SOCIAL EXPECTATIONS AND INTELLIGIBILITY OF ARABIC-ACCENTED SPEECH IN NOISE

Sophia Fiedler[1,2], Caroline Keller[1], Adriana Hanulíková[1]

[1]Deutsches Seminar, Universität Freiburg, Germany
[2]Centre de Linguistique Appliquée, Université Neuchâtel, Switzerland
sophia.fiedler@unine.ch, elanor_loth@web.de, Adriana.Hanulikova@germanistik.uni-freiburg.de

ABSTRACT

Social expectations and stereotypes may benefit or hinder speech intelligibility. Here we examine the conditions under which intelligibility of non-native speech changes as a function of the speaker's nationality. Specifically, we examine whether information about a speaker's country of origin enhances intelligibility of Arabic-accented speech in noise. Participants transcribed utterances spoken by a Syrian man after being given information about the speaker's background. When listeners were told that the speaker came from Syria, they made more transcription errors than when told that the speaker came from Portugal or when the country of origin remained unspecified. This suggests that information about nationality leads to expectations that affect intelligibility, but that these expectations vary for different nationalities. The results underline the necessity to view and model stereotypes and social expectations in a graded way.

Keywords: Arabic-accented speech, nationality, ethnicity, speech intelligibility, social expectations.

1. INTRODUCTION

A substantial body of research shows that speaker attributes, such as gender, ethnicity, social affiliation, or age, influence speech perception and evaluation of a speaker and his/her language [1, 3, 7, 11, 12, 13, 15, 18, 19, 20, 23, 25]. For example, [23] has shown that the perception of t/d-deletion in English is modulated by speaker ethnicity. Listeners were more likely to link a t/d-realization in spoken English to a Caucasian speaker and a t/d-deletion to an Afro-American speaker. Such implicit knowledge and associations with speakers depicted on photographs also affected processing of ambiguous sentences. These stereotype-driven expectations activated phonological knowledge that led to differential processing of ambiguous words. Similar results were obtained for the reduction of word-final nasal -ING [4]. When realized as a single phoneme /n/, speech was perceived as less formal and the speaker as less educated, while the full realization was perceived as more formal and the speaker as more intelligent. To evoke social stereotypes and expectations, it seems sufficient to merely mention a geographic background of the speaker [19]. When asked to choose a matching vowel from a range of resynthesized vowels, participants who were told that the speaker was Canadian chose raised-diphthong tokens, while participants believing that the speaker was from Detroit did not.

The evaluation of speech seems deeply linked with the evaluation of speakers [17, 20]. One of the first studies to show this link [20] applied the matched-guise technique and reported that hearing a native speaker of Standard American English while seeing a photograph of an Asian American woman led to reduced comprehension and more negative comprehensibility and competence ratings compared to the same speaker presented along with a photograph of a Caucasian woman. The results were explained with a bias-based model of social speech processing. Due to biases against non-native speakers, the perceived ethnicity of a speaker led to a negative perception of Standard American speech.

Follow-up studies replicated the general pattern of results but argued against a bias-based approach. For example, [1] reported that speech intelligibility (as measured by the proportion of transcribed words in utterances) dropped when native speakers of Canadian English were presented along with photographs of Asian men compared to Caucasian men. These results were interpreted based on social expectations: Listeners expect hearing foreign-accented speech upon seeing an Asian speaker and intelligibility was enhanced if the expected and perceived speech matched. The authors explain the results within the dual-route framework [24] that links social and linguistic representations.

