

Original Paper

An Experimental Study on Societal Factors Affecting VOT of English Plosives

Lin Xiaoyan¹

¹ School of Foreign Languages, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, China

Received: June 15, 2023

Accepted: July 8, 2023

Online Published: July 25, 2023

doi:10.22158/elsr.v4n3p78

URL: <http://dx.doi.org/10.22158/elsr.v4n3p78>

Abstract

Plosives are integral components of English consonants. In phonetics, English plosives are classified into voiceless plosives /p, t, k/ and /b, d, g/. VOT (voice onset time) was defined as “the time interval between the burst that marks release of the stop closure and the onset of quasi-periodicity that reflects laryngeal vibration”. VOT is a significant acoustic feature and analytic parameter of plosives. Referring to Labov’s experimental model of linguistic variation analysis, this study investigates the influences of societal factors have on the VOT of English plosives.

In this study, 15 English words with word-initial voiceless plosives /p, t, k/ and 15 word-initial voiced plosives /b, d, g/ were selected as reading material; meanwhile, 30 subjects were randomly recruited to read, and audio samples were collected. It is found that the two social factors (gender and regional dialect) selected in this experiment have influences in different degrees on the English plosive VOT of the subjects. The specific results are as follows.

For gender, no significant difference exists between males and females, but the mean VOT of females is longer than that of males, which is basically consistent with previous research results. The underlying reasons of the gender VOT differences inferred by this paper can be physiological and sociophonetical. For regional dialects, the VOT of the subjects were primarily influenced by Southwest Mandarin and Min Dialect, in which the mean value of voiceless plosives was higher and the difference was greater for speakers of Southwest Mandarin, and the mean value of voiced plosives was higher and the difference was greater for speakers of Min Dialect. The results of this empirical study theoretically provide some reference for acoustic researches, and pedagogically, provide some implications for optimizations of English curriculums in university.

Keywords

plosive, English plosive, VOT (voice onset time), acoustic experiment, phonetics study, PRAAT, gender, regional dialect

1. Introduction

The thesis undertakes an experimental study on societal factors affecting VOT of English Plosives. As a brief introduction to the thesis, this chapter consists of four parts: background of the study, significance of the study, objectives of the study and outline of the study.

1.1 Background of the Study

Language is first perceived through its sounds. From the structuralism viewpoints, language consists of three parts: speech sound, grammar and vocabulary, of which speech sound is the material shell of language; thus the study of sounds is of great significance in linguistics, especially in phonetics and phonology. In phonetics, speech sounds, all the sounds produced by humans through their speech organs, are defined as the phonic medium of language and taken as the research objects.

In the IPA chart, the sound segments are grouped into consonants and vowels. Crystal (1997) described that consonants are produced “by a closure in the vocal tract or by a narrowing which is so marked that air cannot escape without an audible friction.” The plosive, one sort of consonant, is formed when the vocal tract is completely closed, occupying an important position in all kinds of consonants because it is the only consonant that all languages have; however, there are great distinctions in the classifications and phonological characteristics of plosives in different languages, and hence effective acoustic parameters are needed to classify plosives. Lisker and Abramson (1964) conducted a cross-language investigation of plosives in 11 languages and defined voice onset time (VOT for short) as “the temporal interval from the release burst of the plosives to the onset of the first formant frequency that reflects glottal vibration.” In their study, all plosives are classified into three groups: voiceless aspirated plosives, voiceless unaspirated plosives and voiced plosives. Since then, VOT has been widely used to differentiate plosive categories across languages, and it has come to be regarded as one of the best acoustic cues for discriminating general plosive categories, especially in word-initial position. For English, voicing, rather than aspiration, is considered as the distinctive features of English plosives which phonologically consist of voiceless plosives /p, t, k/ and voiced plosives /b, d, g/.

It is undoubtedly that the concept and measurement benchmark of VOT foster the development of experimental phonetics especially pronunciation test thereinto. At present, researchers at home and abroad have been committed to the development and application of speech analysis software so as to visualize the implicit characteristics of speech. Among those tools, PRAAT is more than widely employed in the analysis, annotation, processing and synthesis of digital speech signals due to its practicability and maneuverability.

From the traditional perspective of structuralism, language is a homogeneous and orderly internal system on which external factors have few affects. On the contrary, William Labov, an American linguist who has taken the lead in sociolinguistics, believes that language is a heterogeneous system that will produce variation under the influence of societal factors. He drew his conclusions from his four famous studies: the Martha’s Vineyard survey, the New York survey, the Black-English Dialect Study and the Philadelphia Survey. According to Labov, language and society are two independent entities; the

responsibilities for sociolinguists are to connect linguistic facts (phonology, lexicology, syntax...) with social facts (social class, gender, age, occupation...) and to probe their correlations.

Having realized the significance of speech sound research, some linguists have done some relevant studies. In the aspect of plosive research, for instance, Zheng Xianri and Li Yinghao (2007) conducted a study on the comparison of VOT between English and Chinese and the acquisition of English plosives by Han students; Zhang Jinyu (2012) conducted an acoustic analysis of VOT of French plosives among Chinese students who had acquired French and found the influence of “negative transfer”. However, the implications some societal factors can have to speech sounds have not received its deserved attention. Hereby, this paper presents a hypothesis that there are some connections between VOT of English plosives and social factors, which will be statistically demonstrated in the following chapters.

1.2 Significance of the Study

This research is an empirical acoustic study on the basis of interrelated English phonetic theories and presented plosive research results, which fundamentally employs theories and methodologies of sociolinguistics and phonetic analytical technique of PARRT. And the significance of this study will be stated from the theoretical and practical perspectives as follows.

Firstly, the theoretical values. This survey is based on the VOT theoretical model proposed by Lisker and Abramson, which can promote the further study of phonetics in VOT. In addition, still insufficient studies on English plosive VOT at home and abroad have been shown, and hence this research can help bridge this gap. Moreover, there are few studies combining experimental phonetics and sociolinguistics although which both belong to field of linguistics but thoroughly two different branches and research directions. This survey explores the societal factors influencing English plosive VOT, reflecting the innovation of linguistic research and enriching the theoretical vacancy of phonological sociality.

Secondly, the practical significance. This acoustic empirical research conducted with sound segmenting software is based on manipulation, which to some extent reflects scientific spirit of experimental verification. Then, social investigation and the quantitative research method of sociolinguistics is included in the experimental process of this project, which is conducive to exploring the effects of multiple social factors on VOT of English plosives in different degrees. What is more, based on the overall acquisition situation, the study investigates and analyzes the influences of different social factors, helping put forward correlative suggestions, for example, improving pedagogical methods and teaching techniques to enhance the English acquisition levels of Chinese college students.

1.3 Objective of the Study

In a macroscopic sense, as is mentioned above, there are insufficient researches exploring the functions societal factors have on speech sounds, especially on English plosives. Therefore, an experiment is required to be designed to investigate the degrees of functioning of different social factors on English plosives; thereinto, VOT will be employed as an essential acoustic parameter to accomplish accuracy and objectiveness of measurement. Ultimately, the production of plosive by English learners in China and the analysis results will be exhibited in the presented paper.

In a microscopic sense, firstly, consonant, which is secondary to vowel but bearing more than important values in acoustic study, is lack of empirical research. To explore this sound category further, the study aims to extract and compare the acoustic features of one sort of English consonant—plosive of Chinese people under the effects of different social factors. Moreover, as aforementioned, sociolinguistics connects the linguistic facts and societal facts; but the noteworthy thing is that the specific societal facts which can actually function on linguistic facts and their corresponding functional ways, which are the main questions this study aims to answer.

1.4 Outline of the Study

The thesis consists of five chapters to have a comprehensive analysis of the societal factors influencing VOT of English plosives on the basis of phonetics experiment data. And the first chapter, the presented one, serves a brief introduction to the whole thesis.

