Haslam, M., & Zetterholm, E. (2019). The role of consonant clusters in English as a lingua franca intelligibility. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, ISSN 2380-9566, Ames, IA, September 2018 (pp. 276-284). Ames, IA: Iowa State University.

PRESENTATION/POSTER

THE ROLE OF CONSONANT CLUSTERS IN ENGLISH AS A LINGUA FRANCA INTELLIGIBILITY

Mara Haslam, Stockholm University
Elisabeth Zetterholm, Stockholm University

This study attempts to add to the limited body of research on what aspects of English pronunciation affect intelligibility for non-native listeners (users of English as a lingua franca). It addresses the claim from Jenkins' (2000; 2002) Lingua Franca Core that intelligibility will suffer if consonants are deleted from initial consonant clusters or from final clusters in ways that do not fit English phonology, and that addition of extra sounds to a cluster by epenthesis will not adversely affect intelligibility. Monosyllabic words with initial or final consonant clusters produced by talkers of different language backgrounds were played for 11 Swedish listeners, who transcribed what they heard in standard English orthography. Responses were then matched against results of an acoustic analysis. Listeners were more successful overall at identifying the intended word structure than acoustic results would indicate, and this pattern holds for stimuli with final clusters but not those with initial clusters. Deletion of one of the consonants from the cluster was shown to be the most common reason for mismatch. These results partially support the Lingua Franca Core but also demonstrate that the location of epenthesis of a vowel in relation to a consonant cluster affects the likelihood of match or mismatch.

INTRODUCTION

The concept of intelligibility, the ability of "listeners to understand the speaker's intended message" (Derwing & Munro, 2015, p. 1) is important to both researchers and pronunciation teachers. For example, pronunciation teachers would like to know where to focus their efforts: which aspects of pronunciation are most likely to make a difference in understanding? However, intelligibility depends not only on the speaker's pronunciation but also on what the listener finds understandable. Therefore, in order to investigate intelligibility researchers have to think about listeners.

Quite a bit of information is known about what aids intelligibility in English when native speakers are listening. In particular, suprasegmentals have been found to be very important for native-speaker listeners of English (Hahn, 2004; Kang & Pickering, 2011). In the case of English, though, the majority of English users are actually non-native speakers and listeners (Lewis, Simons, & Fennig, 2015). It is therefore likely that many English speakers will find themselves communicating more with non-native speakers than with native speakers, or perhaps not with native speakers at all. It is reasonable to question whether the pronunciation aspects that facilitate intelligibility for native speakers will be the same aspects that facilitate intelligibility for non-native speakers. Pronunciation teachers of English for those who will be using English as a lingua franca are left with few research-based resources upon which to draw.

Jenkins' (2000, 2002) Lingua Franca Core (LFC) is just the sort of resource that many pronunciation teachers are looking for; it is a syllabus attempting to establish exactly which aspects of pronunciation are important for non-native listeners and which are not important or perhaps might even be detrimental to ELF intelligibility. Many of the points of the LFC are different from what is believed about what facilitates intelligibility for native listeners. For example, the LFC does not include word stress for intelligibility in ELF situations. The LFC is comprehensive, covering both segmental and suprasegmental aspects of pronunciation. As the seminal resource for information on ELF intelligibility, the LFC has been relatively widely published in books for teachers and pronunciation textbooks.

However, the LFC suffers from some methodological issues. In addition to the fact that the data upon which the LFC is based is rather limited, the study was based on listening for situations in ELF conversation that triggered repair and then trying to determine which aspect of the pronunciation had caused the misunderstanding. This method has the potential to miss misunderstandings which were not repaired, such as misunderstandings that the participants did not recognize as misunderstandings or misunderstandings that they chose not to repair. For reasons like these, further research into the LFC has been called for (Dauer, 2003; Haslam & Zetterholm, 2016). In particular, perceptual research has the power to more directly establish the effects of certain pronunciation aspects on intelligibility.

