

People with larger social networks show poorer voice recognition



Shiri Lev-Ari 

Quarterly Journal of Experimental Psychology
2022, Vol. 75(3) 450–460
© Experimental Psychology Society 2021



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17470218211030798
qjep.sagepub.com



Abstract

The way we process language is influenced by our experience. We are more likely to attend to features that proved to be useful in the past. Importantly, the size of individuals' social network can influence their experience, and consequently, how they process language. In the case of voice recognition, having a larger social network might provide more variable input and thus enhance the ability to recognise new voices. On the other hand, learning to recognise voices is more demanding and less beneficial for people with a larger social network as they have more speakers to learn yet spend less time with each. This paper tests whether social network size influences voice recognition, and if so, in which direction. Native Dutch speakers listed their social network and performed a voice recognition task. Results showed that people with larger social networks were poorer at learning to recognise voices. Experiment 2 replicated the results with a British sample and English stimuli. Experiment 3 showed that the effect does not generalise to voice recognition in an unfamiliar language suggesting that social network size influences attention to the linguistic rather than non-linguistic markers that differentiate speakers. The studies thus show that our social network size influences our inclination to learn speaker-specific patterns in our environment, and consequently, the development of skills that rely on such learned patterns, such as voice recognition.

Keywords

Social networks; talker identification; voice recognition; communicative needs

Received: 13 April 2021; revised: 16 June 2021; accepted: 17 June 2021

Language processing is directed by our experience. Our experience shows us which aspects in our input are relevant. For example, we can learn from it whether lexical stress can influence meaning in our language, or whether word order carries any grammatical or pragmatic meaning. In addition, our past experience also teaches us how useful each cue is in terms of how often and how reliably it could facilitate processing in the future. Importantly, our social experience can influence our input and thus both what we can learn and the relative utility of learning it. Past research on the effects of social network size has found it to promote linguistic skills such as speech perception (Lev-Ari, 2018), comprehension of evaluative language (Lev-Ari, 2016) and lexical prediction (Lev-Ari, 2019; Lev-Ari & Shao, 2017) by providing more variable input. This paper tests a case where past experience would, on the one hand, provide more variable input that can promote learning, but on the other, would demonstrate lower benefit for learning. In particular, the paper tests whether social network size influences voice recognition, and if so, in which direction.

Linguistic input is multi-dimensional. Each token we hear contains information about the sounds, their duration, intonation, stress, loudness, and so forth. Some of this information is important for language processing whereas some of it might reflect noise, that is, features that are not used contrastively in the language to identify a linguistic unit. Languages differ in which cues are relevant, and an important part of language learning is discovering which of these cues (e.g., stress, pitch, word order) are relevant for language processing and in what way. Therefore, speakers of different languages attend to different cues during language processing. For example, speakers of languages with contrastive stress attend to stress more than

Royal Holloway, University of London, Egham, UK

Corresponding author:

Shiri Lev-Ari, Royal Holloway, University of London, Egham TW20 0EX, UK.

Email: shiri.levari@rhul.ac.uk

speakers of languages in which stress is not contrastive (Dupoux et al., 2008). Similarly, the boundaries between voiced and voiceless stops vary across languages, and speakers of different languages are accordingly differentially sensitive to different regions along the voice onset time continuum with heightened sensitivity around phoneme boundaries in their language compared with non-boundary stretches of the continuum (e.g., Liberman et al., 1957).

The role of linguistic knowledge in processing

The process of voice recognition is a prime example of the role of past experience in language processing. By definition, voice recognition is the ability to link an incoming voice with past experience, and it has been found that the processing of familiar and unfamiliar voices have different neural signatures (Belin et al., 2011; Maguinness et al., 2018). Furthermore, individuals' linguistic experience moderates individuals' ability to recognise voices (e.g., Kadam et al., 2016; Perrachione et al., 2011).

To better understand why that is the case, it is useful to consider how voices are recognised. One of the leading current models of voice recognition is the prototype model (Lavner et al., 2001). According to this model, people construct a representation of an average speaker based on their past experience. Individual speakers are then represented by their deviation from the average speaker on different features. Speakers differ in which features they deviate from the average speaker, and therefore different features are most useful for identifying different speakers. Listeners correspondingly rely on different features to identify different speakers (Lavner et al., 2001; Van Lancker et al., 1985). This suggests that individuals with different experience could theoretically extract different average speakers, as well as find different features to be most useful for identifying a speaker and distinguishing that speaker from other speakers. Indeed, different listeners not only differ in their voice recognition skills but seem to rely on different features to recognise voices (Lavner et al., 2001).

When it comes to social network size, individuals with larger social networks are likely to have a more diverse and representative sample of the population. This could allow them to correctly identify which features best distinguish a speaker from others. For example, individuals with a larger social network might be better able to determine whether producing a vowel with certain formant frequencies is indeed atypical and therefore a good distinguishing factor whereas it might only be under-represented in one's sample but not in the population. This hypothesis is in line with recent findings that social network size boosts speech perception because it provides greater knowledge of the distribution of formant frequencies (Lev-Ari, 2018). Furthermore, having a larger social

network might allow one to determine that even though the formant frequencies of the vowel themselves are not atypical, their co-occurrence with another feature is, as the greater input variability in larger networks can facilitate learning of such conditional probabilities in the input (e.g., Gómez, 2002; Lev-Ari, 2016). Greater ability to identify and use such conditional probabilities can further improve speaker representation and recognition.

