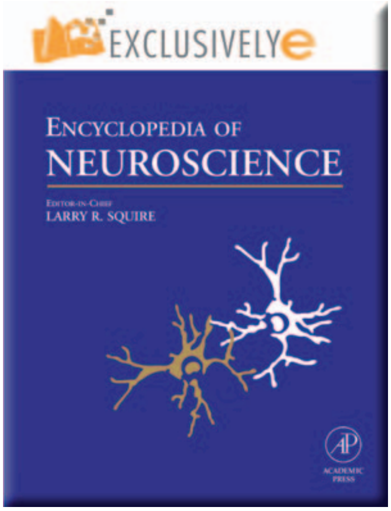

Holt L L (2009) Speech Perception: Adult. In: Squire LR (ed.) Encyclopedia of Neuroscience, volume 9, pp. 219-226. Oxford: Academic Press.

Speech Perception: Adult

L L Holt, Carnegie Mellon University, Pittsburgh, PA, USA

Introduction

As adult listeners, we comprehend the messages conveyed by spoken language with little effort. However, the ease of everyday conversation masks the perceptual and cognitive complexities involved in perceiving speech. Upon examination of the acoustic speech signal, it becomes immediately clear that this everyday skill is a demanding perceptual task.

A spoken syllable may persist in the world for a mere tenth of a second. Yet, as adult listeners, we are able to gather a great deal of information from these fleeting acoustic signals. We may apprehend the physical location of the speaker, the speaker's gender, regional dialect, age, emotional state, or identity. These spatial and indexical factors are conveyed by the acoustic speech signal in parallel with the linguistic message of the speaker. Although the spatial and indexical characteristics of the spoken message are of much interest in their own right, speech perception most commonly refers to an understanding of how the physical patterns of acoustic energy evoke perception of a linguistic message. Thus, the study of speech perception concerns the mapping between a time-varying acoustic signal received from a talker and linguistic meaning.

Units of Speech Perception

Some accounts of speech perception have suggested the possibility that adult listeners map the acoustic signal directly onto lexical entries, the words of the language. However, most models of language comprehension assume that the acoustic speech signal is translated into an intermediate unit that is used to access the words forming the lexicon. There has been much debate regarding the best characterization of these intermediate units and, for the most part, the possibilities have related closely to the natural classes described by linguistic theory, whether distinctive features, phonemes, or syllables. Despite these theoretical distinctions, the vast majority of research in speech perception has focused on the mapping from acoustics to phoneme or phonetic category as in the distinction between /b/ and /p/ in 'bin' versus 'pin.' It is likely that the concepts understood from this research apply similarly to other possible basic units. At all levels thus far investigated in-depth, research has demonstrated that the mapping from acoustics is very complex.

Context in Encoding

The complexities of speech perception begin with the nature of the acoustic signal. A great deal of early research documented the considerable variability inherent in the acoustics of speech. To summarize this broad literature very briefly, there do not appear to be acoustic signatures that uniquely specify phonetic categories, syllables, or words. Thus, listeners are faced with the perceptual challenge of mapping highly variable acoustic signals onto speech units in a many-to-one manner.

The variability inherent in the acoustics of speech arises from multiple sources. One notable source of acoustic variability is due to demands on speakers. Part of the efficiency of spoken language is the fluent nature with which it may be expressed. Speakers produce 10–15 consonants or vowels per second at typical rates of speech, and they may even double this rate in hurried speech. The rapid nature of speech production affects the acoustic signal that reaches the ear of a listener. In fluent speech, speakers articulate sounds such that movements of the vocal tract overlap in space and time across the production of different consonants and vowels. This overlap is known as coarticulation; its result on the acoustic speech signal is to cause the information for speech units to be spread across the acoustic signal. Phonetic elements of a spoken word (e.g., the three sounds in 'bin,' /bIn/) are not presented discretely as are the letters of a written word. Rather, the acoustic information for phonetic categories overlaps within the acoustic signal such that any particular time slice of an acoustic speech signal may contain information about multiple vowels and consonants. For example, the particular acoustic signature of a consonant /g/ as in 'gum' is greatly influenced by whether it is preceded by a word such as 'feel' or 'fear,' as in 'feel gum' and 'fear gum.' To describe this simply, consider the placement of the tongue body in these articulations. In articulating 'feel gum,' the tongue must move quickly from a more anterior position in the /l/ of 'feel' to the more posterior position of the /g/ in 'gum.' The resulting articulation is a compromise between where the tongue has been and where it is going. This is in part a consequence of physical constraints such as mass and inertia on the articulators, but speakers may also actively adjust articulation.

