

The effect of clear speech to foreign-sounding interlocutors on native listeners' perception of intelligibility

Jayanthiny Kangatharan^{a,*}, Maria Uther^b, Fernand Gobet^c

^a Brunel University London, Kingston Lane, London, Uxbridge UB8 3PN, United Kingdom

^b Faculty of Education, Health and Wellbeing, University of Wolverhampton, Walsall Campus, Walsall WS1 3BD, United Kingdom

^c London School of Economics and Political Science, London WC2A 2AE, United Kingdom

ARTICLE INFO

Keywords:

Clear speech
Hyperarticulation
Speech intelligibility
Foreigner-directed speech (FDS)
Speech perception

ABSTRACT

Hyperarticulation is an acoustic modification of the speech stream that has been reliably shown to be naturally part of clear speech. Despite the large number of studies that have investigated the relationship between clear speech properties (e.g. expanded vowel space) and intelligibility, only an associative, rather than a causal relationship between vowel space expansion and speech intelligibility has been documented to this date.

The focus of this study was to more directly explore the impact of vowel hyperarticulation on intelligibility. Specifically, it tested the effect that naturally elicited speech in interaction with native-sounding adults and foreign-sounding adults had on intelligibility. Twenty-one native speakers of English transcribed and provided ratings of typicality and clarity for speech that was produced while speakers were interacting with native-sounding and foreign-sounding adults. Results revealed that speech directed at L2 speakers is more intelligible and rated as being clearer and a more typical representation of speech compared to speech directed at L1 speakers. Implications of the findings are discussed in terms of Lindblom's Hyper- and Hypoarticulation theory (1992).

1. Introduction

Within the vocal tract the peaks of resonances are referred to as formants. They are specified by their center frequencies and bandwidth and correspond with the peaks in a vowel spectrum. Measured in Hertz (Hz), the formant frequencies F1 and F2 specifically are regarded to contribute to the qualitatively distinct sounds between vowels (Raphael et al., 2007). The expansion of vowel space involves the acoustic exaggeration of F1 and F2 (Ladefoged, 1996) and has been shown to be part of clear speech, which includes acoustic-phonetic features such as decreased speech rate and greater sound pressure levels (Chen, 1980; Picheny et al., 1986; Uther et al., 2007). Clear speech occurs when speakers notice a difficulty in the listeners' speech perception because of a hearing-impairment, background noise or different language background (Smiljanić and Bradlow, 2005). Although previous research has indicated clear speech to be more intelligible than conversational speech for normal-hearing listeners in noisy environments, it is not known which acoustic features of clear speech are the most responsible for enhanced intelligibility. Previous research has shown that enlarged vowel space correlates with speech intelligibility (Bond and Moore,

1994; Liu et al., 2003; Monsen, 1976). It was also found that there is a positive correlation between intelligibility and speakers' ability to naturally articulate phonetic contrasts accurately on a segmental level and employ a larger vowel space (Bradlow et al., 1996; Byrd, 1994; Hazan and Markham, 2004).

To date, despite the large number of studies that have investigated the relationship between vowel space and intelligibility, only an associative relationship between vowel space expansion and speech intelligibility has been explored. In particular, no study has directly tested whether speech with vowel space expansion that is produced under naturalistic conditions in a controlled setting (as opposed to acting in imagined scenarios) is more intelligible than speech samples with non-expanded vowel space. Moreover, because correlation does not necessarily indicate causation, conclusions about the role of hyperarticulation on intelligibility in the context of speech directed to native and foreign speakers are tentative. To move beyond correlational data, the present study investigates the extent to which vowel hyperarticulation affects clarity by focusing on the evaluation of the intelligibility of clear speech produced in speech to foreign-sounding and native-sounding interlocutors.

* Corresponding author.

E-mail address: jayanthinykangatharan@gmail.com (J. Kangatharan).

Previous clear speech research mainly focused on studies with artificially modified stimuli or stimuli that were produced under imaginary speaking conditions rather than using naturalistic samples (e.g. Bradlow et al., 2003; Gagné et al., 1994; Helfer, 1998; Krause and Braida, 2002; Picheny et al., 1985). For example, one clear speech study instructed its speakers to produce sentences as if talking to an individual with hearing loss (Ferguson and Kewley-Port, 2007). Only one recent study recorded speech naturally when participants were engaging in an interactive task (Pettinato et al., 2016). However, the study focused on children to find out how age and gender affect vowel space in conversational speech with communicative intent. It therefore did not focus on the impact of naturally elicited clear speech on speech intelligibility at word level. Thus, the purpose of the present study was to ascertain whether speech with hyperarticulation, as elicited in natural speech, leads adult listeners to perceive enhanced intelligibility at word level.

The present study used three different measures of intelligibility with naturally elicited speech samples collected from a previous speech production experiment in which speech directed to foreign-sounding interlocutors and speech directed to native-sounding interlocutors were recorded (Kangatharan et al., 2021). These measures of intelligibility consist of an orthographic transcription task (together with a confidence rating), a typicality rating task and a clarity rating task.

An orthographic transcription task was used to find out if adult native listeners were able to correctly transcribe the naturally elicited acoustic speech stimuli from speech directed either to foreign-sounding interlocutors or native-sounding interlocutors. A confidence rating was included to find out how confident the native listeners were in how accurate their transcription was. A typicality rating was collected to explore how hyperarticulation relates to the listeners' perception of how typical the acoustic-phonetic exaggeration of speech sounds. A clarity rating was used to find out whether speech directed to foreign-sounding interlocutors and native-sounding interlocutors differs in clarity.

