論文の内容の要旨

論文題目　*Evaluating the Effect of an Ethnic Bias on Speech Perception by Non-native Listeners*

（非母語話者の発話知覚における民族バイアスの影響評価）

氏名　カルピニスカ・マジェナ・エルジビエタ

Marzena Elzbieta Karpinska

Speech perception is by its nature a multimodal process involving the integration of both linguistic and social information that are being linked together in order to encode the intended message along with information pertaining to the person who uttered it. In other words, listeners use both these factors in order to map a highly variable acoustic input onto a mental representation. When the social information matches the linguistic information as per listener's expectations (e.g., Caucasian face is paired with a native English utterance) the speech can be processed more easily (Babel & Mellesmoen, 2019; McGowan, 2015). However, any mismatch between these two factors can alter listener's perception of the utterance (Babel & Russell, 2015; Gnevsheva, 2018).

The current study contributes to this discussion by evaluating how social information, in particular speaker's perceived ethnicity, may affect speech perception by non-native English listeners. More specifically, it investigated whether Japanese native speakers, who come from monolingual families, will be affected by speaker's ethnicity when rating accentedness and comprehensibility of native English utterances. Furthermore, the current study also evaluated whether ethnic bias will have an effect on the actual intelligibility of the native English stimuli.

In addition, the effect of ethnic bias was also investigated in relation to the type of stimuli, pictures of the speakers presented with audio files vs. videos of the speakers. Furthermore, the gender of the speaker was also taken into consideration as it seems to be an informative socioindexical cue (Johnson et al., 1999; Strand & Johnson, 1996), yet most previous research on the effect of ethnic bias on speech perception by native English listeners employed only female speakers (e.g., de Weers, 2019; McGowan,2011, 2015; Rubin, 1992; Rubin & Smith, 1990).

This link between the social and linguistic information appears to be strong for native English listeners listening to native and non-native English speech. For instance, native English listeners may have less

difficulties in understanding Chinese-accented English utterances when they are presented with an East Asian face (McGowan, 2015). Here, the social information provides additional information about the speaker and their speech, which facilitates the processing of the message. On the other hand, native English listeners may have more trouble understanding native English utterances when presented with an East Asian guise (Babel &Mellesmoen, 2019; Babel & Russell, 2015). Additionally, they may also rate native English speech as more accented when presented with an East Asian face than when presented with a Caucasian face (Babel & Russell, 2015; Rubin, 1992).

Contrary to what was demonstrated for the native English listeners, Japanese listener's in the current study were *not* affected by the speaker's perceived ethnicity when rating the accentedness and comprehensibility of native English speech. Moreover, speaker's ethnicity did *not* have any effect on listeners' intelligibility. These results indicate that non-native English listeners in the current study were not affected by speaker's ethnicity when listening to native English speech.

Furthermore, there was no significant interaction between the type of stimuli (pictures of the speakers presented with audio files vs. videos) and speaker's ethnicity. This suggests that native Japanese listeners rated accentedness and comprehensibility of native English utterances as well as transcribe these utterances in a comparable way regardless of whether the ethnicity of the speaker was operationalized with picture or video stimuli.

In addition, an interaction between the speaker's ethnicity and speaker's gender was also investigated as previous studies tended to include mainly female speakers (de Weers, 2019; McGowan, 2015; Rubin, 1992; Rubin & Smith, 1990). However, this interaction was also not significant indicating that there was no effect of speaker's perceived ethnicity on speech perception for neither female, not male speakers.

Two competing theories endeavored to explain the effect of ethnicity on speech perception by native English listeners: the reverse linguistic stereotyping (RLS) and the experience-based models. The RLS is a theoretical framework developed by Rubin and his colleagues and first introduced in detail in Kang and Rubin (2009). The RLS assumes that listeners hold a negative bias against Asian English speakers. Because of this bias listeners may “hear” a non-native accent in native English speech. Furthermore, the bias could also cause poorer intelligibility of native English utterances when presented with an Asian face than when presented with a Caucasian face as listeners who have a *negative* bias towards East Asian-looking English speakers may intentionally *choose* to pay *less* attention to the utterance, which will lead to a communicative breakdown (Kang & Rubin, 2009; Lippi-Green,2012; Rubin, 1992).

On the other hand, the experience-based approach, which is derived from exemplar theory (Foulkes, 2010; Foulkes & Hay, 2015), assumes that episodic traces are being stored in memory in order to be activated when presented with a consistent social category (e.g., speaker's ethnicity). This idea was incorporated by Sumner and colleagues into the socially-weighted speech perception model introduced in Sumner et al. (2014). In a model like this, an utterance is being parsed into multiple social and linguistic information, which interact with each other in the process of social weighting. This means that listeners` perception of speech is being shaped by their past experience and societal stereotypes.

Since native English listeners usually associate being American (and hence possibly a native English speaker) with being Caucasian (Babel & Russell, 2015; Devos & Banaji, 2005; Gnevsheva, 2018; Yi et al., 2013, 2014) presenting native English speech with an Asian face may for some listeners create an incongruent condition. This would lead to a mismatch effect, which would then affect speech perception. The socially-weighted speech perception model goes even further providing also explanation for the positive effects in the literature, such as native English listeners having *less* problems understanding Chinese-accented English speech when it is presented with an Asian face (McGowan, 2015). The reason for this effect is the congruent condition created by matching Chinese-accented English utterances with an Asian face. Hence, when listeners’ expectation match the actual linguistic signal, it is easier for them to process the spoken utterance.

