hear is called hearing loss.

In humans and other vertebrates, hearing is performed primarily by the auditory system: mechanical waves, known as vibrations, are detected by the ear and transduced into nerve impulses that are perceived by the brain (primarily in the temporal lobe). Like touch, audition requires sensitivity to the movement of molecules in the world outside the organism. Both hearing and touch are types of mechanosensation.

Hearing mechanism

There are three main components of the human auditory system: the outer ear, the middle ear, and the inner ear.

Outer ear

The outer ear includes the pinna, the visible part of the ear, as well as the ear canal, which terminates at the eardrum, also called the tympanic membrane. The pinna serves to focus sound waves through the ear canal toward the eardrum. Because of the asymmetrical character of the outer ear of most mammals, sound is filtered differently on its way into the ear depending on the location of its origin. This gives these animals the ability to localize sound vertically. The eardrum is an airtight membrane, and when sound waves arrive there, they cause it to vibrate following the waveform of the sound. Cerumen (ear wax) is produced by ceruminous and sebaceous glands in the skin of the human ear canal, protecting the ear canal and tympanic membrane from physical damage and microbial invasion.

Middle ear

The middle ear consists of a small air-filled chamber that is located medial to the eardrum. Within this chamber are the three smallest bones in the body, known collectively as the ossicles which include the malleus, incus, and stapes (also known as the hammer, anvil, and stirrup, respectively). They aid in the transmission of the vibrations from the eardrum into the inner ear, the cochlea. The purpose of the middle ear

ossicles is to overcome the impedance mismatch between air waves and cochlear waves, by providing impedance matching.

Also located in the middle ear are the stapedius muscle and tensor tympani muscle, which protect the hearing mechanism through a stiffening reflex. The stapes transmits sound waves to the inner ear through the oval window, a flexible membrane separating the air-filled middle ear from the fluid-filled inner ear. The round window, another flexible membrane, allows for the smooth displacement of the inner ear fluid caused by the entering sound waves.

Inner ear

The inner ear consists of the cochlea, which is a spiral-shaped, fluid-filled tube. It is divided lengthwise by the organ of Corti, which is the main organ of mechanical to neural transduction. Inside the organ of Corti is the basilar membrane, a structure that vibrates when waves from the middle ear propagate through the cochlear fluid – endolymph. The basilar membrane is tonotopic, so that each frequency has a characteristic place of resonance along it. Characteristic frequencies are high at the basal entrance to the cochlea, and low at the apex. Basilar membrane motion causes depolarization of the hair cells, specialized auditory receptors located within the organ of Corti. While the hair cells do not produce action potentials themselves, they release neurotransmitter at synapses with the fibers of the auditory nerve, which does produce action potentials. In this way, the patterns of oscillations on the basilar membrane are converted to spatiotemporal patterns of firings which transmit information about the sound to the brainstem.

Neuronal

The sound information from the cochlea travels via the auditory nerve to the cochlear nucleus in the brainstem. From there, the signals are projected to the inferior colliculus in the midbraintectum. The inferior colliculus integrates auditory input with limited input from other parts of the brain and is involved in subconscious reflexes such as the auditory startle response.

The inferior colliculus in turn projects to the medial geniculate nucleus, a part of the thalamus where sound information is relayed to the primary auditory cortex in the temporal lobe. Sound is believed to first become consciously experienced at the primary auditory cortex. Around the primary auditory cortex lies Wernickes area, a cortical area involved in interpreting sounds that is necessary to understand spoken words.

Disturbances (such as stroke or trauma) at any of these levels can cause hearing problems, especially if the disturbance is bilateral. In some instances it can also lead to auditory hallucinations or more complex difficulties in perceiving sound.

Hearing tests

Hearing can be measured by behavioral tests using an audiometer. Electrophysiological tests of hearing can provide accurate measurements of hearing thresholds even in unconscious subjects. Such tests include auditory brainstem evoked potentials (ABR), otoacoustic emissions (OAE) and

electrocochleography (ECochG). Technical advances in these tests have allowed hearing screening for infants to become widespread.

Hearing can be measured by mobile applications which includes audiological hearing test function or hearing aid application.

These applications allow the user to measure hearing thresholds at different frequencies (audiogram). Despite possible errors in measurements, hearing losscan be detected.

Hearing loss

There are several different types of hearing loss: conductive hearing loss, sensorineural hearing loss and mixed types.

There are defined degrees of hearing loss:

- **Mild hearing loss** - People with mild hearing loss have difficulties keeping up with conversations, especially in noisy surroundings. The most quiet sounds that people with mild hearing loss can hear with their better ear are between 25 and 40 dB HL.
- **Moderate hearing loss** - People with moderate hearing loss have difficulty keeping up with conversations when they are not using a hearing aid. On average, the most quiet sounds heard by people with moderate hearing loss with their better ear are between 40 and 70 dB HL.
- **Severe hearing loss** - People with severe hearing loss depend on powerful hearing aid. However, they often rely on lip-reading even when they are using hearing

aids. The most quiet sounds heard by people with severe hearing loss with their better ear are between 70 and 95 dB HL.

- **Profound hearing loss** - People with profound hearing loss are very hard of hearing and they mostly rely on lip-reading and sign language. The most quiet sounds heard by people with profound hearing loss with their better ear are from 95 dB HL or more.

Causes

- Heredity
- Congenital conditions
- Presbycusis
- Acquired
- Noise-induced hearing loss
- Ototoxic drugs and chemicals
- Infection

Prevention

Hearing protection is the use of devices designed to prevent noise-induced hearing loss (NIHL), a type of post-lingual hearing impairment. The various means used to prevent hearing loss generally focus on reducing the levels of noise to which people are exposed. One way this is done is through environmental modifications such as acoustic quieting, which may be achieved with as basic a measure as lining a room with curtains, or as complex a measure as employing an anechoic chamber, which absorbs nearly all sound. Another means is the use of devices such as earplugs, which are inserted into

the ear canal to block noise, or earmuffs, objects designed to cover a person's ears entirely.

Management

The loss of hearing, when it is caused by neural loss, cannot presently be cured. Instead, its effects can be mitigated by the use of audioprosthetic devices, i.e. hearing assistive devices such as hearing aids and cochlear implants. In a clinical setting, this management is offered by otologists and audiologists.

Relation to health

Hearing loss is associated with Alzheimer's disease and dementia with a greater degree of hearing loss tied to a higher risk. There is also an association between type 2 diabetes and hearing loss.

Hearing underwater

Hearing threshold and the ability to localize sound sources are reduced underwater in humans, but not in aquatic animals, including whales, seals, and fish which have ears adapted to process water-borne sound.

In vertebrates

Not all sounds are normally audible to all animals. Each species has a range of normal hearing for both amplitude and frequency. Many animals use sound to communicate with each other, and hearing in these species is particularly important

for survival and reproduction. In species that use sound as a primary means of communication, hearing is typically most acute for the range of pitches produced in calls and speech.

Frequency range

Frequencies capable of being heard by humans are called audio or sonic. The range is typically considered to be between 20 Hz and 20,000 Hz. Frequencies higher than audio are referred to as ultrasonic, while frequencies below audio are referred to as infrasonic. Some bats use ultrasound for echolocation while in flight. Dogs are able to hear ultrasound, which is the principle of 'silent' dog whistles. Snakes sense infrasound through their jaws, and baleen whales, giraffes, dolphins and elephants use it for communication. Some fish have the ability to hear more sensitively due to a well-developed, bony connection between the ear and their swim bladder. This "aid to the deaf" for fishes appears in some species such as carp and herring.

In invertebrates

Even though they don't have ears, invertebrates have developed other structures and systems to decode vibrations traveling through the air, or "sound." Charles Henry Turner (zoologist) was the first scientist to formally show this phenomenon through rigorously controlled experiments in ants. Turner ruled out the detection of ground vibration and suggested that other insects likely have auditory systems as well.

Many insects detect sound through the way air vibrations deflect hairs along their body. Some insects have even developed specialized hairs tuned to detecting particular

frequencies, such as certain caterpillar species that have evolved hair with properties such that it resonates most with the sound of buzzing wasps, thus warning them of the presence of natural enemies.

