Experimental Design and Implementation of English Phonemic Categorical Perception

Lei Guo, Hao Yang, Hongjie Wang

Key Laboratory of China's Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, Lanzhou, Gansu, 730000, China

Keywords: E-prime, Perception, Speech sample synthesis, English

Abstract: Based on the research results of traditional phonetics, this paper proposes a set of calculation parameters for speech perception categories including the calculation of recognition parameters, the calculation of distinguishing experimental results, and the degree of correspondence between recognition results and distinguishing results. Adopt perceptual methods to study the speech understanding and cognitive style of second language speech learners, and we introduce the latest methods and techniques of modern phonetics into the study of college students' English phonetics. The experimental design not only proposes new methods for phonetic research, but also provides an effective reference for phonetic signal processing research and speech acoustics engineering research.

1. Introduction

Categorical perception is an important model of phonological perception. Listeners are generally less sensitive to differences within a category, but are more likely to detect differences between categories. In contrast to categorical perception, continuous perception means that subjects cannot perceive speech stimuli as discrete and limited categories.

The study of category perception of speech started in Haskins laboratory. The classic task paradigm of speech category perception is the recognition experiment and differentiation experiment proposed by Liberman in 1957. The identification experiment requires the subjects to judge the voice stimuli they hear, and to make a forced choice from the given answers. The main function of the recognition test is to ascertain the degree of confusion that listeners may have when making choices in the recognition experiment. This requires that the same set of stimuli be played randomly several times for the subjects to judge. The classic paradigm for distinguishing experiments requires that after hearing the three stimulus sounds A, B, and X, the subjects make a forced alternative choice to determine whether the stimulus X is A or B. The purpose of the discrimination experiment was to determine whether the listener could perceive differences between a sample continuum of isometric variations in a set of acoustic variables.[1]

Theoretically, if the number of measurement errors is excluded, the distinction function and the recognition function should fit perfectly, but early researchers have made an important finding that the actual distinction function in category perception is usually higher than that obtained by the recognition function Predictive distinction function. The research methods of category perception were mainly used in psychological research and natural sound research in the early days. After the Second World War, American psychologists began to study human voice using perception experiments, which set off linguistics and experimental psychology. Most Indo-European languages are toneless languages, so the experimental objects of early category perception were mainly concentrated on Indo-European languages and other well-known minority languages.[2] The research scope is relatively narrow, limited to the characteristics of syllables such as vowel consonants. the study. Since the 1960s, some people have begun to shift their research focus to the characteristics of supersonic segments such as tones.

  DOI: 10.25236/icembe.2020.046

2. Experimental Content of the English Phoneme Category

The study of the perception of modern phonetics is a higher form and stage in the development of phonetics. In order to further explore the pronunciation problems of minority phonetics students' English phonemes, we conducted a study from the perspective of listeners, and used the experimental methods of recognition and hearing to determine the speech categories and boundaries from a series of continuously-synthesized segments.[3]

2.1 Research on English Vowel Phoneme Perception

In the international phonetic vowel phoneme, the same phoneme may be similar or far apart in different languages. In second language phonetic learning, some vowel pronunciation errors are common, such as learners often use similar vowels in their native language to replace vowel phonemes in second language, confusion of long and short vowels, and double vowels into unit sounds. At present, most researches on second language speech acquisition are focused on output, and there are few studies on speech perception, which is one of the important attributions of vowel pronunciation errors. Specific research includes the following three aspects:

(1) Perceptual study of vowel formant structure. The main acoustic characteristics of vowels are realized by formant parameters. With the help of Praat voice software, the vowel formants, bandwidth, fundamental frequency and amplitude parameters are adjusted to synthesize multiple samples. In the hearing experiment, analyze the degree of vowels perception and the vowels category. What parameters are relevant to the English level in listening? Which are less relevant?

(2) Perceptual study of the relationship between vowel formants and fundamental frequency. By adjusting the fundamental frequency parameters, the effect of the fundamental frequency on the vowel sounds in the pronunciation of English vowels is tested. When the energy of the fundamental frequency is continuously increased, will it affect the energy distribution of the vowel spectrum structure?[4]

(3) Research on the perception of vowel sound length. There are 5 long vowels and 7 short vowels in English monophthongs. Each group of phonemes contains two features: length and sound quality. By listening sense experiment, we can determine which feature is the distinguishing feature when subjects perceive these vowel phonemes.In addition to the difference in sound quality, the 8 diphthongs of English also differ in length. In the listening and discrimination experiment, the two parameters of duration and formants were changed to find out which feature have the greatest contribution for the perception of vowel sounds.

