Tulaja, L. (2020). Exploring acceptability: L1 judgements of L2 Danish learners' errors. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, Northern Arizona University, September 2019 (pp. 197–206). Ames, IA: Iowa State University.

EXPLORING ACCEPTABILITY: L1 JUDGEMENTS OF L2 DANISH LEARNERS' ERRORS

Lisa Tulaja, Kiel University

Second language (L2) pronunciation research often focuses on accentedness and comprehensibility. In contrast, little attention has been given so far to the dimension of acceptability (Thomson, 2018). Acceptability of an L2 speaker's pronunciation is a crucial aspect in the communication between native speakers (L1) and L2 speakers. It displays L1 speakers' evaluation of their L2 communication partners and determines their further (speech) behavior. To gain deeper insights into this dimension, this study focuses on Danish as an L2. L2 Danish learners often experience L1 speakers switching into a lingua franca in L1-L2 communication. This is an undesired behavior for L2 learners, especially when they want to practice their target language. Yet, the switching behavior can function as a measure for the evaluation of acceptability. This study explores the dimension by operationalizing acceptability with the switching tendency of L1 Danish speakers. The analysis focuses on two aims. First, it compares L1 ratings of typical L1 German L2 Danish learners' errors to determine those errors that are most critical with regard to comprehensibility, accentedness and acceptability. These results provide evidence-based implications for teaching purposes and can be helpful to design curricula and learning objectives. Second, the analyses aim towards gaining more insight into the dimensions' interrelations. The results of the correlation analyses correspond to previous findings that acceptability and comprehensibility correlate to a higher degree than acceptability and accentedness.

INTRODUCTION

When learning a second language (L2), it is important to achieve a certain level of pronunciation that relates to the overall aim of learning an L2, which is to be able to interact in the target language and consequently to be understood. For assessing L2 speech, Munro & Derwing (1995) distinguish between the dimensions of intelligibility and comprehensibility. Intelligibility describes the actual understanding of an utterance, whereas comprehensibility is the ease to understand an utterance. Numerous studies have focused on these dimensions, finding empirical evidence for their independence from the dimension of accentedness, with accentedness being the perceived difference of speech patterns (Derwing & Munro, 2015). With the Principle of Intelligibility becoming the main goal in teaching L2 pronunciation, related studies mainly investigate phonetic features' impact on intelligibility and comprehensibility. The findings provide evidence-based insight into which phonetic features are important to teach (Jenkins, 2000; Kang et al., 2016; Saito, 2011; Saito et al., 2017).

Levis (2016) states that intelligibility and comprehensibility are not the only important factors for successful interaction in an L2. He mentions several studies (for example Miller, 2003; Buckingham, 2014) that show how intelligible L2 speech can still provoke negative evaluations by L1 speakers, which can lead to undesired experiences for L2 learners. For example, they may be evaluated as not suitable for a job position or even as not a suitable communication partner in the target language. Thomson (2018) refers to those kinds of negative evaluations of L2 accent as

acceptability: a highly subjective while still very real dimension of pronunciation. The dimension of acceptability is investigated in several previous studies.

First, Settinieri (2011) looks into acceptability as "a social implication of non-normative speech". In her study, German L1 speakers rated different accents on direct scales in terms of how positive or negative they considered the accent to be (1 = negative; 7 = positive). However, it is not entirely clear what raters are referring to when evaluating an accent as positive or negative. It remains uncertain whether a negative evaluation implies a social reaction or if a negative judgement simply refers to L2 speech being hard to understand, which would instead represent the dimension of comprehensibility.

Chang Li-Ann (2008) adds context as a component of evaluation to her definition of acceptability. The rating scale ranges from 1 = "I consider this sort of English unacceptable for international communication" to 7 = "I consider this sort of English acceptable for international communication". The analysis focuses on whether raters from different origins vary in their acceptability ratings for different accents. Results show that raters from different origins evaluate accents differently in terms of acceptability.