In the example, the result is that the articulators do not reach quite the same position in articulating the /g/ in 'gum' when it is preceded by 'feel' as when it is preceded by 'fear.' As a result, the acoustic signature of the /g/ in 'gum' is quite different across these two contexts. One issue for speech perception is how the perceptual system recognizes the functional equivalence of these two acoustically distinct /g/ sounds.

In addition to coarticulation, other sources also contribute to variability of the acoustic speech signal. Rate of speech, speaker idiosyncrasies, dialect, accent, emotion, gender, and room reverberance influence the acoustics of an utterance. One of the most significant challenges for understanding speech perception is that despite the radical effect these diverse sources of variability have on the acoustic signal, listeners somehow identify speakers' intended meaning. There has been considerable research investigating how listeners accomplish this feat. At the broadest level, it appears that listeners overcome the perceptual obstacles of acoustic variability by perceiving speech in a wholly context-dependent manner. The same speech signal may be perceived as belonging to different phonetic categories when it is heard in different contexts. Consider the example discussed previously. When listeners are presented with a set of speech stimuli that have been acoustically manipulated to vary perceptually along a /g/ to /d/ series as in 'gum' to 'dumb,' the endpoints of the series are perceived consistently as /g/ and /d/. The intermediate stimuli are perceptually ambiguous, being identified as both /g/ and /d/ some of the time. The boundary is steep, as is typical of identification of consonants. This pattern of perception is illustrated in Figure 1 by the green line.

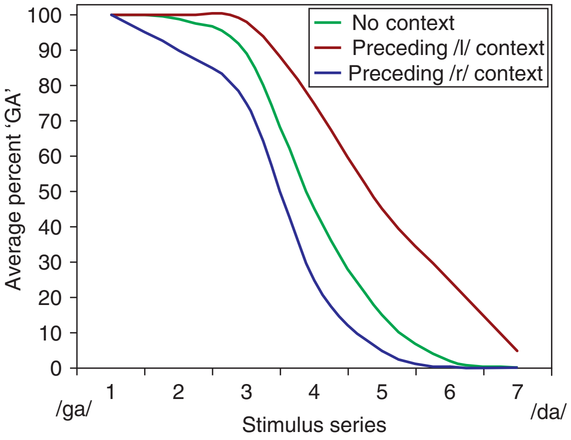

Figure 1 Listners identification of consonants varying perceptually from 'ga' to 'da' changes as a function of the speech that precedes it. A preceding 'l' shifs identification toward more 'ga' identifications whereas a preceding 'r' shifs identification of the same sounds toward 'da'. Data after Mann VA (1980) Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics* 28: 407–412.

Many experiments have demonstrated that listeners' perception of such sounds is influenced greatly by the preceding context. Preceded by /l/ as in 'feel,' listeners tend to label the sounds along the /g/ to /d/ series more often as /g/ as in 'gum' (Figure 1, red line). The same sounds are more often labeled as /d/ when /r/ is the context stimulus (Figure 1, blue line). Thus, physically identical acoustic speech stimuli are perceived very differently as a function of adjacent context.