As mentioned above, there is evidence that social network size can influence speech perception and voice recognition is related to speech perception. People are better at understanding speech in noise when it is produced by speakers they had been familiarised with even if they had never heard the speaker produce those words (Holmes et al., 2021; Nygaard et al., 1994). Correspondingly, voice recognition improves with the addition of linguistic elements: voice recognition is worst with reversed speech, better with pseudo-words, even better with real words and best with meaningful passages (Goggin et al., 1991). Voice recognition in a native language has also been found to be poorer among individuals with dyslexia, and among these individuals, to be better the better their phonological skills are (Perrachione et al., 2011). The importance of linguistic experience is particularly evident in the superior performance listeners have at recognising speakers in their own language versus a foreign language, a phenomenon known as the Language Familiarity Effect (e.g., Goggin et al., 1991; Köster & Schiller, 1997; Perrachione et al., 2011). Thus, English-speaking listeners recognise English speakers better than German speakers, yet German-speaking listeners recognise German speakers better than English speakers (Goggin et al., 1991). Moreover, people show better recognition of the same bilingual speaker when the speaker speaks the native language of the listener rather than a foreign one (Goggin et al., 1991). The superior performance in a familiar language is argued to be due to phonological knowledge of the language (Fleming et al., 2014; Johnson et al., 2018; Perrachione et al., 2011). The process of language learning involves learning the distribution of phonetic features in the language as well as learning to distinguish linguistically conditioned variability from speaker-conditioned variability (Rost & McMurray, 2010). That is, past experience guides listeners' attention to the relevant features that allow voice recognition.

At the same time, voice recognition also relies on non-linguistic properties. Thus, people can recognise speakers even when presented with signals that have been stripped of their linguistic properties, such as reversed speech (Sheffert et al., 2002), pre-verbal infants can recognise voices, although their performance is better in the language they are acquiring (Johnson et al., 2011), and even though people's voice recognition in a foreign language is poorer, it is still above chance, and not dependent on phonological skills (Perrachione et al., 2011). Neural evidence also points to some independence between speech perception and

voice recognition as individuals with aphasia may or may not show impairments in voice recognition, depending on the location of their lesion (Belin et al., 2011).

The role of motivation in processing

Past experience can guide our attention not only by teaching us which properties are relevant (e.g., stress, pitch, variation in VOT in phoneme boundary regions vs non-boundary regions) but also by allowing us to evaluate the utility of encoding and relying on each property. Indeed, the reason we fail to attend to relevant features in a second language is that our experience taught us that they are of little utility in our first language. Thus, we might not attend to differences between vowels if they are allophonic in our language, but we would if they are contrastive. Social factors can also influence the utility of cues. For example, English-learning infants maintain the distinction between phonological contrasts in Mandarin that are irrelevant for English if they regularly interact with a Mandarin speaker, but not if they only view the same Mandarin speaker on TV (Kuhl et al., 2003), potentially because the relevance of a person on TV might seem lower. Children and infants also acquire the accent of the environment they grow up in rather than the accent of their caregivers (Floccia et al., 2012), potentially also because the accent of their environment could be seen as having greater utility. Perhaps most telling is the finding that individuals are more likely to encode speaker-specific information if they think that the speaker is from their ingroup (same university) rather than an outgroup (a different university; Iacozza et al., 2020).

There is some evidence that voice recognition can also be influenced by attention or motivation. For example, bilinguals seem to be better at voice recognition, and it has been suggested that this might be partly due to their greater social perception (Levi, 2018). It has also been found that even though people are better at voice recognition in their native language (The Language Familiarity Effect), they show greater voice change *blindness* in their native language, presumably because they devote more resources to semantic processing, leaving fewer for encoding of indexical information (Neuhoff et al., 2014).

Individuals' social network size might influence individuals' attention to the indexical cues in speakers' speech and the motivation to learn them. Learning speaker-specific speech patterns would be more demanding for individuals with larger social networks as they would have more voices to learn. Furthermore, the utility of learning each of these voices would be reduced as people with larger social networks are likely to interact less with each member of their network. Therefore, past experience with a large social network might discourage individuals from attempting to encode speakers' idiosyncratic features that could assist in their recognition.

To summarise, the size of individuals' social network could influence their voice recognition skills in two different ways. On the one hand, having a larger social network provides individuals with more variable and representative input that can allow them to know better which features are most informative for distinguishing a speaker from others. It can therefore improve their voice recognition ability. On the other hand, having a larger social network might discourage individuals from attending to the idiosyncratic aspects of someone's speech and from attempting to encode and store them. The studies reported here test whether individuals' social network size influences voice recognition, and if so, whether its effect is positive or detrimental. Experiment 1 tests this question with a Dutch sample, Experiment 2 replicates the results with a British sample, and Experiment 3 tests whether the effect of social network size on voice recognition is driven by influencing encoding of linguistic or non-linguistic cues.

Experiment 1

To test whether social network size influences voice recognition, native Dutch speakers provided information about their social network size and then were tested on their ability to learn to recognise unfamiliar native Dutch speakers.