Some insects possess a tympanal organ. These are "eardrums", that cover air filled chambers on the legs. Similar to the hearing process with vertebrates, the eardrums react to sonar waves. Receptors that are placed on the inside translate the oscillation into electric signals and send them to the brain. Several groups of flying insects that are preyed upon by echolocatingbats can perceive the ultrasound emissions this way and reflexively practice ultrasound avoidance.

Neuronal encoding of sound

The **neuronal encoding of sound** is the representation of auditorysensation and perception in the nervous system.

This article explores the basic physiological principles of sound perception, and traces hearing mechanisms from sound as pressure waves in air to the transduction of these waves into electrical impulses (action potentials) along auditory nerve fibers, and further processing in the brain.

Introduction

The complexities of contemporary neuroscienceare continually redefined. Thus what is known now of the auditory system has changed in the recent times and thus conceivably in the next two years or so, much of this will change.

This article is structured in a format that starts with a small exploration of what sound is followed by the general anatomy of the ear which in turn will finally give way to explaining the encoding mechanism of the engineering marvel that is the ear.

This article traces the route that sound waves first take from generation at an unknown source to their integration and perception by the auditory cortex.

Basic physics of sound

Sound waves are what physicists call longitudinal waves, which consist of propagating regions of high pressure (compression) and corresponding regions of low pressure (rarefaction).

Waveform

Waveform is a description of the general shape of the sound wave. Waveforms are sometimes described by the sum of sinusoids, via Fourier analysis.

Amplitude

Amplitude is the size (magnitude) of the pressure variations in a sound wave, and primarily determines the loudness with which the sound is perceived. In a sinusoidal function such as

, C represents the amplitude of the sound wave.

Frequency and wavelength

The frequency of a sound is defined as the number of repetitions of its waveform per second, and is measured in hertz; frequency is inversely proportional to wavelength (in a medium of uniform propagation velocity, such as sound in air). The wavelength of a sound is the distance between any two consecutive matching points on the waveform.

The audible frequency range for young humans is about 20 Hz to 20 kHz. Hearing of higher frequencies decreases with age, limiting to about 16 kHz for adults, and even down to 3 kHz for elders.

Anatomy of the ear

Given the simple physics of sound, the anatomy and physiology of hearing can be studied in greater detail.

Outer ear

The Outer ear consists of the pinna or auricle (visible parts including ear lobes and concha), and the auditory meatus (the passageway for sound).

The fundamental function of this part of the ear is to gather sound energy and deliver it to the eardrum. Resonances of the external ear selectively boost sound pressure with frequency in the range 2–5 kHz.

The pinna as a result of its asymmetrical structure is able to provide further cues about the elevation from which the sound

originated. The vertical asymmetry of the pinna selectively amplifies sounds of higher frequency from high elevation thereby providing spatial information by virtue of its mechanical design.

Middle ear

The middle ear plays a crucial role in the auditory process, as it essentially converts pressure variations in air to perturbations in the fluids of the inner ear. In other words, it is the mechanical transfer function that allows for efficient transfer of collected sound energy between two different media. The three small bones that are responsible for this complex process are the malleus, the incus, and the stapes, collectively known as the ear ossicles. The impedance matching is done through via lever ratios and the ratio of areas of the tympanic membrane and the footplate of the stapes, creating a transformer-like mechanism. Furthermore, the ossiclesare arranged in such a manner as to resonate at 700–800 Hz while at the same time protecting the inner ear from excessive energy. A certain degree of top-down control is present at the middle ear level primarily through two muscles present in this anatomical region: the tensor tympani and the stapedius. These two muscles can restrain the ossiclesso as to reduce the amount of energy that is transmitted into the inner ear in loud surroundings.

Inner ear

The cochlea of the inner ear, a marvel of physiological engineering, acts as both a frequency analyzer and nonlinear

acoustic amplifier. The cochlea has over 32,000 hair cells. Outer hair cells primarily provide amplification of traveling waves that are induced by sound energy, while inner hair cells detect the motion of those waves and excite the (Type I) neurons of the auditory nerve.

The basal end of the cochlea, where sounds enter from the middle ear, encodes the higher end of the audible frequency range while the apical end of the cochlea encodes the lower end of the frequency range. This tonotopy plays a crucial role in hearing, as it allows for spectral separation of sounds. A cross section of the cochlea will reveal an anatomical structure with three main chambers (scalavestibuli, scala media, and scala tympani). At the apical end of the cochlea, at an opening known as the helicotrema, the scalavestibuli merges with the scala tympani. The fluid found in these two cochlear chambers is perilymph, while scala media, or the cochlear duct, is filled with endolymph.

Transduction

Auditory hair cells

The auditory hair cells in the cochlea are at the core of the auditory system's special functionality (similar hair cells are located in the semicircular canals). Their primary function is mechanotransduction, or conversion between mechanical and neural signals. The relatively small number of the auditory hair cells is surprising when compared to other sensory cells such as the rods and cones of the visual system. Thus the loss of a lower number (in the order of thousands) of auditory hair cells can be devastating while the loss of a larger number of

retinal cells (in the order to hundreds of thousands) will not be as bad from a sensory standpoint.

Cochlear hair cells are organized as inner hair cells and outer hair cells; inner and outer refer to relative position from the axis of the cochlear spiral. The inner hair cells are the primary sensory receptors and a significant amount of the sensory input to the auditory cortex occurs from these hair cells. Outer hair cells on the other hand boost the mechanical signal by using electromechanical feedback.

Mechanotransduction

The apical surface of each cochlear hair cell contains a hair bundle. Each hair bundle contains approximately 300 fine projections known as stereocilia, formed by actin cytoskeletal elements. The stereocilia in a hair bundle are arranged in multiple rows of different heights. In addition to the stereocilia, a true ciliary structure known as the kinocilium exists and is believed to play a role in hair cell degeneration that is caused by exposure to high frequencies.

A stereocilium is able to bend at its point of attachment to the apical surface of the hair cell. The actin filaments that form the core of a stereocilium are highly interlinked and cross linked with fibrin, and are therefore stiff and inflexible at positions other than the base. When stereocilia in the tallest row are deflected in the positive-stimulus direction, the shorter rows of stereocilia are also deflected. These simultaneous deflections occur due to filaments called tip links that attach the side of each taller stereocilium to the top of the shorter stereocilium in the adjacent row. When the tallest

stereociliaare deflected, tension is produced in the tip links and causes the stereocilia in the other rows to deflect as well. At the lower end of each tip link is one or more mechano-electrical transduction (MET) channels, which are opened by tension in the tip links. These MET channels are cation-selective transduction channels that allow potassium and calcium ions to enter the hair cell from the endolymph that bathes its apical end.

The influx of cations, particularly potassium, through the open MET channels causes the membrane potential of the hair cell to depolarize.

This depolarization opens voltage-gated calcium channels to allow the further influx of calcium. This results in an increase in the calcium concentration, which triggers the exocytosis of neurotransmitter vesicles at ribbon synapses at the basolateral surface of the hair cell. The release of neurotransmitter at a ribbon synapse, in turn, generates an action potential in the connected auditory-nerve fiber. Hyperpolarization of the hair cell, which occurs when potassium leaves the cell, is also important, as it stops the influx of calcium and therefore stops the fusion of vesicles at the ribbon synapses. Thus, as elsewhere in the body, the transduction is dependent on the concentration and distribution of ions. The perilymph that is found in the scala tympani has a low potassium concentration, whereas the endolymph found in the scala media has a high potassium concentration and an electrical potential of about 80 millivolts compared to the perilymph. Mechanotransduction by stereocilia is highly sensitive and able to detect perturbations as small as fluid fluctuations of 0.3 nanometers,

and can convert this mechanical stimulation into an electrical nerve impulse in about 10 microseconds.

Nerve fibers from the cochlea

There are two types of afferent neurons found in the cochlear nerve: Type I and Type II. Each type of neuron has specific cell selectivity within the cochlea. The mechanism that determines the selectivity of each type of neuron for a specific hair cell has been proposed by two diametrically opposed theories in neuroscience known as the peripheral instruction hypothesis and the cell autonomous instruction hypothesis. The peripheral instruction hypothesis states that phenotypic differentiation between the two neurons are not made until after these undifferentiated neurons attach to hair cells which in turn will dictate the differentiation pathway. The cell autonomous instruction hypothesis states that differentiation into Type I and Type II neurons occur following the last phase of mitotic division but preceding innervations. Both types of neuron participate in the encoding of sound for transmission to the brain.