The association between such phonetic context effects and coarticulation is intimate. Notice that the direction of these phonetic context effects in perception is opposite the coarticulatory effects on speech articulation described previously. In the utterance 'feel gum,' the /g/ is coarticulated such that it is articulated at a more anterior position than it is in isolation. This causes its acoustics to be more similar to the anterior articulation typical of /d/ in isolation; coarticulation of 'feel gum' results in a more /d/-like articulation of /g/. However, perception pulls in the opposing direction. A perceptually ambiguous /g/–/d/ speech token is pulled perceptually toward /g/ when preceded by an /l/ precursor, as in 'feel.' Thus, this contrastive perception works in opposition of coarticulation, in essence compensating for the assimilatory coarticulatory effects that draw neighboring utterances to be more similar to one another. Coarticulation causes /g/ to be acoustically more similar to /d/ in the context of a preceding /l/, whereas speech perception of consonants in the context of /l/ is shifted so that syllables are perceived more often as 'g' than 'd.' Context effects of this sort are ubiquitous in speech perception and have been documented for many speech segments. Moreover, they are observed in infants; therefore, neither experience producing coarticulated speech nor a well-developed lexicon appears to be prerequisite. Perhaps even more surprising, the influence of /l/ and /r/ on consonant perception is evident even among adult listeners for whom /l/ and /r/ are not phonologically distinct. Japanese listeners unable to accurately distinguish English /l/ from /r/ exhibit a phonetic context effect of the same magnitude as native English listeners. Listeners need not be able to classify the context syllables into distinct native phonetic categories (or even be able to hear a difference between precursors) for them to have an influence on perception of following speech.

Of interest in understanding the mechanisms that give rise to such effects, Japanese quail (*Coturnix coturnix japonica*) trained to peck a lighted key in response to presentation of /g/ endpoints of a /g/ to /d/ stimulus series pecked more vigorously to novel ambiguous midseries speech stimuli when they were preceded by /l/. A second set of birds trained to peck

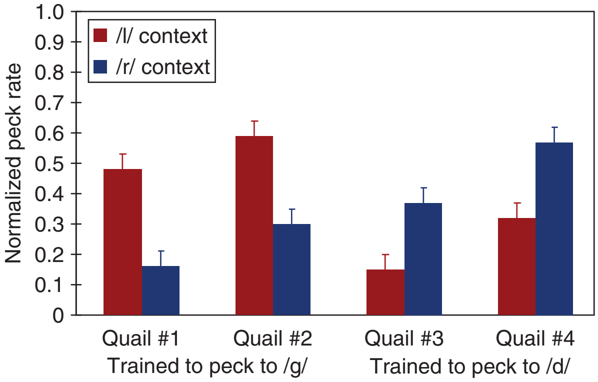

Figure 2 Ovail trained to peck to 'g' for 'd' sounds exhibit context-dependent pecking when 'l' and 'r' precede the target sounds. The influence of context is the same as that plotted for human listeners in **Figure 1**. Data from Lotto AJ, Kluender KR, and Holt LL (1997) Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America* 102: 1134–1140.

in response to /d/ responded more robustly to the same novel stimuli when they were preceded by /r/. Thus, as illustrated in Figure 2, birds exhibit context-dependent shifts in responses contingent on a speech precursor, and the directionality of this effect is the same as that observed for human adults and infants. The extension of phonetic context effects to a non-human species suggests that mechanisms specific to human speech may not be necessary to accommodate the complex variability present in speech. To pose the results in a general way, context sounds with higher frequency acoustic energy (e.g., /l/) shift perception of the following syllable toward the alternative with greater low-frequency energy, /g/. Thus, processes by which the auditory system exaggerates change in the acoustic signal may be sufficient to explain perceptual compensation for coarticulation. This conclusion is supported by research demonstrating that adult human listeners shift phonetic categorization responses not only as a function of neighboring speech contexts but also as a function of nonspeech tones, chirps, and noises that precede or follow speech. In the case of human and nonhuman perception of speech and nonspeech contexts, speech perception appears to be relative to and contrastive with the acoustics of context sounds, whether speech or nonspeech.