Method

Participants. Sixty-four native Dutch speakers participated in the experiment for pay. One participant failed to complete the experiment on time, and the social network questionnaire of another participant was lost due to a computer error. The results are therefore reported for the remaining 62 participants. Most studies on voice recognition examine the effect of categorical factors rather than individual differences. Therefore, sample size was determined a-priori to be around 60, which is larger than the samples that most previous studies used (e.g., Fleming et al., 2014; Johnson et al., 2018; Köster & Schiller, 1997; Sheffert et al., 2002) and the same as in another study that examined the effect of social network size on speech perception using the same statistical methods (Lev-Ari, 2018). The study followed the ethical approval procedure of the Max Planck Institute for Psycholinguistics.

Materials

Social network questionnaire. The social network questionnaire used in this experiment is the same online questionnaire used in previous studies investigating effects of social network size (e.g., Lev-Ari, 2019). Participants were asked to list all the people they regularly interact with for at least 5 minutes each week. They were asked to only include native speakers above the age of 12. For each speaker, participants indicated age and relation to them, and the patterns of interaction of that speaker with

other people in the network. In addition, participants indicated how many hours they spend interacting with these people each week. Participants' social network size was defined as the number of interaction partners they listed in the questionnaire. The number of Hours of interaction was also coded to control for amount of input and ensure that any effects of social network size are not due to differences in amount of input.

Voice recognition. Four female native Dutch speakers recorded 12 monosyllabic Dutch words each. Six of these words were used during both learning and test stages, and the other six words were used only during the test, to examine generalisation. The words in the learning and generalisation sets contained the same vowels. Two words of each set contained each of the following vowels: /a/, /ei/, and /œy/ (See the online Supplementary Material A for the full list of stimuli). Four colourful geometrical shapes with eyes and a mouth served as avatars for the speakers.

Procedure. All participants answered the social network questionnaire first, and then performed the voice recognition task. The voice recognition task started with six familiarisation rounds, one per word. In each round, participants heard the same word from each of the four speakers in random order. The avatar of the speaker appeared in the middle of the screen 500 ms before word onset, and remained for 4 s in total. After all four tokens have been presented, participants were tested on their ability to recognise the speakers. Each of the four tokens was presented in random order. The avatars of all four speakers appeared on the screen with numbers below them. Participants indicated which of the four speakers had said the word. Responses were self-paced without a time limit. Participants received feedback, but progressed to the next round regardless of the accuracy of their performance. The order of the six familiarisation rounds was randomised per participant.

After participants completed all six familiarisation rounds, they started the training rounds. In each round, all 24 tokens (six words by each of the four speakers) appeared in random order. Participants always saw all four numbered avatars on the screen and indicated which speaker had produced the token they had just heard. Participants responded at their own pace without a time limit. They received feedback on their responses. If participants responded correctly to at least 18 out of the 24 trials, learning ended and participants progressed to the test. If their accuracy was lower, they received another training round for a maximum of 10 rounds. If participants failed to reach criterion after 10 rounds, they still progressed to test. The number of training rounds ranged from 1 to 10, with an average of 6.5.

During test participants heard all 48 tokens (6 new words and 6 trained words from each of the four speakers) in random order. The screen always showed all four numbered avatars. Participants indicated which speaker produced the word by typing the number associated with the

corresponding avatar. Responses were self-paced without a time limit. Participants did not receive feedback on their responses.

Results and discussion

Participants' Social Network Size ranged from 5 to 61 ($M=21.4$, $SD=13.24$) and the number of hours of interaction ranged from 5 to 80 ($M=31.25$, $SD=16.8$). To examine the effect of social network size on voice recognition, a logistic mixed effects analysis was run on accuracy during the test round using the lme4 package (Bates et al., 2010) in R (R Core Team, 2020). The model included Participants, Word, and Speaker as random factors, and Social Network Size (scaled), Hours (scaled) and Word Type (trained, novel), as well as the interactions of Word Type with Social Network Size and with Hours as fixed effects. Here and in all analyses, we included the maximal random structure that still converged. When models showed singular fit or failed to converge, slopes that had a correlation of $|1|$ with the intercept or other slopes were removed one by one. If the model still failed to converge, the slopes that contributed the least to the model were removed one by one until the model successfully converged. In this case, slopes were not included as they led to singular fit. Results showed an effect of Word Type such that voice recognition was better for trained words than for novel words ($\beta=-0.55$, $SE=0.23$, $z=-2.45$, $p<.02$; See Supplementary Material B for the full table of results). Results also showed an effect of Social Network Size at the reference level (trained words), such that the larger participants' social network, the *worse* they performed in the voice recognition task ($\beta=-0.28$, $SE=0.09$, $z=-3.21$, $p<.01$). There was also a just significant interaction between Social Network Size and Word Type that suggested that the effect of Social Network Size might be smaller for novel words than it is for trained words ($\beta=.17$, $SE=.09$, $z=1.96$, $p<.05$; See Figure 1). Note that chance performance is 25%, so the potentially smaller effect of social network size for novel words is unlikely to be due to floor effects, as performance was still better than chance. The number of hours of interaction did not influence voice recognition ability nor did it interact with Word Type (all z 's < 1). For all experiments, alternative analyses were also run in which models were pruned to remove n.s. interactions and predictors one by one. The results of all pruned models are reported in the online Supplementary Material B as well. In this experiment, the removal of n.s. predictors and interactions led the interaction of Word Type and Social Network Size to become only marginal ($\beta=0.13$, $SE=0.08$, $z=1.72$, $p<.09$).