Type I neurons

Type I neurons innervate inner hair cells. There is significantly greater convergence of this type of neuron towards the basal end in comparison with the apical end. A radial fiber bundle acts as an intermediary between Type I neurons and inner hair cells. The ratio of innervation that is seen between Type I neurons and inner hair cells is 1:1 which results in high signal transmission fidelity and resolution.

Type II neurons

Type II neurons on the other hand innervate outer hair cells. However, there is significantly greater convergence of this type of neuron towards the apex end in comparison with the basal end. A 1:30-60 ratio of innervation is seen between Type II neurons and outer hair cells which in turn make these neurons ideal for electromechanical feedback. Type II neurons can be physiologically manipulated to innervate inner hair cells provided outer hair cells have been destroyed either through mechanical damage or by chemical damage induced by drugs such as gentamicin.

Brainstem and midbrain

The auditory nervous system includes many stages of information processing between the ear and cortex.

Auditory cortex

Primary auditory neurons carry action potentials from the cochlea into the transmission pathway shown in the adjacent image. Multiple relay stations act as integration and processing centers. The signals reach the first level of cortical processing at the primary auditory cortex (A1), in the superior temporal gyrus of the temporal lobe. Most areas up to and including A1 are tonotopically mapped (that is, frequencies are kept in an ordered arrangement). However, A1 participates in coding more complex and abstract aspects of auditory stimuli without coding well the frequency content, including the presence of a distinct sound or its echoes. Like lower regions,

this region of the brain has combination-sensitive neurons that have nonlinear responses to stimuli.

Recent studies conducted in bats and other mammals have revealed that the ability to process and interpret modulation in frequencies primarily occurs in the superior and middle temporal gyri of the temporal lobe. Lateralization of brain function exists in the cortex, with the processing of speech in the left cerebral hemisphere and environmental sounds in the right hemisphere of the auditory cortex.

Music, with its influence on emotions, is also processed in the right hemisphere of the auditory cortex. While the reason for such localization is not quite understood, lateralization in this instance does not imply exclusivity as both hemispheres do participate in the processing, but one hemisphere tends to play a more significant role than the other.

Recent ideas

- Alternation in encoding mechanisms have been noticed as one progresses through the auditory cortex. Encoding shifts from synchronous responses in the cochlear nucleus and later becomes dependent on rate encoding in the inferior colliculus.
- Despite advances in gene therapy that allow for the alteration of the expression of genes that affect audition, such as ATOH1, and the use of viral vectors for such end, the micro-mechanical and neuronal complexities that surrounds the inner ear hair cells, artificial regeneration in vitro remains a distant reality.

- Recent studies suggest that the auditory cortex may not be as involved in top down processing as was previous thought. In studies conducted on primates for tasks that required the discrimination of acoustic flutter, Lemus found that the auditory cortex played only a sensory role and had nothing to do with the cognition of the task at hand.
- Due to the presence of the tonotopic maps in the auditory cortex at an early age, it has been assumed that cortical reorganization had little to do with the establishment of these maps, but these maps are subject to plasticity. The cortex seems to perform a more complex processing than spectral analysis or even spectro-temporal analysis.

Prosody (linguistics)

In linguistics, **prosody** (/ˈprɒsədi,ˈprɒzədi/) is concerned with those elements of speech that are not individual phonetic segments (vowels and consonants) but are properties of syllables and larger units of speech, including linguistic functions such as intonation, stress, and rhythm. Such elements are known as **suprasegmentals**.

Prosody may reflect various features of the speaker or the utterance: the emotional state of the speaker; the form of the utterance (statement, question, or command); the presence of irony or sarcasm; emphasis, contrast, and focus. It may otherwise reflect other elements of language that may not be encoded by grammar or by choice of vocabulary.

Attributes of prosody

In the study of prosodic aspects of speech, it is usual to distinguish between auditory measures (subjective impressions produced in the mind of the listener) and objective measures (physical properties of the sound wave and physiological characteristics of articulation that may be measured objectively). Auditory (subjective) and objective (acoustic and articulatory) measures of prosody do not correspond in a linear way. Most studies of prosody have been based on auditory analysis using auditory scales.

There is no agreed number of prosodic variables. In auditory terms, the major variables are:

- the pitch of the voice (varying between low and high)
- length of sounds (varying between short and long)
- loudness, or prominence (varying between soft and loud)
- timbre or voice quality (quality of sound)

In acoustic terms, these correspond reasonably closely to:

- fundamental frequency (measured in hertz, or cycles per second)
- duration (measured in time units such as milliseconds or seconds)
- intensity, or sound pressure level (measured in decibels)
- spectral characteristics (distribution of energy at different parts of the audible frequency range)

Different combinations of these variables are exploited in the linguistic functions of intonation and stress, as well as other prosodic features such as rhythm and tempo. Additional prosodic variables have been studied, including voice quality and pausing. The behavior of the prosodic variables can be studied either as contours across the prosodic unit or by the behavior of boundaries.

Phonology

Prosodic features are said to be suprasegmental, since they are properties of units of speech larger than the individual segment (though exceptionally it may happen that a single segment may constitute a syllable, and thus even a whole utterance, e.g. "Ah!"). It is necessary to distinguish between the personal, background characteristics that belong to an individual's voice (for example, their habitual pitch range) and the independently variable prosodic features that are used contrastively to communicate meaning (for example, the use of changes in pitch to indicate the difference between statements and questions). Personal characteristics are not linguistically significant. It is not possible to say with any accuracy which aspects of prosody are found in all languages and which are specific to a particular language or dialect.

Intonation

Some writers (e.g., O'Connor and Arnold) have described intonation entirely in terms of pitch, while others (e.g., Crystal) propose that what is referred to as "intonation" is, in fact, an amalgam of several prosodic variables. The form of English intonation is often said to be based on three aspects:

- The division of speech into units
- The highlighting of particular words and syllables
- The choice of pitch movement (e.g., fall or rise)

These are sometimes known as *tonality*, *tonicity* and *tone* (and collectively as "the three T's").

An additional pitch-related variation is *pitch range*; speakers are capable of speaking with a wide range of pitch (this is usually associated with excitement), while at other times with a narrow range. English has been said to make use of changes in *key*; shifting one's intonation into the higher or lower part of one's pitch range is believed to be meaningful in certain contexts.

Stress

From the perceptual point of view, stress functions as the means of making a syllable prominent; stress may be studied in relation to individual words (named "word stress" or lexical stress) or in relation to larger units of speech (traditionally referred to as "sentence stress" but more appropriately named "prosodic stress"). Stressed syllables are made prominent by several variables, by themselves or in combination. Stress is typically associated with the following:

- pitch prominence, that is, a pitch level that is different from that of neighbouring syllables, or a pitch movement
- increased length (duration)
- increased loudness (dynamics)
- differences in timbre: in English and some other languages, stress is associated with aspects of vowel

quality (whose acoustic correlate is the formant frequencies or spectrum of the vowel). Unstressed vowels tend to be centralized relative to stressed vowels, which are normally more peripheral in quality

These cues to stress are not equally powerful. Cruttenden, for example, writes "Perceptual experiments have clearly shown that, in English at any rate, the three features (pitch, length and loudness) form a scale of importance in bringing syllables into prominence, pitch being the most efficacious, and loudness the least so".

When pitch prominence is the major factor, the resulting prominence is often called*accent* rather than stress.

There is considerable variation from language to language concerning the role of stress in identifying words or in interpreting grammar and syntax.

Tempo

Rhythm

Although rhythm is not a prosodic variable in the way that pitch or loudness are, it is usual to treat a language's characteristic rhythm as a part of its prosodic phonology. It has often been asserted that languages exhibit regularity in the timing of successive units of speech, a regularity referred to as isochrony, and that every language may be assigned one of three rhythmical types: stress-timed (where the durations of the intervals between stressed syllables is relatively constant), syllable-timed (where the durations of successive syllables are

relatively constant) and mora-timed (where the durations of successive morae are relatively constant). As explained in the isochrony article, this claim has not been supported by scientific evidence.