Related to these studies, Lee et al. (2019) investigate "listeners' perceived (subjective) attitude toward some particular aspects of speech" and they refer to this dimension as pleasantness instead of acceptability. In their study, raters evaluated speech samples on a 9-point rating scale from 1 (very unpleasant) to 9 (very pleasant). In the analysis, the dimension of pleasantness is correlated with comprehensibility and accentedness ratings. Results show that pleasantness ratings were more closely connected to comprehensibility than to accentedness ratings.

All three studies illustrate that the dimension of acceptability somehow needs to get disentangled from both comprehensibility and accentedness. This distinction should also be implemented in rating settings to avoid participants referring to comprehensibility or accentedness when they should actually be rating acceptability. One way to do this is to operationalize the dimension of acceptability by verbalizing the social impact a negative evaluation would have. Pilott (2016) integrates the social impact of L2 speech into his measures of acceptability. He measures acceptability by investigating whether L1 employers evaluate L2 speakers accent as acceptable for employment purposes. In this study, acceptability is closely linked to L2 learners' ultimate goals. Ratings are taken on a scale from 1 (not acceptable) to 9 (very acceptable). Pilott (2016) compares the rating scores of acceptability with scores of comprehensibility, intelligibility and accentedness. The highest correlations were found between comprehensibility and acceptability.

The current study

In the current study, Danish L1 speakers rated L2 speech samples of German learners on the dimensions comprehensibility, accentedness and acceptability. Danish was chosen as a case study because the data that is used in this article was originally embedded into a larger study (Tulaja 2020), with the main goal to improve the conditions of L2 Danish pronunciation teaching in German schools. Danish is taught as an L2 in schools in Northern Germany and teachers report a lack of teaching materials specifically for pronunciation. Hence, the main interest of this study was a rather practical issue and set out to find those phonetic features that are crucial for German L2 Danish learners. The findings provide evidence-based insights for teaching implications.

Danish offers an interesting focus for studies investigating acceptability: Danish L1 speakers are said to switch relatively quickly into a lingua franca when they perceive an L2 accent. This narrative can be supported by two facts. First, Danish L1 speakers show a low degree of tolerance for variation from the standard variety in general (Kristiansen, 2003). Second, Danish is only spoken by approximately five million speakers and L1 speakers usually do not assume that foreigners speak their language. The tendency of switching despite well-functioning comprehensibility is an interesting focus for studies that investigate the dimension of acceptability. Hence, the decisions on design and operationalization of this study can be of interest for research on other second languages as well. In this study, we present an experimental design to investigate the dimension acceptability by integrating L1 speakers' reaction to L2 speech in L1-L2 communication as a measure. We measure acceptability by L1 speakers' tendency to switch into a lingua franca, taking this switching behavior as an indicator that the L1 rater does not evaluate the L2 speaker as an acceptable communication partner in the target language.

Research Questions

The study aimed to gain insight into two different aspects of German L2 Danish pronunciation. First, we aimed to gain more knowledge about error gravity of typical pronunciation errors of German L2 Danish learners in order to give advice for teaching purposes. Second, we wanted to gain deeper insights into the dimension of acceptability. The two research questions are:

1. How different do L1 speakers rate L2 speech samples with typical errors with respect to different dimensions of pronunciation? Are there error types that are rated significantly more negatively than others?
2. How is acceptability related to comprehensibility and accentedness? Do L1 Danish speakers show switching behavior into a Lingua Franca despite well-functioning comprehensibility?

METHODS

To answer the research questions an L1 rating was carried out. L1 Danish speakers evaluated speech stimuli of German L2 Danish learners' typical errors in three dimensions: comprehensibility, acceptability and accentedness.

Stimuli

To find differences in the typical errors that could be interesting for teaching purposes, each stimulus focused on a different phonetic feature. The original dataset (Tulaja, 2020) included 23 phonetic features tested in three different stimulus types: an erroneous L2 example, a control L1 example and an advanced L2 example without error. In this article, only the ratings for the erroneous stimuli of five phonetic features of different segmental categories are considered.