If reliable, the interaction between Social Network Size and Word Type would suggest that individuals with smaller networks may attempt to encode speakers' characteristics more, but that they fail to identify speakers' identifying characteristics, and instead, learn features that are

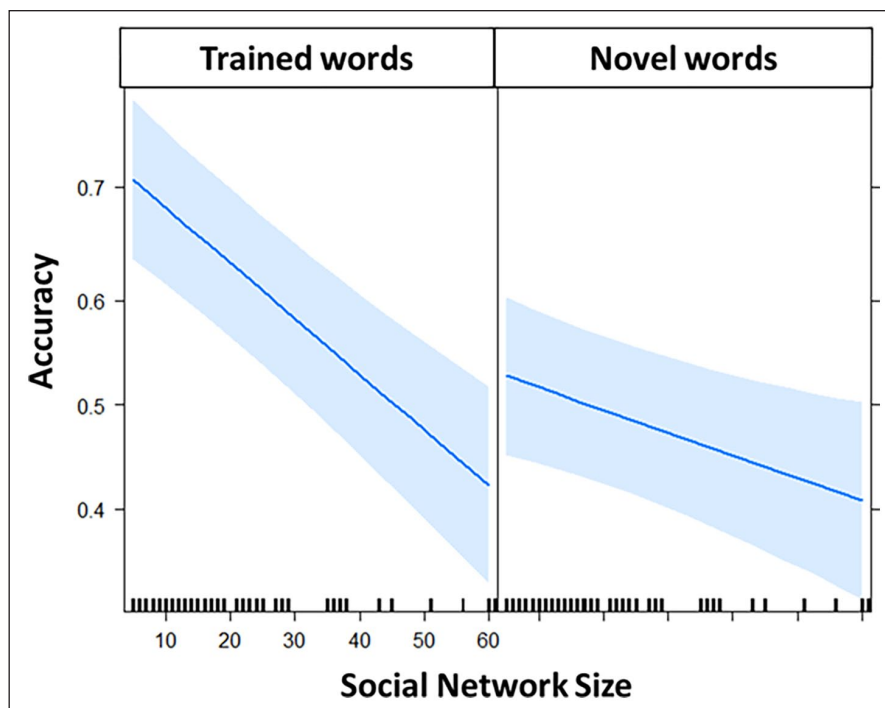


Figure 1. The effect of Social Network Size on voice recognition accuracy as dependent on Word Type (trained, novel). Light blue bands indicate standard error.

word-specific. Consequently, their learning does not generalise well to novel words and their ability to recognise the speakers in general is more limited. The introduction reviewed research that suggests that those with smaller social networks might be more motivated to learn to recognise voices but that they might also be less able to do so, because the reduced variability in the input that they receive could reduce their ability to learn the distribution of phonetic features in the language and learn contingencies in the input and identify speakers' characteristics (e.g., Gómez, 2002; Lev-Ari, 2018; Rost & McMurray, 2010). That said, it is preliminary to draw these conclusions from the current results. While the effect reached conventional level of significance in the main analysis, it was barely so ($p = .049$), it was only marginal in the alternative analysis reported in the Supplementary Material B, and as reported later, did not replicate in Experiment 2.

The results of Experiment 1 reveal that having a larger social network is associated with worse learning of others' voices. As this is a novel finding and it was not clear a-priori in which direction the effect would be, Experiment 2 repeated the experiment using a different language sample and different stimuli to ascertain the results' reliability.

Experiment 2

Method

Participants. Ninety-nine participants were recruited. All participants were students at a British university in the outskirts

of London and received credit for their participation. As it was not possible to limit participation only to native speakers and data was predicted to be noisier in this online version, recruitment remained open until the end of term with the aim of recruiting as many participants as possible. The study followed the ethical approval procedure of Royal Holloway, University of London.

Materials

Social Network Questionnaire. The Social Network Questionnaire had similar instructions to those in Experiment 1 except that it was run on a different experimental platform and most of the additional questions from Experiment 1 that were never analysed were not included. Participants were asked to report the size of their social network on one screen and list the network on the following screen. Unfortunately, most of the data from the second screen was lost due to software issues so analyses are based on the information provided in the first screen.

Voice recognition task. Twelve monosyllabic words were recorded by three native speakers of British English, all undergraduate students at the university.¹ As in Experiment 1, six of the words were used in both training and test, and six only at test (See Supplementary Material A for the full list of stimuli).

Procedure. The experiment was conducted online via Gorilla (<https://app.gorilla.sc>). Participants first listed their social network using the same instructions as in Experiment 1 and

indicated what their native language was. The voice recognition task itself was the same as in Experiment 1 with the following exception: there were three rather than four speakers, because a pilot study indicated that performance was lower in this sample, potentially because of the online nature of the study. Therefore, each training round included 18 rather than 24 trials and the test round included 36 rather than 48 trials. The criterion was set at 15 correct responses out of 18. In addition, if participants did not provide any response within 8 s, the experiment progressed to the next trial.

Results and discussion

Before analysing the results, all responses that were faster than the duration of the audio file were removed. This led to the exclusion of 350 responses (9.8%). Furthermore, if participants responded before the audio file completed playing on > 20% of trials, the remaining responses from these participants were excluded as well. This led to the exclusion of 15 participants. Analyses were therefore conducted on the remaining 84 participants. Participants' social network ranged from 1 to 30 ($M=10.25$, $SD=6.84$). Sixty-nine of the participants were native speakers and 15 were non-native speakers. These were native speakers of Cantonese, German, Greek, Hindu, Italian, Lithuanian, Montenegrin, Polish, Portuguese, Romanian, Spanish, and Urdu.