Pause

Voiced or unvoiced, the pause is a form of interruption to articulatory continuity such as an open or terminal juncture. Conversation analysis commonly notes pause length. Distinguishing auditory hesitation from silent pauses is one challenge. Contrasting junctures within and without word chunks can aid in identifying pauses.

There are a variety of "filled" pause types. Formulaic language pause fillers include "Like", "Er" and "Uhm", and paralinguistic expressive respiratory pauses include the sigh and gasp.

Although related to breathing, pauses may contain contrastive linguistic content, as in the periods between individual words in Englishadvertisingvoice-overcopy sometimes placed to denote high information content, e.g. "Quality. Service. Value."

Chunking

Pausing or its lack contributes to the perception of word groups, or chunks. Examples include the phrase, phraseme, constituent or interjection. Chunks commonly highlight lexical items or fixed expressionidioms. Chunking prosody is present on any complete utterance and may correspond to a syntactic category, but not necessarily. The well-known English chunk "Know what I mean?" sounds like a single word ("No-whuta-meen?") due to blurring or rushing the articulation of adjacent

word syllables, thereby changing the potential open junctures between words into closed junctures.

Cognitive aspects

Intonation is said to have a number of perceptually significant functions in English and other languages, contributing to the recognition and comprehension of speech.

Grammar

It is believed that prosody assists listeners in parsing continuous speech and in the recognition of words, providing cues to syntactic structure, grammatical boundaries and sentence type. Boundaries between intonation units are often associated with grammatical or syntactic boundaries; these are marked by such prosodic features as pauses and slowing of tempo, as well as "pitch reset" where the speaker's pitch level returns to the level typical of the onset of a new intonation unit. In this way potential ambiguities may be resolved. For example, the sentence “They invited Bob and Bill and Al got rejected” is ambiguous when written, although addition of a written comma after either "Bob" or "Bill" will remove the sentence's ambiguity. But when the sentence is read aloud, prosodic cues like pauses (dividing the sentence into chunks) and changes in intonation will reduce or remove the ambiguity. Moving the intonational boundary in cases such as the above example will tend to change the interpretation of the sentence. This result has been found in studies performed in both English and Bulgarian. Research in English word recognition has demonstrated an important role for prosody.

Focus

Intonation and stress work together to highlight important words or syllables for contrast and focus. This is sometimes referred to as the *accentual function* of prosody. A well-known example is the ambiguous sentence "I never said she stole my money", where there are seven meaning changes depending on which of the seven words is vocally highlighted.

Discourse

Prosody plays a role in the regulation of conversational interaction and in signaling discourse structure. David Brazil and his associates studied how intonation can indicate whether information is new or already established; whether a speaker is dominant or not in a conversation; and when a speaker is inviting the listener to make a contribution to the conversation.

Emotion

Prosody is also important in signalling emotions and attitudes. When this is involuntary (as when the voice is affected by anxiety or fear), the prosodic information is not linguistically significant. However, when the speaker varies her speech intentionally, for example to indicate sarcasm, this usually involves the use of prosodic features. The most useful prosodic feature in detecting sarcasm is a reduction in the mean fundamental frequency relative to other speech for humor, neutrality, or sincerity. While prosodic cues are important in indicating sarcasm, context clues and shared knowledge are also important.

Emotional prosody was considered by Charles Darwin in *The Descent of Man* to predate the evolution of human language: "Even monkeys express strong feelings in different tones – anger and impatience by low, – fear and pain by high notes." Native speakers listening to actors reading emotionally neutral text while projecting emotions correctly recognized happiness 62% of the time, anger 95%, surprise 91%, sadness 81%, and neutral tone 76%. When a database of this speech was processed by computer, segmental features allowed better than 90% recognition of happiness and anger, while suprasegmental prosodic features allowed only 44%–49% recognition. The reverse was true for surprise, which was recognized only 69% of the time by segmental features and 96% of the time by suprasegmental prosody. In typical conversation (no actor voice involved), the recognition of emotion may be quite low, of the order of 50%, hampering the complex interrelationship function of speech advocated by some authors. However, even if emotional expression through prosody cannot always be consciously recognized, tone of voice may continue to have subconscious effects in conversation. This sort of expression stems not from linguistic or semantic effects, and can thus be isolated from traditional linguistic content. Aptitude of the average person to decode conversational implicature of emotional prosody has been found to be slightly less accurate than traditional facial expression discrimination ability; however, specific ability to decode varies by emotion. These emotional have been determined to be ubiquitous across cultures, as they are utilized and understood across cultures. Various emotions, and their general experimental identification rates, are as follows:

- Anger and sadness: High rate of accurate identification

- Fear and happiness: Medium rate of accurate identification
- Disgust: Poor rate of accurate identification

The prosody of an utterance is used by listeners to guide decisions about the emotional affect of the situation. Whether a person decodes the prosody as positive, negative, or neutral plays a role in the way a person decodes a facial expression accompanying an utterance. As the facial expression becomes closer to neutral, the prosodic interpretation influences the interpretation of the facial expression. A study by Marc D. Pell revealed that 600 ms of prosodic information is necessary for listeners to be able to identify the affective tone of the utterance. At lengths below this, there was not enough information for listeners to process the emotional context of the utterance.

Child language

Unique prosodic features have been noted in infant-directed speech (IDS) - also known as baby talk, child-directed speech (CDS), or "motherese". Adults, especially caregivers, speaking to young children tend to imitate childlike speech by using higher and more variable pitch, as well as an exaggerated stress. These prosodic characteristics are thought to assist children in acquiring phonemes, segmenting words, and recognizing phrasal boundaries. And though there is no evidence to indicate that infant-directed speech is necessary for language acquisition, these specific prosodic features have been observed in many different languages.

Aprosodia

An aprosodia is an acquired or developmental impairment in comprehending or generating the emotion conveyed in spoken language. Aprosodyis often accompanied by the inability to properly utilize variations in speech, particularly with deficits in ability to accurately modulate pitch, loudness, intonation, and rhythm of word formation. This is seen sometimes in persons with Asperger syndrome.

Brain regions involved

Producing these nonverbal elements requires intact motor areas of the face, mouth, tongue, and throat. This area is associated with Brodmann areas44 and 45 (Broca's area) of the left frontal lobe. Damage to areas 44/45, specifically on the right hemisphere, produces motor aprosodia, with the nonverbal elements of speech being disturbed (facial expression, tone, rhythm of voice).

Understanding these nonverbal elements requires an intact and properly functioning right-hemisphere perisylvian area, particularly Brodmann area 22 (not to be confused with the corresponding area in the *left* hemisphere, which contains Wernicke's area). Damage to the right inferior frontal gyrus causes a diminished ability to convey emotion or emphasis by voice or gesture, and damage to right superior temporal gyrus causes problems comprehending emotion or emphasis in the voice or gestures of others. The right Brodmann area 22 aids in the interpretation of prosody, and damage causes sensory aprosodia, with the patient unable to comprehend changes in voice and body language.

Chapter 7

Phonetic Transcription

Phonetic transcription (also known as **phonetic script** or **phonetic notation**) is the visual representation of speech sounds (or phones) by means of symbols. The most common type of phonetic transcription uses a phonetic alphabet, such as the International Phonetic Alphabet.

Versus orthography

The pronunciation of words in all languages changes over time. However, their written forms (orthography) are often not modified to take account of such changes, and do not accurately represent the pronunciation. Pronunciation can also vary greatly among dialects of a language. Standard orthography in some languages, such as English and Tibetan, is often irregular and makes it difficult to predict pronunciation from spelling. For example, the words *bough*, *chough*, *cough*, *though* and *through* do not rhyme in English even though their spellings might suggest otherwise. Other languages, such as Spanish and Italian have a more consistent (but still imperfect) relationship between orthography and pronunciation, while a few languages may claim to have a fully phonemic spelling system (a phonemic orthography).

For most languages, phonetic transcription makes it possible to show pronunciation with something much nearer to a one-to-one relationship between sound and symbol than is possible with the language's orthography. Phonetic transcription allows

one to step outside orthography, examine differences in pronunciation between dialects within a given language and identify changes in pronunciation that may take place over time.

