Effect of Familiarity with Nonnative Accents on Listening Comprehension across

Proficiency Levels

Mohammadreza Dalman

Northern Arizona University

Abstract

This study investigated whether the inclusion of second language (L2) accent varieties in listening

tests would create an unequal advantage for listeners who share the speakers' first language (L1).

Participants were divided into two levels of proficiency. The lower proficiency group was

comprised of 9 L2 learners, 5 of whom had Chinese L1 background. The higher proficiency group

included 11 L2 learners, 4 of whom were Chinese L1 learners of English. Participants in both

levels of proficiency listened to two conversations and two lectures adopted from TOEFL iBT.

One conversation and one lecture were re-recorded by ChineseL1 speakers of English and the

other conversation and the lecture were delivered by a native American speaker. The results of this

study revealed that there is a potential advantage for shared L1 since Chinese L1 listeners

outperformed other listeners with non-Chinese L1 background in the listening section delivered

by a Chinese L1 speakers of English. Additionally, lower proficiency learners took more advantage

from this familiarity.

Effect of Familiarity with Nonnative Accents on Listening Comprehension
across Proficiency Levels

Background

With the spread of globalization and large-scale interaction between people from different parts of the world, hearing people who speak English with a variety of second language (L2) accents is a rule rather than an exception. With the emergence of English as an international language (EIL), the ESL/EFL distinction that polarizes contexts of language use became more blurred. This new perspective has presented some challenges to the field of language testing since test developers should carefully consider the potential range of varieties that examinees might encounter in the target language use domain (Harding, 2011; Major, Fitzmaurice, Bunta, & Balasubramanian, 2002). To address this growing trend, the listening sections of some high stakes English proficiency tests like TOEFL (Test of English as a Foreign Language) and IELTS (International English Language Testing System) are including parts presented by speakers with non-orthodox accent varieties. In the TOEFL iBT, in addition to North American varieties, there could be some academic lectures presented by speakers having British or Australian accents (see ETS, 2005). Likewise, IELTS is including North American accent in addition to British, Australian, and New Zealander varieties (see Cambridge ESOL, 2008).

However, these varieties of English are not truly representative of those that L2 listeners will potentially encounter in the real world. As Canagarajah (2006) pointed out, in most of the target language use domains like business and academic, listeners, in high probability, face not only a wide range of native speaker accents, but also a cline of L2 varieties. On the basis of this

rationale, if a test of listening comprehension is to be authentic and an accurate representation of listening construct, one of the primary requirements will be to include accented English. Despite the importance of accented English inclusion in enhancing authenticity of listening comprehension tests, there are some concerns which curtail its feasibility.

One of the main concerns would be the test bias having been given rise to by the accented language. The test bias, as a confounding variable, provides an unequal opportunity for a group of listeners who share the speaker's L1 to outperform others who do not while listening to that speaker in a test (Major et al. 2002; Harding, 2011). In other words, the test bias could be considered as a threat to the test construct validity since the test does not truly measure the listeners' knowledge of the construct, and as a result, the test score obtained by the listener will be questionable. A critical question that might arise, with respect to this fact, is whether test developers should enhance the authenticity of listening tests at the expense of jeopardizing the test construct validity. This is a controversial question which can be answered only after robust evidence collected through adopting rigorous empirical methodology proves this shared advantage.

To the best of the author's knowledge, to date, the results of empirical research having investigated the potential for shared L1 advantage, as elaborated in the next part, have not been conclusive and few, if any, study has addressed the effect of speakers listeners shared L1 across proficiency levels. Given these facts, the objectives of this study are twofold: first, it explores if those L2 listeners who share the speaker's L1 perform better in the listening comprehension test compared with those who do not . Second, it investigates if the effect of shared L1 is consistent across proficiency levels.

Research Questions

This study is founded on the assumption that the familiarity of listeners with speakers' accent would have an impact on their comprehension of the listening stimuli. It is also hypothesized that listeners' level of proficiency would moderate the potential benefits of shared L1 advantage. To shed light on these issues, this study seeks to answer the following two research questions.

1. Do Chinese L2 learners of English perform significantly better than non-Chinese L2 learners on the listening section of the TOEFL iBT delivered by a Chinese L1 speaker? Is there any interaction effect between the listeners' L1 and their levels of proficiency?

2. Is there any significant difference between the performance of Chinese and non-Chinese L2 learners on the listening section of the TOEFL iBT presented by a native American speaker? Is there any interaction effect between the listeners' L1 and their levels of proficiency?