To test whether Social Network Size predicts voice recognition, a logistic mixed effects model was conducted. The model included Social Network Size (scaled), Word Type (trained, novel), Native Status (NS, NNS) and the interaction of Social Network Size and Word Type as fixed effects and Accuracy as the dependent measure.² The random structure included intercepts for Participants and Items and a by-Items slope for Social Network Size.³ Results revealed a negative effect of Social Network Size ($\beta=-0.26$, $SE=0.09$, $z=-2.89$, $p<.01$; See Figure 2 and Supplementary Material B), replicating the results of Experiment 1. The analysis also showed an effect of Native Status ($\beta=0.80$, $SE=0.20$, $z=4.0$, $p<.001$) reflecting the fact that native speakers performed better than non-native speakers. This finding is in line with prior findings on the Language Familiarity Effect and reflects the role of linguistic knowledge in voice recognition (e.g., Goggin et al., 1991; Köster & Schiller, 1997; Perrachione et al., 2011). An alternative analysis in which n.s. interactions and predictors were removed yielded the same results and is reported in Supplementary Material B. As a cautionary step, to ensure our results are not influenced by the mix of native and non-native speakers in the same analysis, we also analysed separately the results of the native and non-native speakers. Both groups showed the same negative effect of Social Network Size (Native speakers: $\beta=-0.24$, $SE=0.09$, $z=-2.52$, $p<.02$; Non-native speakers: $\beta=-0.37$, $SE=0.18$, $z=-2.0$, $p<.05$). The results of Experiment 2

thus replicated those of Experiment 1 with speakers of a new language and a new set of stimuli.

The results of Experiments 1 and 2 indicate that Social Network Size influences voice recognition, and that its influence is detrimental, potentially because interacting with a larger social network reduces the utility of encoding speakers' idiosyncratic patterns that enable voice recognition. This raises the question of what listeners with smaller social networks encode when they process speech. In particular, voice recognition depends on both linguistic and non-linguistic properties. Thus, individuals can recognise voices in reversed speech, which removes many phonetic cues but maintains acoustic cues, but they can also recognise voices in sinewave speech, which removes acoustic cues while maintaining phonetic information (e.g., Sheffert et al., 2002). Therefore, the better performance of individuals with smaller social networks could be due to better encoding of linguistic properties (e.g., specific vowel articulation), better encoding of non-linguistic properties (e.g., voice timbre), or both. To test that, Experiment 3 tested the effect of social network size on voice recognition in a foreign language.

Individuals are better at voice recognition in their native language than in a foreign language (e.g., Goggin et al., 1991; Köster & Schiller, 1997; Perrachione et al., 2011). The reason for that is that individuals can rely on both linguistic and non-linguistic cues in their native language but only non-linguistic cues are available when recognising voices in a foreign language. Therefore, phonological skills correlate with voice recognition in a native language, but not in a foreign language (Perrachione et al., 2011). If social network size still leads to better performance even in a foreign language, then the better performance of individuals with smaller social networks is at least partly due to better encoding of non-linguistic properties. In contrast, if social network size does not influence voice recognition in a foreign language, it would suggest that the better performance of individuals with smaller social networks found in Experiments 1 and 2 might be due to better encoding of linguistic properties.

Experiment 3

Experiment 3 tests whether social network size influences voice recognition in a foreign language. In this experiment, native speakers of British English performed the same voice recognition task as in Experiment 1 with the same Dutch speakers. The relation of their performance to their social network size was measured. The study followed the ethical approval procedure of Royal Holloway, University of London.

Method

Participants. One-hundred-and-seventy-three participants completed the study via the online platform Prolific (<https://www.prolific.co/>). Nine participants were excluded for