A basic principle of phonetic transcription is that it should be applicable to all languages, and its symbols should denote the same phonetic properties whatever the language being transcribed. It follows that a transcription devised for one individual language or group of languages is not a phonetic transcription but an orthography.

Narrow versus broad transcription

Phonetic transcription may be used to transcribe the phonemes of a language, or it may go further and specify their precise phonetic realization. In all systems of transcription there is a distinction between **broad transcription** and **narrow transcription**. Broad transcription indicates only the most noticeable phonetic features of an utterance, whereas narrow transcription encodes more information about the phonetic characteristics of the allophones in the utterance. The difference between broad and narrow is a continuum, but the difference between phonemic and phonetic transcription is usually treated as a binary distinction. Phonemic transcription is a particular form of broad transcription which disregards all allophonic difference; as the name implies, it is not really a phonetic transcription at all (though at times it may coincide with one), but a representation of phonemic structure. A transcription which includes some allophonic detail but is closely linked to the phonemic structure of an utterance is called an **allophonic** transcription.

The advantage of the narrow transcription is that it can help learners to produce exactly the right sound, and allows linguists to make detailed analyses of language variation. The disadvantage is that a narrow transcription is rarely representative of all speakers of a language. While most Americans, Canadians and Australians would pronounce the /t/ of *little* as a tap[ɾ], many speakers in southern England would pronounce /t/ as [ʔ] (a glottal stop; t-glottalization) and/or the second /l/ as a vowel resembling [ʊ] (L-vocalization), possibly yielding [ˈlɪʔʊ].

A further disadvantage of narrow transcription is that it involves a larger number of symbols and diacritics that may be unfamiliar to non-specialists. The advantage of broad transcription is that it usually allows statements to be made which apply across a more diverse language community. It is thus more appropriate for the pronunciation data in foreign language dictionaries, which may discuss phonetic details in the preface but rarely give them for each entry. A rule of thumb in many linguistics contexts is therefore to use a narrow transcription when it is necessary for the point being made, but a broad transcription whenever possible.

Types of notational systems

Most phonetic transcription is based on the assumption that linguistic sounds are segmentable into discrete units that can be represented by symbols. Many different types of transcription, or "notation", have been tried out: these may be divided into *Alphabetic* (which are based on the same principle as that which governs ordinary alphabetic writing, namely that of using one single simple symbol to represent each sound),

and *Analphabetic* (notations which are *not* alphabetic) which represent each sound by a composite symbol made up of a number of signs put together.

Alphabetic

The International Phonetic Alphabet (IPA) is the most widely used and well-known of present-day phonetic alphabets, and has a long history. It was created in the nineteenth century by European language teachers and linguists. It soon developed beyond its original purpose as a tool of foreign language pedagogy and is now also used extensively as a practical alphabet of phoneticians and linguists. It is found in many dictionaries, where it is used to indicate the pronunciation of words, but most American dictionaries for native English-speakers, e.g., *American Heritage Dictionary of the English Language, Random House Dictionary of the English Language, Webster's Third New International Dictionary*, avoid phonetic transcription and instead employ *respelling* systems based on the English alphabet, with diacritical marks over the vowels and stress marks. (See Pronunciation respelling for English for a generic version.)

Another commonly encountered alphabetic tradition was originally created by American linguists for the transcription of Native American and European languages, and is still commonly used by linguists of Slavic, Indic, Semitic, Uralic (here known as the Uralic Phonetic Alphabet) and Caucasian languages. This is often labeled the Americanist phonetic alphabet despite having been widely used for languages outside the Americas. The principal difference between these alphabets and the IPA is that the specially created characters of the IPA

are abandoned in favour of already existing typewriter characters with diacritics (e.g. many characters are borrowed from Eastern European orthographies) or digraphs. Examples of this transcription may be seen in Pike's *Phonemics* and in many of the papers reprinted in Joos's*Readings in Linguistics 1*. In the days before it was possible to create phonetic fonts for computer printers and computerized typesetting, this system allowed material to be typed on existing typewriters to create printable material.

There are also extended versions of the IPA, for example: Ext-IPA, VoQS, and Luciano Canepari's*IPA*.

Aspects of alphabetic transcription

The International Phonetic Association recommends that a phonetic transcription should be enclosed in square brackets "[ ]". A transcription that specifically denotes only phonological contrasts may be enclosed in slashes "/ /" instead. If one is unsure, it is best to use brackets since by setting off a transcription with slashes, one makes a theoretical claim that every symbol phonemically contrasts for the language being transcribed.

For phonetic transcriptions, there is flexibility in how closely sounds may be transcribed. A transcription that gives only a basic idea of the sounds of a language in the broadest terms is called a *broad transcription*; in some cases, it may be equivalent to a phonemic transcription (only without any theoretical claims). A close transcription, indicating precise details of the sounds, is called a *narrow transcription*. They are

not binary choices but the ends of a continuum, with many possibilities in between. All are enclosed in brackets.

For example, in some dialects the English word *pretzel* in a narrow transcription would be [ˈpɻ̊ʷɛʔts.ɬ], which notes several phonetic features that may not be evident even to a native speaker. An example of a broad transcription is [ˈpɻ̊ɛts.ɬ], which indicates only some of the features that are easier to hear. A yet broader transcription would be [ˈpɹɛts.l] in which every symbol represents an unambiguous speech sound but without going into any unnecessary detail. None of those transcriptions makes any claims about the phonemic status of the sounds. Instead, they represent certain ways in which it is possible to produce the sounds that make up the word.

There are also several possibilities in how to transcribe the word phonemically, but here, the differences are generally of not precision but analysis. For example, *pretzel* could be /ˈprɛts.l̩/ or /ˈprɛts.əl/. The latter transcription suggests that there are two vowels in the word even if they cannot both be heard, but the former suggests that there is only one.

Strictly speaking, it is not possible to have a distinction between "broad" and "narrow" within phonemic transcription, since the symbols chosen represent only sounds that have been shown to be distinctive. However, the symbols themselves may be more or less explicit about their phonetic realization. A frequently cited example is the symbol chosen for the English consonant at the beginning of the words 'rue', 'rye', 'red': this is frequently transcribed as /r/, despite the symbol suggesting an association with the IPA symbol [r] which is used for a tongue-tip trill. It is equally possible within a phonemic

transcription to use the symbol /ɹ/, which in IPA usage refers to an alveolar approximant; this is the more common realization for English pronunciation in America and England. Phonemic symbols will frequently be chosen to avoid diacritics as much as possible, under a 'one sound one symbol' policy, or may even be restricted to the ASCII symbols of a typical keyboard, as in the SAMPA alphabet. For example, the English word *church*may be transcribed as /tʃɝːtʃ/, a close approximation of its actual pronunciation, or more abstractly as /crc/, which is easier to type. Phonemic symbols should always be backed up by an explanation of their use and meaning, especially when they are as divergent from actual pronunciation as /crc/.

Occasionally a transcription will be enclosed in pipes ("| |"). This goes beyond phonology into morphological analysis. For example, the words *pets* and *beds* could be transcribed phonetically as [pʰɛʔts] and [bˌɛdˌz̬] (in a fairly narrow transcription), and phonemically as /pɛts/ and /bɛdz/. Because /s/ and /z/ are separate phonemes in English, they receive separate symbols in the phonemic analysis. However, a native English speaker would recognize that underneath this, they represent the same plural ending. This can be indicated with the pipe notation. If the plural ending is thought to be essentially an *s*, as English spelling would suggest, the words can be transcribed |pɛts| and |bɛds|. If it is essentially a *z*, these would be |pɛtz| and |bɛdz|.

To avoid confusion with IPA symbols, it may be desirable to specify when native orthography is being used, so that, for example, the English word *jet* is not read as "yet". This is done with angle brackets or *chevrons*: ⟨jet⟩. It is also common to

italicize such words, but the chevrons indicate specifically that they are in the original language's orthography, and not in English transliteration.