Chapter 2 is basically a literature review on basic concepts, theoretical framework, empirical research and summary. To put it on detail, it includes the definition and classification of plosives, the introduction to English plosives, the illustration of voice onset time (VOT) and Labov's theories. Results of existing researches will be exhibited as well.

Chapter 3 describes the methodology of this research. Thereinto, research questions, quantitative research method and experimental designs will be demonstrated. Specifically, the designs include subjects, materials and procedures.

In Chapter 4, the experiment results will be focused. The data will be analyzed and further discussed on the basis of two sets of variations, namely two different societal factors: gender and regional dialect.

The last chapter provides the conclusion of the entire research, where major findings will be presented. The limitations of this study and suggestions for the future researches will also be elaborated.

2. Literature Review

This chapter will firstly introduce the basic concepts of the study including plosives and English plosives. Then, the theoretical frameworks adapted in this study will be presented; to put it specifically, the introduction of voice onset time and language transfer theory will be included. Lastly, the previous empirical studies concerning plosives, both at home and abroad, will be reviewed as well.

2.1 Basic Items

2.1.1 Plosives

Plosives, a sort of consonants formed when the vocal tract is completely closed, play an essential role in phonetics. Some phoneticians have pointed out that “plosives are the only kind of consonants that occur in all languages” (Ladefoged & Maddieson, 1996) and “plosives show the characteristics of consonants best”; hence, studies on plosives are of great significance.

2.1.1.1 Definitions of Plosives

Plentiful definitions of plosives have so far been given.

Sloat, Taylor and Hoard (1978) defined that “plosives are made by a complete blockage of the airstream at some points in the vocal tract. The blockage does not produce the sounds alone; rather it is the formation and the release of blockage which produces sounds”.

Liu Zhengyi (1990) described that “plosives are formed by a complete closure of the air-passage and then the air abruptly issuing on the release of the closure”.

Wang Guizhen (2002) indicated that the plosives are the sounds which were made by completely hindering the airflow at some points in the mouth, and releasing it into the sound that follows.

Gussenhoven and Jacobs (2011) also stated that “plosives are formed by a complete closure in the speech tract, behind which the air from the lung is compressed until the closure is suddenly released and the air explodes outwards”.

As all shown in the aforementioned definitions, when the obstruction is created by the speech organs in total or complete, the speech is produced with the obstruction audibly released, and the air passing out again. On this condition, the plosives are produced.

2.1.1.2 Articulation of Plosives

Roach (2000) indicated that the general articulation process as follows: first, one or two articulations are moved against each other so as to form a stricture that allows no air to escape from the vocal tract. Then, after this stricture has been formed and air has been compressed behind it, it is released, that is, the air is allowed to escape.

The corresponding articulation phases of plosives to the above process are indicated by Zhou Weijing (2018). These phases are closure, hold and release. Then, the following figure is conducive to illustrating the whole articulation of plosives comprising three phases.

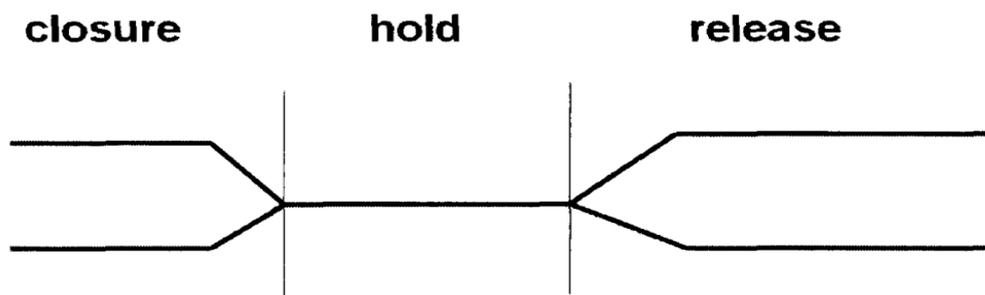


Figure 1. The Articulation Process of Plosives (Zhou Weijing, 2018, p. 103)

The first phase is the closure phase, during which the articulating organs move together to form the obstruction. In this stage, one articulator is moved against another, or two articulators are moved against each other to form a total stricture allowing no air to escape from the oral tract.

The second phase is the hold phase. It is a stage during which the lung actions compress the air behind the closure. The compressed air may or may not be accompanied by voicing, that is, the vibration of the vocal folds.

The third phase is the release phase, during which the organs forming the closure rapidly allow the compressed air to abruptly escape. In this stage, the air escape will produce noise loud enough to be heard, namely, an explosion, and hence the name “plosives” are obtained.

2.1.2 English Plosives

2.1.2.1 Classifications of English Plosives

English consonants can be classified in two ways: in terms of manner of articulation and in terms of place of articulation; whereas, although plosives are consonants, they can be categorized from more different perspectives, owing to their phonetic characteristics.

In terms of manner of articulation, all English plosives belong to obstruents, meaning complete closure without airstream escaping from the mouth (Cruttenden, 2001). But, with respect to place of articulation, English plosives can be divided into three parts: bilabial plosives, alveolar plosives and velar plosives (Roach, 2000). Thereinto, /p/ and /b/ are bilabial plosives, because the lips will be pressed together while articulating; /t/ and /d/ are alveolar plosives, on account that during the articulation the tongue blade will press against the alveolar ridge; /k/ and /g/ are velar plosives, and the back of the tongue will press against the position where the hard palate ends and the soft palate begins.

In line with the force of articulation, English plosives comprise fortis and lenis (Cruttenden, 2001). While English plosives are pronounced, /p, t, k/ tend to be articulated with more muscular energy than /b, d, g/, so as stronger breath effort; thus, /p, t, k/ are demonstrated as fortis and /b, d, g/ are lenis.

According to the work of vocal folds, English plosives can be classified into voiceless plosives and voiced plosives (Zhao, 2007). The former are /p, t, k/ because they accompanied no vibration from the larynx; /b, d, g/ belong to the latter, and the vocal cords vibrate while articulating these three plosives.

Table 1. The Classification of English Plosives

Plosives	Place of Articulation	Work of Vocal Cords	Force of Articulation
/p/	Bilabial	Voiceless	Fortis
/b/		Voiced	Lenis
/t/	Alveolar	Voiceless	Fortis
/d/		Voiced	Lenis
/k/	Velar	Voiceless	Fortis
/g/		Voiced	Lenis

2.1.2.2 Phonetic Features of English Plosives

It is worth emphasizing that voicing and aspiration are two essential phonetic characteristics of plosives in phonetics and phonology, which can be shown in line with manners of articulation mentioned above. Voicing refers to the articulatory process during which the vocal fold vibrates. Cruttenden (2001) stated that "...at any place of articulation, a consonantal articulation may involve the vibration of the vocal folds, that is, may be voiceless or voiced. In this situation, voicing comes into being". In English plosives, voicing is defined as the distinctive features, that is, in the same phonetic context; the concrete meaning of words can be distinguished by voicing. For instance, in minimal pairs like "big" and "pig", "tap" and "gap", the voiced plosive /b/ (/t/) and the voiceless plosives /p/ (/g/) form phonetic contrast, namely, they only differentiate in the voicing, but they somehow show the different meaning of two words.

As for aspiration, it refers to the strong burst of airflow accompanying the release or the closure of some obstruents in the case of preaspiration. Ladefoged and Johnson (2015) defined aspiration as "aspiration is a period of voicelessness after the plosive articulation and before the start of the voicing for the vowel". Aspiration is not the distinctive features for English plosives, it is to say, the meanings of words will not change whether the plosives aspirate or not. Raise "speak" as an example, aspiration of /p/ is absent when pronouncing because it is preceded by s in a stressed syllable; whereas, whether the /p/ is aspirated or unaspirated, the meanings of the original word have no change, because the aspirated [p^h] and the unaspirated [p] are simultaneously the allophones, namely, the phonetic variants of the phoneme /p/. English and Chinese belong to two different phonological categories. As aforementioned, the distinctive feature of English is voicing rather than aspiration; nonetheless, in Chinese, aspiration is the distinctive feature, not voicing.