Thus, the aim of the current study was to determine whether hyperarticulated speech as produced in speech to foreign-sounding interlocutors improves intelligibility for listeners at the word level as compared to native-sounding interlocutors. This would help evaluate whether hyperarticulated speech is beneficial to native listeners and therefore contributes to an enhanced intelligibility of speech in English. Based on previous research (Bradlow and Bent, 2002; Ferguson and Kewley-Port, 2007; Smiljanić and Bradlow, 2011), it was hypothesized that adult native listeners of English will correctly identify more words from native speakers' speech when directed toward foreign-sounding interlocutors than when directed toward native-sounding interlocutors. This is because speech to foreign-sounding interlocutors might be spoken more clearly than speech to native-sounding interlocutors due to the perceived linguistic needs of foreign-sounding interlocutors (Burnham et al., 2002; Munro and Derwing, 1999; Uther et al., 2007; Uther et al., 2012). Consequently, native listeners will feel more confident in their accuracy of transcribing native speakers' speech when it is directed toward foreign-sounding interlocutors than when it is directed toward native-sounding interlocutors. It was also hypothesized that native listeners will perceive speech directed to foreign-sounding interlocutors as clearer, but not more typical than speech that is directed to native-sounding interlocutors. Moreover, hyperarticulated speech could instead be expected to be perceived as *less* typical as generally people do not hyperarticulate in everyday language unless accommodating for an interlocutor perceived to have a specific need. This would be in line with the Hyper- and Hypoarticulation (H&H) theory, according to which adults modify their speech to maximize discriminability to provide the listener with sufficient information to make speech clearer (Lindblom, 1992). For example, clear speech was used in speech to hearing-impaired listeners in quiet and not casual speech (Picheny et al., 1985) and to normal-hearing listeners in noise (Uchanski et al., 1996). The hypothesis on clarity can also be argued to theoretically agree with the notion of audience design that considers speakers to change speech for different listeners and to design their

speech style accordingly (Bell, 2001). Similarly, there is a view that 'linguistic convergence' in which speakers take on the interlocutors' way of speaking occurs as a way of decreasing social distance and raising communicative effectiveness. This view is consistent with the communication accommodation theory (CAT) (Giles, 2009). This theory suggests that hyperarticulation in native speakers' speech to foreign-sounding interlocutors can be seen as native speakers' attempt to arrive at linguistic convergence with the foreign-sounding interlocutor.

2. Method

This study had two parts: a speech production experiment to elicit spontaneous speech produced when doing a 'Spot the Difference' task with different types of interlocutors as described previously in Kangatharan et al. (2021), and a perceptual evaluation experiment using target words extracted from the speech production experiment. Both the speech production experiment and the perceptual evaluation experiment were approved by the Ethics committee at the Psychology Department at Brunel University. All participants filled in a consent form prior to participating and were debriefed following participation.

2.1. Speech production experiment

This speech production experiment was conducted in a previous study (Kangatharan et al., 2021), and details are reported again here in the present study to describe how the speech tokens used for the speech perception experiment were obtained. The speech production experiment used a 2 (interlocutor's accent: native, foreign) × 2 (interlocutor's physical appearance: native, foreign) × 3 (three target vowels: /a:/, /u:/ and /i:/) mixed design with four different types of interlocutors: NLNS (native-looking and native-sounding), NLFS (native-looking and foreign-sounding), FLNS (foreign-looking and native-sounding), and FLFS (foreign-looking and foreign-sounding). The extent of hyperarticulation in the target words in which one of the three target vowels was present was the dependent variable (see Kangatharan et al., 2021). In the speech production study, fifty-two female White British speakers were instructed to interact with an interlocutor from one of the four different interlocutor types.

After a White British English speaker and an interlocutor had been seated opposite each other, they were each given a folder with three pictures, each showing a different scene. There were 13 differences between the picture the speaker received and the picture their partner interlocutor received. The task was for the speaker and the interlocutor to talk with each other to verbally detect the differences between their pictures during an half an hour audio-recorded interaction. In the beginning, 150 target words from one of thirteen target vowels were recorded from native speakers in the speech production study. The vowels /a:/, /i:/, /u:/, /e/ and /ɒ/ were chosen from the target words "car", "beach", "blue", "pink", "red" and "shop" as they contained a minimum of one sample per participant.

Statistical analyses of the speech samples with the target vowels /a:/, /i:/, /u:/ revealed a significant effect of speech in comparison to appearance among the four different interlocutor types (see Kangatharan et al., 2021, for details). Based on this finding, the speech samples from the foreign-sounding interlocutor conditions NLFS and FLFS were combined and the speech samples from the native-sounding interlocutor conditions NLNS and FLNS were combined for the Listening Experiment to find out whether hyperarticulated speech (as produced in speech to foreign-sounding interlocutors) improves clarity for listeners at the word level as compared to native-sounding interlocutors.

2.2. Listening experiment

2.2.1. Participants

The listeners were 21 native speakers of English (16 female, 5 male;

mean age 19) from the Southeast London area, who declared to have regular hearing and orthographic abilities. They were undergraduate students who were recruited from university. They received course credits for their participation.