Thus, there is evidence that some of the perceptual challenges of the variability within the acoustic speech signal may be met with general perceptual processes that produce contrast. Experiments investigating listeners' sensitivity to distributional regularity in sentence-length acoustic contexts indicate that such contrastive processing may also provide a means of adjusting perception in response to variability in the acoustic speech signal introduced by different speakers. Overall, these results indicate that listeners retain information about context stimuli across multiple timescales and adjust speech perception in relation to this information. The exact nature of the mechanisms by which listeners retain context across multiple timescales remains to be determined, but empirical findings suggest that the mechanisms may rely on perceptual processing general to speech and nonspeech acoustic signals (and, indeed, available to even quail) rather than processing specialized to accommodate acoustic complexities of the speech signal.

Speech and Expertise

A great deal of research has been directed at identifying invariants, unique acoustic cues indicative of a particular phonetic category. However, despite these attempts, it appears that phonetic categories, by and large, are signaled by multiple cues that vary with the context of the spoken utterance, with no single cue being necessary or sufficient to signal phonetic identity. The case of English /b/ versus /p/ is illustrative: 16 distinct acoustic cues have been cataloged as differentiating these two categories in syllable-initial position. Recalling discussion in the previous section, it is also the case that acoustic cues signaling a particular phonetic category are quite dependent on the context in which the acoustic signal is presented. However, the existence of multiple acoustic cues to category identity does not imply that all of these cues are equally informative to listeners. For example, both spectral and temporal cues differentiate the English vowels /i/ and /I/, but listeners rely much more on the spectral cue than the temporal cue in categorizing these vowels. The precise nature of perceptual weighting of acoustic cues appears to depend, at least in part, on experience with the distribution of cues within the native language. There is evidence that the developmental trajectory of this learning is protracted; preschool-aged children apply very different perceptual weights to acoustic cues than do adults. Insight into how listeners learn to weight acoustic cues differentially may come from studies of other complex sounds, such as music and novel nonspeech categories, for which there is also evidence that listeners learn to weight acoustic cues based on cue distributions in the input.

The details of the acoustic distributions of the native language community play a very significant role in adult speech perception. In the first year of life, well before infants are speaking, they are already beginning to perceive speech in a native language-specific manner. As mentioned previously, this attunement to the native language continues into later childhood before reaching adult-like performance and may, some argue, require learning to read to produce the kinds of phonetic categorization typically studied in investigations of adult speech perception. Among adults, perception of the acoustic speech signal is very significantly influenced by experience with the native language. Identical acoustic signals may be perceived very differently by listeners with different language or even dialect backgrounds. The classic example of this was alluded to previously. Japanese does not make a phonological distinction between the English sounds /r/ and /l/. Experience with the structure of Japanese shapes the way that adult Japanese native listeners hear speech such that many Japanese listeners are unable to distinguish English /r/ from /l/ even following significant training. Expertise with a native language has a strong influence on the manner by which speech is perceived. The relative interaction of native language experience with perception of nonnative sounds depends significantly on the degree to which the nonnative sounds overlap with native sounds in perceptual space. Understanding these phenomena and investigating the degree to which adult listeners may or may not exhibit plasticity in learning second-language sounds are active areas of research.