2.2 Theoretical Framework

2.2.1 VOT

Voice Onset Time (VOT), is a temporal feature of plosive productions and one of the most commonly used acoustic cues in phonetic analysis.

Lisker and Abramson (1964) tested four American English speakers in an acoustic experiment and published the article *A cross-language study of voicing in initial stops: Acoustical measurements* to describe the characteristics of plosives using the concept of voice onset time (VOT) which was defined as "the time interval between the burst that marks release of the stop closure and the onset of quasi-

periodicity that reflects laryngeal vibration”. Since then, a considerable number of studies of other languages have been undertaken, and VOT has come to be considered as one of the most important methods for examination of plosive voicing time (especially in word-initial position). They also suggested that each plosive category falls into one of three ranges respectively: -125 to -75 ms (millisecond), 0 to $+25$ ms, and $+60$ to $+100$ ms. In accordance with Lisker and Abramson’s categorization, both English and Mandarin fall into the group of two-category languages, occupying the same range along the VOT continuum, that is, 0 to $+25$ ms for $[p, t, k]$ and $+60$ to $+100$ ms for $[p^h, t^h, k^h]$. The VOT value of plosives thus becomes the only scale being able to achieve plosive classifications according to voicing contrast. Afterwards, Cho and Ladefoged (1999) classified more specifically the range for voiceless aspirated and unaspirated concentrating particularly on velar plosives across 18 languages. They distinguished four categories, which they dubbed unaspirated (around 30 ms), slightly aspirated (around 50 ms), aspirated (around 90 ms) and highly aspirated (over 90 ms).

MacKay’s research (1984) showed that the VOTs of plosives have positive and negative values. When voicing begins subsequent to the burst of airflow, VOT is positive and called “voicing lead”; when voicing precedes the burst of air, VOT is negative and called “voicing lag”. In cases when the two co-occur, VOT is zero. Keating (1987) further divided voicing lag plosives into “long lag” and “short lag” plosives, with the VOT of the former greater than 35 ms and of the latter between 20-35 ms; thus, plosives then fall into three categories: voiced, voiceless unaspirated, and voiceless aspirated. She (1987) also developed theory of “polarization” indicating that in a language, the VOT values of different plosives should be in different ranges, to achieve different plosive categories whose manifestation is the VOT central values of different plosives should be separate from other else.

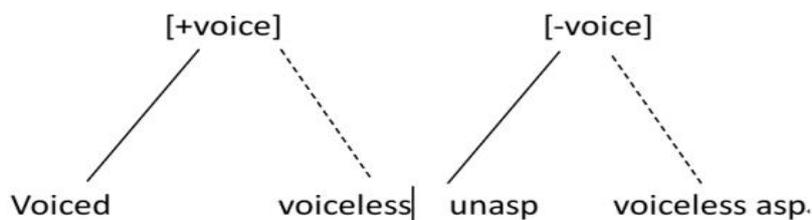


Figure 2. The Categories of Plosives (Keating, 1987, p. 309)

2.2.2 Language Transfer Theory

The word “transfer” is originally a psychological term. Odlin defined transfer as “a sort of influence, which comes from the similarity and the difference between L1 and L2”. The concept of language transfer was formally proposed in the comparative analysis theory that emerged in the 1940s. Since the 1950s, various interpretations of the concept of language transfer were indicated by many famous scholars such as Lado, Stockwell, Schachter and Ellis. Generally, they thought in the process of second language acquisition, learners’ habits of using their first language will have direct effects on the acquisition of second language, which plays a positive role in promotion or negative in interference; those which have positive influences on SLA are called “positive transfer”, and the counterpart is called “negative transfer”. For instance, Chinese and English are distinct from each other no matter in pronunciation, semantics or sentence patterns, and the English students whose first language is Chinese will presumably encounter various difficulties, which is the negative transfer of mother language that has interference with their English learning.

2.3 Empirical Research

2.3.1 Empirical Research at Home

The acoustical researches of plosives started from abroad and gradually arose in the phonetics field in China. Pertinent researches in China involves multilingual (not limited to English) experiment and multi-field (such as language comparative study, language acquisition and language teaching).

In terms of bilingual comparative experiment, Zheng Xianri and Li Yinghao (2007) conducted a study comparing of VOTs between English and Chinese, and investigating the acquisition of English plosives by Han Students. It was found that there were presumably language transfers (transfer the Chinese pronunciation patterns to that of English) during the generation of English voiced plosives which are dissimilar from Chinese. Wu Gang (2012) also experimentally compared VOTs of English and Chinese plosives, and shed light upon some enlightenment for phonetic instructions.

For the acoustic experiments concerning other languages, Zhang Jinyu (2012) conducted acoustic analysis of plosive VOT with Chinese students who had learned French, indicating that there were some errors that the VOTs were excessively small and tend to be affected by language negative transfer of the unaspiration voiceless plosives in Chinese. Su Jiajia (2012) conducted a research on the VOT variations of Korean plosives based on speech corpus, propounding suggestions for Korean teaching in China. For the acoustic experiment of Chinese English learners, a phonetic analysis was conducted by Wang Maolin (2009) to study the English word-middle plosives of Chinese learners. Dai Zhengzheng (2012) studied the plosive VOT value of advanced English learners in China. Gao Yujuan and Zhang Mengmeng (2018) carried out an experimental study on the acquisition of English plosives by native Chinese speakers, verifying Flege’s learning model theory that similar plosives are more difficult to acquire for advanced learners, and even “fossilization” appears.

Societal factors will be regarded as independent variables in this study to investigate the differences in plosive VOT among English learners in China. Whereas, it is found that the current studies on social

factors affecting VOT in China primarily concentrate on regional dialect differences, a bulk of which discussing the language transfer. Jiang Yuyu (2015) took English learners using Wu Dialect and Jianghuai Mandarin as experimental subjects and conducted a comparative study on VOT between vowels, revealing that the VOTs of voiceless plosives in stressed syllables produced by the subjects were distinct from that of native speakers. An acoustic study on the English plosive production of Dalian students conducted by Wei Zhaozhao (2018), and found that differentiation from that of English native speakers existed as well, with significant pronunciation deviation mainly affected by dialectical negative transfers. Zhao Chenyang (2019) carried out a study on fricative VOT generated by English learners in Harbin for the similar purpose as the above. It is noteworthy that studies on other societal factors such as gender, age and university major, are insufficient.

2.3.2 Empirical Research Abroad

Compared with domestic studies, foreign studies started earlier and are more abundant. Different social factors are taken as variables in the acoustic study.

Take gender factor as an example. Bradford L. Swartz (1992) started to survey the gender differences in VOT. In a videoen-doscopic study conducted by Cooke, a tendency was found that for males to produce voicing onset sooner than females, indicating that the distance between the vocal processes of males is less than females. John Ryalls, Allison Zipprer and Penelope Baldauff (1997) conducted an investigation of the effects of gender and race on VOT, finding that the mean VOT values of female were higher than that of male. Whiteside and Marshall (2001) published an article in *Phonetica* discussing the influences of gender differences on VOT. Morris, McCrea and Herring (2008) studied the VOT difference between male and female on isolated syllables, and published their findings in *Journal of Phonetics*. Robb, Gilbert and Lerman (2005) investigate the influences of gender and social environment on VOT. Eunjin (2011) studied the gender difference of pause VOT in Korean. Bradford L. Swartz (2014) investigated gender differences in VOT, and so on.