2.2.2. Materials and apparatus

After the recordings were made in the speech production experiment, target words were extracted from the sound files and were equated for root-mean-square amplitude. These word-length target files formed the samples that were rated by the native English listeners in the listening experiment. The stimuli were presented on a laptop with the e-prime software installed (Schneider et al., 2002a; Schneider et al., 2002b) via headphones (Sennheiser HD429) at a comfortable listening volume. All participants' responses, which were submitted using the computer keyboard, were recorded via the e-run software application.

2.2.3. Design

The speech samples from the foreign-sounding interlocutor conditions NLFS and FLFS in the speech production experiment were combined to jointly present the foreign-sounding interlocutor condition in the current Listening Experiment. The speech samples from the native-sounding interlocutor conditions NLNS and FLNS in the speech production experiment were combined to present the native-sounding interlocutor condition in the current Listening Experiment. Ten speech stimuli were randomly chosen from each interlocutor condition (native-sounding and foreign-sounding conditions) for each of the three vowels /a:/, /i:/ and /u:/ as target vowels and for each of the three vowels /i/, /e/ and /ɒ/ as distractors. This means that, overall, 120 stimuli were presented in random order, with each stimulus being rated and transcribed by each native listener.

2.2.4. Procedure

The participants completed three tasks: (a) a transcription task, together with confidence rating, (b) a typicality rating task; and (c) a clarity rating task. During each task, the participants listened to 120 audio stimuli. The measures were obtained by participants either typing a word or pressing one of six buttons with the numbers 1, 2, 3, 4, 5 and 6 on a Likert scale format (from 1 = very confident/typical/clear to 6 = not confident/typical/clear at all). Each task was presented through a separate e-prime program. Stimuli were presented in a random order. Altogether, the three sessions lasted approximately 35 min and were completed on the same day.

The first session included a transcription task, together with providing a confidence rating. Participants carefully listened to every word stimulus and typed out what they heard using the keyboard within the space that was displayed on the screen. The orthographic transcription task was considered a measurement of intelligibility of speakers' speech (Giolas and Epstein, 1963; Tikofsky and Tikofsky, 1964; Yorkston and Beukelman, 1981; Garcia and Cannito, 1996; Hustad, 2008). Participants then indicated on a Likert scale from 0 to 6 how confident they were in how accurate their transcription was (1 = very confident; 6 = not very confident at all). The presentation of a novel word stimulus occurred 500 milliseconds following listeners' rating of their confidence in their transcription. An arrow that appeared for 200 milliseconds indicated the presentation of the next word.

While completing the typicality rating task, participants viewed the words written on the screen for 300 milliseconds, and after another 300 milliseconds, listened to speech corresponding to the written words. They then rated the words for typicality. A new word stimulus was presented 500 milliseconds after listeners had finished rating the typicality (1 = typical; 6 = not typical at all) of the previous word. The presentation of the next word was indicated by an arrow that appeared for 200 milliseconds.

While completing the clarity task, participants viewed the word written on the screen for 300 milliseconds and, after another 300 milliseconds, listened to speech corresponding to the word that was

presented visually. They then rated the words for clarity on a scale from 1 to 6 (1 = very clear; 6 = not clear at all). A new word stimulus was presented 500 milliseconds after listeners had finished rating the typicality of the previous word. The presentation of the next word was indicated by an arrow that appeared for 200 milliseconds.

2.2.5. Data analysis

Based on the results of the speech production study that showed a significant effect of speech and not of appearance (Kangatharan et al., 2021), the experimental conditions NLNS and FLNS were combined to present the native-sounding interlocutor condition and the experimental conditions NLFS and FLFS the foreign-sounding interlocutor condition in the present Listening Experiment.

Based on previous research (e.g. Bradlow and Bent, 2002; Bradlow and Alexander, 2007; Smiljanić and Bradlow, 2011; Munro and Derwing, 1999; Lane, 1963), the scoring of the transcription task accepted only fully transcribed words that were identified correctly and did not accept any word candidates that were recognized half-way. This was done so as not to be unclear about whether near-hits resulted because of typing errors or because listeners actually did not recognize the presented word.

The transcription, confidence, typicality and clarity data were analysed with paired-samples t-tests. Four different paired samples t-tests were used because the four measures were not related. The Shapiro-Wilk normality test showed that the data did not violate the assumption of normality (transcription: $W = 0.944$, $p = .24$; confidence: $W = 0.954$, $p = .373$; typicality: $W = 0.983$, $p = .954$; clarity: $W = 0.964$, $p = .581$).

3. Results

3.1. Transcription accuracy and confidence

The mean confidence rating was not significantly different between the foreign-speaker directed speech and the native-speaker directed speech. The mean transcription accuracy as shown in Fig. 1 was higher for speech samples directed to foreign-sounding interlocutors than native-sounding interlocutors. This difference was statistically significant ($t(20) = 2.426$, $p < .05$, $r = 0.520$). This result is in line with the hypothesis that listeners would show higher transcription accuracy for speech directed to foreign than native-sounding interlocutors.

3.2. Typicality

As shown in Fig. 2, native speakers rated speech samples directed to foreign sounding interlocutors as more typical of representing the word in the English language than the speech samples directed to native sounding interlocutors. This difference was statistically significant ($t(20) = -5.036$, $p < .05$, $r = 0.722$). This is not in line with the hypothesis that there would be no differences in native listeners' rating of words to foreign-sounding interlocutors and native-sounding interlocutors.

