

Study on the motor theory of speech perception



**ASSIGN
BUSTER**

The motor theory of speech perception by Liberman et al (1967) deals with the recognition of phonemes which are either from an acoustic and/or visual speech signal, which are considered as the phonetic gestures of a speaker.

This theory suggests that a listener refers incoming signals back to the articulatory instructions, which articulators would receive from a listener.

This in result produces a set of same sequences. This theory points out the large variance in the acoustic signal. It also points out that there is a high invariance of the motor commands to the articulatory organs.

The encoding of this linguistic information is articulatory, so the process of decoding in the auditory system must be analogous. So it is clear that good production depends on normal perception.

Motor theorists believe that before phonetic and non phonetic processing has taken place, the gating of phonetically processed signals into specialised neural units must occur, which means that non - phonetic processing is abandoned when the signal is then identified as speech.

So according to this theory speech is received serially, by the ear. This must be processed in a parallel function, to make sure that any other cues do not communicate to any phonemic boundaries.

One main claim of this theory is that speech is "special", which bridges the gap between acoustic data and linguistic levels, "Special" in the sense that perception of sounds of processing. This is both innate and through language which humans process. Another claim of this theory is that it claims

evolutionary adaptations of mammalian motor system made speech possible for humans.

There are many views which say speech is special, but this theory of speech perception is seen to be very controversial, as support and a lot of criticism is given to this model. One main question which still remains is really how “special” speech is? Also how much of what’s heard is in the signal and how much is constructed in the listeners mind.

The main criticism by many psychologists for this theory is the use of both active and passive theories and that none of the claims of the theory have been proved and backed with enough evidence. However a main advantage of this theory is that it has the advantage that motor commands to the vocal mechanism are more amenable to psychological study. Also that this theory has added further to knowledge on speech perception.

Cohort theory

Another theory of speech perception is an active model of word identification being the cohort theory by Marslen - Wilson & Tyler (1980). The Cohort theory was designed to account for the perception of individual words, all the words in memory that have similar phonemes are activated in proportion to how similar they are to the extracted phoneme(s). After the Cohort is activated, other acoustic or phonetic information operates to eliminate all the candidates except the most appropriate one. Word perception involves a mixture of bottom up and top down process. So when the first part of a word is heard, all of the words the listener knows, starting with that sound sequence are activated. These set of words are known as the ‘word initial

cohort'. Gradually elimination occurs for members of the cohort for a range of different reasons. They may be unsuccessful in matching certain parts of the sequence of sounds, or are unable to match semantic or other context.

There is some experimental support for the theory provided by Marslen-Wilson and Tyler (1980). Their participants engaged in listening to sentences which involved them to identify target words as quickly as possible. The sentences were all unrelated words either being normal to random sentences. In another condition the targets were identical to a given word. The prediction of this was that target detection would be faster with normal sentences in comparison to random sentences. This prediction proved to be correct, as contextual information in normal sentences allow top-down processes to eliminate words from the word initial cohort. This study showed support to the theory and that contextual information is used at an early stage of processing.

This theory however has been edited as the original theory attached to much importance to the first phoneme of the word. The modified version of the theory's advantage is that it had more flexibility than the original theory. The cohort theory has made a lot of contribution to speech perception, as it is backed up by a lot of research, it is also a main theory used in understanding speech perception in comparison to the motor theory which is seen to be very controversial. Although there is a lot of praise for this theory, there are still many questions which remain unanswered such as; how do we recognise non-words and also how is multi-word input perceived? So there are a few questions which still need to be addressed. One main deficiency of this model is that it has too much sensitivity to speech rate. So as any theory

<https://assignbuster.com/study-on-the-motor-theory-of-speech-perception/>

there are both advantages and disadvantages which have been or still need to be addressed.

Trace theory

Another well known model in speech perception is the trace model by McClelland and Elmans (1986). This model consists of units which are broken down into three interacting levels, these are; feature level, phoneme level and the word level. Each of the levels are similar to nodes, as are all well interconnected. Activation in one node activates all other nodes, as they are all connected. Here information processing takes place through excitatory interactions of a large number of simple processing units. In the lowest level of the model being the feature level the nodes represent the phonetic features, in the second level the phoneme level; the nodes represent the phonetic segments. The last level of the model, word level here the nodes represent words when a particular level of activation is reached, the nodes are fired, which indicates that a feature, phoneme or word is present (Moore, 1997). At the feature level there are detectors for each dimension of speech sounds, there are detectors for every word at the word level. Activation is passed along nodes which are connected when a node is fired. Excitatory links between nodes exists at different levels; this may cause nodes to fire at the next level. This then allows nodes which are highly activated to compete with those nodes which have less activity, so therefore this leaves one node capturing all of the activity.

This model is widely used in speech perception, this model is able to recover distortion of a words beginning, and it can also use activations of phoneme units to adjust connection strengths. This determines features which will

<https://assignbuster.com/study-on-the-motor-theory-of-speech-perception/>

activate which phoneme. This model has its weaknesses as it predicts interaction of top down and bottom processes, but it has shown that those actually operate independently. It also exaggerates importance of top down processing. Also top down processing effects depends on degraded stimuli than is predicted by the model. This model does not explain timing of speech sounds or individual differences in speech rate. Finally, one main criticism of this model is that it doesn't learn, so overall this model plays a big role in speech production, but with it there are many criticisms. However this theory is one of the main and wide used theories of speech perception.

McGurk effect

The McGurk effect is looking at the role of vision in speech. Here, a video was shown of a woman saying an alternate one of four sounds, the sounds consisted of; " bah", " kah", " gah" or " pah". McGurk found that people, who watched the videos, sometimes fail to correctly recall what was actually said. So, if " gah" was mouthed, but dubbed with a voice saying " bah", people would usually think they heard " dah", but the correct sound " bah" was heard, only when they had their backs to the video. Another finding was that to hear the correct vocal sounds could not be forced. When viewers were told they were being given the wrong visual information. These findings show that visual information is integrated into our perception of speech unconsciously. Also that our speech function makes use of all different types of relevant information, not taking into account modality.