In terms of age factor, Katrin Stolten, Niclas and Kenneth (2014) studied the influences of ages and speaking rates on voiceless pauses of L2 learners. Judit Bona (2015) discussed the correlations between age and VOT.

There are also a few relevant studies concerning regional factors. For instance, EL Zen (2020) studied the influences of Indonesians with different regional language backgrounds on the production of plosive VOT in English.

2.4 Summary

In summary, the acoustical researches on VOT plosives are plentiful. The domestic studies have involved multiple languages, and the effects of societal factors in the speech productions are further studied in the researches abroad. Whereas, although many previous studies have examined the VOT of plosives and some influential factors, few studies focus on the acoustic measurement of phonetic variations caused by some societal factors such as different dialects in China, majors in university, and so on. On the basis on

theoretical results at home and abroad, this study will combine the empirical research on English plosive VOT with societal factors to enrich the existing researches theoretically and practically.

3. Methodology

In this chapter, research methodology will be addressed. Specifically, experiment subjects, instruments and materials, procedure and analysis methods will be introduced.

3.1 Subjects

This research studies the correlative societal factors to the VOT of English plosives with high openness, and hence no specific limitations on the conditions of subjects are required. After collection, examination and screening, 30 valid samples were remained; it is to say, in all the recruited participants, there are 30 speech samples which are eligible for acoustic experiments and analysis. These 30 subjects are all Chinese EFL (English as a Foreign Language) learners, whose times of English-learning fall in the range of 5 to 20 years, including 15 males and 15 females. Generally speaking, they all come from diverse regions in mainland China, master distinct dialects and have different economic backgrounds, which is in accordance with the randomness of social survey. What should be underscored is that EFL learners whose major is English in college or university are excluded in this research in order to minimize the influences of pertinent phonetic learning or training and certify the veracity of experiment.

3.2 Instruments and Material

3.2.1 Analytic Instruments

In this study, acoustic methods are adopted to analyze the production of English plosives by Chinese EFL learners. The implicit features of speech sounds are difficult to be revealed by sole listening, but they can be represented through visualization technology. PRAAT (named “Doing Phonetics by Computer”), an effective acoustic analytic tool, is employed to extract and annotate the sample data displayed in the spectrogram and waveform of sounds. Compared with other acoustic analytic software, PRAAT is free of charge with a small space (less than 10m), strong universality, easy manipulation and timely update. It is mainly used by phoneticians for acoustic analysis, annotation, processing and synthesis of digital speech signals, and the generation of various speech graphs and text reports.

Additionally, SPSS (Statistical Product Service Solutions) will be employed to assist analysis processing, in which descriptive and correlational analyses were conducted to evaluate the influences of different societal factors on VOT of English plosives.

3.2.2 Reading Material

As for the reading material which is displayed in the following table, regarding the feasibility, monosyllable words with plosives in the initial position are chosen in this study. It needs to be emphasized that only word-initial plosives are studied in this survey due to some arguments contending that “VOT focuses narrowly on the timing of voicing in word-initial plosives and does not take into account plosives in word-final and word-medial positions” (Docherty, 1992).

In the respect of the phonological structure, all of the elected words will start with a plosive, keeping a vowel as the nuclear and a consonant as the coda. The plosives are respectively the voiceless plosives /p/, /t/, /k/ and the voiced plosives /b/, /d/, /g/; and the pure vowels chosen are /i:/, /æ/, /ʌ/, /ɔ:/, /ɑ:/, representing the front, middle and back vowels in several, in accordance with the place of articulation of vowels, one perspective of vowel classification, in order to preclude the influence of vowel environment. Additionally, definite article “a” is placed before all the words, so as to eliminate extra interference in identifications of plosives like rapid speech rate or noise; therefore, the “a” has no grammatical effect but phonetically separative function, namely, some of these phrases have no semantic meanings.

Table 2. Reading Material

VP PP	/p/	/t/	/k/	/b/	/d/	/g/
/i:/	a peak	a team	a key	a beef	a deed	a geek
/æ/	a pat	a tap	a cat	a bat	a dad	a gap
/ʌ/	a pub	a tub	a cup	a bug	a duck	a gut
/ɔ:/	a port	a torch	a call	a ball	a door	a gorse
/ɑ:/	a park	a tar	a car	a bar	a dark	a garb

VP: vowel phoneme

PP: plosive phoneme

3.2.3 Questionnaire

Since this is a survey of the influences of social factors, voice sampling is solely one of the integral steps, and it is critical that the social conditions of the testees should be investigated.

In this experiment, the basic information of the subjects is collected by the means of questionnaire in which questions are all set to survey the societal backgrounds, and the questionnaire results will correspond with the recordings individually. On this basis, the effects of social factors on VOT of plosives will be analyzed. The questions include: age, gender, place of residence, annual family income, and EFL duration.

3.3 Procedure

This experiment generally encompasses the following three steps.

First of all, after material selection, question setting and questionnaire design, the questionnaires are distributed, and the subjects are recruited to complete them.

In the second place, after opening the questionnaire, the subjects first answered the questions in several, and then read the materials subsequently given and record. It should be noted that due to the limitation of insufficient conditions like displacement and equipment lack, professional recording software such as PRAAT and *Cool Edited* cannot be adopted in the experiment, namely, the sampling will be accomplished by using the recording tools of the subjects' mobile phone per se. Before recording, the subjects were

required to keep an opportune distance (5-10 cm) from the microphone of the phone, averting air blasting and keeping the volume and speed moderate to ensure the sound quality and clarity of the recording for subsequent analysis. Upon recording, the subjects upload the audio and submit it to complete a sample collection.

Afterwards, a certain number of collected samples are integrated. The off-specification samples are screened, and the qualified and analyzable ones are converted and imported into PRAAT in turn. The spectrogram of the audio of each subjects are generated; then, the VOT value is calculated and marked. The two social factors selected as dependent variables in this experiment are gender and regional dialect. Lastly, on preliminary annotations, group the subjects and the samples in light of the above factors for further analysis.

3.4 Data Analysis Methods

Generally, quantitative and qualitative analyses are adopted as research paradigms of this survey. The integral part of the former is the determination of independent variables (different societal factors) and dependent variables (selected plosives). Qualitative analysis aimed at the comprehensive analysis of the research objects, combining methods of induction and deduction, to reveal the essence nexus of the two variables.

As aforementioned, PRAAT and SPSS are employed as analytic tools in this study. Firstly, PRAAT is chiefly used for spectrogram generations and data annotations. On account of the definition and classification of VOT, the VOT values corresponding to each voiceless plosives and voiced plosives are separately calculated and recorded. While pronouncing each unit of the reading material, the vocal cord vibration caused by the front vowel phoneme /ə/ will be followed by the word-initial plosives, and the process embodies in the spectrogram as the duration from the end of the formant to the beginning of the subsequent first spike. After observation, the VOT values of voiceless and voiced plosives are individually recorded, marked and extracted.

Upon data recording, the results are grouped in accordance with the independent variables, namely, different societal factors (gender, region and financial background); then, the mean values of VOT of each group will be calculated for respective contrast analysis; for instance, in the respect of the region factors, the subjects are grouped according to their resident regions and mastered dialects, and the inference of its effects will be made. At last, correlation analysis of the comparative results of each societal factor will be conducted using SPSS to complete the final evaluation of the degree of influences.

4. Results and Discussions

This chapter is to present the experimental results of VOT of English plosives of the 30 subjects and the analysis and discussion, to probe the influences the societal factors (gender and regional dialect) have on the English plosive VOT.

4.1 Results

The measurements of VOT of English plosives for the 30 subjects in this experiment are displayed in the following table. As mentioned above, it is the voicing contrast that distinguishes the lexical meanings of words in English, which determines the different measurement methods and criteria for voiceless plosives and voiced plosives; therefore, these two sorts of plosives will be gauged and shown separately. The unit is ms (millisecond).