While this research does not rule out the RLS, I argued that experience-based models, or more specifically the socially-weighted speech perception model, may offer a possibly better explanation that can account for both the results of the current experiment and the results of previous studies. The experience-based

approach suggests that listeners will rely on their personal experience when incorporating social information to the process of speech perception.

Hence, the null effect of speaker's ethnicity observed in the current study could suggest that the Japanese participants simply had different experience, that is, they had had experience interacting with both East Asian-looking and Caucasian-looking native English speakers. This idea can be partially supported by the fact that almost all listeners in the current study were recruited at the University of Tokyo, Komaba Campus, where in 2015 there was about the same number of students from English speaking countries in Asia, for instance Philippines or Singapore, as from the English speaking countries from other parts of the world, for instance the US or the UK (The University of Tokyo, n.d.). This idea that experience would shape the speech perception of the Japanese listeners in the current experiment is supported by the socially-weighted model of speech perception (Sumner et al., 2014).

Alternatively, the lack of significant results in terms of the effect of speaker's ethnicity in the current study could suggest that non-native speakers may rely on the socioindexical information, in particular, the ethnicity of the speaker much *less* than native speakers do. This would create a need to refine the socially-weighted model of speech perception as if the socioindexical information is linked together with linguistic information in the socially-weighted process of speech perception, then it is possible that this link may not be strong enough for non-native listeners, at least listeners in the current study. Hence, they may not make any assumptions about the speaker and the way he or she would speak, that is they may not expect native English speech from a Caucasian speaker and non-native English speech from an Asian speaker. To confirm this theory, more research is required with native English speakers included in the same research design in order to assure that the null effect of speaker's ethnicity observed in the current study was due to a difference between native and non-native listeners and not due to the research design. Should an effect of speaker's ethnicity be observed for native English listeners, this would suggest that non-native listeners are not affected by social information, at least by the speaker's ethnicity, to the same extent as native listeners. A finding like this would also suggest a need to develop a new *refined* model of socially-weighted speech perception, which would account equally for both native and non-native listeners.

There is almost three decades of research regarding the effect of speaker's perceived ethnicity on speech perception by native listeners. Hence, it would then be unwise to generalize the results of this study to *all* non-native listeners and certainly, research including non-native listeners from other countries is needed in order to gain a deeper understanding of how and when socioindexical cues may affect the speech perception by non-native listeners.

Moreover, it would be equally interesting to see whether native Japanese listeners would act in the same way as native English listeners seem to when evaluating native Japanese utterances presented with an Asian guise and a Caucasian guise. Should there be an effect of speaker's ethnicity for native Japanese speech evaluated by native Japanese listeners, it would suggest that native Japanese listeners rely on the socioindexical cues in the same way as native English listeners do, when listening to their own native language. This would potentially provide evidence in favor of the socially-weighted speech perception model and also provide more generalizability to it as the same effect of social information would be observed for language other than English.

References

Babel, M., & Mellesmoen, G. (2019). Perceptual adaptation to stereotyped accents in audio-visual speech. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences* (pp.1044–1048). Melbourne, Australia: Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. T*he Journal of the Acoustical Society of America*, *137*(5), 2823–2833. Retrieved from https://doi.org/10.1121/1.4919317
Devos, T., & Banaji, R. M. (2005). American = White? *Journal of Personality and social Psychology*, *88*(3), 447–466.

de Weers, N. (2019, August). The interaction between (un)expected speaker ethnicity and accent combinations on response times. Paper presented at New Sounds 2019, Tokyo.

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology*, *1*(1), 5–39.

Foulkes, P., & Hay, J. B. (2015). The emergence of sociophonetic structure. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 292–313). New York: John Wiley & Sons, Inc. Retrieved from https://doi.org/10.1002/9781118346136.ch137

Gnevsheva, K. (2018). The expectation mismatch effect in accentedness perception Asian and Caucasian non-native speakers of English. *Linguistics*, *56*(3), 581–598. Retrieved from https://doi.org/10.1515/ling-2018-0006

Johnson, K., Strand, E. A., & DImperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. Retrieved from https://doi.org/10.1006/jpho.1999.0100

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, *28*(4), 441–456. Retrieved from https://doi.org/10.1177/0261927x09341950

Lippi-Green, R.(2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed.). London; New York: Routledge.

McGowan, K. B. (2011).The Role of Socioindexical Expectation in Speech Perception (Unpublished doctoral dissertation). The University of Michigan.

McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech*, 58(4), 502–521.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. Research in Higher Education, 33(4), 511–531.

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates perceptions of nonnative English-speaking teaching assistants. I*nternational Journal of Intercultural Relations*, *14*(3), 337–353. Retrieved from https://doi.org/10.1016/0147-1767(90)90019-s

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology, results of the 3rd KONVENS Conference, Bielefeld, Germany, 8 October 1996* (pp. 14–26). de Gruyter.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology*, *4*, 1–13. Retrieved from https://doi.org/10.3389/fpsyg.2013.01015

The University of Tokyo. (n.d.).The University of Tokyo, Komaba prospectus 2015-2016. Retrieved 2020-02-28, from http://www.c.u-tokyo.ac.jp/info/about/booklet-gazette/prospectus/2015/prospectus2015E.pdf

Yi, H.-G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America*, *134*(5), EL387–EL393. Retrieved fromhttps://doi.org/10.1121/1.4822320

Yi, H.-G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, *8*.