

Word Recognition

Humans have the remarkable capacity to express their thoughts, beliefs, and intentions through a physical medium (sounds, gestures, or pictograms) to share with others. Language's expressive power resides in the ability to express any novel thought by combining elements (e.g., cow, purple, ticklish) into a sequence that conveys that thought (e.g., a ticklish purple cow). Thus, a critical component of language comprehension consists of recognizing the presence of these elements, the words, in the speaker's discourse in order to retrieve their individual meanings and combine them to derive an overall interpretation of the discourse. Research on word recognition has focused on understanding how people categorize a physical (auditory or visual) token as one of the many words they know. This entry first reviews the factors that affect the perceptual choice inherent to the recognition of spoken or printed words. It then discusses how people represent the forms of words they know in order to recognize them in speech or print. Finally, the entry briefly discusses how people's linguistic knowledge influences their evaluation of sensory information.

Word recognition as a perceptual choice

People recognize the words of their language remarkably quickly and accurately, whether written or spoken. This ease conceals the complexity of the process, however. The roughly 50,000 words that an adult knows are formed by combining sound or letter elements drawn from a much smaller set (e.g., 26 letters and between 40 and 45 distinct sounds in English, depending on the dialect). As a result, words resemble one another to a large degree. The physical difference between the words *cat* and *cot* is subtle but critical

to attend to if the correct meaning is to be retrieved. Thus, recognition entails selecting the best matching hypothesis among alternatives.

Importantly, these alternatives exert an influence on the outcome. Under some conditions, words are harder to recognize if they have many “neighbors,” that is, if they resemble many other words. For instance, while reading, people spend more time fixating on printed words with many neighbors than words with few, suggesting that the former take longer to identify. Likewise, spoken words that sound similar to other words are recognized more slowly, and less accurately when presented in noise, than spoken words with fewer neighbors. These findings are often taken as indicating that recognition is a competitive process, where the spoken or written stimulus is compared to every hypothesis simultaneously and where the best-matching hypothesis is identified with respect to the overall support accrued by the alternatives. Under some conditions, however, similarity to many other words facilitates a word’s recognition. For example, written words that share all but one of their letters with many other words (e.g., *cave*, similar to at least 16 other words such as *cake*, *have*, *gave*, *care*) are recognized faster than words with fewer such neighbors (e.g., *next*, similar to only 4 other words, *newt*, *neat*, *nest*, and *text*). That neighbors can facilitate and impede a word’s recognition has been taken as reflecting the existence of two separate mechanisms by which neighbors affect a word’s recognition. Neighbors provide mutually incompatible interpretations of the same portion of the stimulus, the part that differs among them. The more alternatives, the harder the perceptual choice. But neighbors also share some elements, and the perception of these parts is enhanced by the support provided by their occurrence in

multiple words. Whether neighbors facilitate or impede the overall outcome depends on which influence outweighs the other.

The ease of recognizing a word is also affected by how often it has been encountered before, with frequent words being recognized more accurately and faster than rarer words. In studying this effect, frequency of occurrence is estimated by the number of times each word occurs in a large sample of spoken and written language. The so-called *word-frequency effect* has sometimes been explained as reflecting the ease with which an operation is performed given the number of times it has been performed in the past. This account is supported by the neural mechanism that underlies learning, whereby frequent exposures to a given stimulus strengthen the connections between the population of neurons that encode the physical attributes of the stimulus and those that encode its interpretation. However, the word-frequency effect can also be attributed to the influence of biases (or priors) on perception. This view assumes that the physical attributes of a word take time to apprehend and decode and provide only probabilistic cues, and people are generally making decision under some uncertainty. Under these conditions, an optimal strategy for deciding whether or not a given word is present in the stimulus is to consider the probability of encountering the word at all. Frequent words have a high base rate, and people's guesses reflect this rate. Frequent words are recognized more accurately than rare words because, based on noisy information from the stimulus, the best bet is to assume that the word is a frequent one. Rare words are recognized more slowly or less accurately because, initially at least, people favor their more frequent alternatives.

Finally, the process of recognizing spoken and printed words differs because of differences in their medium. A written word is a sequence of elements spatially organized, and the reader controls the uptake of sensory information. By contrast, a spoken word is a sequence of elements temporally organized, and the order and the speed at which sensory information is received is not under the listener's control. Speech is a complex, transient, and rapidly changing signal. Because humans' sensory memory is limited, speech must be evaluated and interpreted incrementally rather than, say, phrase by phrase. Real-time processing is at the core of William Marslen-Wilson's seminal work on the recognition of spoken words. According to Marslen-Wilson's "cohort" model, the first sounds of a word determine a cohort of hypotheses that are compatible with this early information. Subsequent information serves to prune the hypotheses that are no longer supported by the signal. The point in time at which a spoken word is recognized can be precisely identified as the point where the input has excluded all candidates but one (i.e., the word's *uniqueness point*). Although Marslen-Wilson revised his theory to allow phonetic input to provide gradient, rather than all-or-none, support to word candidates, the assumption that word recognition makes immediate use of the signal as it unfolds has prevailed.

How are word forms represented?

In order to decide which word they see or hear, people compare sensory information to an internal representation of the forms that words take. How should this internal representation be characterized? Although there is no definitive answer to this question, issues pertaining to the processing of visual or auditory information have

helped identify the central properties that internal representation must possess. Here, I review some of these issues.

How do people judge which word they see or hear based on the sensory information available? One possibility is that they compare the visual or auditory stimulus to each auditory or visual abstract image or template associated with the words they know, and determine which template provides the best match. However, considering the number of words and their similarity, templates would be difficult to tell apart. An alternative consists of identifying components of words, which represent a much smaller set, and establish their arrangement. This information distinguishes any word from its alternatives.