Table 3. The Results of VOT of Voiceless Plosives (individual mean values)

110.7	75.3	108.2	83.4	107.1	106.2	95.7	93.9	99.7	79.9
82.5	83.3	119.8	92.5	70.9	94	92.5	54.9	87.6	119.7
104.4	102.3	99.8	99.9	84.7	108.4	125	136.9	81.1	82.8

Table 4. The Results of VOT of Voiced Plosives (individual mean values)

29.3	16	17.3	15.6	14.6	15.6	18.6	13.5	13.3	10.5
19	16	15.9	17.1	16.9	15.7	18.3	17.1	20.6	21.1
18.3	15.7	20.5	15.9	14.3	20.6	15.1	17.6	13.5	14

Two societal factors were extracted from the questionnaire questions: gender and regional dialect. And in this section, the influences of these two factors on the VOT of English plosives of all the subjects will be shown. SPSS will be employed as the analytical tool in this study, in which the methods of analysis will be used mainly, encompasses independent sample T test and ANOVA.

4.1.1 Gender

Since gender is a factor with few variables, namely, biologically, there are merely two kinds of gender, male or female, independent sample T test is used for analysis.

4.1.1.1 Voiceless Plosives

Table 5. The Results of VOT of Voiceless Plosives

variable name	variable value	sample size	mean value	standard deviation	tvalue	p value	difference value of the mean value	Cohen's d value
mean value of English voiceless plosive	male	15	93.94	14.505				

VOT

female	15	98.267	19.986				
total	30	96.103	17.298	-0.679	0.503	4.327	0.248

Chart illustrations: The table above shows the results of independent sample T test, including the results of mean \pm standard deviation, T-test results, significance p-value and Cohen's D value of effect size. Each item should be estimated whether it is less than 0.05 or not. If it is significant, the null hypothesis is rejected, indicating that there factually is significant difference between the two groups of data (the difference can be analyzed according to mean \pm standard deviation); otherwise, it indicates that no difference exists between the two groups.

The p value of F test was $0.503 \geq 0.05$, so the statistical result was not significant, indicating that there was no significant difference in the mean VOT of individual voiceless plosives between male and female. The Cohen's D value is 0.248, explaining a difference range (0.20, 0.50 and 0.80 correspond to small, medium and large critical points respectively).

The mean values of VOT of male and female of voiceless plosives are 93.94ms and 98.267ms.

In terms of mean VOT values and standard deviation, the VOT of English voiceless plosives is longer in women than in men, and the divergence is greater in woman as well.

4.1.1.2 Voiced Plosives

Table 6. The Results of VOT of Voiced Plosives

variable name	variable value	sample size	mean value	standard deviation	t value	p value	difference value of the mean value	Cohen's d value
mean value of English voiced plosive VOT	male	15	16.613	4.113				
	female	15	17.22	2.601				
	total	30	16.917	3.395	-0.483	0.633	0.607	0.176

The p value of F test is $0.633 \geq 0.05$, indicating that no significant discrepancy exists between male and female in the average VOT of individual voiced plosives. The Cohen's D value is 0.176, demonstrating a small difference.

The mean values of VOT of male and female of voiced plosives are 16.613ms and 17.22ms, respectively. On average, VOT of female is slightly longer than VOT of male; but in terms of standard deviation, the data distribution for men is more discrete, namely the difference is larger.

Combining the measurement results of VOT of voiceless plosives and voiced plosives, the consequence that female has a longer VOT of English plosives than male can be seen.

4.1.2 Regional Dialect

To study Regional dialect factors, that is, to investigate whether regional dialects have an impact on plosive productions and which kinds of dialects have a greater influence. In this experiment, the selection of subjects is disordered. On the basis on the questionnaire questions "household registration location" and "most commonly used dialect", six dialect types that 30 subjects have mastered and have potential influences on their speech output are integrated.

In accordance with the divisions for historical and geographical reasons, there are seven dialect regions in China: mandarin dialect or official dialect region, Wu Dialect region, Gan Dialect region, Xiang Dialect region, Min Dialect region, Cantonese region and Hakka Dialect region. Each dialect area is divided into several sub-dialect areas; for instance, mandarin dialects can be subdivided into northeast Mandarin, Jiaoliao Mandarin, Beijing Mandarin, Jilu Mandarin, Central Plains Mandarin, Jianghuai Mandarin, Lanyin Mandarin, and southwest Mandarin.

In the analysis, the codes of different sorts of dialects are as follows: Min Dialect is 1, Jianghuai Mandarin and Wu Dialect are 2, Southwest mandarin is 3, Lanyin Mandarin is 4, Zhongyuan Mandarin is 5, and Cantonese is 6. The codes are in line with the number of people, decreasing from 1 to 6.

In this factor analysis, one-way ANOVA was employed to compare the mean values and standard deviations of plosive VOT to determine which regional dialects have relatively significant influences on plosive VOT, and to analyze the correlated reasons for the effects. Each item should be analyzed to check whether it is less than 0.05.

4.1.2.1 Voiceless Plosives

Table 7. The Results of VOT of Voiceless Plosives

variable name	variable value	sample size	mean value	standard deviation	F value	p value
	1.0	13	96.838	17.823		
mean value of English voiceless	2.0	6	87.233	11.755		

plosive VOT	6.0	2	95.350	18.173		
	4.0	3	104.333	4.038		
	5.0	2	83.300	17.536		
	3.0	4	107.625	25.575		
total	30		96.103	17.298	1.037	0.419

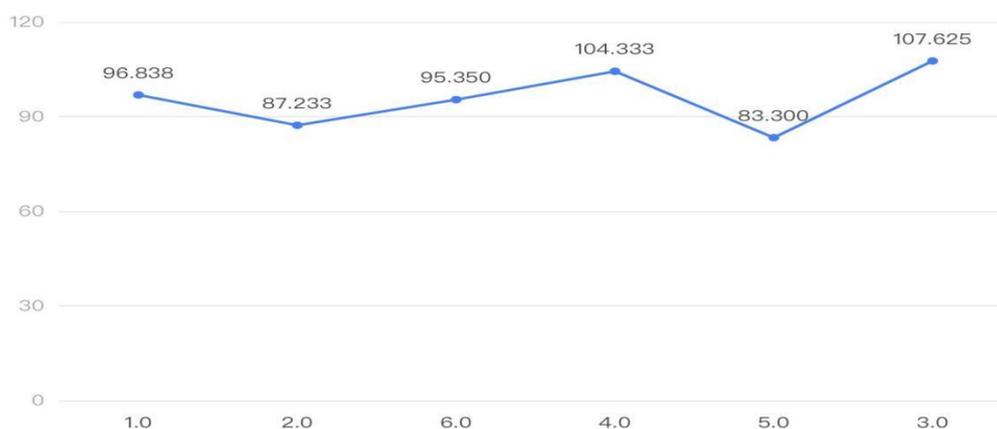


Figure 3. The Results of VOT of voiceless Plosives

The p value of variance analysis is $0.419 > 0.05$, indicating that the statistical result is not significant, that is, there is no significant difference in the mean VOT of individual voiceless plosives in these six groups of different regional dialects.

From the average, the mean values of VOT of English voiceless plosives from Groups 1 to Group 6 are respectively 96.838ms, 87.233ms, 95.350ms, 104.333ms, 83.300ms, 107.625ms; the VOT of Group 3 is the longest, followed by Group 4 and then the Group 1.

In terms of standard deviations, the value of Group 3 is the largest, 25.575ms, it is to say, the dispersion degree is the highest and the difference is the largest within the groups, subsequent to which is the value of Group 6 (18.173ms). The smallest standard deviation value is in 4, 4.038ms, indicating the smallest discrepancy.