A similar framework was suggested by [18], who reports the opposite effect for non-native speech. When the ethnicity of the ostensible speaker was Chinese, and her photograph was presented along with Chinese-accented speech, intelligibility was enhanced relative to the same speech presented along a photograph of a Caucasian speaker. [18] used a matched-guise task and a between-subject design. Three groups of participants were assigned to one of three conditions: While listening to the same Chinese-accented speaker of English in noise,

they either saw the picture of a Chinese speaker (congruent condition), a Caucasian speaker (incongruent condition), or an uninformative silhouette (control condition). Speech intelligibility (the proportion of correct transcriptions of the last word of each sentence) was higher in the congruent condition than the other two conditions. This result was interpreted in terms of matching social expectations triggered by a Chinese speaker. [18] argued that the results do not support a bias-based view [17, 20], according to which listeners reduce their attention to the speech signal or consciously misunderstand it due to stereotypes and biases. [18] argues for an exemplar-based model of socially indexed speech perception. Following this account, activation of a given social category will result in the activation of episodic traces that are consistent with or linked to the given social category. Thus, seeing a Chinese speaker should activate Chinese-accented speech and thus enhance intelligibility.

It remains unclear, however, how generalizable these effects are to other social categories and non-native accents. [18] used Asian-accented speech in his study, and listeners were familiar with this accent, at least to some extent, since Asians are a large immigrant group in the United States of America. The question arises how intelligibility is affected when listeners' familiarity with a given foreign-accent and a given ethnicity or nationality is limited or less detailed.

In the present study, we apply a similar design as [18] with two crucial differences: We provided participants only verbal information about where the non-native speaker came from (no photographs were shown in this study), and we created three congruent conditions with non-native speech that matches a non-native speaker. The mismatch that we implemented resulted only from the labeled nationality of the non-native speaker. We presented Arabic-accented German in three guises: a matching nationality of the speaker (Syria), a mismatching nationality of the speaker (Portugal), and an unspecified nationality of the speaker (a learner of German as a foreign language).

There were at least three possible outcomes for how nationality might affect speech intelligibility. Following the exemplar-based approach, the label "non-native speaker" will activate episodic traces that are consistent with the social category of a non-native speaker. This could lead to two possible outcomes. Listeners might perceive a general match (congruency) between the foreign nationality and the foreign accent in all conditions, leading to a comparable intelligibility performance (null result). Alternatively, listeners might recognize the accent to be Arabic and hence perceive a match between

accent and speaker in the Syrian guise and a mismatch in the Portuguese guise. In this case, the intelligibility of the Syrian speaker should increase relative to the Portuguese speaker, in line with [18].

The third option would be that, since the knowledge about social categories is relatively detailed [3, 13, 23], listeners link the labels Syria and Portugal to the category "non-native speakers" but that the associations for each of these labels differ along several dimensions, resulting in differences in speech intelligibility.

One of these dimensions concerns cultural proximity. A speaker from Syria is likely to be perceived as culturally more distant while Portugal is culturally less distant to Germans. Cultural proximity is interpreted in the literature in distinct ways; here we refer to perceived differences between two cultures with an emphasis on like-mindedness, geographic background, culture, language, and religion [for a review, see 21]. Given the prominence of Syria in the media at the time of testing (January 2016, just after the migration wave), we expected that listeners would more likely associate the label Syria with non-EU members, refugees, armed conflicts, migration, Arabic, and beginner learners of German. Portugal on the other hand is an EU-member and would signal cultural and historical proximity as well as a popular vacation destination. Since social expectations, prior knowledge, attitudes, and prejudice play an important role in speaker evaluation [1, 6, 7, 8, 11, 14, 15, 16, 18], perceived cultural proximity might lead to different expectations concerning German proficiency. Thus a Portuguese speaker would likely be expected to be more proficient in German, not only due to a long European tradition of teaching German as a foreign language, but also due to the in-group status as an EU-member. These expectations could then lead to increased intelligibility of the Portuguese speaker compared to the Syrian speaker.

It is important to note that we expected that German listeners would not have detailed episodic traces for accented German spoken by a Syrian or a Portuguese speaker. In 2015, Syria became the most important migration group coming to Germany [22], when about 158,657 Syrian refugees were registered within a few months. In the same year, only 133,929 Portuguese immigrants have lived in Germany. They have a different migrant status due to their membership in the EU and have been a small but relatively stable group in German society during the five years preceding this study. Therefore, we assumed that listeners would not easily identify an Arabic-accented or Portuguese-accented German. Since ability to determine a speaker's first language improves with increasing familiarity [5], we

reasoned that listeners would not be consciously aware of the mismatch between the label Portuguese speaker and the Arabic accent.