The stimuli were taken from L2 learners' read aloud speech. The learners had been studying Danish for three years at school. The tested errors were each embedded in one sentence. Error 1 represents the erroneous substitution as *[l] for the "soft *d*" ([ð̩]), which is a typical problem for

L2 Danish learners. Error 2 represents problems with the Danish vowel system, which is quite complex because of the many narrow differences in vowel qualities (Höder, 2016). For example for vowel [ɐ̯], German learners tend to produce the more open variant *[ɛ]. Error 3 represents a substitution in the diphthongs. In this study, we chose the diphthong [ɤ̯i̯], which German learners typically produce as *[øi̯]. In general, there is a much wider range of (phonetic) polyphthongs in Danish than in German, which also include triphthongs (Error 4). Error 5 represents missing lenition processes. In Danish, voiceless plosives behind the root vowel are pronounced voiced. German L2 Danish learners tend to produce a voiceless version. In the stimulus, the target word *Mette* ([mɐ̯**d̥**ə], female Danish name) is produced as [mɐ̯**tʰ**ə].

Operationalization of the rating dimensions

In the rating setting, L1 Danish speakers evaluated speech stimuli in three dimensions: comprehensibility, accentedness, and acceptability. We define comprehensibility and accentedness according to Derwing and Munro (2015). Comprehensibility is the listeners' feeling of the ease or difficulty in understanding a speech sample. Accentedness refers to the degree to which a speech sample differs from the local variety. The definition for acceptability builds on the ideas of previous studies discussed above. We define acceptability as a social implication of non-normative speech (Settinieri, 2011), dependent on context (Chang Li-Ann, 2008; Pilott, 2016) and causing a reaction of L1 raters beyond communication (Pilott, 2016). We operationalized acceptability as the tendency to switch into a lingua franca. Acceptability was tested with a 5-point agree-disagree Likert scale connected to the statement 'I would speak a different language than Danish to him/her' with a scale from 1 (disagree totally) to 5 (agree totally). We adopted this number, because agree-disagree scales are best varied between 5 and 7 categories (Krosnick & Fabrigar, 2012). We also redesigned Munro & Derwing's (1995) original items for comprehensibility and accentedness into agree-disagree Likert items, because all items of a questionnaire should be structured in the same way. The statement for comprehensibility was 'It is difficult to understand what he/she is saying'. Accentedness was tested with the statement 'He/she speaks with an accent'.

L1 raters

Danish L1 raters evaluated German L2 Danish learners' speech samples. In total, 192 Danish L1 speakers participated as raters in the study. However, since only raters who stated that Danish was their first language were taken into account for the analyses, the data from 11 raters had to be excluded. The remaining 181 participants (43 male, 136 female, 2 no answer) were between 14 and 76 years old, with a median of 25 years. However, not all of the raters completed the full set of items: Depending on the error type, valid responses in all three rating dimensions were only available from 110 – 114 participants.

We chose an online questionnaire to get a higher number of participants and a broader range than the typical set of university students. Forty raters completed the rating in a laboratory while the rest completed the rating online. We tested whether the rating scores between online and laboratory conditions differed. The two-sample t-test showed no significant difference in both groups ($M_{(laboratory)}$ = 2.10; $M_{(online)}$ = 1.96; $t(904) = -1.80$, $p = .07$). Therefore, we analyzed the results of both groups together.

Rating procedure

For the ratings, we used *SoSci Survey* (Leiner, 2018), which is an online tool that is free to use in non-commercial settings and conforms with the German data privacy act. The L2 speech stimuli were presented to the L1 raters via headphones. Each stimulus was played to the L1 raters four times with a pause of two seconds in between each repetition. When the raters had answered to all Likert items for the three variables, the next stimulus would appear. Raters were asked to rate the stimulus in whole and not to focus on errors.

For the original study, raters rated all items across the three different stimuli types (see section Stimuli), which took thirty to forty-five minutes. Raters were allowed to take breaks in between. The order of presentation of the stimuli was random to avoid sequence effects. This paper only reports the results for the five phonetic features in the erroneous version of the stimuli.