This is a significant challenge because adult listeners are experts in perceiving speech, having spent the greater part of their waking hours engaged in perception of speech and being exposed to the speech signal since their later prenatal months. Exposure to the acoustic systematicities of a language is thought to organize perception in terms of phonetic categories or equivalence classes. Due to the acoustic variability of speech, there is no one-to-one mapping between acoustics and phonetic category. The mapping of a set of variable acoustic signatures onto a common phonetic label (or syllable or word), as described previously, is an example of categorization; physically distinct acoustic signals may be treated by listeners as instances of the same phonetic category. These categories appear to have graded internal structure such that utterances of a particular vowel or consonant may be perceived to be more or less representative of the category. Despite general agreement on these observations, it has proven difficult to determine the precise learning mechanisms by which this organization occurs because adults, and even young infants, possess a great deal of phonetic experience. Without characterization or control of the acoustic distributions of this experience, it is difficult to tease apart how experience with different distributions of speech input may shape phonetic categorization. It is possible to observe that experience with different native languages produces very different phonetic organizations, but the means by which this occurs is not well understood. Several empirical inroads to these issues have been made by investigating learning in nonhuman animal models of speech perception and by investigating the boundaries and constraints on human adults' general ability to categorize complex nonspeech sounds. The findings from this growing literature suggest that listeners (both human and nonhuman) are able to extract regularities from distributions of complex acoustic signals such that categorical responses may be made. For example, European starlings trained to respond to the vowels /i/ and /I/ (as in 'beet' and 'bit') respond to the vowels in a graded manner, as do human adults. The strength of the starling responses to individual vowels correlates very strongly with native English human adults' ratings of the 'goodness' of these vowels. Thus, a nonhuman species is capable of learning to respond categorically to speech stimuli and does so in a way that mirrors the graded internal structure of human adult perception.

The virtue of nonhuman animal models is that there is a greater possibility of experimental control over the histories of experience such that firm conclusions about the course of learning may be drawn. A complementary approach that shares this virtue is training human adults to categorize novel nonspeech acoustic stimuli such that the constraints imposed by auditory categorization, in general, may be understood. Investigations of nonhuman animal models and adult human novel sound category learning may play an important role in understanding phonetic categorization because of the experimental control afforded. A challenge will be to balance this control with the ecological validity of studies of infant phonetic category acquisition and nonnative adult listeners' phonetic categorization to work toward an understanding of how adult perceivers become expert with the speech signals of their language community and how this expertise shapes speech perception.

Categorical Perception

Categorical perception is perhaps the best-known pattern of speech perception. The speech series illustrated in Figure 1 serves as an example. Although the speech sounds along the series vary gradually in their acoustic characteristics, listeners' pattern of identification changes abruptly, not gradually; this sharp identification function is one of three hallmarks of categorical perception. A second defining characteristic of categorical perception is the pattern of discrimination across the series. When listeners discriminate pairs of stimuli drawn from the series, the resulting discrimination function is discontinuous. Discrimination is nearly perfect for stimuli that lie on opposite sides of the sharp identification boundary, whereas it is very poor for pairs of stimuli that are equally acoustically distinct but lie on the same side of the identification boundary. The final characteristic of categorical perception is that identification performance predicts discrimination performance; speech sounds that are identified with the same label are difficult to discriminate, whereas those identified with different labels are discriminated.

Categorical perception was formerly thought to be a peculiarity of speech perception. However, categorical perception has since been observed for perception of human faces and facial expressions, music intervals, and artificial stimuli that participants learn to categorize in laboratory tasks. It is observed in the behavior of nonhuman animals as well. Rather than a speech-specific phenomenon, categorical perception is a far more general characteristic of how perceptual systems respond to experience with regularities in the environment.

Influences from Other Sources of Information

As described previously, the mapping between acoustic speech signal and intermediate representation (acoustics to phonetics, in most characterizations) is complex. It is also the case that this mapping may be influenced by higher-order linguistic information, such as the knowledge of words within the language. An ambiguous sound that might be perceived as /g/ or /k/, for example, is more likely to be identified as /g/ if it is followed by 'ift.' The same sound is perceived as /k/ if it is followed by 'iss.' The direction of this influence is to shift phonetic categorization toward the alternative that forms a real English word. The influence of lexical information on phonetic perception has been established in a variety of paradigms, and the mechanisms by which these effects arise are hotly debated. Specifically, there is significant controversy between interactive accounts, which posit that such effects arise via feedback from the lexicon that directly influences the mapping from acoustics to phonetics, and strictly bottom-up autonomous accounts, whereby the influence of lexical information on phonetic responses emerges only at a later decision stage and does not directly influence the mapping from acoustics to phonetics (Figure 3). As in other domains of cognitive science that are modeled by these two approaches, distinguishing between interactive and autonomous accounts in speech perception has been difficult because the two models make many similar predictions. However, there is a domain in which the predictions of the models depart quite radically. Interactive accounts posit that lexical information may influence prelexical processing directly via feedback from the lexicon to prelexical processing. In autonomous models, this kind of direct modulation is impossible because information from phonetic and lexical processing is combined only at a later decision stage. If lexical information influences a prelexical phenomenon, it may be taken as evidence for interactive models.