Method

Twenty participants from an American university took part in the study. All the twenty participants were studying English as a second language in an intensive course and represented a range of L1 backgrounds. The purpose of the intensive program designed by the university was to improve the language skills of the learners and prepare them for taking academic courses. Based on the placement test having been administered to the learners by the center, they had been divided into two levels of proficiency and had been assigned into two separate classes accordingly. The lower proficiency group was comprised of 9 L2 learners, 5 of whom had Chinese L1 background and among the other 4, 2 were Korean and 2 were Arab L1 learners of English. The higher proficiency group included 11 L2 learners, 4 of whom were Chinese L1 learners of English and the other 7 were 3 Spanish and 4 Arab L1 learners of English. Since the

learners had already been tested for the proficiency, no further placement test was administered to them. Those who were in the lower proficiency group had the approximate TOEFL iBT score of 45-56 and those who were in the high proficiency group had the approximate TOEFL iBT score of 56-69.

To measure the listeners' performance on the listening comprehension test, the listening part of the TOEFL iBT (ETS, 2005) was used. The internal consistency of the test, as measured by Cronbach alpha was $\alpha=0.76$. The listening section of the TOEFL iBT is comprised of two sub-section. In each sub-section there is one conversation and two academic lectures. Each conversation is followed by 5 questions and each lecture is followed by 6 questions. The questions measure a wide range of skills: the ability to understand main ideas, details, making inference, and the purpose of the speaker. While the listeners are listening to the input, they do not have any access to the questions. They are allowed to take note while they are listening and can use their notes to answer the questions when the conversation or the lecture is finished. Two conversations and two lectures (all from the same test) were randomly selected to be used in this study. As a result of consultation with the instructors of the two classes, the researcher understood that learners in both classes were not familiar with purpose questions. Therefore, to make sure that the questions measure the skills that the learners had already been taught and practiced, the researcher eliminated the last question from each conversation and each lecture that tested the listeners' ability to answer purpose questions. Hence the conversations and the lectures used in this study included 8 and 10 questions respectively.

Two speakers with Chinese L1 background were recruited to re-record one conversation and one lecture of the TOEFL iBT listening test. The speakers were selected among those who met the following three criteria adopted from Harding (2011). (1) speakers had equivalent levels

of general intelligibility; (2) speakers were not perceived to be unreasonably difficult to understand; (3) L2 speakers had accents which were identifiably L2 varieties. To achieve this aim, a pool of 6 Chinese L1 speakers of English recorded a conversation and a lecture of the TOEFL iBT. The recorded sounds were played for 2 native speakers of English, 1 Spanish and one Korean L1 learners of English. They were prompted to listen to the recorded sounds, transcribe them and rate their intelligibility on a Likert scale from 1 (least intelligible) to 5 (most intelligible). As a result of this rigorous process, Two Chinese L1 speakers of English were selected to re-record one conversation and one lecture selected fro the listening section of the TOEFL iBT.

Results

The first research question aimed to explore if the performance of the Chinese L2 learners of English differed significantly from those of non-Chinese L2 learners on the listening section delivered by a Chinese L1 speaker and whether there was an interaction effect between Listeners' L1 and their levels of proficiency. To test this hypothesis, first the results of descriptive statistics were consulted. The results showed that Chinese L1 low proficiency listeners (M=7.2, SD=0.83) outperformed the Chinese L1 high proficiency listeners (M=5, SD=0.81). However, the trend was not the same for non-Chinese listeners since the higher proficiency listeners (M=5.2, SD=0.83) outperformed the lower proficiency listeners (M=3.6, SD=1.03). To compare the mean difference between the performance of these two groups of listeners an independent sample t test was conducted . However, before conducting the analysis, the normality assumptions were tested. The assumptions were met as the skew and kurtosis levels were estimated at 0.27 and 1.1 for "level" variable and -0.14 and -1.06 for "scores" variable, respectively. The results of t test analysis revealed that Chinese L1 listeners performed

significantly better than non-Chinese L1 listeners ($t(7)$=3.96, p=0.005<0.025). To further

measure if there was an interaction effect between learners L1 and levels of proficiency, a two-

way analysis of variance (ANOVA) was conducted. Prior to conducting a two-way ANOVA,

the results of Levene's test of equality of variance was considered and the results showed that

there was no significant difference between the two groups (p=0.86>0.05). As the result of the

analysis showed that in addition to the listeners' L1, their level of proficiency affected their

performance on the listening test, too ($F(1,16)$=21.1, p=0.00<0.05). This result further indicated.

Therefore, the null hypothesis stating that Chinese and non-Chinese L1 listeners perform the

same on the listening section of the TOEFL iBT delivered by a Chinese L1 speaker and there is

no interaction effect between the listeners' L1 and their levels of proficiency is rejected.

The second research question concerned whether Chinese and non-Chinese L1 listeners

performed significantly differently on the section of the TOEFL iBT listening comprehension

presented by a native American speaker. The results of descriptive statistics demonstrated that

Chinese L1 higher proficiency listeners performed better (M=5.20, SD=0.83) than Chinese L1

lower proficiency listeners (M=5.00, SD=0.81). Similarly, non-Chinese higher proficiency

listeners outperformed (M=7.60, SD=1.14) non-Chinese lower proficiency listeners (M=5.50,

SD=1.04). To further find out if the mean difference between the performances of these two

groups of listeners was statistically significant, an independent sample t test was conducted.