RESULTS

The results show L1 ratings on the dimensions comprehensibility, accentedness and acceptability for five typical errors of German L2 Danish learners. The analysis focuses on two research questions. Firstly, we analyzed differences between ratings of the errors to determine those that were rated significantly more negatively than others. Secondly, we calculated correlations between the dimensions to gain deeper insights into their interrelations.

Table 1 shows the mean scores of the ratings in the three dimensions (accentedness, comprehensibility and acceptability) and the SD for the five typical errors.

Table 5
Mean scores for comprehensibility, accentedness and acceptability ratings in five different errors.

Error	Comprehensibility *M* (SD)	Accentedness *M* (SD)	Acceptability *M* (SD)
1 Soft *d* [ð̩] (*[l])	3.33 (1.33)	4.74 (.58)	2.33 (1.25)
2 Vowel quality [e̞] (*[ɛ])	2.09 (1.11)	4.73 (.58)	1.72 (.96)
3 Diphthongs [ɐ̯i̯] (*[øi̯])	2.88 (1.25)	4.74 (.58)	2.26 (1.18)
4 Triphthongs	2.96 (1.29)	4.71 (.65)	2.16 (1.12)
5 Intervocalic lenition	2.41 (1.24)	4.74 (.65)	1.99 (1.02)

Note. Comprehensibility measure: 'It is hard to understand what he/she is saying.'; Accentedness measure: 'He/she speaks with an accent.'; Acceptability measure: 'I would speak a different language than Danish to him/her.'; 1 = disagree totally, 5 = agree totally.

For each of three dimensions, a one-way ANOVA was run to test if there was significant difference between the average ratings of errors. Statistical difference at the $p < .05$ level was shown for the mean scores of comprehensibility [$F(4, 558) = 17$, $p < .001$, partial eta squared = .109] and acceptability [$F(4, 558) = 5.528$, $p < .001$, partial eta squared = .038]. No statistical difference could be found between the error scores of the dimension accentedness [$F(4, 558) = .042$, $p = .99$, partial eta squared = .000].

Tukey post-hoc analysis of comprehensibility revealed that errors 1, 3 and 4 were rated significantly worse than errors 2 and 5. Errors 2 and 5 did not differ in their comprehensibility ratings and form the baseline of the error gravity ratings of this data set. Error 3 and 4 were rated moderately worse. Error 1 was rated significantly worse than all the other errors.

Tukey post-hoc analysis for acceptability revealed no significance between the errors 1, 3, 4 and 5. However, there were significant differences between errors 1 and 2 ($p < .001$) (.61, 95 % CI [-1.02, 0.20]), 2 and 3 ($p = .002$) (.54, 95 % CI [0.14, 0.95]), and between errors 2 and 4 ($p < .02$) (.44, 95 % CI [.03, .84]). Fig. 1 facilitates the comparisons of the errors.