2. EXPERIMENT

2.1. Methods

2.1.1. Participants

Participants were 48 native speakers of German (mean age 24.5, range 20-30, 27 women) with no reported hearing difficulties. They were students and postgraduates at the University of Freiburg and volunteered to participate.

2.1.2. Materials

A set of 20 grammatically correct sentences (comparable to previous studies) was constructed (three practice trials and 17 critical sentences). The sentences consisted of six to nine words and were descriptive of the speaker and his native country (e.g. *Mein Haus stand am Ende der Straße.* "My house was at the end of the street", *Meine Großeltern wohnten im nächsten Dorf.* "My grandparents lived in the neighboring village"). A native speaker of Arabic from Syria was audio-recorded while reading the sentences. He learned German for approximately ten months at the time of the recordings and had an audible non-native accent. The audio-recordings were made in a home for refugees. All sentences were embedded in background noise that was created with Praat (formula randomGauss(0,0.1), noise 80dB and sentences 70dB). In addition, a questionnaire was prepared containing demographic questions, a self-evaluation, and questions of experience with living abroad, experience with foreign-accented German, and involvement in helping or teaching refugees. Speaker ratings on a scale from 0 to 5 were also included to assess sympathy, trustworthiness, comprehensibility, accent, and proximity of their own culture to the speaker's culture. The endpoint labels were provided individually for each question.

2.1.3. Procedure

Participants were tested in three groups of 16 in a quiet room.[1] They were seated at a table and received a sheet of paper and a pen. They were told that the experiment tested comprehensibility of foreign students of "German as a foreign language" and that they would hear one student who had been in Germany for 10 months and had been learning German ever since. The only difference between the three groups was the information provided about the nationality of the foreign student. The first group was told that the learner came from Syria. The second group was told that he came from Portugal. The third group was told that the student was from a foreign country.

Participants were presented with each utterance over loudspeaker one at a time. After each utterance, participants were asked to write down what the speaker said on a structured sheet indicating the number of each sentence. There was no time pressure for responses. After the transcription task was completed, participants completed the questionnaire.

2.2. Results

2.2.1. Transcription accuracy

Word responses were coded in a binary fashion as correct or incorrect (in line with previous studies). Typographical errors (e.g. *Großer Bruder* "big brother" instead of *großer Bruder*) and number errors (e.g. Bruder "brother" instead of "Brüder" "brothers") were rated as correct.
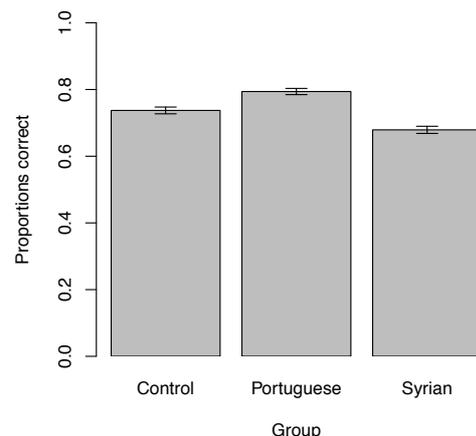

Figure 1: Mean proportion of correct transcriptions and standard errors in the three groups.

As seen in Figure 1, transcription accuracy was lower in the Syrian condition compared to the other two conditions. There were 31.6 % errors in the Syrian group, 20.6 % errors in the Portuguese group, and 24.6% errors in the control group. A logistic mixed effects regression model (R package lme4, [2]) for binary responses was fitted to the data. Transcription accuracy was taken as the dependent variable. *Group* (Syria, Portugal, Control) was entered as a fixed effect. Participants and items were modeled with random intercepts. Random slopes were also included in order to capture effects of *Group* on items. Compared to the Syrian group, there was a significant increase in transcription accuracy in the Portuguese group ($\beta = 0.72$, $z = 3.21$, $p = 0.0013$) and in the control group ($\beta = 0.42$, $z =$

2.22, $p = 0.0265$). There was no difference in accuracy between the control group and the Portuguese group ($\beta = 0.30$, $z = 1.48$, $p = 0.139$).