4.1.2.2 Voiced Plosives

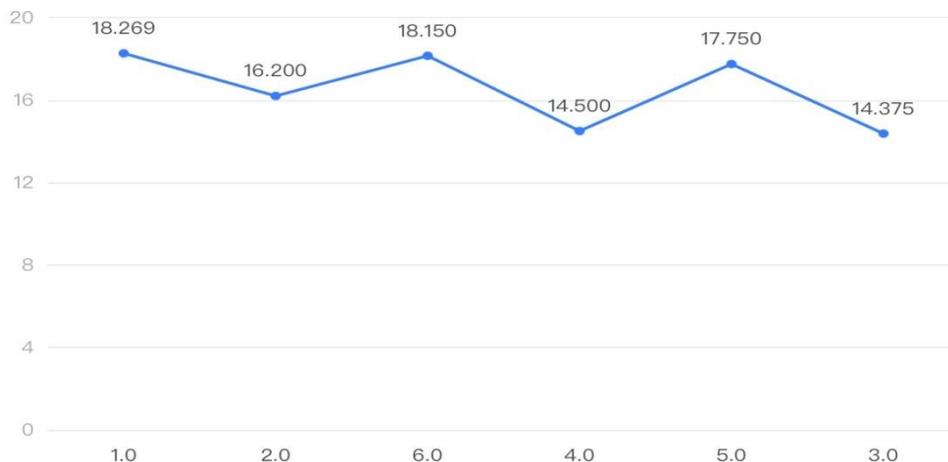


Figure 4. The Results of VOT of Voiceless Plosives

Table 8. The Results of VOT of Voiced Plosives

variable name	variable value	sample size	mean value	standard deviation	F value	p value
mean value of English voiced plosive VOT	1.0	13	18.269	4.013		
	2.0	6	16.200	2.545		
	6.0	2	18.150	1.202		
	4.0	3	14.500	1.153		
	5.0	2	17.750	1.202		
	3.0	4	14.375	3.083		
	total	30	16.917	3.395	1.380	0.267

The p value of ANOVA is $0.267 > 0.05$, by the same token, insignificant result is shown, demonstrating that no significant difference in the average VOT of individual voiced plosives in these regional dialect groups is detected.

From the average, the mean VOT values of voiced plosives of Group 1 to Group 6 are respectively 18.269ms, 16.200ms, 18.150ms, 14.500ms, 17.750ms and 14.375ms. The value of Group 1 is the longest,

Group 6 comes the second, and then is Group 5. The value of Group 3 is the shortest. In terms of standard deviation, the value of Group 1 is the largest, 4.013ms, that is, the dispersion degree is the highest; and the subsequent one is the value of Group 3, which is 3.083ms.

In general, in terms of English voiceless plosives, the mean VOT value of Group 3, the Southwest mandarin, is the longest and the difference is the largest. In the aspect of voiced plosives, Group 1 has the longest mean VOT value and the major divergence; the values of Group 2, however, perform equivalently in VOT of voiceless plosives as well as voiced plosives, with small mean value and subtle difference. Then, the following analysis will focus on the group 3 and the group 1 which have greater influences on English plosive productions.

4.2 Discussions

4.2.1 Gender

In terms of English voiceless plosive VOT, from the perspective of T test, the p value is 0.503ms, indicating that there lays small difference in the average VOT of individual voiceless plosives between male and female. In the respect of mean values and standard deviations, the VOT of voiceless plosives in females is longer than that in males, and the difference is larger.

In terms of English voiced plosive VOT, the p value is 0.633ms, indicating that the difference of VOT between male and female in voiced plosives is simply small. On average, VOT of female is slightly longer than VOT of male; in the aspect of standard deviations; nevertheless, the data distribution for male is more discrete.

The above data shows that the VOT of no matter voiceless plosives or voiced plosives is longer in females than in males, which is consistent with some previous studies on the effects of gender on VOT. For instance, Swartz (1992) found that significantly longer VOT values produced by female speakers for voiced or voiceless alveolar plosives than male speakers did on the same speech production task. Although there lays a variety of gender-linked acoustic diversities in speech, a satisfactory explanation concerning the reasons why a temporal measure like VOT is supposed to be different across genders has not been offered. More researches into these issues become warranted; meanwhile, some reasonable assumptions can be also made, which are as follows.

Physiologically, gender differences in respiratory function, laryngeal and supralaryngeal behavior for productions of speech have been documented. Presumably, fundamental biological discrepancies are reflected by the gender differences figured out in VOT production in the diverse speech subsystems. Take female productions of prevocalic plosives as an example, it is associated with greater oral pressure compared to productions of male. Additionally, a characteristic feature of female phonation noted by Bless and Abbs, serves as a posterior glottal opening or chink which occurs within the vocal fold closing duration and results in a leak of airstream into the supralaryngeal space. Within the context of female VOT productions, there are presumably more air escapes into the supralaryngeal space prior to vowel productions, leading in greater oral pressure for plosives compared to male addressers. The greater oral pressure results in longer intervals between plosive burst and voicing onset (Michael, 2005).

Although biological differences can be the underlying consequences of gender differences in VOT productions, there can be some sociophonetic factors that presumably contribute to VOT differences as well. For instance, VOT is correlated negatively with increases in speech rates or word durations; it is to say, slower speaking and word producing with a longer duration can result in longer VOT values, especially for voiceless plosives. Thereby, gender differences would also be conspicuous in VOT productions if they exist in speaking rates. Cheshire (2005) stated that females and males differ in their general pronunciation patterns, with females have the inclination to employ more carefully articulated speeches and adopt this speaking style, and thus, for the sex-related VOT patterns, it is possible that speaking rate differences and word duration are the other responsible reasons in addition to biological discrepancies.

4.2.2 Regional Dialect

The above data shows that in terms of regional factors, the p values of VOT of voiceless and voiced plosives are 0.419 and 0.267, being both larger than 0.05, indicating that there is no significant difference among the dialect groups. Nonetheless, what is noteworthy is that, in the process of comparative analysis, it can be found that the mean VOT of plosives in Southwest Mandarin group and Min dialect group are relatively longer, and the corresponding dispersion degrees, namely, the differences are greater; hence, the pronunciation characteristics of the aforementioned two dialects and the potential reasons for their correlated influences will be analyzed in the following part.

It should be emphasized that China dialects are variants of standard Chinese. Chinese belongs to Sino-Tibetan language family, and English is an Indo-European language; thereby, their phonological systems are distinct. In terms of consonants, although Lisker classifies both Chinese and English as second-category languages according to the parameter VOT, they factually have different distinguishing features. Specifically, there is no voiced consonant in Chinese, of which the word meanings are determined by aspiration or not, and [b], [d], [g] in Chinese is estimated to be close to [p], [t], [k] in English. For English, the distinguishing feature is the voicing contrast, rather than aspiration contrast. Studies have shown that the interference of Chinese initials to English consonant productions is greater than that of finals to English vowel pronunciations.

4.2.2.1 Southwest Mandarin

Southwest Mandarin is a branch of Mandarin Chinese, being mainly distributed in Sichuan, Chongqing, Yunnan and Guizhou areas, as well as most of Hubei, northwestern Hunan, northern Guangxi and southern Shaanxi, with some dialect islands in other regions. The definition of Southwest Mandarin given in *The Atlas of Chinese Language* is “the Chinese dialect used in southwest China and the surrounding areas thereof, with entering tone assigned to a tone value similar to the one used commonly in Chengdu, Wuhan, Chongqing, Changde, Guiyang, Kunming, Guilin”.

English has 28 consonant phonemes, and Southwest Mandarin has 28 to 34 consonants. Although some initial consonants are equivalent to English consonants in written form, they are not thoroughly corresponding; and their pronunciation characteristics are also distinct (Zhu, 2010).