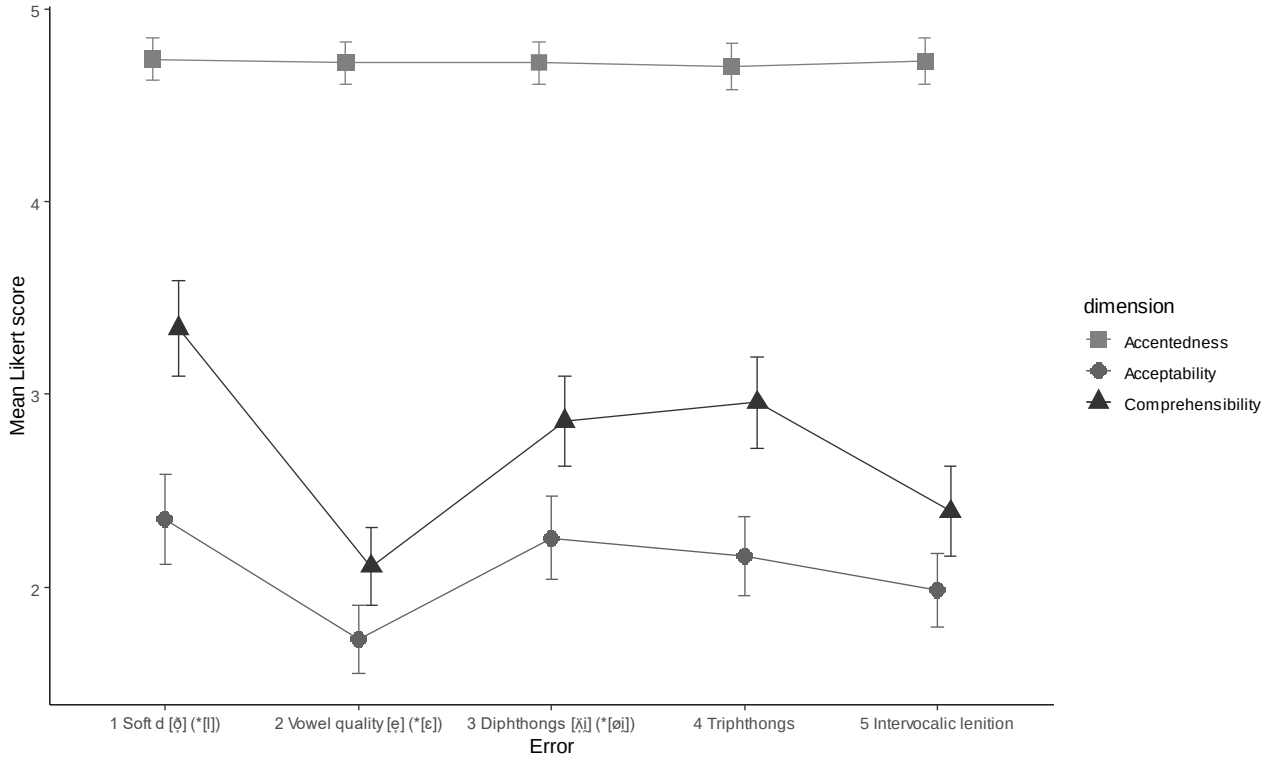

Figure 4 Mean Likert scores in all three dimensions with 95 % CI.

To answer the second research question of how acceptability is related to the other dimensions, we calculated Spearman's rho correlation tests. The results show a significant correlation between acceptability and comprehensibility ($r_s = .559$, $p < .001$, $N = 561$). Between acceptability and accentedness, we found also a significant, however, weaker correlation ($r_s = .116$, $p = .005$, $N = 561$). We also ran a mixed-effects model of acceptability ratings with the *lmer* function in *R* (Kuznetsova et al., 2017). Accentedness and comprehensibility were set as fixed effects, while intercepts for raters and stimuli were set as random effects. The results are shown in Table 2.[4] For comprehensibility, we received significant effects. The effects for accentedness, however, did not prove to be significant.

[4] This method of illustration is adopted from Lee et al. (2019).

Table 6 *Output of the mixed effects regression model of acceptability with comprehensibility and accentedness.*

Predictors	Estimates	*SE*	*T*	*p*
(Intercept)	.572	.304	1.884	.06
Comprehensibility	.417	.029	14.135	< .001
Accentedness	.076	.065	1.184	.237
Observations	561			

The relationship between acceptability and the other two dimensions is depicted in Fig. 2.