2.2.2. Questionnaire

Mean ratings for each group are provided in Table 1. We applied linear regression as well as a Wilcoxon Signed-ranks test and found no significant differences between the three groups. Even though descriptively there was a tendency to rate the cultural background of the Syrian speaker as more distant compared to the other two groups, this difference failed to reach significance, likely due to a relatively small number of data points.

	Syria	Portugal	Control
Comprehensibility (5 = very good)	3.3 (0.9)	2.9 (0.9)	2.8 (1.2)
Accentedness (5 = strong)	3.5 (1.0)	3.5 (0.7)	3.3 (1.0)
Sympathy (5 = high)	3.9 (0.7)	3.6 (1.0)	3.8 (0.9)
Trustworthiness (5 = high)	3.9 (0.8)	3.8 (0.8)	3.6 (0.7)
Cultural proximity (5 = strong)	2.5 (1.2)	3.0 (1.0)	2.7 (1.1)

Table 1: Mean speaker ratings for each group. Standard deviation is provided in brackets.

3. DISCUSSION

Using a matched-guise task, we presented Arabic-accented German speech in noise and asked participants to write down what the speaker said. When participants were told that the speaker was Syrian (match between speaker and accent), they performed worse than when told that the speaker was Portuguese (mismatch between speaker and accent) or when no specific country of origin was provided.

There are several possible explanations for why the Syrian guise was least intelligible. One possibility is that participants' expectations towards the Syrian's German proficiency were lower than towards the Portuguese speaker or the speaker in the neutral condition. The Syrian refugees' situation had been present on daily basis in the German media in 2015 and 2016. Syrian people were therefore likely to be perceived as a new group of immigrants in need of linguistic integration and this might have led to lower expectations towards their linguistic skills.

Second, it is possible that the cultural proximity also modulated expectations and assessments of proficiency. As an EU-member, Portugal is more likely to have an in-group status, while Syria is not part of Europe and likely to have an out-group status. Even though the ratings showed a tendency to

perceive the Syrian speaker as culturally more distant from Germans than the Portuguese speaker, they did not reach significance. Nevertheless, factors such as cultural proximity and in-group and out-group membership may lead to different expectations towards linguistic skills and affect intelligibility. While the subjective ratings in this study were positive overall and do not confirm negative biases towards different groups, the possibility that they influence speech intelligibility cannot be excluded. According to previous studies, stigmatized groups and their language are often rated negatively [16, 17]. In contrast, the rating data here did not show any significant patterns that would mirror the intelligibility data. While differences in subjective ratings and functional data such as intelligibility are not necessarily surprising [9, 10], rating data cab vary and even be shifted within an experiment depending on various factors (e.g. static vs. moving pictures) [26]. Future studies should examine additional nationalities and in-group/out-group memberships to show the exact nature of such linguistic expectations.

Finally, it is important to note that we only included one speaker in a between-subject design. While this was intended to keep the design (as well as the number of participants) similar to [18], it could be argued that a within-subject design and more speakers would yield a different result. However, given that social effects on speech intelligibility or evaluation have been replicated in both between- and within-subject designs and with one or more speakers [1, 11, 18], this factor alone probably cannot account for the present findings.

Taken together, the results indicate that speech intelligibility is affected by factors that are not exclusively linked to biases towards non-native speakers, as suggested by [20], or to episodic traces and a simple congruency between a non-native speech and a non-native speaker [18]. It seems that the effect of non-linguistic information on speech processing is based on more complex knowledge about social categories that cannot be reduced to the notion of 'familiarity' since [18] showed that it doesn't influence the intelligibility rate [but see 16]. The fact that information about a non-native speaker's nationality co-determined speech intelligibility suggests that a more nuanced view of social expectations, as well as in-group and out-group memberships is needed. A framework that links social categories and linguistic representations may account for these results, but it remains to be addressed which specific social categories are activated that alter processing of an otherwise identical speech signal.