Iconic

In *iconic* phonetic notation, the shapes of the phonetic characters are designed so that they visually represent the position of articulators in the vocal tract. This is unlike alphabetic notation, where the correspondence between character shape and articulator position is arbitrary. This notation is potentially more flexible than alphabetic notation in showing more shades of pronunciation (MacMahon 1996:838–841). An example of iconic phonetic notation is the Visible Speech system, created by Scottish phonetician Alexander Melville Bell (Ellis 1869:15).

Analphabetic

Another type of phonetic notation that is more precise than alphabetic notation is *analphabetic* phonetic notation. Instead of both the alphabetic and iconic notational types' general principle of using one symbol per sound, analphabetic notation uses long sequences of symbols to precisely describe the component features of an articulatory gesture (MacMahon1996:842–844).

This type of notation is reminiscent of the notation used in chemical formulas to denote the composition of chemical compounds. Although more descriptive than alphabetic notation, analphabetic notation is less practical for many purposes (e.g. for descriptive linguists doing fieldwork or for

speech pathologists impressionistically transcribing speech disorders). As a result, this type of notation is uncommon.

Two examples of this type were developed by the Danish Otto Jespersen (1889) and American Kenneth Pike (1943). Pike's system, which is part of a larger goal of scientific description of phonetics, is particularly interesting in its challenge against the descriptive method of the phoneticians who created alphabetic systems like the IPA. An example of Pike's system can be demonstrated by the following. A syllabicvoicedalveolar nasal consonant (/n̩/ in IPA) is notated as

- *M*aIlDe*C*VoeIpvnnAP*p*a*a*t*d*tl*t*n*r*ansnsfS*p*v*a*v*d*tlv*t*n*r*anss*s*f T*p*g*a*g*d*tlwv*t*itv*r*ansn*s*fSrp*F*Ss

In Pike's notation there are 5 main components (which are indicated using the example above):

- *M* – manner of production (i.e., *M*aIlDe)
- *C* – manner of controlling (i.e., *C*VoeIpvnn)
- description of stricture (i.e., AP*p*a*a*t*d*tl*t*n*r*ansnsfS*p*v*a*v*d*tlv*t*n*r*anss*s*fT*p*g*a*g*d*tlwv*t*itv*r*an sn*s*f)
- *S* – segment type (i.e., *S*rp)
- *F* – phonetic function (i.e., *F*Ss)

Chapter 8

Phonologies of the World's Languages

Abkhaz phonology

Abkhaz is a language of the Northwest Caucasian family which, like the other Northwest Caucasian languages, is very rich in consonants. Abkhaz has a large consonantal inventory that contrasts 58 consonants in the literary Abzhywa dialect, coupled with just two phonemic vowels (Chirikba 2003:18–20).

Abkhaz has three major dialects: Abzhywa, Bzyp and Sadz, which differ mainly in phonology, with the lexical differences being due to contact with neighbouring languages.

Phonemes preceded by an **asterisk** (*) are found in the Bzyp and Sadz dialects of Abkhaz, but not in Abzhywa; phonemes preceded by a **dagger** (†) are unique to the Bzyp dialect. The total number of consonant phonemes in Abkhaz is, therefore, 58 in the Abzhywa dialect, 60 in the Sadz dialect, and 67 in Bzyp.

The obstruents are characterised by a three-fold contrast between voiced, aspirated voiceless and glottalised forms; both the aspirated and glottalised forms are not strong, unless they are being emphasised by the speaker. The glottal stop may be analysed as a separate phoneme by some, since it can be distinguish certain pairs as áaj 'yes', and ʔaj 'no', and it can

also be an allophonic variant of [q’] in intervocalic positions. Some speakers also pronounce the word /aˈp’a/ with a [f’], but it is not encountered anywhere else.

The consonants highlighted in red are the 4 kinds of labialisation found in Abkhaz. For this reason most Abkhaz linguists prefer using ʷ to represent them in general instead of the standard IPA symbol. The [w]-type is found with the velar stops and uvular stops and fricatives. The labial-palatal rounding involves the alveolar, pharyngeal and palatal fricatives. The one found in the dental-alveolar affricates and fricatives is described as an endo-labiodental articulation. The [p]-type is found in the dental stops, where there is full bilabial closure.

The non-pharyngealised dorsal fricatives of Abkhaz may be realised as either velar or uvular depending upon the context in which they are found; here, they have been ranged with the uvulars. Also, while the labialised palatal approximant/ɥ/ is here placed with the approximants, it is actually the reflex of a labialisedvoiced pharyngeal fricative, preserved in Abaza, and a legacy of this phoneme's origin is a slight constriction of the pharynx for some speakers, resulting in the phonetic realisation[ɥˤ].

Vowels

Abkhaz has only two distinctive vowels: an open vowel /a/ and a close vowel /ɨ ~ ə/. These basic vowels have a wide range of allophones in different consonantal environments, with allophones [e] and [i] respectively next to palatals, [o] and [u] next to labials, and [ø] and [y] next to labiopalatals. /a/ also

has a long variant /aː/, which is the reflex of old sequences of */ʕa/ or */aʕ/, preserved in Abaza.

Dialects

The Sadz dialect has distinctive consonant gemination; for example, Sadz Abkhaz contrasts /a.χʷa/ ('ashes') vs. /a.χʷːa/ ('worm'), where Abzhywa and Bzyp Abkhaz have only the one form /a.χʷa/ for both; it seems that many Sadz singletons reflect positions where a consonant has been dropped from the beginning of a cluster in the Proto-Northwest Caucasian form (compare Ubykh/tχʷa/ 'ashes').

Some scholars (for instance, Chirikba 2003) prefer to count the Sadz consonant inventory at well over 100 (thus forming the largest consonant inventory in the Caucasus, outstripping Ubykh's 80–84) by treating the geminated consonants as a set in their own right. (Note, however, that this practice is not usual in counting the consonant inventory of a language.)

The Bzyp consonant inventory appears to have been the fundamental inventory of Proto-Abkhaz, with the inventories of Abzhywa and Sadzbeing reduced from this total, rather than the Bzyp series being innovative. Plain alveolopalatal affricates and fricatives have merged with their corresponding alveolars in Abzhywa and Sadz Abkhaz (compare Bzyp/a.t͡ɕ’a.ra/ 'to know' vs. Abzhywa/a.t͡s’a.ra/), and in Abzhywa the labialisedalveolopalatal fricatives have merged with the corresponding postalveolars (compare Bzyp/a.ɕʷa.ra/ 'to measure' vs. Abzhywa/a.ʃʷa.ra/).

Acehnese phonology

Acehnese, the language spoken by the Acehnese people of Aceh, Indonesia, has a large vowel inventory, with ten oral monophthong vowels, twelve oral diphthongs, seven nasal monophthong vowels, and five nasal diphthongs.

Vowels

Native-speaking linguists divide vowels in Acehnese into several categories: oral monophthongs, oral diphthongs (which are further divided into the ones ending with /ə/ and with /i/), nasal monophthongs, and nasal diphthongs.

Adyghe phonology

- Adyghe is a language of the Northwest Caucasian family which, like the other Northwest Caucasian languages, is very rich in consonants, featuring many labialized and ejective consonants. Adyghe is phonologically more complex than Kabardian, having the retroflex consonants and their labialized forms.

Stress

Stress in Adyghe is phonemic, that it is unpredictable. The lexical stress tends to fall on one of two last syllables of the word stem. Longer words can also have multiple stress patterns, as in below:

Orthography: чэлэцъикор

Stress 1: чэлэ**цъи**кор

Stress 2: чэлэцъи**кор**

Stress 3: **чэ**лэцъикор

Stress 4: **чэлэцъи**кор

Stress 5: **чэ**лэцъи**кор**

Blue: Primary stress

Green: Secondary stress

However, the functional load of stress is extremely low, but yet there are pairs that differ optionally.

Afrikaans phonology

Afrikaans has a similar phonology to other West Germanic languages, especially Dutch.

The phonetic quality of the close vowels

- /y/ tends to be merged with /i/ into [i].
- /u/ is weakly rounded and could be more narrowly transcribed as [u̜] or [ɯ̟]. Thus, it is sometimes transcribed /ɯ/.

The phonetic quality of the mid vowels

- /ɛ, ɛː, ɔ, ɔː/ vary between mid [ɛ̝, ɛ̝ː, ɔ̝, ɔ̝ː]or close-mid [e, eː, o, oː].