3.3. Clarity

As shown in Fig. 3, on a scale from 1 (very clear) to 6 (not clear at all), native speakers rated speech samples directed to foreign-sounding interlocutors as clearer than the speech samples directed to native-sounding interlocutors. This difference was statistically significant ($t(20) = -6.066$, $p < .05$, $r = 0.764$). This finding supports the hypothesis that native listeners would rate words to foreign-sounding interlocutors as clearer than the words to native-sounding interlocutors.

A simple linear regression was conducted on the clarity rating to understand whether changes observed in the level of clarity were associated with changes in vowel space. Used as predictors, F1 and F2 accounted for 19.2% of the variance in the clarity rating. The F-ratio was 2.963, which was close to significance ($p = .070$). The regression coefficients were -0.265 for F1 and 0.241 for F2.

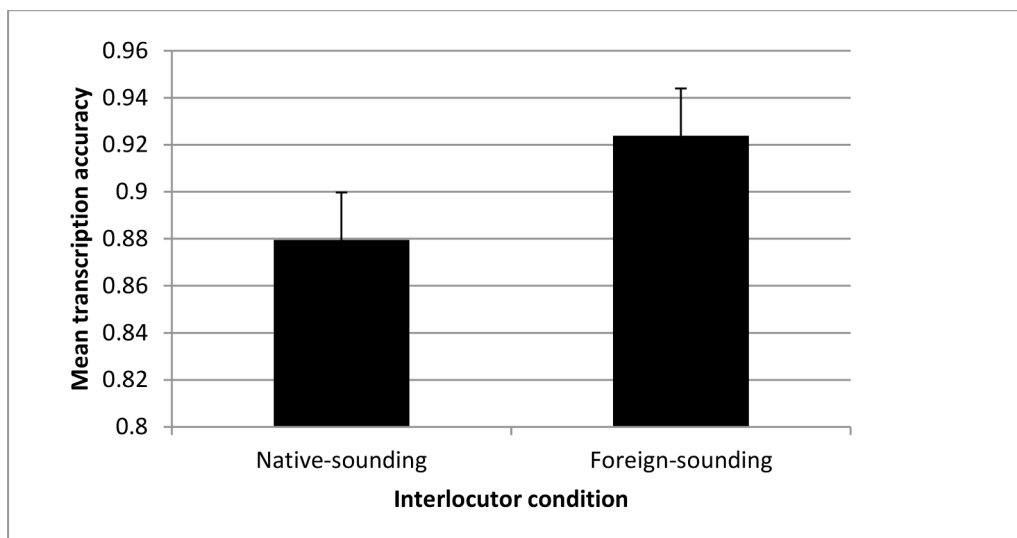


Fig. 1. Mean transcription accuracy for the native-sounding and the foreign-sounding interlocutor conditions. Error bars show ± 1 standard error from the mean.

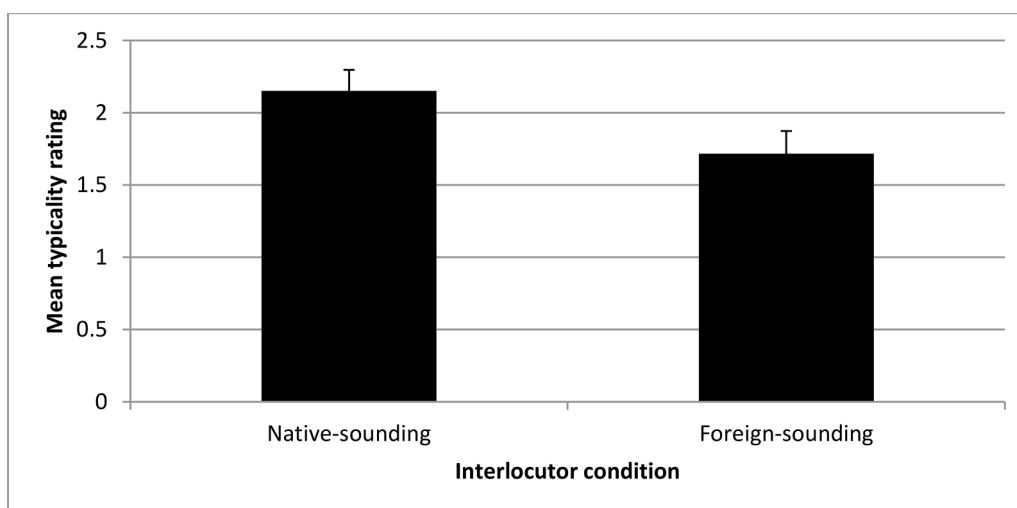


Fig. 2. Mean typicality rating for the native-sounding and the foreign-sounding interlocutor conditions on a scale from 1 (very typical) to 6 (not typical at all). Error bars show ± 1 standard error from the mean.

4. Discussion

The goal of the study was to determine whether hyperarticulated speech as produced in speech to foreign-sounding interlocutors improves intelligibility for listeners at the word level as compared to native-sounding interlocutors. Therefore, the speech samples that had been produced by native English speakers in speech to foreign-sounding and native-sounding interlocutors in a previous speech production experiment (Kangatharan et al., 2021) were scored at word level from transcriptions provided by twenty-one native speakers of English for intelligibility.