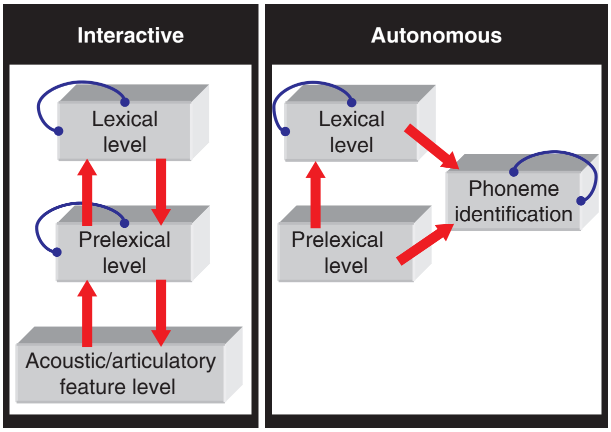

Figure 3 Schematic diagrams of interactive and autonomous models of speech perception.

There is one empirical paradigm in which this proposition has been tested, retested, and much debated. The phonetic context effects described previously are widely considered to arise from the mapping from acoustics to phonetic categories; that is, they are prelexical in the sense that they do not require information from the lexicon and operate at a level below that of words. Several lines of research have investigated whether lexical knowledge of the words of English influences phonetic context effects and thus whether lexical information modulates a prelexical effect, as predicted by interactive models. At the prelexical level, a preceding /s/ or /ʃ/ (as in 'sip' vs. 'ship') influences phonetic categorization of subsequent sounds as /t/ or /k/. Preceded by /s/, listeners more often categorize the speech as /k/, whereas preceded by /ʃ/, the same sounds are more often categorized as /t/. Of interest regarding the issue of the role of lexical processing in influencing prelexical effects, a single ambiguous speech sound between /s/ and /ʃ/ is heard as /ʃ/ if it is preceded by 'fooli__' but as /s/ if it is preceded by 'Christma__.' Knowledge of the English language, which includes the words 'foolish' and 'Christmas' but not 'fooliss' or 'Christmash,' serves to disambiguate the acoustically ambiguous speech sounds. Of critical importance, however, is the effect of this on listeners' perception of subsequent sounds as /t/ or /k/. When the ambiguous /s/–/ʃ/ sound is presented in the word contexts and precedes the /t/ to /k/ sounds, there is a shift in listeners' phonetic categorization such that preceded by 'Christma__' listeners respond /k/ more often, whereas preceded by 'fooli__' the same sound is labeled as /t/. Thus, there is a phonetic context effect on the second sound that is driven by the lexical information, as predicted by interactive models. There has been considerable debate over findings such as these, resulting in replications and alternative conclusions more consistent with autonomous models. Recent investigations have tipped the scales back in favor of an interactive account because these effects have held up to empirical scrutiny.