In terms of initials, one of the major characteristics of Southwest Mandarin is that there is no distinction between blade-alveolar and cacuminal, in some cases, cacuminal even can be regarded to be absent. /zh/, /ch/ and /sh/ can be all supplanted by /z/, /c/ and /s/. For example, “sichuan” is pronounced as “sicuan”; “zhi chi” is pronounced as “zi ci”; “li shi” is pronounced as “li si” (Zheng, Li, & Qing, 2012, p. 92). In a precious interview conducted in dialect, the interviewees described themselves as rarely having tongue tip roll-up in daily communication, with a generally low tongue position and relaxed facial muscles. Similar pronunciation is also shown in the initial /r/. In Southwestern Mandarin, the tongue is almost not roll-up during the pronunciation of /r/, which then sounds similar to /z/. In addition, in this sort of dialect, most /b/ and /p/ can be distinguishable, but there are a few exceptions; for instance, generally, “pubian” will be often pronounced as “pupian” (Zhu, 2010).

Thereby, it can be inferred that this pronunciation habit of initials in Southwest Mandarin, namely, the inclination of turning voiceless consonants into voiced ones, has a potential influence on the pronunciation of voiceless plosives. The phenomenon is consistent with the results of this experiment which shows that the VOT of English voiceless plosives in Southwest Mandarin is the longest, while that of voiced plosives is the shortest.

4.2.2.2 Min Dialect

Min Dialect, also dubbed as Min Chinese, is primarily distributed in Fujian, southern Zhejiang, Taiwan, Guangdong, Hainan and some regions overseas.

Li Rong (1989) pointed out that the ancient total voiced consonants in Fujian Min Dialect have gradually realized devocalization; nowadays, whereas, during the pronunciation of plosives and fricative, “aspiration is absent in more characters”, which contradicts the convention that many ancient total voiced consonants are aspirated in most Fujian languages.

Li Xiaofan and Xiang Mengbing (2009) also pointed out that “in ancient Min Dialect, the total voiced consonants have become devoicing, and now most of the plosives and affricates have no aspiration, with few have aspiration.” This indicates that speakers of Min Dialect tend to make sound unaspirated and of vocal cord vibration. Take Hokkien (southern Fujian dialect) for example, in accordance with “there was no dentilabial consonant sound in ancient times”, the dentilabial initial sound /f/ was pronounced as /p/ or /b/ in ancient ages (Wu, 2014). This phonetic convention is maintained until the present in Hokkien vernacular. For instance, “fan” in standard Chinese is pronounced as “ben” in southern Fujian dialect. To give another example, in Chaozhou dialect which belongs to Min Dialect, /b/ and /g/ phonemes are accompanied by pre-nasalization during the pronunciation (Liu, 2009, p. 76). On account of the articulatory position of the lips, /b/ phonemes are supposed to be articulated as [ʰb]; and due to the velarization, phoneme /g/ should be added a diacritic and recorded as [ᵑg] in narrow transcription (Liu, 2009, p. 76). In general, the tendency of voicing and nasalization will presumably lengthen the VOT of voiced plosives of Min Dialect speakers; in other words, the vocal cords vibration duration will be longer while pronouncing voiced plosives.

Additionally, phoneme simplification occurs in Southeast Asian English under the influences of Min dialect. The most significant factor in word localization is phoneme recombination which can affect the whole lexical system. In this case, the plosives subsequent to the consonant in Southeast Asian English are often left unvoiced. In places like Myanmar and the Philippines, even voiceless and voiced consonants are not distinguished, which is in line with the so-called “economic principle” that the ancestors of Min Dialect region, who migrated to Southeast Asia, simplified English as they learned it (Chen, 2006, p. 124). This also explains the relatively longer VOT of English voiced plosives of the subjects who come from the Min Dialect region in this experiment.

5. Conclusions

This chapter is to introduce the main findings of this experiment, some theoretical and pedagogical implications, and the limitations this study has.

5.1 Findings

The major findings yielded from this study can be summarized as follows.

5.1.1 Gender Factor

As the statistics showed in chapter four, there is little difference between VOT of English plosives of male and female; however, it is noteworthy that female VOT is longer than male VOT, which is consistent with some previous studies. The underlying causes are twofold.

Firstly, there are fundamental biological differences in the various speech subsystems. The physiological structure of female phonation is a posterior glottal opening which occurs during vocal fold closing; hence, it is possible that more air escapes into the supralaryngeal space prior to vowel production compared to male speakers.

Secondly, psychosocially, there are suggestions that sex differences in VOT may be due to a sociophonetic influence, whereby the differences can reflect conscious or unconscious manipulations of speech patterns to overtly or tacitly to convey gender identity, one of which is the speaking rate. Presumably, women will prefer slower speech and lengthened word duration in order to make utterances more intelligible and make themselves more dignified, and maintain their gender identities and images, both of which can help produce longer VOTs.

5.1.2 Dialectal Factor

Languages do not exist in isolation but in interaction, of which language transfer is a manifestation, and an integral factor affecting second language acquisition. Phonology of Chinese is distinct from that of English; consequently, in SLA, English learners will tend to be affected by their dialects, transfer the dialectal features to English and neglect the differences between two phonetic systems, thus affecting the speech output.

In the light of statistics in chapter four, based on the regional division of Chinese dialects, to some extent, some influences dialects have on the VOT of plosives can be found. Insofar this experiment, the effects of Southwest Mandarin and Min Dialect on English VOT is mainly analyzed. The results show that the

voiceless plosive VOT is the longest and the counterpart VOT is the shortest. Owing to tendency of decuminal and devocalization, the production of English voiceless plosives can be presumably affected; that is, to lengthen the time from voiceless sound blasting to vocal cord vibration.

Secondly, the average VOT value of voiced plosives is the highest in the subjects using Min Dialect. Combined with the dialectical pronunciation characteristics, it can be seen that the speakers tend to make unaspirated sounds or vocal cord vibration sounds. Inclination of pre-nasalization can lengthen vocal cord vibration time as well, resulting in a longer VOT of voiced plosives. Additionally, influenced by immigration and phonetic changes, phoneme simplification can lead the sounds less voiceless and more voiced.

5.2 Implications

The implications of this study are primarily at the theoretical and pedagogical aspects.

Theoretically, from the perspective of research design, firstly the acoustic parameter VOT is adopted in this study to investigate the group differences of English plosives, which can provide some theoretical references for acoustic researches, and also offer some ideas for the future researches on English plosives. In the second place, this study introduces social factors investigation into phonetics research, combining theoretical knowledge and research methods of phonetics and sociology, which can provide innovation for the subsequent studies. In addition, this experiment, with reference to Labov's experimental model, can boost the generalization of sociolinguistic theories and research methodologies.

Pedagogically, from the research results, there are primarily the following three implications. First, in English teaching, different kinds of students should be coped with flexibly. Instructors are supposed to carry out targeted teaching and training of students. The results show that although significant, some discrepancies factually exist in the plosive VOTs of students of different genders and dialects. In these cases, teachers should not cut across the board and should maximize the effects of English teaching according to the specific situation. Secondly, affected by language transfer, some errors or mistakes presumably occur during the English learning of university students. In phonetics teaching, instructors can adopt the theories of contrastive analysis, interlanguage and continuum to maximize the effects of language positive transfer and for the other, actively correct English pronunciation of students and aggrandize comprehensible phonetic input, minimizing the impact of negative language transfer. Some practical tools such as PRAAT can be applied as well, to present of the acoustic features of English segments to students and moreover, correct students' articulations. Third, the English curriculum designs should be optimized. Universities should attach attention to English teaching; specifically, more reasonable English class settings and designs, and larger allocation of English instructors should be offered.