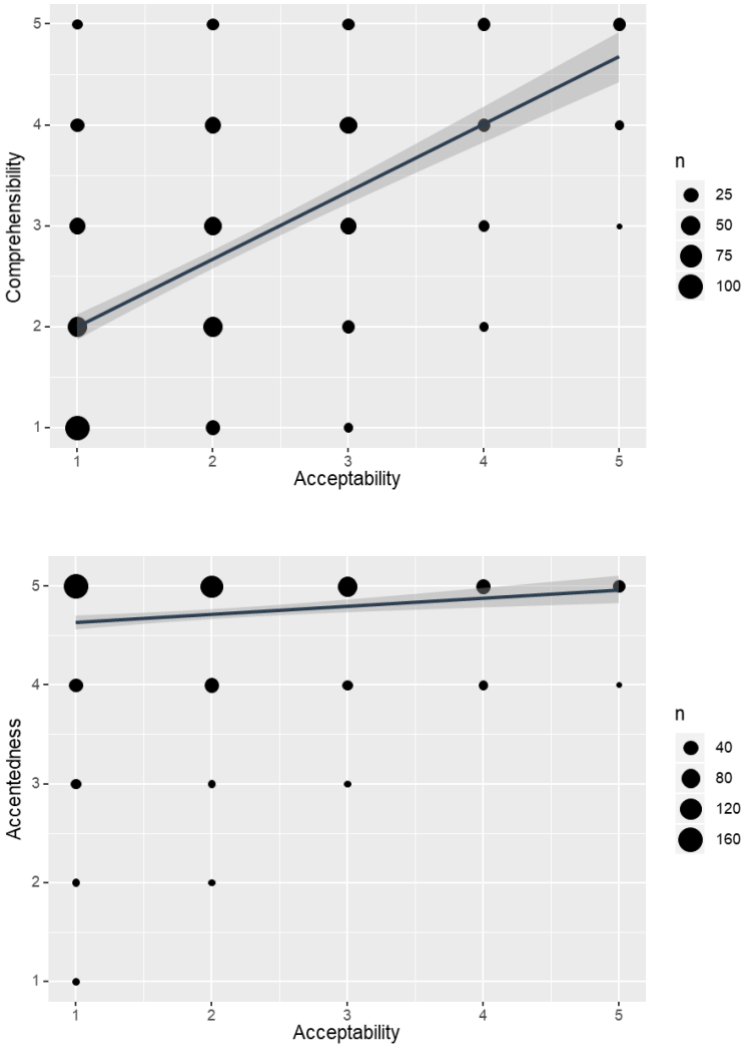

Figure 5. Correlation plots for acceptability by comprehensibility and accentedness. The shaded area indicates a 95 % confidence-interval.

DISCUSSION

This study investigated L1 ratings of L2 pronunciation errors by German learners. The analysis focused on two aims. The first aim was to find differences between errors, in order to gain empirical-based knowledge for teaching purposes. The second aim was to get further insights into the relationship between acceptability and the other dimensions of pronunciation.

The results for the first research question show that the dimensions were not equally useful for highlighting differences between errors. The ratings in accentedness showed no significant differences between errors. Ratings in comprehensibility and acceptability, however, both yielded valuable insights into error differences. Following these results, teachers should focus on the soft *d* (error 1), the diphthongs (error 3) and the triphthongs (error 4). In comparison, the graduations in the vowel qualities (error 2) as well as the lenition processes (error 5) seem to play a less important role and could therefore be a focus at a later stage of pronunciation acquisition. These results are important empirical evidence on which teachers can base their decisions in the future.

In the second step of our analysis, we examined acceptability in relation to comprehensibility and accentedness. We found that the connection between acceptability and comprehensibility was stronger than between acceptability and accentedness. These findings support results of previous studies (Chang Li-Ann, 2008; Pilott, 2016; Lee et al., 2019). The results of the mixed effects model did not show a significant effect of accentedness. This may be due to the fact that in the present study accent on average was rated constantly across the stimuli. The results for the correlation calculation lead to the unambiguous interpretation that more comprehensible speech is more likely to be evaluated as acceptable. Therefore, comprehensibility might be a prerequisite for acceptability.

Acceptability rating scores were, on average, lower than the corresponding scores on the comprehensibility scale (see the respective curves in Fig. 1). This means that Danish L1 speakers tend to continue communication in Danish despite comparatively low comprehensibility (Lower scores indicate greater disagreement with the statements 'I would speak a different language than Danish to him/her.' and 'It is hard to understand what he/she is saying.', respectively). Hence, the anecdotal presumption that Danes in L1-L2 communication would switch into a lingua franca when perceiving an accent cannot be confirmed in the scope of this study. However, subsequent descriptive analyses of the interactional data revealed that some of the raters behave as assumed and show a willingness to switch despite comprehension. Depending on the error, 8 – 16 % of the raters rated acceptability at least one point higher than comprehensibility.[5] This percentage cannot be described as a general data trend, but it suggests that the switching behavior should be re-examined in a further study under more controlled conditions.