In the domain of written words, where most of the research has been conducted on English by presenting printed as opposed to handwritten, words, each letter is a symbol that is neatly separated from the surrounding letters and reproduced identically across words (although its image on the retina changes as a function of its size, spatial location, and other low-level factors). Thus, letters are discrete elements that are juxtaposed to form words. Letter recognition seems a necessary component of printed word recognition. The situation is quite different in the auditory domain. The articulation of a speech sound is greatly influenced by the sounds that precede and follow it. For example, the shape and position of the lips prior to articulating the sound *b* change dramatically as a function of the vowel that follows it, as in *bat* or *boot*. This contextual influence on the articulation of speech sounds is reflected in their acoustic realization. It has been proven difficult to identify acoustic properties that instances of the same speech sound produced across contexts all share. Consequently, whether the recognition of a

spoken word involves the identification of its subcomponents is a highly debated question. The perception of cursive hand writing has not been studied much but raises similar issues as speech perception.

Because elements are shared among many different words, establishing their order is critical to correct identification. The letter sequence “top” should be interpreted as the word *top* and not *opt*, even though *opt* contains the same elements. Thus, the comparison between sensory information and word-form representations requires an alignment point. For the recognition of visual words, the spaces surrounding a string of letters offer such alignment points. Each letter in the stimulus is viewed as occupying a given slot in space whose position is established with respect to the beginning of the sequence (the left space). The match between the stimulus and a given word hypothesis is evaluated by comparing each letter slot independently. Accordingly, *opt* and *top* have no aligned letters, and cannot be confused. However, this view of alignment fails to account for readers’ ability to recognize some words with transposed letters (e.g., “jugde” for *judge*) as easily as their intact counterparts, or the strong perceived similarity between words like *pluck* and *luck*, which, under the left-aligned scheme, share no overlap.

The alignment issue is even more salient in speech because the alignment between the sensory information and particular word hypotheses must be assessed on a temporal, as opposed to spatial, dimension. For example, the sensory information that results from the phrase *the catalogue* is compatible with the words *cat*, *a*, *log*, *cattle*, and *catalogue*, among others. (Rules of syntax and semantics, as well as fine-grained acoustic details, help disambiguate the utterance’s interpretation.) However, these word hypotheses entertain different relationships with one another, and establishing the relationship among

these word candidates requires a precise alignment with the stimulus. *Cat*, *cattle*, and *catalogue* are mutually incompatible hypotheses because they account for some of the same portion of the input, roughly the syllable “cat”. Said slightly differently, the syllable “cat” in the speech may be attributed to the word *cat*, *cattle*, or *catalogue*, but to only one of them. Likewise, *catalogue* and *log* are mutually incompatible hypotheses because they are different interpretations of the same portion of the signal, the syllable “log”. By contrast, *cat* and *log* are mutually compatible because they each account for a different part of the stimulus. Thus, the perceptual system must represent the temporal dimension at a fine-grained level.

How does prior knowledge influence evaluation of sensory information?

Word recognition relies on assessing a perceptual stimulus in the light of prior knowledge—the words of one’s language. The interdependence between sensory information and prior knowledge was demonstrated in one of the earliest empirical studies of word recognition, conducted by James Cattell at the end of the 19th century. Under tachistoscopic conditions, where written stimuli are presented for a very brief period of time, Cattell showed that letters were identified more accurately when presented in words than in random letter sequences. This finding was later extended by Gerald Reicher, who showed that individual letters were identified faster when occurring in a word context than in isolation. This latter result is especially puzzling if one takes letter recognition as a prerequisite to word recognition. The so-called *word-superiority effect* has been extensively studied, with demonstrations that it extends to the auditory modality. For example, listeners tend to interpret a sound that is ambiguous between the sounds “g” and “k” as “g” if followed by “ift”, and as “k” if followed by “iss”. Two

competing accounts for the effect have been offered. According to one account, knowledge of which letter sequences form words can penetrate the mechanisms of perception and thereby enhance, or even reshape, the perception of the letters. This view was implemented by James McClelland and David Rumelhart in a cascaded and interactive model of word recognition. According to these models, sensory information provides initial support to some letter hypotheses, which in turn provide support for the words that contain them; these hypotheses then feed back their support to their letter constituents. This process takes place very rapidly and before a single percept has been isolated. Thus, letter recognition can affect and be affected by word recognition until the process reaches a stable state. Importantly, the feedback from words to letters reflects the influence of knowledge on perception. An alternative account of the word-superiority effect offered by Dominic Massaro and colleagues maintains independence between the sources of information. The sensory information associated with each letter is integrated with sensory information extracted from the context, i.e., the other letters. The intersection of these two sources of information constrains the set of possible letters at each position in the word more than each source independently. Identification reflects the integration of both sources of information, without assuming that one affects the other. Distinguishing these two competing accounts empirically has proved difficult.

Delphine Dahan

See also *Eye Movements and Reading*; *Speech perception*; *Top-Down and Bottom-Up Processing*.

Further Readings

Cattell, J. M. (1886). The time taken up by cerebral operations. *Mind*, *11*, 220-242.

Dahan, D., & Magnuson, J. S. (2006). Spoken-word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 249-283). Amsterdam, The Netherlands: Elsevier.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71-102.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: the joint influence of stimulus and context in perception. *Cognitive Psychology*, *23*, 558-614.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1-86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407.

Rastle, K. (2007). Visual word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp 71-87). Oxford, U.K.: Oxford University Press.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274-280.