A difference in native English listeners' transcription accuracy of speech to foreign-sounding interlocutors and native-sounding interlocutors was observed. This result appears to support the hypothesis that native English listeners would transcribe more accurately words that had been articulated in speech to foreign-sounding interlocutors than to native-sounding interlocutors. It is in line with the finding of a difference in native English listeners' clarity rating of speech directed to foreign-sounding interlocutors and native-sounding interlocutors. It also appears to support the hypothesis that native English listeners would rate the speech directed to foreign-sounding interlocutors as clearer than

speech directed to native-sounding interlocutors. This finding suggests that the acoustically expanded vowel space as observed in hyperarticulated speech directed to foreign-sounding interlocutors might have contributed to a higher transcription accuracy by native English listeners due to enhanced speech clarity. This also appears to be in line with previous research (e.g. Burnham et al., 2002; Uther et al., 2007) and extends prior evidence of vowel space expansion contributing to vowel intelligibility (e.g. Ferguson and Kewley-Port, 2007). However, vowel space was not observed to be a good predictor of intelligibility. This result is not in line with previous studies where vowel space expansion was found to correlate with speech intelligibility (e.g. Bond and Moore, 1994; Bradlow et al., 1996; Hazan and Markham, 2004; Picheny et al., 1986) and does not support the H&H theory according to which hyperarticulated speech allows more accurate phonetic units to be more easily perceived as being acoustically distinct (Lindblom, 1990). This finding could be explained by there not being one 'go-to' acoustic-phonetic characteristic that facilitates intelligible processing of speech at word level. It would need to be explored how this characteristic interacts with other acoustic-phonetic features to allow for better intelligibility in clear speech at word level.

Although the appropriate spoken word identification and therefore

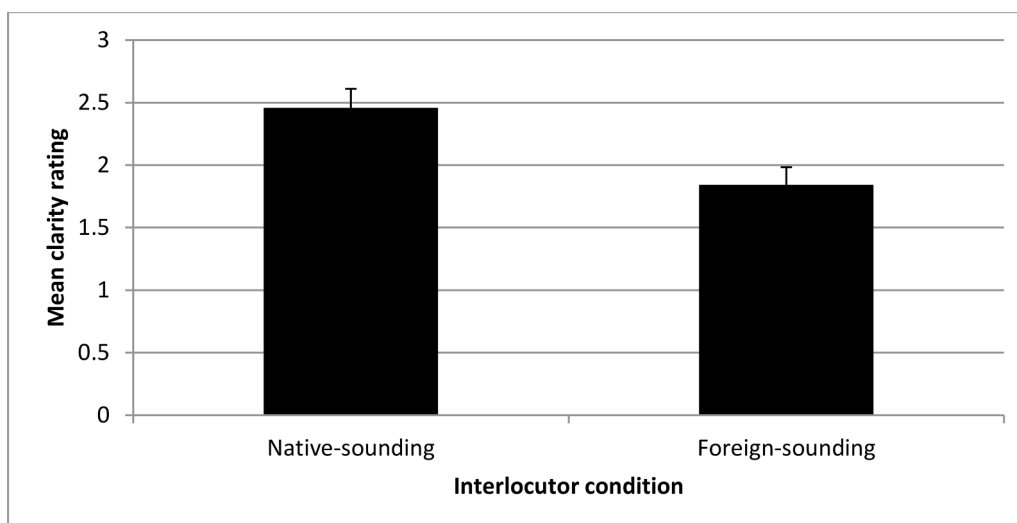


Fig. 3. Mean clarity rating for native-sounding and foreign-sounding interlocutor conditions on a scale from 1 (very clear) to 6 (not clear at all). Error bars show ± 1 standard error from the mean.

speech intelligibility had previously been found to be affected by how accurate vowels are articulated (Liu et al., 2005; Weismer et al., 2001), other acoustic-phonetic features have also been highlighted in clear speech perception such as increased vowel length and word length (Bond and Moore, 1994; Scarborough et al., 2007). For example, increased word length was observed to be part of clear speech (Bond and Moore, 1994; Piazza et al., 2021) and increased vowel length was reported in speech to foreigners (Ashby, 2004; Sankowska et al., 2011). It could therefore be stated that it is a combination of vowel space together with other acoustic-phonetic features that might contribute to clear speech. This can be considered to be very likely for the vowels /a:/, /i:/ and /u:/ that were used as target vowels in the present study because, in Received Pronunciation – that is the standard accent of the English language in the UK – these tense vowels are longer in duration than short/lax vowels and are also elongated in clear speech (Uchanski, 1988, 1992). Future speech perception research could therefore focus on distinguishing between the contributions made by the different acoustic-phonetic characteristics such as vowel space, word length and vowel length. For example, native listeners could rate naturally elicited speech samples for intelligibility in three experimental conditions: The first condition would contain only expanded vowel space and neither elongated vowel length nor increased word length; the second experimental condition would include only elongated vowel length; and the third experimental condition would include only elongated word length. A baseline condition could also be included where the same stimuli are rated, but do not include any of those three acoustic-phonetic characteristics.

Two limitations could have contributed to the finding that vowel space was not a good predictor of intelligibility: a small sample size and the nature of speech materials used. Specifically, the speech samples were produced in a spontaneous speech environment, meaning that over the duration of the half an hour interaction between the White British English speakers and interlocutors, the words might have been articulated with less extreme vowels. This means that there was no experimentally instructed control over the quality of the speech produced by the speakers in the present study in contrast to previous studies where speakers articulated speech in a shorter amount of time and in the absence of an interlocutor and were instructed to produce speech as if speaking to a hearing-impaired listener (e.g. Ferguson and Kewley-Port, 2002) or to read out in clear speech (e.g. Bradlow and Bent, 2002; Ferguson, 2004; Smiljanić and Bradlow, 2005).