4. REFERENCES

[1] Babel, M., Russel, J. 2015. Expectations and speech intelligibility. *J. Acoust. Soc. Am.* 137/5, 2823-2833.

[2] Bates, D., Maecheler, M., Bolker, B., Walker, S. 2016. *lmer4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-12. http://CRAN.R-project.org/package=lme4.

[3] Brown, K. 1992. American college student attitudes toward non-native instructors. *Multilingua* 11, 249-265.

[4] Campbell-Kibler, K. 2005. *Listener perceptions of sociolinguistic variables: The case of (ING)*. Doctoral dissertation, Department of Linguistics, Stanford University.

[5] Derwing, T. M., Munro, M. J. 1997. Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition* 19(1), 1–16.

[6] Fiske, S. 2000. Stereotyping, prejudice, and discrimination at the seam between the centuries: evolution, culture, mind, and brain. *European Journal of Social Psychology* 30, 299-322.

[7] Fraser, H. 2003. Issues in transcription: factors affecting the reliability of transcripts as evidence in legal cases. *Forensic Linguistics* 10(2), 203-226.

[8] Glock, S., Krolak-Schwerdt, S. 2013. Does nationality matter? The impact of stereotypical expectations on student teachers' judgments. *Social Psychology of Education* 16(1), 111–127.

[9] Hanulíková, A. 2019. Perceived ethnicity effects on speech processing across the life span: a comparison between subjective and objective measures, Poster presented at *Aging and Cognition*, Zürich.

[10] Hanulíková, A. 2018. Effects of talker identity on speech comprehension across the lifespan, Talk presented at *Socially situated language processing (SSLP 2018)*, ZAS, Berlin.

[11] Hanulíková, A. 2018. The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard* 4(1), 2199-174X.

[12] Hanulíková, A. et al. 2012. When one person's mistake is another's standard usage: The Effect of Foreign Accent on Syntactic Processing. *J. of Cog. Neuroscience* 24 (4), 878–887.

[13] Hay, J., Nolan A., Drager K. 2006. From *fush* to *feesh*. Exemplar priming in speech perception. *The Linguistic Review* 23, 351-379.

[14] Hugenberg, K., Sacco D. 2008. Social categorization and stereotyping: How social categorization biases perception and face memory. *Social and Personality Psychology Compass* 2(2), 1052-1072.

[15] Lattner, S., Friederici, A. 2003. Talker's voice and gender stereotype in human auditory sentence processing: Evidence from event-related brain potentials. *Neuroscience Letters* 339, 191–194.

[16] Lindemann, Stephanie 2005. Who speaks "broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics* 15(2), 187-212.

[17] Lippi-Green, R. 2012. *English with an accent: Language, ideology and discrimination in the United States*. London/NewYork: Routledge.

[18] McGowan, K. 2015. Social expectations improve speech perception in noise. *Language and Speech* 58(4), 502-521.

[19] Niedzielski, N. 1999. The Effect of Social Information on the Perception of Sociolinguistic Variables. *J. of Language & Social Psychology* 18(1), 62-85.

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

[21] Spencer-Oatey, H. 1996. Reconsidering power and distance. *Journal of Pragmatics* 26(1), 1-24.

[22] Statistisches Bundesamt. 2016. *Ausländische Bevölkerung nach ausgewählten Staatsangehörigkeiten*. www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Tabellen/auslaendische-bevoelkerung-staatsangehoerigkeit-jahre.html

[23] Staum Casasanto, L. 2008. Does social information influence sentence processing? In Love, B. C., McRae, K., Sloutsky, V. M. (eds.), *Proceedings CogSci Society*, Austin, TX, 799-804.

[24] 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. *Front. Psychol.* 4, 1015, 1–13.

[25] Van Berkum, J. et al. 2008. The neural integration of speaker and message. *J. of Cog. Neuroscience* 20(4), 580-591.

[26] Zeng, Y., Samuel, A. G. 2017. Does seeing an Asian face make speech sound more accented? *Atten. Percept Psychophys* 79, 1841-1859.