However, prior to conducting t test, the normality assumptions were determined. As the results

of descriptive statistics showed, the normality assumptions were met since the skew and kurtosis

levels were estimated as -0.21 and -1.60 for the "level" variable and 0.69 and 0.03 for "score"

variable, respectively. The results of independent sample t test indicated that the mean difference

between the performances of the two groups was not statistically significant ($t(18)$=-2.41, p=

0.027>0.025). In the next part of the analysis, to find out if there was an interaction effect between the listener's L1 and their levels of proficiency, a two was ANOVA was conducted. Before conducting two way ANOVA, the results of Levene's test of equality of equality of error variances was examined. As the results indicated, two groups were homogeneous, (p=0.67>0.05). The results of ANOVA demonstrated that there was an interaction effect between listeners' L1 and their levels of proficiency ($F(1,16)$= 6.68, p=0.02<0.05). Hence, we fail to reject the part of the null hypothesis stating that there is no statistically significant difference between the performance of Chinese and non-Chinese L1 listeners. However, we reject the other part stating that there is no interaction between the L1 and the levels of proficiency of the listeners.

Relevance to the PIE and Second Language Learning

As the results of this study suggest, low proficiency listeners are more likely to be affected by L2 accent varieties than high proficiency listeners. If listeners and speakers have the same L1 background, this effect will be in the form of a positive influence that benefits the listeners and help them perform better compared to the time that they would listen to a listening input delivered by native American speakers. However, if the listeners and the speakers are not coming from the same L1 background, this effect will be in the form of a negative influence that disadvantage these groups of listeners. Nevertheless, as the listeners' level of proficiency increases, this effect, either positive or negative minimizes. Therefore, the inclusion of L2 accent varieties is not recommended for designing listening tests for low proficiency listeners. However, since TOEFL iBT is designed for those whose listening proficiency is higher intermediate, the inclusion of L2 accent varieties could be justified since not only does it not drastically benefit or disadvantage a group of listeners, but it also encourages TOEFL iBT instructors to include L2

accent varieties in their teaching materials which would both prepare listeners for taking the test

and understanding those whose accents deviate from what is established as native varieties. A

word of caution, however, is that this study only explored the case of Chinese L1 listeners. To

reach a more conclusive result, more studies with larger sample sizes and other L1 accent

varieties need to be done.

References

Abeywickrama, P. (2013). Why not non-native varieties of English as listening

      comprehension test input? *RELC Journal, 44*(1), 59-74.

Adank, P., Evans, B., Stuart-Smith, J., & Scott, S. (2009). Comprehension of familiar

      and unfamiliar native accents under adverse listening conditions. *Journal of

      Experimental Psychology, 35*, 520-529.

Adank, P., Janse, E. (2010). Comprehension of a novel accent by young and older

      listeners. *Psychology and Aging, 25*, 736-740.

Anderson-Hsieh, J. & Koehler, K. (1988). The effect of foreign accent and speaking rate

      on native speaker comprehension. *Language Learning, 38*, 561-613.

Bradlow, A. R. and T. Bent. 2008. Perceptual adaptation to non-native speech. *Cognition

      106* , 707–9.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit.

      Journal of the Acoustical Society of America, 114, 1600-1610.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives:

      Testing English as an international language. *Language Assessment Quarterly*,

      3, 229–242.

Derwing, T. M., Munro, M. J. (1997). Accent, intelligibility, and comprehensibility:

      Evidence from four L1s. *Studies in Second Language Acquisition, 20*, 1-16.

Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching:

      32 A research-based approach. *TESOL Quarterly, 39*, 379-397.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2

      speech. *Studies in Second Language Acquisition*, *28*, 111-131.

ETS [Educational Testing Service] (2005). *TOEFL iBT Tips: How to prepare for the next generation TOEFL test and communicate with comfort*. Princeton, NJ: Educational Testing Service.

Gass, S. M., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34*, 65-89.

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing*, *29*, 163-180.

IELTS [International English Language Testing System] (2007). *IELTS Handbook*. British Council, IDP: IELTS Australia, University of Cambridge ESOL Examinations.

Imai, S., Flege, J. E., & Walley, A. (2003). Spoken word recognition of accented and unaccented speech: Lexical factors affecting native and non-native listeners. In *Proceedings of the International Congress on Phonetic Science* (pp. 845-848).

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, *36*, 173-90.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111-131.

Ockey, G. J., & French, R. (2016). From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics, 37*, 693-715. doi: 10.1093/applin/amu060

Van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, *111*, 1906-1916.