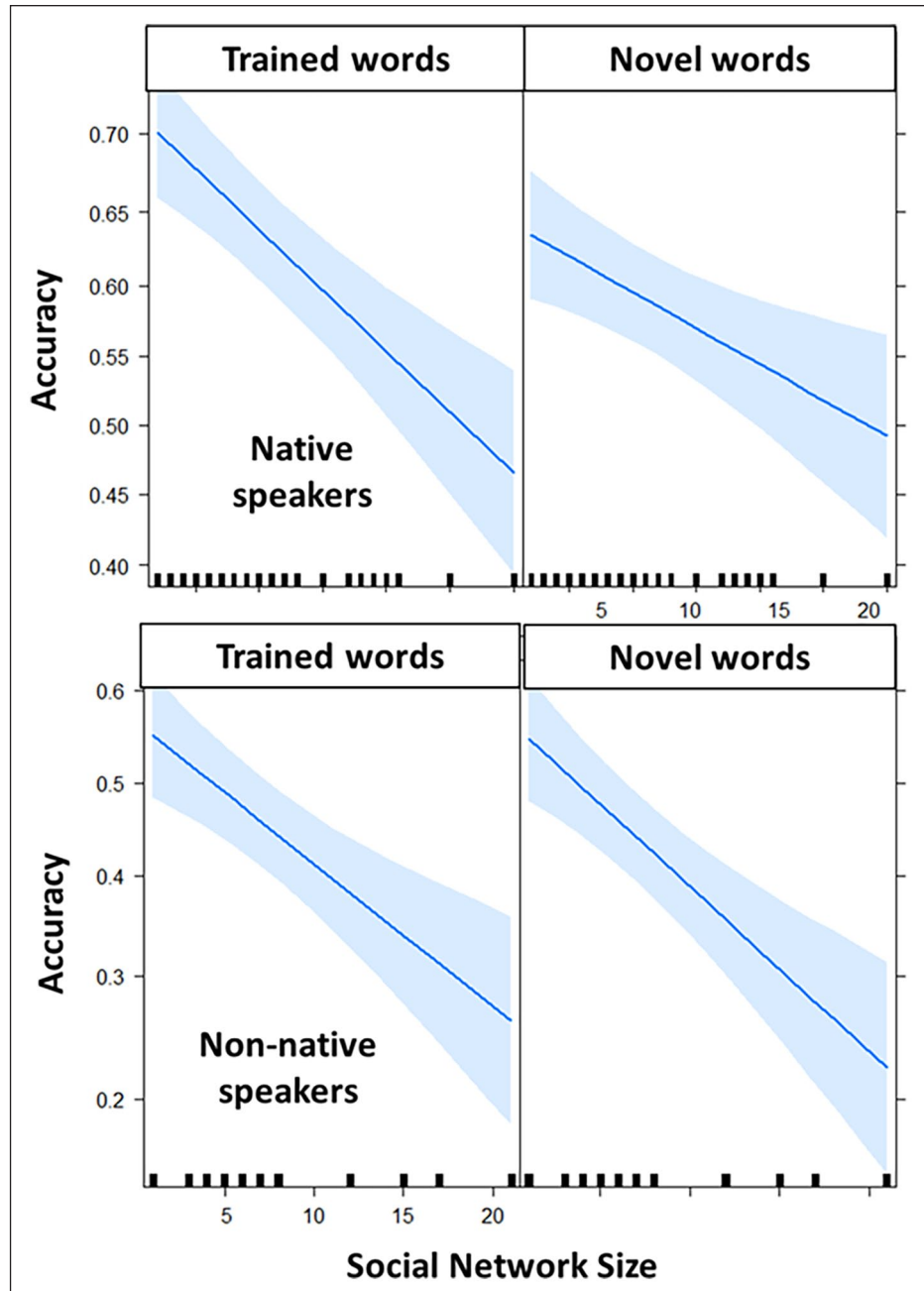


Figure 2. The effect of Social Network Size on accuracy in voice recognition in trained and novel words among native speakers (top panels) and non-native speakers (bottom panel). Light blue bands indicate standard error.

reporting speaking Dutch or not reporting whether they spoke Dutch. Twenty-nine participants were excluded for not reporting their social network size or for reporting social networks sizes that differ by 10 people or more in their responses to the two social network size questions (see below). Finally, 14 participants responded before the audio completed playing on > 20% of the trials or did not provide any response and let all trials time out. Results are reported for the remaining 121 participants. As in Experiment 2, a relatively large sample size for a voice recognition study

was recruited as it was expected that the online version would lead to noisier data than a lab experiment.

Materials

Social network questionnaire. The social network questionnaire was similar to the one used in Experiment 2. Participants were asked to report the size of their social network on one screen and list the network on the following screen. This led to cases in which participants reported different number of interaction partners on the two

screens. When the difference between the 2 reports was 10 or greater, participants were excluded. When the estimated network sizes differed by < 10 , the network size from the second detailed screen was used as it seemed more reliable (e.g., fewer round numbers).

Voice recognition task. The stimuli for this task were the same as for Experiment 1.

Procedure. The procedure in Experiment 3 was the same as in Experiment 1 except that it was run online via PsyToolkit (<https://www.psychtoolkit.org/>) rather than in the lab.

Results

To investigate the role of social network size on voice recognition in a foreign language, a logistic mixed model analysis was run on participants' accuracy of voice recognition. Network size ranged from 1 to 52 ($M=15.17$, $SD=9.31$) and number of hours of interaction ranged from 1 to 100 ($M=30.95$, $SD=23.22$). The model included Participants, Word, and Speaker as random variables, and Social Network Size (scaled), Hours (scaled), Word Type (trained, novel), and the interactions of Word Type with Social Network Size and Hours as fixed factors. The random structure included by-speaker slope for Word Type. Other slopes were not included as the model failed to converge otherwise. Results did not reveal any effects though there was a marginal effect of Word Type ($\beta=0.18$, $SE=0.09$, $z=1.9$, $p < .06$; See Supplementary Material B for the full results). Importantly, Social Network Size did not predict performance on its own or in interaction with Word Type (both z s $< |1|$). An alternative analysis in which n.s. interactions and predictors were removed also yielded a marginal effect of Word Type and no other effects or interaction. It is reported in Supplementary Material B. Similarly, analysing the data only with the English native speakers ($N=113$) yielded the same marginal effect of Word Type ($\beta=0.16$, $SE=0.09$, $z=1.76$, $p < .08$) and no effects or interactions with Social Network Size ($z < |1.1|$).

Experiment 3 tested whether the better performance of individuals with smaller social networks in Experiments 1 and 2 is due to greater attention to linguistic cues, non-linguistic cues or both. If individuals with smaller social networks pay closer attention to non-linguistic cues, they should be better at voice recognition in a foreign language as well. Experiment 3 shows that this is not the case. As Experiment 3 used the same stimuli as Experiment 1 and was online like Experiment 2, the absence of an effect of social network size cannot be due to insufficient sensitivity in the stimuli or procedure. While one cannot draw strong inferences from null results, the absence of an effect in Experiment 3 suggests that the better performance of individuals with smaller social networks in Experiments 1 and 2 is due to their reliance on linguistic cues rather than non-linguistic cues in voice recognition.