- According to some scholars, the stressed allophone of /ə/ is actually closer than mid ([ɨ]). However, other scholars do not distinguish between stressed and unstressed schwas. This article uses the symbol [ə] regardless of the exact height of the vowel.
- The central /ə, əː/, not the front /ɛ, ɛː/ are the unrounded counterparts of /œ, œː/. Phonetically, /ə, əː, œ, œː/ have been variously described as mid [ə, əː, ɘ, ɘː] and open-mid [ɜ, ɜː, ɞ, ɞː].
- /œ, œː/ are rather weakly rounded, and many speakers merge /œ/ with /ə/ into [ə], even in formal speech. The merger has been noted in colloquial speech since the 1920s.

The phonetic quality of the open vowels

- In some words such as *vanaand*/faˈnɑːnt/ 'this evening; tonight', unstressed ⟨a⟩ is actually a schwa [ə], not [a].
- /a/ is open near-front[a̟], but older sources describe it as near-open central [ɐ] and open central [ä].
- /ɑː/ is either open near-back [ɑ̟ː] or open back [ɑː]. Especially in stressed positions, the back realization may be rounded [ɒː], and sometimes it may be even as high as the /ɔː/ phoneme. The rounded realization is associated with younger white speakers, especially female speakers of northern accents.

Other notes

- As phonemes, /iː/ and /uː/ occur only in the words *spieël*/spiːl/ 'mirror' and *koeël*/kuːl/ 'bullet', which

used to be pronounced with sequences /i.ə/ and /u.ə/ respectively. In other cases, [iː] and [uː] occur as allophones of /i/ and /u/ respectively before /r/.

- Close vowels are phonetically long before /r/.
- /ɛ/ contrasts with /ɛː/ only in the minimal pair *pers*/pɛrs/ 'press' – *pêrs*/pɛːrs/ 'purple'.
- Before the sequences /rt, rd, rs/, the /ɛ–ɛː/ and /ɔ–ɔː/ contrasts are neutralized in favour of the long variants /ɛː/ and /ɔː/, respectively.
- /əː/ occurs only in the word *wîe* 'wedges', which is realized as either [ˈvəːə] or [ˈvəːɦə] (with a weak [ɦ]).
- The orthographic sequence ⟨ûe⟩ is realised as either [œː.ə] or [œː.ɦə] (with a weak [ɦ]).
- /œː, ɔː/ occur only in a few words.
- As a phoneme, /æ/ occurs only in some loanwords from English, such as *pêl*/pæːl/ 'pal', or as a dialectal allophone of /ɛ/ before /k, χ, l, r/, most commonly in the former Transvaal and Free State provinces.
- /a/has been variously transcribed with ⟨a⟩, ⟨ɐ⟩ and ⟨ɑ⟩. This article uses ⟨a⟩.
- /ɑː/has been variously transcribed with ⟨ɑː⟩ and ⟨aː⟩. This article uses the former symbol.
- In some words, such as *hamer*, short /a/ is in free variation with long /ɑː/ despite the fact that the spelling suggests the latter. In some words, such as *laat* (vb. 'let'), the pronunciation with short /a/ occurs only in colloquial language, to distinguish from homophones (*laat,* adj. 'late'). In some other words, such as *aan* 'on', the pronunciation with short /a/ is already a part of the standard language. The shortening of /ɑː/has been noted as early as 1927.

- The orthographic sequence ⟨ae⟩can be pronounced as either [ɑː] or [ɑːɦə] (with a weak [ɦ]).

Nasalized vowels

In some instances of the postvocalic sequence /ns/, /n/ is realized as nasalisation (and lengthening, if the vowel is short) of the preceding monophthong, which is stronger in some speakers than others, but there also are speakers retaining [n] as well as the original length of the preceding vowel.

- The sequence /ans/ in words such as *dans* is realised as [ãːs]. In monosyllabic words, that is the norm.
- The sequence /ɑːns/ in more common words (such as *Afrikaans*) is realized as either [ɑ̃ːs] or [ɑːns]. In less common words (such as *Italiaans*, meaning *Italian*), [ɑːns] is the usual pronunciation.
- The sequence /ɛns/ in words such as *mens* (meaning "human") is realized as [ɛ̃ːs].
- The sequence /œns/ in words such as *guns* (meaning "favour") is realised more often as [œns] than as [œ̃ːs]. For speakers with the /œ–ə/ merger, these transcriptions are to be read as [əns] and [ə̃ːs], respectively.
- The sequence /ɔns/ in words such as *spons* is realised as [ɔ̃ːs].

Collins &Mees (2003) analyze the pre-/s/ sequences /an, ɛn, ɔn/ as *phonemic* short vowels /ɑ̃, ɛ̃, ɔ̃/ and note that this process of nasalising the vowel and deleting the nasal occurs in many dialects of Dutch as well, such as The Hague dialect.

/ɪø, ɪə, ʊə/

- According to Lass (1987), the first elements of [ʏə, ɪə, ʊə] are close-mid, more narrowly transcribed [ë, ë, ö] or [ɪ̝, ɪ̝, ʊ̝]. According to De Villiers (1976), the onsets of [ɪə, ʊə] are near-close [ɪ, ʊ]. For simplicity, both variants will be written simply as [ɪø, ɪə, ʊə]. [ɪ, ʊ] are commonly used for centralized close-mid vowels anyway - see near-close near-front unrounded vowel and near-close near-back rounded vowel.
- Some sources prescribe monophthongal[øː, eː, oː] realizations of these; that is at least partially outdated:
- There is not a complete agreement about the realisation of /ɪø/:
- According to Lass (1987), it is realised as either rising [ɪ̯ø] or falling [ɪø̯], with the former being more common. The unrounded onset is a rather recent development and is not described by older sources. The monophthongalrealisation[øː] is virtually nonexistent.
- According to Donaldson (1993), it is realised as [øə]. Its onset is sometimes unrounded, which can cause it to merge with /eə/.
- There is not a complete agreement about the realisation of /ɪə, ʊə/
- According to Lass (1987), they may be realised in four ways:
- Falling diphthongs. Their first element may be short [ɪə̯, ʊə̯] or somewhat lengthened [ɪːə̯, ʊːə̯].
- Rising diphthongs [ɪ̯ə, ʊ̯ə]. These variants do not seem to appear word-finally. The sequence /ɦʊə/ is commonly realised as [ɦʊ̯ə] or, more often, [ɦʊ̯ə̤], with /ɦ/realised as breathy voice on the diphthong.

- Phonetically disyllabic sequences of two short monophthongs [ɪ.ə, ʊ.ə], which may occur in all environments.
- Monophthongs, either short [ɪ, ʊ] or somewhat lengthened [ɪˑ, ʊˑ]. The monophthongalrealisations occur in less stressed words as well as in stressed syllables in words that have more than one syllable. In the latter case, they are in free variation with all of the three diphthongal realisations. In case of /ʊə/, the monophthongal[ʊ] also appears in unstressed word-final syllables.
- According to Donaldson (1993), they are realized as either [eə, oə] or [iə, uə].
- /ɪə/ also occurs in words spelled with ⟨eë⟩, like *reël*/ˈrɪəl/ 'rule'. Historically, these were pronounced with a disyllabic sequence /eː.ə/ and so *reël* used to be pronounced /ˈreː.əl/.
- There is not a complete agreement about the dialectal realisation of /ɪə, ʊə/ in the Boland area:
- According to Lass (1987), they are centralized close-mid monophthongs [ɪ, ʊ], which do not merge with /i/ and /u/.
- According to Donaldson (1993) and De Villiers, they are close monophthongs, long [iː, uː] according to Donaldson (1993), short [i, u] according to De Villiers.

Other diphthongs

- The scholar DaanWissing argues that /əi̯/ is not a phonetically correct transcription and that /æɛ̯/ is more accurate. In his analysis, he found that [æɛ̯]

makes for 65% of the realisations, the other 35% being monophthongal, [ə], [æ] and [ɛ].