2.2 Research on English Consonant Phoneme Perception

Compared with vowels, consonants have more number and pronunciation types. Owren and Cardillo (2006) believe that consonants have more information, and consonants are more important to the understanding and intelligibility of pronunciation.The research on the perception of consonant phonemes is carried out from the following three aspects.

(1)The perception of stop sounds. In the perception of stop sounds, through the direction of the second formant of the vowel sound followed the stop sounds, that is, changing the track frequency synthesis sample, through changing the audio frequency composite sample, we can test the learners' perception boundary of the six stop phonemes in English. The voice start time (VOT) is closely related to the voiced and unvoiced perception of stop sounds. Three voiced stop sounds in English are gradually synthesized into the voiced stop sounds of the same part, and the hearing recognition experiment is conducted to find the categorical perception of students based on VOT parameters.[5]

(2) The perception of frictives. There are 9 fricatives in English. In addition to [h], the other can form a pair of voiced and voiceless sounds.The acoustic features on the diagram are mainly the chaotic lines. By adjusting the duration, tone intensity, frequency peak of chaotic line, chaotic line starting frequency, and the lower limit frequency, we can synthesize multiple samples to determine which parameter is an important phonetic feature for subjects to recognize English fricatives.

(3) Perception of nasals and approximants.There are 3 nasals and 3 approximants in English. The

place of articulation involves from the lips to the soft palate. The pronunciations of approximants [w] and [j] are similar to vowels [u:] and [i:]. In the sound-recognition experiment, adjust the formant, duration, amplitude and fundamental frequency parameters to determine which parameter is the distinguishing feature, and test the difference between English nasals and approximants perception among learners of different English levels.

2.3 Research on English Word Stress Perception

There are four basic characteristics of the first stress in English words, longer sound length, stronger sound strength, higher pitch and full vowels. Select two-syllable words, three-syllable words, and four-syllable words from the corpus respectively, design recognition and differentiation experiments, through recognition experiments, test learners of different English proficiency in English word stress, through differentiation experiments, analyze the contribution of vowel phonemes and consonant phonemes in the perception of accented syllables, and explore which specific parameters are important features of stress hearing recognition for the listener listening to English word stress. Finally find the perceptual boundaries between the light and heavy syllables in English words.

3. Experimental Method of English Phoneme Category

3.1 Signal Recording and Acquisition

Using Adobe Audition to collect sound and process audio files, in order to facilitate the later use of E-Prime2.0 software for listening recognition experiments, the sampling rate of the collected voice signal is reduced to 11025Hz and saved in the corresponding folder. After recording, select a sound with stable pronunciation as the original sample of speech synthesis, and use Praat and PSOLA algorithm for speech synthesis.

3.2 Speech Sample Synthesis and e-Prime Listening Experiment

The software used in the experiment is E-prime 2.0, which is divided into two parts: recognition experiment and differentiation experiment.

Identify the experiment. There are 11 sound samples in each group of continuum, and each sound sample appears twice, so the subjects need to make 22 responses in a group of continuum recognition experiments. 11 sound samples were randomly presented to the subjects. After each sample was played twice in succession, the subjects were asked to determine which sound they heard within 5 seconds.[6] The first sound was LEFTARROW, and the second sound was RIGHTARROW. The choice for this experiment is a forced alternative.

Distinguish experiments. This experiment uses the AB form, that is, after playing two sounds, the subject judges whether the two sounds are the same. Each two sound combinations have four combinations: AA, AB, BA, BB. There are 11 identical pairs in each continuum discrimination experiment: 1-1, 2-2, 3-3, 4-4, 5-5, 6-6, 7-7, 8-8, 9-9 , 10-10, 11-11. Different pairs from 1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11, 11-9, 10-8, 9-7 , 8-6, 7-5, 6-4, 5-3, 4-2, 3-1, a total of 18. All sound combinations will appear twice randomly, so each opposite group has 58 sample pairs (11*2+18*2=58). The selection interface will appear on the screen after the playback of each sample pair. The subject must make a judgment within 5 seconds, and the selection is also a forced two-choice one form.