One difference between the first two experiment and the third one is the similarity of speakers and listeners. In Experiment 1 most participants were university students, often female, and the speakers were women in their early 20s. In Experiment 2, all participants were first year undergraduate students, mostly female, and the speakers were female undergraduate students as well. In contrast, the participants in Experiment 3 were recruited more widely and therefore might have shown greater variation in age and gender. It is therefore possible that the effect of social network size is stronger in cases where speakers and listeners are similar in their demographic characteristics and are more likely to be included in each other's social network. Future research should examine this possibility further.

General discussion

The experiments in this paper show that the size of individuals' social networks is negatively associated with their voice recognition ability. This result is in line with the hypothesis that interacting with a larger social network reduces the utility of encoding speaker-specific patterns. One potential reason for that is that individuals who regularly interact with more people have more speakers to encode. Furthermore, assuming that such individuals spend less time with each interaction partner, a pattern that was found in these studies,⁴ the benefit of encoding that person's voice might be reduced.

One limitation of the experiments is that social network size was not manipulated. Therefore, it is possible that it is not social network size that influences voice recognition but something that correlates with it. The experiments in this paper controlled for amount of interaction but not for other factors that might correlate with social network size. Nevertheless, while there is no evidence for the causal role of social network size, the results suggest that in the real world people with smaller social networks would be better at voice recognition even if potentially because of other factors that could correlate with social network size.

Another limitation is that in Experiment 2 only the overall report of the social network size was collected for all participants, whereas information about the number of hours of interaction and the detailed list of network contacts were lost for most of the participants, leading to their exclusion from analysis. While it is impossible to know for certain the effect that the missing information could have had, the number of hours of interaction did not influence voice recognition either on its own or in interaction with Word Type. Therefore, it is unlikely that it would have had an effect here. The detailed information about the social network was mostly used in Experiment 3 to exclude participants whose responses to the general and detailed social network question meaningfully differed. These participants are likely to be inattentive and non-compliant. The inability to detect and exclude them in Experiment 2 means that it is possible that some participants with unreliable

network size estimates and inattentive voice recognition performance were included. The results of Experiment 2 might therefore include some additional noise. There is no reason to assume though that these participants would systematically distort the results, and therefore the fact that Experiment 2 still had clear effects that replicate Experiment 1 suggest that the reported effects are robust.

Voice recognition relies on some properties that are used for speech perception, such as formant frequencies range in the articulation of different vowels, but also on features that are independent of speech perception, such as voice timbre. Thus, individuals show better speech perception with familiar voices (Holmes et al., 2021; Nygaard et al., 1994), and correspondingly, better voice recognition the better their phonological skills (Kadam et al., 2016; Perrachione et al., 2011). Individuals also exhibit better voice recognition the more linguistic content there is in the speech (e.g., paragraphs vs words vs pseudowords vs reversed speech, Goggin et al., 1991) and can recognise speakers even in sinewave speech, which removes most of the acoustic non-linguistic information (Sheffert et al., 2002). At the same time, individuals are still able to recognise voices when the linguistic information has been removed, such as reversed speech (Sheffert et al., 2002) or in a foreign language which they are not familiar with (e.g., Goggin et al., 1991; Köster & Schiller, 1997; Perrachione et al., 2011). There are also individuals with aphasia with unimpaired voice recognition suggesting some independence between the abilities (Belin et al., 2011). The results of the experiments in this paper suggest that the benefit that having a smaller social network confers is due to greater encoding of linguistic properties since individuals with smaller social networks performed better in their native language (Experiments 1 and 2) but not in a foreign language (Experiment 3) even though the stimuli in Experiment 3 were the same as those used in Experiment 1 and the native language of participants was the same as in Experiment 2.

One caveat, discussed earlier, is that the participants in Experiments 1 and 2 were similar to the speakers in age, education, and often gender. In contrast, the participants in Experiment 3 were potentially more varied in their demographic background, and thus potentially less similar to the speakers. The effect of social network size in Experiments 1 and 2 is hypothesised to be due to the motivation of encoding speaker-specific characteristics. It is possible that when the speakers are less similar to the members of one's social network, this effect is reduced. In addition, the greater heterogeneity of the participants in Experiment 3 might have also obscured any effect of social network size by introducing other variables that might influence voice recognition. The finding that the benefit of having a smaller social network is due to better encoding of linguistic rather than non-linguistic features should therefore be considered as preliminary and further investigated in future research.

This paper is not the first to investigate the effect of social network size on linguistic performance. Past research

has highlighted the positive effect of having a larger social network on linguistic skills. It shows that individuals with larger social networks are better at speech perception (Lev-Ari, 2018), comprehension of evaluative language (Lev-Ari, 2016), and lexical prediction (Lev-Ari, 2019; Lev-Ari & Shao, 2017). In these cases, the size of the social network influenced the distributional properties of the input that individuals received. In particular, individuals with larger social networks received more variable input which led to more robust learning. One might have predicted that receiving more variable input would also help individuals learn better how to identify and use speaker-conditioned variation, and thus be better at voice recognition. In contrast, the studies in this paper show that this is not the case. The experiments in this paper do not test directly why individuals with larger social networks show poorer voice recognition but one hypothesis is that while past experience might have provided individuals with larger social network better input to learn these speaker-conditioned patterns, it also taught them that learning these patterns confers little benefit and is therefore not very useful.