In summary, this study contributes to the field of research with a so far unexplored L1-L2 combination. We found valuable insights for teaching purposes in Danish as an L2 and showed an attempt to operationalize the dimension acceptability by measuring L1 speakers switching to a lingua franca in L1-L2 communication, which is a real problem for L2 learners. This approach

[5] Please note that due to the polarity of the rating scale of this study the higher values indicate worse rating.

opens the field for further research that is needed to investigate the gap between the subjective self-assessment of raters and their actual switching behavior.

ACKNOWLEDGMENTS

I am very grateful to Gunnar Wendt, who has supported me with tireless effort in the analysis process of my data. I also want to thank Ocke-Schwen Bohn and Elina Tergujeff for encouraging me in my research.

ABOUT THE AUTHOR

Dr. Lisa Tulaja is a lecturer and research associate at the Institute for Scandinavian Studies of Kiel University. She is currently teaching in teacher education for Danish as a foreign language. Her main research interests include perception of accented L2 speech and cognitive methods in L2 pronunciation teaching.

Contact information:
Lisa Tulaja
Kiel University
Institute for Scandinavian Studies, Frisian and General Linguistics
Leibnizstraße 8, 24118 Kiel
Email: l.tulaja@isfas.uni-kiel.de

REFERENCES

Buckingham, L. (2014). Attitudes to English teachers' accents in the Gulf. *International Journal of Applied Linguistics*, *24*(1), 50–73.

Chang Li-Ann, D. (2008). *Understand me or not? Accent, acceptability and intelligibility in International English. The case of Singapore.* Singapur <http://scholarbank.nus.edu.sg/handle/10635/13223>.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. Philadelphia: John Benjamins Publishing Company.

Höder, S. (2016). Dänische Phonetik im Kontrast zu norddeutschen Ausgangsvarietäten beim schulischen Zweitspracherwerb. *Kieler Arbeiten zur skandinavistischen Linguistik* 1.

Jenkins, J. (2000). *The phonology of English as an international language: new models, new norms, new goals*. Oxford: Oxford University Press.

Kang, O., Vo, S., & Moran, M. (2016). Perceptual judgments of accented speech by listeners from different first language backgrounds. *TESL-EJ*, *20*(1), 1-24.

Kristiansen, T. (2003). Language attitudes and language politics in Denmark. *International Journal of the Sociology of Language*, *159*, 57-71.

Krosnick, J. A., & Fabrigar, L. R. (2012). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.) *Survey measurement and process quality* (pp. 141–164). Hoboken: Wiley.

Kuznetsova A., Brockhoff, P. B., & Christensen R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Lee, J., Kim, D. J., & Park, H. (2019). Native listener's evaluations of pleasantness, foreign accent, comprehensibility, and fluency in the speech of accented talkers. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 168-178). Ames, IA.

Leiner, D. (2018). SoSci Survey. Version 3.1.06-i. Munich <http://www.soscisur-vey.com>. [Computer software].

Levis, J. M. (2016). Accent in second language pronunciation research and teaching. *Journal of Second Language Pronunciation, 2*(2), 153–159.

Miller, J. (2003). *Audible difference: ESL and social identity in schools*. Clevedoon: Multilingual Matters.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

Pilott, M. (2016). *Migrant pronunciation: What do employers find acceptable?* Wellington: Victoria University of Wellington, Diss.

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology, *Language Awareness*, *20*(1), 45-59.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439-462.

Settinieri, J. (2011). Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht, 16*, 66–80.

Thomson, R. (2018). Measurement of accentedness, intelligibility, and comprehensibility. In: Kang, O. & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). London/New York: Routledge.

Tulaja, L. (2020). *Dänische L2-Aussprache von Lernern mit Deutsch als Ausgangssprache. Fehler und Fehlerschwere.* Kiel: Christian-Albrechts-Universität zu Kiel, Diss.