In addition, previous research has indicated that speech stimuli produced more than once in a natural conversation might be of lower

acoustic emphasis than novel referents (Prince, 1981; Watson et al., 2010). This could have possibly led to the generation of words with less extreme vowels because the speech stimuli that were used in the orthographic transcription task contained forms of speech stimuli that were mentioned a second and more times (Pettinato et al., 2016). This in turn can be considered to be reflected in the absence of a differential perception of speech directed to native-sounding and foreign-sounding interlocutors in participants' confidence ratings. This could therefore also explain why English native raters were not more confident in their transcription of speech directed to foreign-accented individuals than their transcription of speech directed to native-accented individuals. Moreover, producing the speech stimuli more than once over a period of a 30-minute interaction could have led to semantic satiation and speaker fatigue that could have shaped the acoustic generation of speech stimuli (Jacobovitz, 1965).

This weakness in the experimental design could be addressed by reducing the time taken to complete the interaction task. For example, speakers and interlocutors could be instructed to talk for only a brief period of time, such as five minutes, to find differences between their pictures about one scenario. When completing the task for a second picture for another five minutes, a different scenario could be used that makes use of some of the same words as the first picture to facilitate the elicitation of the same target words in a refreshed way. This approach would allow speakers to articulate these words in a revitalised manner. The completion of the task with a third picture on a different scenario would follow for another five minutes and overall this would encourage speakers to articulate these words with a stable level of alertness. It would also be useful for future research to introduce a short break in between each problem-solving task.

One could argue that if clear speech had been produced in more experimentally controlled settings, this could have yielded a significant correlation between vowel space and intelligibility. Such a finding would have been explained by the H&H theory and to some extent via the communication accommodation theory (CAT), because hyper-articulation in native speakers' speech to foreign-sounding interlocutors can be seen as an effort by native speakers to reach linguistic convergence with the foreign-sounding interlocutor (Giles, 2009). According to CAT, convergence involves adopting the interlocutors' way of speaking and this can be elicited by the speakers' desire to decrease social distance and raise communicative effectiveness (Giles, 1973). Native speakers' aim for communicative efficacy can lead to changes in speech that are characteristic of 'foreigner talk' (FT; Ferguson, 1971) such as the use of high-frequency words as well as changes in pronunciation that

according to CAT are considered to consequently occur as modifications in speech (Giles and Coupland, 1991; Zuengler, 1991). Such a finding could also be consistent with the idea of audience design, according to which the speech style as designed by speakers can change for a variety of listeners (Bell, 2001). However, accommodation theory focuses mainly on elements of speech that are paralinguistic in nature, such as length of utterance or speech rate, and not segmental aspects such as formant frequencies (Bell, 2006), and audience design places emphasis solely on the practical nature of language style in the creation of identity (Bell, 2006). The H&H theory therefore is a better-suited theory to consider the link between vowel space and intelligibility.

Contrary to expectations, there was a difference in native English listeners' typicality rating in how typical the speech directed to native-sounding interlocutors and foreign-sounding interlocutors sounded. It could be argued that this finding is in line with Lindblom's (1992) H&H theory, according to which speakers adjust their speech using information about their interlocutors and specifically, use hyperspeech by articulating speech as accurately as possible to enhance communicative efficiency (Lindblom, 1990). Speech production can, in this regard, be considered a listener-oriented modification with the goal to ensure sufficient contrast in the speech output. Because speech can vary on a continuum from hypospeech to hyperspeech, the speech directed at foreign-sounding interlocutors is likely to be hyperarticulated in contrast to speech directed to native-sounding interlocutors, which in turn would be more 'typical' of native English production. By the same token, speech directed to native-sounding interlocutors can be regarded representative of hypospeech that is an economical mode of speaking, in which speech is only differentiated to a certain extent so that speakers do not overly economize their speech and become unintelligible. Hypo-speech can therefore be seen in contrast to hyperspeech as more linguistically divergent from typical English production patterns (Lindblom, 1990). The data on typicality rating could therefore be considered in line with Lindblom's (1992) H&H theory.

A notable limitation is that the present study made use of simple target words. Future research can therefore use the hyperarticulated samples as embedded within a sentence context, which might make it easier for native English listeners to notice a change in the articulation of speech because hyperarticulation within a sentence might be more discernable to a listener than when presented in a simple word format. Given that speech communication naturally occurs within sentence and broader communicative contexts, it would be more ecologically valid and interesting to consider longer speech samples in future studies. Another limitation is that the study also presented the stimuli three times for each task, which involves a potential confound of repeated exposures. Future research could therefore conduct a speech intelligibility study in which the stimuli are presented once and participants are asked to transcribe and rate each stimulus for typicality and clarity after listening to each stimulus once.