Another source of influence on speech perception comes from the visual modality. Although speech perception is most often considered from the perspective of the acoustic signal, visual information from a speaker's face can also greatly influence perception of speech. The classic example of this is the McGurk effect, whereby an acoustic /ba/ and a visual /ga/ collaborate to produce a percept of /da/, unique from either unimodal source. A great deal of research illustrates the advantages of perceiving acoustic speech with synchronous, matched visual information of the face producing the speech. Speech presented in noise, for example, is more intelligible when a face producing the speech is visible. Some of the same issues and debates present in understanding why lexical influences on speech perception emerge in considering how the perceptual system integrates optic and acoustic information for speech perception; there are both autonomous and interactive models of the phenomenon. In addition, an important theoretical question that is under investigation is the degree to which auditory–visual integration for speech is a result of the two types of information sharing a common source or whether it is a result of learning about the natural covariance of these visual and auditory signals. This question is important in distinguishing among the theoretical accounts of speech perception. The question is very difficult to address because acoustic speech is intrinsically linked with the visual signals from articulating faces, but adult listeners also have vast experience with the covariation between articulating faces and acoustic speech. This has been investigated by examining whether listeners are able to learn to associate novel visual signals, not in any way associated with faces or with production of speech, with acoustic–phonetic categories. With sufficient training, listeners do, in fact, learn this relationship and, with learning, begin to integrate the novel visual information for the phonetic categories with acoustic speech. The phonetic category membership of an ambiguous acoustic stimulus is disambiguated by the newly learned visual signals. Moreover, the learning results in benefits to intelligibility of acoustic speech in noise when a newly learned visual signal associated with the phonetic category is paired with it. Further work is necessary to determine the relationship of speech expertise with auditory–visual integration in speech perception.

Theoretical Approaches

The challenges of speech perception have been addressed by multiple theoretical accounts of how listeners accommodate the complexities of the speech signal. Landmark studies detailing the relationship between speech acoustics and phonetic perception revealed the complex mapping. These data convinced many that this mapping was far too complex for acoustics to provide a plausible object for speech perception; the objects of speech perception were thought to need to be more or less invariant with phonetic categories. From this tenet arose the motor theory of speech perception. Given the radical variability in the mapping from acoustics to phonetic categories, the motor theory holds that the only tenable object of perception is the intended articulatory gesture of the speaker, as exemplified by the neuromotor command to the articulators. By this view, the objects of speech perception are the articulatory events of a speaker rather than the acoustic or auditory events presented to the listener. Another important aspect of the motor theory is that speech perception is uniquely human and cannot be ascribed to general auditory processing or perceptual learning. Instead, it is said to rely on a modular decoder that is part of the innately specified biological specialization for language, distinct from other forms of auditory processing.

Another theoretical approach, direct realism, shares with motor theory the claim that the objects of speech perception are articulatory rather than acoustic events. However, distinct from motor theory, direct realism posits that the articulatory objects of speech perception are the phonetically structured vocal tract movements, or gestures, and not the neuromotor commands. Direct realism is also very distinct from motor theory in that it denies that specialized processes are necessary to account for speech perception. Rather, following in the Gibsonian tradition of direct perception in the visual modality, this theory asserts that the articulatory gestures of a speaker that shape the acoustic signal serve as information for the listener to directly recover these gestures. This act of perception is not mediated by processes of inference or hypothesis testing but, rather, is direct in the sense that the gestures are apprehended without cognitive mediation from the rich information present in the acoustic signal that specifies a speaker's phonetic gestures. This theory is realist in the sense that perceivers are thought to recover the actual physical properties of the articulatory gestures from the acoustic signal.

The general approach to speech perception is distinguished from motor theory in that it does not invoke specialized mechanisms or modules to explain speech perception. Rather, its working hypothesis is that acoustic speech sounds are perceived with the same mechanisms of auditory perception and cognition that have evolved to handle other classes of complex environmental sounds. Insofar as speech is perceived as a multimodal event, the objects of perception in other modalities are hypothesized to also be general and not specific to speech. The general approach assumes that mapping from signal to meaning is not mediated by the perception of articulatory gestures but, rather, involves mapping the complex structure of the acoustic signal to regularities learned through experience with the distributions of the ambient language community. This general theoretical perspective embraces general perceptual and cognitive mechanisms, not specific to speech but neither limited to solely low-level sensory processing and psychophysics. The account is ‘general’ in the sense that it suggests that the broad perceptual/cognitive processing of the central nervous system and also the considerable feedback that higher centers have to lower levels of processing are brought to bear in perceiving spoken language.