5.3 Limitations

Although this study provides some implications, there are still limitations.

Firstly, the sample size of this experiment is not large enough, with the number of 30. More eligible samples are needed if much efficiency and reliability are required. In this study, the validity of the main

findings needs substantiating by future research which can extend the scope of participants and gain much accuracy.

Secondly, some errors may exist in the sampling and measurement. In terms of sampling, due to the limitations of epidemic and geography, experiments can only be carried out on the Internet; therefore, professional recording software cannot be employed. And, the recording equipment and methods of each subject may be different, which can bring some errors and influence analysis. In the respect of measurement, other environmental factors can be the inferences to the subjects, and some samples have obvious striations in the spectrogram. More professional equipment should be employed in future studies to minimize experimental errors.

Thirdly, on account of the constraint on conditions, other potential influential factors are not taken into consideration in this survey. The independent variables of this study are two social factors: gender and regional dialect; however, the results may be interfered by other subjective factors, such as individual speaking habits and diverse learning levels. Afterwards, the scope of factors can be extended to conduct phonetic experiments with higher objectivity and pertinence.

References

- Bradford, L. S. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills*, 7(5), 983-999.
- Bruce, P. (1964). The analysis of word sounds by young children. *British Journal of Educational Psychology*, 34, 158-170. <https://doi.org/10.1111/j.2044-8279.1964.tb00620.x>
- Chen, H. H. (2006). Fujian Southeast Asian English Varieties and the Dialects of Min and Yue. *Journal of Huaqiao University (Philosophy and Social Sciences Edition)*, 1, 123-124.
- Cruttenden, A. (2001). *Gimson's Pronunciation of English*. Beijing: Foreign Language Teaching and Research Press.
- Crystal, D. (1997). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Dai, Z. Z. (2012). VOT Values of Plosives Produced by Advanced Chinese English learners. *Journal of Xidian University (Social Science Edition)*, 6, 102-106.
- Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents*. New York: Foris. <https://doi.org/10.1515/9783110872637>
- Eunjin, O. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39(1), 59-67. <https://doi.org/10.1016/j.wocn.2010.11.002>
- Flege, E. J. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of Acoustical Society of America*, 89(1), 395-411. <https://doi.org/10.1121/1.400473>
- Gao, Y. J., & Zhang, M. M. (2018). An Experimental Study on Chinese Students' Acquisition of English Plosives. *Journal of Northeast Normal University (Philosophy and Social Sciences Edition)*, 6, 70-76.

- Gussenhoven, C., & Jacobs, H. P. (2011). *Great Britain: Hodder Education* (3th ed.).
- Jiang, Y. Y. (2015). A Comparative Study of Inter-vowel Plosive VOT among English Learners in Different Dialect Areas in China—Taking Wu Dialect and Jianghuai Dialect Area as Examples. *Journal of Taiyuan Urban Vocational College*, 9, 187-189.
- Keating. (1984). A phonetic and phonological representation of stop consonant voicing. *Language*, 2, 286-318. <https://doi.org/10.2307/413642>
- Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford/New York: Pergamon Press.
- Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics*. Beijing University Press.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 3, 384-422. <https://doi.org/10.1080/00437956.1964.11659830>
- Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language & Speech*, 10, 1-28. <https://doi.org/10.1177/002383096701000101>
- Li, R. (1989). Division of Chinese Dialects. *Dialects*, 4, 241-259.
- Li, X. F., & Xiang, M. B. (2009). *Basic Course of Chinese Dialectology*. Beijing: Peking University Press.
- Liu, Z. (2009). Chaozhou Dialects as a Negative Transfer in the Acquisition of English Phonetics—A Comparative Study of Phonemes of English Consonants and the Initial Consonants in Chaozhou Dialect. *Journal of Zhanjiang Normal College*, 30(5), 75-77.
- Liu, Z. Y. (1999). *English Phonetics*. Nanjing: Jiangsu Education Press.
- Mackay, R. A. (1987). *Phonetics: The Science of Speech Production*. Boston: Little Brown.
- Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetic*, 36(2), 308-317. <https://doi.org/10.1016/j.wocn.2007.06.003>
- Roach, P. (2000). *English Phonetics and Phonology: A Practical Course*. Beijing: Foreign Language Teaching and Research Press.
- Robb, M., Gilbert, H., & Lerman, J. (2005). Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica Logopaedia*, 57, 125-133. <https://doi.org/10.1159/000084133>
- Ryalls, J., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on Voice Onset Time. *Journal of Speech, Language, and Hearing Research: JSLHR*, 40(3), 642-645. <https://doi.org/10.1044/jslhr.4003.642>
- Simpson, A. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3(2), 621-640. <https://doi.org/10.1111/j.1749-818X.2009.00125.x>
- Sloat, C., Taylor, S. H., & Hoard, J. E. (1978). *Introduction to Phonology*. Prentice Hall, Inc., Englewood Cliffs, N.J.

- Stlten, K., Abrahamsson, N., & Hyltenstam, K. (2014). Effects of age of learning on voice onset time: Categorical perception of Swedish stops by near-native L2 speakers. *Language and Speech-Teddington*, 57(4), 425-450. <https://doi.org/10.1177/0023830913508760>
- Su, J. J. (2012). Variability in VOTs of Korean Stop: A Speech Corpus Study. *Foreign Language Teaching and Research*, 53(5), 656-668.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, 27, 276-286. <https://doi.org/10.1177/002221949402700503>
- Vickie, Y., & Elizabeth, W. P. (2014). Effects of age, sex and syllable number on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech*, 58(2), 152-167. <https://doi.org/10.1177/0023830914522994>
- Wang, G. Z. (2002). *English Phonetics Course*. Beijing: Higher Education Press.
- Wang, M. L. (2009). A Study of Chinese Learners' Pronunciation of English Stop at Intervocalic Position. *Modern Chinese*, 32(2), 186-194.
- Wei, Z. Z. (2018). *An Acoustic Study on the English Plosive Production by Dalian English*. Changchun: Jilin University.
- Whiteside, S. P., & Marshall, J. (2001). Developmental trends in voice onset time: some evidence for sex differences. *Phonetica*, 58, 196-210. <https://doi.org/10.1159/000056199>
- Wu, G. (2012). Comparison of English and Chinese Plosive VOT and the Enlightenment to Phonetic Teaching. *China Education Innovation Herald*, 19, 93.
- Wu, F. N. (2014). Negative Transfer of Minnan Dialect to English Phonetic Learning and University English Teaching. *Journal of HuBei TV University*, 34(6), 140-141.
- Zhang, J. Y. (2012). An VOT Analysis of Chinese Students Acquiring French Plosives. *Journal of International Chinese Studies*, 28, 174-181.
- Zhao, C. Y. (2019). *An Acoustic Study on Affricates Produced by Harbin English Learners*. Changchun: Jilin University. <https://doi.org/10.1109/O-COCOSDA46868.2019.9060844>
- Zhao, Z. D. (2007). *Phonology*. Shanghai: Shanghai Foreign Language Education Press.
- Zheng, X. R., & Li, Y. H. (2007). Comparison of Voiced Onset Time (VOT) of English and Chinese Plosives and Study of English Plosive Acquisition by Han Students. *Journal of Changchun Normal University* (Humanities and Social Sciences Edition), 26(1), 92-95.
- Zheng, Y. R., Li, R., & Qin, T. (2012). A Brief Analysis of the Influences of Southwestern Mandarin on English Listening Learning. *Youth*, 12, 92-94.
- Zhou, W. J. (2008). *Outline of English Phonetics*. Anhui: Anhui University Press.
- Zhu, J. (2010). Investigation and Countermeasures of the Impacts of Different Dialect Backgrounds on College Students' English Phonetic Learning. *Journal of Gannan Medical University*, 30(5), 758-759.