3.3 Calculation of Listening Parameters

Identification parameter calculation: including identification rate, identification boundary and identification width. Recognition rate is the percentage of the participants who recognized the sound as Option A or Option B. After calculating the recognition rates of the two speeches in all the opposite groups, it is necessary to further establish a regression model through Logit transformation to fit the continuous distribution of the starting pitch steps and the corresponding speech recognition rates, and finally can obtain a more accurate category boundary . Let X be the stimulation step in a continuous body of an opposite group, b0 is the regression constant, b1 is the regression coefficient of X, and the formula of the recognition function is:

$$\ln[PA/(1-PA)]=b0+b1X \quad \text{(Formula 1)}$$

When the recognition rate is 50%, the value of X is the position of the recognition boundary. The width of the recognition boundary is the linear distance between the recognition rate of 25% and the recognition rate of 75%. At the same time, b1 can be used as a reference for the steepness of the identification curve at the boundary. The narrower the identification boundary, the larger b1 means that the steeper the identification curve at the identification boundary and the more obvious the boundary.

The calculation to distinguish the experimental results: the formula proposed by Xu (2006) is used:

$$P=P(\text{“S”}|S)*P(S)+ P(\text{“D”}|D)*P(D) \quad \text{(Formula 2)}$$

P(“S”|S) is the probability that the same sample pair is judged to be the same, P(“D”|D) is the probability that different sample pairs are judged to be different, P(S) and P(D) are the proportions of the same and different sample pairs in the whole differentiation experiment.

Correspondence between recognition results and discrimination results: The degree of correspondence between recognition results and discrimination results can be judged by predicting the correlation between the discrimination curve and the measured discrimination curve. The formula for calculating the discrimination ratio PA-B of the stimulus to A-B is: $P_{A\text{-}B}$=Number of times that all subjects were correctly distinguished/(Number of subjects*Number of stimulus repetitions in this group).

Since the stimulus has four presentation sequences A-A, A-B, B-A, and B-B for A-B, the discrimination rate P(A,B) of this stimulation group is the average of the correct discrimination rates of the four presentation sequences:

$$P_{(A,B)}=(P_{A\text{-}A}+P_{A\text{-}B}+P_{B\text{-}A}+P_{B\text{-}B})/4 \quad \text{(Formula 3)}$$

According to the recognition rate calculated above, the correct differentiation rate of each stimulus group in the experiment can be predicted. The correct differentiation rate P* of each stimulus group can be calculated as follows:

$$P^*=[1+(PA-PB)2]/2 \quad \text{(Formula 4)}$$

P_A and P_B are data separated by two steps in the recognition experiment. There are 11 sound samples in a continuous body, the first P_A is the recognition rate of sample 1, the first P_B is the recognition rate of sample 3, and so on, a total of 9 prediction distinction data are obtained.

4. Conclusion

This paper studies the English learners' understanding and cognition of the speech with accent from the listener's perspective through perception method, using scientific instruments to collect speech signals, by using modern speech science and technology to process the collected signals to extract the speech parameters, and then establish an acoustic database. Introduce the latest methods and techniques of modern phonetics into the study of English pronunciation of ethnic minority students, analyze the speaker's phonetic errors, explore the root causes of phonetic difficulties, design a speech perception experiment from the perspective of the listener, and explore speech perception and output mechanism, in order to put forward effective teaching strategies to provide new technologies and new methods for the study of minority English phonetics.

Acknowledgment

The successful of this paper was subsidized by Central Universities Fundamental Research Funds for the project: A study of English phonemes of Lhasa Tibetan college students based on perception

(31920190110), and Ethnic Research Project of the State People's Committee: A Comparative Study on English Learning among College Students from Minority Nationalities (2019-GMD-066).

References

[1] Ellis, R. The Study of Second Language Acquisition, Oxford: Oxford University Press, 1994, pp.37-45.

[2] Altmann, H. The Perception and Production of Second Language Stress: A Cross-linguistic Experimental Study, University of Delaware, 2006.

[3] Debaene, M. Perceiving and Producing Native and Non-native Vowels- An Experimental Study on the Effects of First Language Regional Variation, University of Ghent, 2013.

[4] Lin, C.Y. Perception and Production of Five English Front Vowels by College Students. English Language Teaching, vol.11, no.2, pp.70-77, 2014.

[5] Mlinar, R. Pronunciation of English Dipthongs by Speakers of Serbian: Acoustic Characteristics, University of Novi Sad, 2011.

[6] Strange, W. Automatic Selective Perception (ASP) of First and Second Language Speech: a Working Model. Journal of Phonetics, no.12, pp.7-11, 2011.