5. Conclusion

This study addressed the research question of the effect of vowel hyperarticulation produced when interacting with foreign-sounding adults on intelligibility compared to speech directed to native-sounding adults. Results from this study indicated that native listeners transcribed words from speech directed at L2 speakers more accurately than words from speech directed at L1 speakers, although they were not significantly more confident in their transcriptions. They also rated words from speech directed at L2 speakers significantly clearer and more typical than words from speech directed at L1 speakers. The findings suggest that vowel hyperarticulation as elicited together with other acoustic-phonetic features in foreigner-directed speech could prove beneficial to improving at word level the intelligibility of clear speech that is produced in communication with foreign-sounding interlocutors. To provide clear evidence in support of Lindblom's Hyper-and Hypo-articulation (H&H) theory and to understand to what extent vowel

hyperarticulation could be used for didactic purposes, future research is needed that can also address some limitations existent in the present study, including the simple nature of the speech material and the limited statistical power, that is low statistical sensitivity caused by small sample size.

Compliance with ethical standards

Funding

The studies were funded by an Isambard Scholarship from Brunel University.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the studies.

CRedit authorship contribution statement

Jayanthiny Kangatharan: Methodology, Software, Formal analysis, Investigation, Resources, Project administration, Writing – original draft. **Maria Uther:** Conceptualization, Supervision, Writing – review & editing. **Fernand Gobet:** Validation, Supervision, Writing – review & editing.

Declaration of Competing Interest

Author Jayanthiny Kangatharan declares that she has no conflict of interest. Author Maria Uther declares that she has no conflict of interest. Author Fernand Gobet declares that he has no conflict of interest.

Data availability

Data will be made available on request.

References

- Ashby, S. (2004). Is infant-directed speech hyperspeech? An acoustical analysis of speech to infants and other accommodative speech styles. PhD Thesis, University College Dublin, Ireland.
- Bell, A., 2001. Back in style: reworking audience design. P. Eckert, & J. Rickford Style and Sociolinguistics Variation. Cambridge University Press, Cambridge.
- Bell, A., 2006. Speech accommodation theory and audience design. K. Brown Encyclopaedia of Language & Linguistics. Cambridge University Press, New York.
- Bond, Z.S., Moore, T.J., 1994. A note on the acoustic-phonetic characteristics of inadvertently clear speech. Speech Commun. 14, 325–337.
- Bradlow, A.R., Alexander, J., 2007. Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Am. 112, 2339–2349.
- Bradlow, A.R., Bent, T., 2002. The clear speech effect for non-native listeners. J. Acoust. Soc. Am. 112, 272–284.
- Bradlow, A.R., Torretta, G.M., Pisoni, D.B., 1996. Intelligibility of normal speech. I. global and fine-grained acoustic-phonetic talker characteristics. Speech Commun. 20, 255–272.
- Bradlow, A.R., Kraus, N., Hayes, E., 2003. Speaking clearly for children with learning disabilities: sentence perception in noise. J. Speech Lang. Hear. Res. 46, 80–97.
- Burnham, D., Kitamura, C., Vollmer-Conna, U., 2002. What's new, pussycat? On talking to babies and animals. Science 296, 1435.
- Byrd, D. (1994). Relations of sex and dialect to reduction speech communication, 15, 39–54.
- Chen, F. (1980). Acoustic characteristics and intelligibility of clear and conversational speech at the segmental level (Unpublished Master's thesis). Massachusetts Institute of Technology, Cambridge.