A great deal of empirical research has called into question the necessity of specialized, modular mechanisms for speech perception. Nonhuman animals exhibit many of the hallmark perceptual phenomena previously thought to be indicative of specialized processing by humans for speech. Moreover, nonspeech acoustic contexts influence human adult speech categorization, a finding that would not be expected if speech processing is modular. There remains a very active debate among the proponents of direct realism and the general approach. However, both theories make predictions about the expected patterns of speech perception, so empirical research promises to increase understanding of speech perception and either further refine these accounts or suggest new thinking.

Neural Processing

One source that many believe will provide theory refinement is the study of the neural processing underlying speech perception. A great deal of research has been directed to understanding the neural bases of speech perception, and many of the efforts have been applied to the theoretical distinctions discussed previously. Research using neuroimaging techniques, for example, has compared auditory processing of speech versus processing of environmental sounds matched along multiple acoustic and perceptual dimensions, including rhythm, content, and duration. Common brain regions are activated by both stimulus types, but of possible theoretical interest, some regions are activated more by speech than by environmental sounds. Some have argued that these sorts of data demonstrate that human adults have specialized neural mechanisms for perceiving speech. However, it is worthwhile to keep in mind that these brain regions may take part in nonlinguistic processing but simply be more engaged by speech stimuli than other nonspeech sound sources. In fact, further investigation of the regions that are activated more by speech than nonlinguistic sound stimuli has found that these areas are also activated by pitch, melody, environmental sound, and/or nonauditory conceptual processing. In summary, caution must be used in interpreting brain regions as indicative of speech-specific processing based on limited controls. The differential activation of brain regions by speech versus nonspeech acoustic signals may arise from differential demands on auditory and conceptual processes brought to bear by speech versus nonspeech signals or by the vast differences in experience listeners have with speech compared to the nonspeech control stimuli.

See also: Connectionist Models of Language Processing; Sentence Comprehension; Speech Perception: Development; Speech Perception: Neural Encoding; Speech Perception: Cortical Processing; Statistical Learning of Language; Word Recognition.

Further Reading

Diehl RL, Lotto AJ, and Holt LL (2004) Speech perception. *Annual Review of Psychology* 55: 149–179.

Flege JE (2002) Interactions between the native and second-language phonetic systems. In: Burmeister TPP and Rohde A (eds.) *An Integrated View of Language Development: Papers in Honor of Henning Wode*, pp. 217–224. Trier, Germany: Wissenschaftlicher Verlag.

Fowler CA (1986) An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics* 14: 3–28.

Ganong WF (1980) Phonetic categorisation in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6: 110–125.

Hickok G and Poeppel D (2000) Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 4: 131–138.

Holt LL (2005) Temporally non-adjacent non-linguistic sounds affect speech categorization. *Psychological Science* 16: 305–312.

Johnson K (2003) *Acoustic & Auditory Phonetics*. Malden, MA: Blackwell.

Liberman AM (1996) *Speech: A Special Code*. Cambridge, MA: MIT Press.

Lotto AJ, Kluender KR, and Holt LL (1997) Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America* 102: 1134–1140.

Mann VA (1980) Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics* 28: 407–412.

McClelland J, Mirman D, and Holt LL (2006) Are there interactive processes in speech perception? *Trends in Cognitive Science* 10: 363–369.

McGurk H and McDonald J (1976) Hearing lips and seeing voices. *Nature* 263: 747–748.

Pickett JM (1999) *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology.* Needham Heights, MA: Allyn & Bacon.

Pisoni DB and Remez RE (eds.) (2005) *The Handbook of Speech Perception.* Malden, MA: Blackwell.

Price C, Guillaume T, and Griffiths T (2005) Speech-specific auditory processing: Where is it? *Trends in Cognitive Science* 9: 271–276.