Voice recognition is not the only skill whose utility might depend on social network size. Individuals with larger social networks might be similarly discouraged from learning other speaker-specific patterns at other linguistic levels, such as lexical choices. As with the case of voice recognition, the more people one interacts with, the more speaker-specific patterns one has to learn, and the lower benefit there is to learning each pattern. All prior studies that found a positive effect of having a larger social network focused on understanding novel speakers and forming correct group-level expectations. It might be the case that social network size modulates attention to speaker-specific vs speaker-independent features, and therefore, individuals with smaller social networks might be better at learning speaker-specific patterns. For example, individuals with larger social networks might be better at predicting the speech of the average person or a person of a specific social group (Lev-Ari, 2019; Lev-Ari & Shao, 2017) but individuals with smaller social networks might be better at predicting the speech of specific speakers. This could also influence patterns of alignment in communication. Future research should further examine the way in which social experience modulates what is attended to and encoded during interaction, as well as the consequences for how information is stored, including whether it is stored in a speaker-specific manner, a speaker-independent manner, or discarded. This study provides the first step in showing how rich social experience can *reduce* individuals' likelihood of learning patterns in the input they receive.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

ORCID iD

Shiri Lev-Ari  <https://orcid.org/0000-0002-3784-5184>

Data accessibility statement



The data and materials from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/wtb5f/>

Supplementary material

The supplementary material is available at qjep.sagepub.com.

Notes

1. The speakers were originally recorded for a different experiment 2 years earlier. Therefore, the participants, who were all first year students, did not know them.
2. Number of hours was not included because software errors led to loss of this information from most participants. This factor, however, did not predict performance or interact with any other factor in Experiment 1.
3. The random structure in this case is different from the random structure in Experiment 1, since inclusion of three random variables (Participant, Word, and Speaker) led to singular fit. In contrast, when Speaker and Word were combined into a single random variable, the Audiofile, the model converged. The results of the model with singular fit show the same effects of Social Network Size ($\beta = -0.25$, $SE = 0.08$, $z = -3.02$, $p < .01$) and Native status ($\beta = 0.77$, $SE = 0.19$, $z = 4.03$, $p < .001$). They also show a marginal effect of Word Type ($\beta = -0.14$, $SE = 0.08$, $z = -1.81$, $p < .07$).
4. The correlation between social network size and number of hours per interaction partner was negative and significant in both experiments that include data on number of hours of interaction: $r = -0.42$ ($p < .001$) in Experiment 1, and $r = -0.28$ ($p < .01$) in Experiment 3.

References

- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2010). *lme4: Mixed-effects modeling with R*. <https://cran.r-project.org/web/packages/lme4>
- Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725.
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition*, *106*(2), 682–706.
- Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences, USA*, *111*(38), 13795–13798.
- Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, *124*(1), 95–100.
- Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.
- Holmes, E., To, G., & Johnsrude, I. S. (2021). How long does it take for a voice to become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training. *Psychological Science*, *32*, 903–915.
- Iacozza, S., Meyer, A. S., & Lev-Ari, S. (2020). How in-group bias influences the level of detail of speaker-specific information encoded in novel lexical representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 894–906.
- Johnson, E. K., Bruggeman, L., & Cutler, A. (2018). Abstraction and the (misnamed) language familiarity effect. *Cognitive Science*, *42*(2), 633–645.
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*(5), 1002–1011.
- Kadam, M. A., Orena, A. J., Theodore, R. M., & Polka, L. (2016). Reading ability influences native and non-native voice recognition, even for unimpaired readers. *The Journal of the Acoustical Society of America*, *139*(1), EL6–EL12.
- Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, *4*, 18–28.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101.
- Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The prototype model in speaker identification by human listeners. *International Journal of Speech Technology*, *4*(1), 63–74.
- Lev-Ari, S. (2016). How the size of our social network influences our semantic skills. *Cognitive Science*, *40*, 2050–2064. <https://doi.org/10.1111/cogs.12317>
- Lev-Ari, S. (2018). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology*, *71*(10), 2249–2260. <https://doi.org/10.1177/1747021817739865>
- Lev-Ari, S. (2019). People with larger social networks are better at predicting what someone will say but not how they will say it. *Language, Cognition, and Neuroscience*, *34*(1), 101–114. <https://doi.org/10.1080/23273798.2018.1508733>
- Lev-Ari, S., & Shao, Z. (2017). How social network heterogeneity facilitates lexical access and lexical prediction. *Memory & Cognition*, *45*(3), 528–538. <https://doi.org/10.3758/s13421-016-0675-y>
- Levi, S. (2018). Another bilingual advantage? Perception of talker-voice information. *Bilingualism: Language and Cognition*, *21*(3), 523–536.
- Lieberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. C. (1957). The discrimination of speech sounds within

- and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.
- Maguinness, C., Roswadowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*, 116, 179–193.
- Neuhoff, J. G., Schott, S. A., Kropf, A. J., & Neuhoff, E. M. (2014). Familiarity, expertise, and change detection: Change deafness is worse in your native language. *Perception*, 43(2–3), 219–222.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42–46.
- Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, 333, 595.
- R Core Team. (2020). *A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org>
- Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, 15(6), 608–635.
- Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance*, 28(6), 1447–1469.
- Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters Part I: Recognition of backward voices. *Journal of Phonetics*, 13(1), 19–38.