- Ferguson, C., 1971. Absence of copula and the notion of simplicity: a study of normal speech, baby talk, foreigner talk and pidgins. D. Hymes Pidginization and Creolization of Languages. Cambridge University Press, Cambridge, pp. 141–150.
- Ferguson, S.H., Kewley-Port, D., 2002. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 112, 259–271.
- Ferguson, S.H., 2004. Talker differences in clear and conversational speech: vowel intelligibility for normal-hearing listeners. *J. Acoust. Soc. Am.* 116, 2365–2373.
- Ferguson, S.H., Kewley-Port, D., 2007. Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J. Speech Lang. Hear. Res.* 50, 1241–1255.
- Gagné, J.P., Masterson, V., Munhall, K.G., Bilida, N., Queregensser, C., 1994. Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech. *J. Acad. Rehabil. Audiol.* 27, 135–158.
- Garcia, J., Cannito, M., 1996. Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *J. Speech. Hear. Res.* 39 (4), 750–760.
- Giles, H., 1973. Accent mobility: a model and some data. *Anthropol. Linguist.* 15, 87–109.
- Giles, H., 2009. Communication accommodation theory. Reis, H. & Sprecher, S. *Encyclopaedia of Human Relations*. SAGE Knowledge.
- Giles, H., Coupland, J., Coupland, N., 1991. Accommodation theory: communication, context, and consequence. H. Giles, J. Coupland, & N. Coupland *Contexts of Accommodation*. Cambridge University Press, New York.
- Giolas, T., Epstein, A., 1963. Comparative intelligibility of word lists and continuous discourse. *J. Speech. Hear. Res.* 6, 349–358.
- Hazan, V., Markham, D., 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. *J. Am. Acad. Audiol.* 116, 3108–3118.
- Helfer, K.S., 1998. Auditory and auditory-visual recognition of clear and conversational speech by older adults. *J. Am. Acad. Audiol.* 9, 234–242.
- Hustad, K.C., 2008. The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *J. Speech Lang. Hear. Res.* 51, 562–573.
- Jacobovitz, L., 1965. Repetition of auditorily presented information. *Psychol. Rep.* 17, 113–114.
- Kangatharan, J., Uther, M., Gobet, F., 2021. The effect of hyperarticulation on speech comprehension under adverse listening conditions. *Psychol. Res.*
- Krause, J.C., Braid, L.D., 2002. Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility. *J. Acoust. Soc. Am.* 112, 2165–2172.
- Ladefoged, P., 1996. *Elements of Acoustic Phonetics*. University of Chicago Press, Chicago.
- Lane, H., 1963. Foreign accent and speech distortion. *J. Acoust. Soc. Am.* 35, 451–453.
- Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H&H theory. W. J. Hardcastle, & A. Marchal *Speech Production and Speech Modeling*. Kluwer Academic, Amsterdam, The Netherlands, pp. 403–439.
- Lindblom, B., 1992. Phonological units as adaptive emergents of lexical development. C. A. Ferguson, L. Menn, & C. Stoel Gammon *Phonological development: Models, research, Implications*. York, Timonium, MD, pp. 131–163.
- Liu, H.M., Kuhl, P.K., Tsao, F.M., 2003. An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* 6, F1–F10.
- Liu, H.M., Tsao, F.M., Kuhl, P.K., 2005. The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Acoust. Soc. Am.* 117, 3879–3889.
- Monsen, R.B., 1976. Normal and reduced phonological space: the productions of English vowels by deaf adolescents. *J. Phon.* 4, 189–198.
- Munro, M.J., Derwing, T.M., 1999. Foreign accent comprehensibility and intelligibility in the speech of second language learners. *Lang. Learn.* 49, 285–310.
- Pettinato, M., Tuomainen, O., Granlund, S., Hazan, V., 2016. Vowel space area in later childhood and adolescence: effects of age, sex and ease of communication. *J. Phon.* 54, 1–14.
- Piazza, G., Martin, C. & Kalashnikova, M. (2021). The acoustic features and didactic function of Foreigner Directed Speech: a literature review. Retrieved from <https://psyarxiv.com/scnke/download/?format=pdf>.
- Picheny, M.A., Durlach, N.I., Braid, L.D., 1985. Speaking clearly for the hard of hearing. I: intelligibility differences between clear and conversational speech. *J. Speech. Hear. Res.* 28, 96–103.
- Picheny, M.A., Durlach, N.I., Braid, L.D., 1986. Speaking clearly for the hard of hearing II: acoustic characteristics of clear and conversational speech. *J. Speech Hear. Res.* 29, 434–446.
- Prince, E., 1981. *Toward a taxonomy of given-new information*. P. Cole *Radical Pragmatics*. Academic Press, New York, NY, pp. 223–255.
- Raphael, L.J., Borden, G.J., Harris, K.S., 2007. *Speech Science Primer*. Lippincott, Williams & Wilkins, Baltimore, MD.
- Sankowska, J., Garcia Lecumberri, M.L., Cooke, M., 2011. Interaction of intrinsic vowel and consonant durational correlates with foreigner directed speech. *Poznan Stud. Contemp. Linguist.* 47, 109–119.
- Schneider, W., Eschman, A., Zuccolotto, A., 2002a. *E-Prime User's Guide*. Psychology Software Tools Inc, Pittsburgh.
- Schneider, W., Eschman, A., Zuccolotto, A., 2002b. *E-Prime Reference Guide*. Psychology Software Tools Inc, Pittsburgh.
- Smiljanić, R., Bradlow, A., 2005. Production and perception of clear speech in Croatian and English. *J. Acoust. Soc. Am.* 118, 1677–1688.
- Smiljanić, R., Bradlow, A., 2011. Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: intelligibility and accentedness. *J. Acoust. Soc. Am.* 130, 4020–4032.
- Tikofsky, R.S., Tikofsky, R.P., 1964. Intelligibility as a measure of dysarthric speech. *J. Speech. Hear. Res.* 7, 25–333. <https://doi.org/10.1044/jsr.0704.325>.
- Uchanski, R.M. (1988). *Spectral and temporal contributions to speech clarity for hearing impaired listener*. PhD Thesis, Massachusetts Institute of Technology, USA.
- Uchanski, R.M., Millier, K.M., Reed, C.M., Braid, L.D., 1992. Effects of Token Variability on Resolution for Vowel Sounds. *The Auditory Processing of Speech: From Sounds to Words*. Mouton de Gruyter, Berlin, pp. 291–302. M. E. H. Schouten.
- Uchanski, R.M., Choi, S., Braid, L., Durlach, N., 1996. Speaking clearly for the hard of hearing IV: further studies of the role of speaking rate. *J. Speech. Hear. Res.* 39, 494–509.
- Uther, M., Knoll, M.A., Burnham, D., 2007. Do you speak E-NG-I-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Commun.* 49, 2–7.
- Watson, D., Arnold, J.E., Tanenhaus, M.K., 2010. Tic Tac TOE: effects of predictability and importance on acoustic prominence in language production. *Cogn.* 114 (3), 462–463.
- Weismer, G., Laures, J.S., Jeng, J., Kent, R.D., Kent, J.F., 2001. The effect of speaking rate manipulations on speech timing, formant frequencies, and perceptual judgments in dysarthric speakers with amyotrophic lateral sclerosis. *Folia Phoniatr. Logop.* 52, 201–219.
- Yorkston, K., Beukelman, D., 1981. *Assessment of Intelligibility of Dysarthric Speech*. C. C. Publications, Tigard, OR.
- Zuengler, J., 1991. Accommodating in native-nonnative interactions: going beyond the 'What' to the 'Why' in second-language research. H. Giles, J. Coupland, & N. Coupland *Contexts of Accommodation*. Cambridge University Press, New York.