- Most often, /œɪ̯/ has an unrounded offset. For some speakers, the onset is also unrounded. That can cause /œɪ̯/ to merge with /əɪ̯/, which is considered non-standard.
- /ɔɪ̯, aɪ̯/ occur mainly in loanwords.
- Older sources describe /œu/ as a narrow back diphthong [ou]. However, newer sources describe its onset as more front. For example, Lass (1984), states that the onset of /œu/ is central [ɵu].
- In some words which, in English, are pronounced with /əʊ̯/, the Afrikaans equivalent tends to be pronounced with /œʊ̯/, rather than /ʊə/. That happens because Afrikaans /œʊ̯/ is more similar to the usual South African realization of English /əʊ̯/.

Long diphthongs

The long diphthongs (or 'double vowels') are phonemically sequences of a free vowel and a non-syllabic equivalent of /i/ or /u/: [iu, ui, oːi, eu, ɑːi]. Both [iu] and [eu] tend to be pronounced as [iu], but they are spelled differently: the former as ⟨ieu⟩, the latter as ⟨eeu⟩.

'False' diphthongs

In diminutives ending in /ki/ formed to monosyllabic nouns, the vowels /u, ɪə, ʊə, ɛ, ə, œ, ɔ, a, ɑː/ are realised as closing diphthongs [ui, ei, oi, ɛi, əi, œi, ɔi, ai, ɑːi].

In the same environment, the sequences /ɛn, ən, œn, ɔn, an/ are realized as [ɛiɲ, əiɲ, œiɲ, ɔiɲ, aiɲ], i.e. as closing diphthongs followed by palatal nasal.

- The suffixes ⟨-aad⟩ and ⟨-aat⟩ (phonemically /ɑːd/ and /ɑːt/, respectively) and the diminutive suffix /ki/ are realised as [ɑːki] (with a monophthong), rather than [ɑːiki].
- In practice, the diphthong [əi] is realised the same as the phonemic diphthong /əi/.
- [œi], when it has arisen from diphthongisation of [œ], differs from the phonemic diphthong /œi/ by having a slightly different onset, although the exact nature of that difference is unclear. This means that *puntjie* 'point' sounds somewhat different than*puintjie* 'rubble'.

Obstruents

- All obstruents at the ends of words are devoiced so that, for instance, a final /d/ is realised as [t].
- /p, b/ are bilabial, whereas /f, v/ are labiodental.
- According to some authors, /v/ is actually an approximant [ʋ].
- /p, t, k, tʃ/ are unaspirated.
- /k/ may be somewhat more front before front vowels; the fronted allophone of /k/ also occurs in diminutives ending in *-djie* and *-tjie*.
- /dʒ, z/ occur only in loanwords.
- /χ/ is most often uvular, either a fricative, [χ] or a voiceless trill [ʀ̥], the latter especially in initial position before a stressed vowel. The uvular fricative is also used by many speakers of white South African English

as a realisation of the marginal English phoneme /x/. In Afrikaans, velar [x] may be used in a few "hyper-posh" varieties, and it may also, rarely, occur as an allophone before front vowels in speakers with otherwise uvular [χ].

- /ɡ/ occurs mostly in loanwords, but also occurs as an allophone of /χ/ at the end of an inflected root where G is preceded by a short vowel and /r/ and succeeded by a schwa such as in *berg(e)* ('mountain', /bæːrχ, ˈbæ(ː)rɡə/).
- /w/ occurs frequently as an allophone of /v/ after other obstruents, such as in *kwaad* ('angry').

Sonorants

- /m/ is bilabial.
- /n/ merges with /m/ before labial consonants. Phonetically, this merged consonant is realized as bilabial [m] before /p, b/, and labiodental [ɱ] before /f, v/.
- /n/ merges with /ŋ/ before dorsals (/k, χ/).
- /l/ is velarised[ɬ] in all positions, especially noticeably non-prevocalically.
- /r/ is usually an alveolar trill [r] or tap [ɾ]. In some parts of the former Cape Province, it is realiseduvularly, either as a trill [ʀ] or a fricative [ʁ]. The uvular trill may also be pronounced as a tap [ʀ̆].

American Sign Language phonology

Sign languages such as American Sign Language (ASL) are characterized by phonological processes analogous to, yet

dissimilar from, those of oral languages. Although there is a qualitative difference from oral languages in that sign-language phonemesare not based on sound, and are spatial in addition to being temporal, they fulfill the same role as phonemes in oral languages.

Basically, three types of signs are distinguished: one-handed signs, symmetric two-handed signs (i.e. signs in which both hands are active and perform the same or a similar action), and asymmetric two-handed signs (i.e. signs in which one hand is active [the 'dominant' or 'strong' hand] and one hand is held static [the 'non-dominant' or 'weak' hand]). The non-dominant hand in asymmetric signs often functions as the location of the sign. Almost all simple signs in ASL are monosyllabic.

Phonemes and features

Signs consist of units smaller than the sign. These are often subdivided into *parameters*: handshapes with a particular orientation, that may perform some type of movement, in a particular location on the body or in the "signing space", and non-manual signals. These may include movement of the eyebrows, the cheeks, the nose,

the head, the torso, and the eyes. Parameter values are often equalled to spoken language phonemes, although sign language phonemes allow more simultaneity in their realization than phonemes in spoken languages. Phonemes in signed languages, as in oral languages, consist of features. For instance, the /B/ and /G/ handshapes are distinguished by the number of selected fingers: [all] versus [one].

Most phonological research focuses on the handshape. A problem in most studies of handshape is the fact that often elements of a manual alphabet are borrowed into signs, although not all of these elements are part of the sign language's phoneme inventory (Battison 1978). Also, allophones are sometimes considered separate phonemes. The first inventory of ASL handshapes contained 19 phonemes (or *cheremes*, Stokoe, 1960). Later phonological models focus on handshape features rather than on handshapes (Liddell & Johnson 1984, Sandler 1989, Hulst, 1993, Brentari 1998, Van der Kooij 2002).

In some phonological models, movement is a phonological prime (Liddell & Johnson 1984, Perlmutter 1992, Brentari 1998). Other models consider movement as redundant, as it is predictable from the locations, hand orientations and handshape features at the start and end of a sign (Hulst, 1993, Van der Kooij, 2002). Models in which movement is a prime usually distinguish *path* movement (i.e. movement of the hand[s] through space) and *internal* movement (i.e. an opening or closing movement of the hand, a hand rotation, or finger wiggling).

Allophony and assimilation

Each phoneme may have multiple allophones, i.e. different realizations of the same phoneme. For example, in the /B/ handshape, the bending of the selected fingers may vary from straight to bent at the lowest joint, and the position of the thumb may vary from stretched at the side of the hand to folded in the palm of the hand. Allophony may be free, but is also often conditioned by the context of the phoneme. Thus,

the /B/ handshape will be flexed in a sign in which the fingertips touch the body, and the thumb will be folded in the palm in signs where the radial side of the hand touches the body or the other hand.

Assimilation of sign phonemes to signs in the context is a common process in ASL. For example, the point of contact for signs like THINK, normally at the forehead, may be articulated at a lower location if the location in the following sign is below the cheek. Other assimilation processes concern the number of selected fingers in a sign, that may adapt to that of the previous or following sign. Also, has been observed that one-handed signs are articulated with two hands when followed by a two-handed signs.

Phonotactics

As yet, little is known about ASL phonotactic constraints (or those in other signed languages). The Symmetry and Dominance Conditions (Battison 1978) are sometimes assumed to be phonotactic constraints.

The Symmetry Condition requires both hands in a symmetric two-handed sign to have the same or a mirrored configuration, orientation, and movement. The Dominance Condition requires that only one hand in a twohanded sign moves if the hands do not have the same handshape specifications, *and* that the non-dominant hand has an unmarked handshape. However, since these conditions seem to apply in more and more signed languages as cross-linguistic research increases, it is doubtful whether these should be considered as specific to ASL phonotactics.

Suprasegmentals

Like most signed languages, ASL has an analogue to speaking loudly and whispering in oral language. "Loud" signs are larger and more separated, sometimes even with one-handed signs being produced with both hands. "Whispered" signs are smaller, off-center, and sometimes (partially) blocked from sight to unintended onlookers by the speaker's body or a piece of clothing. In fast signing, in particular in context, sign movements are smaller and there may be less repetition. Signs occurring at the end of a phrase may show repetition or may be held ("phrase-final lengthening").