Haslam and Zetterholm (2016) is a first attempt to evaluate one of the aspects of the LFC using perceptual methodology by testing the LFC's claim that aspiration is required on fortis consonants in initial position in stressed syllables. In the 2016 study perceptual results were compared with acoustic analysis to more directly establish the relationship between acoustic factors and intelligibility for ELF listeners. Results of the study showed that the actual picture of ELF intelligibility is more complicated than the LFC predicted for this situation. Listeners quite successfully identified the target words, regardless of acoustic characteristics. In more detailed acoustic analysis, while the fortis consonants /t/ and /k/ partially followed the LFC's predictions, results for /p/ showed a completely different pattern. Therefore, these results demonstrate the need for further investigation into the aspects of the LFC for deeper understanding of the picture of ELF intelligibility.

The present study is another investigation into the points of the LFC, specifically the claims about consonant clusters. The LFC requires the following for ELF intelligibility:

"no omission of sounds in word-initial clusters, eg. in promise, string;
"Omission in middle and final clusters only permissible according to L1 English rules of syllable structure, e.g. factsheet can be pronounced 'facsheet' but not 'fatsheet' or 'facteet';…
"Addition is acceptable, for example, 'product' pronounced [pər'ɑdʌkʊtɔ] was intelligible to NNS interlocutors, whereas omission was not, for example, 'product' pronounced /'pɑdʌk/." (Jenkins, 2002, p. 97)

In the present study, perceptual methodology in combination with acoustic analysis was used to address the following questions:

(1) What strategies do ELF speakers use to pronounce consonant clusters (i.e. deletion, epenthesis, etc.) in initial and final positions?

(2) Do these strategies facilitate or hinder intelligibility for ELF listeners?

METHOD

76 tokens representing 66 unique monosyllabic words with consonant clusters in either initial position (47 words) or final position (36 words) were selected. Recordings of these words were selected from the Wildcat corpus (Bradlow et al., 2007; Van Engen, et al., 2010), specifically from a task where non-native speakers of English who did not share a native language discussed differences between two similar pictures. There were 21 different talkers with the following native languages: Chinese (n=5), Korean (n=8), Persian (n=1), Italian (n=1), Japanese (n=1), Marathi/Hindi (n=1), Russian (n=1), Spanish (n=2), and Thai (n=1).

Listeners were native speakers of Swedish and therefore did not share an L1 with any of the talkers. There were 11 listeners who reported a range of English proficiency from Basic to Advanced.

Listeners were asked to complete a computerized dictation task. For each item, the listener heard a stimulus recording. He/she was then asked to type in what word he/she thought he/she had heard using normal English orthography. Responses were coded for CV structure according to the normal spelling rules of English (e.g. "black" -> CCVC). CV structure of the target word was also coded.

Acoustic analysis was also completed on the stimuli for CV structure using Praat software (Boersma, 2001). Based on the acoustic analysis, each stimulus was assigned a CV structure. CV structures of the responses were then compared to the CV structures from the acoustic analysis and the CV structures of the target words to identify match or mismatch. If the consonant cluster existed in both structures in the targeted position (i.e. initial position or final position), this was counted as match. Therefore, if the target word were "black" (CCVC) and the response were "brag" (CCVC) this would be counted as a match because the CV structure is the same. Mean proportions of match between listeners' responses and the acoustic analysis were calculated as well as mean proportions of match between listeners' responses and the CV structure of the target word. When a mismatch occurred, the item was also coded for type of mismatch.

RESULTS

For all stimuli, the mean proportion of match between listeners' responses and the acoustic analysis was 0.669 (SD=0.03321), while the mean proportion of match between listeners' responses and the target word was 0.7193 (SD=0.08387). That is, listeners were more successful at identifying the CV structure of the target word than the acoustic analysis would suggest. These results are presented in Figure 1. Results were submitted to paired-samples t-test and the difference between the two means was found to be significant ($t(9)=-2.281$, $p=0.049$).

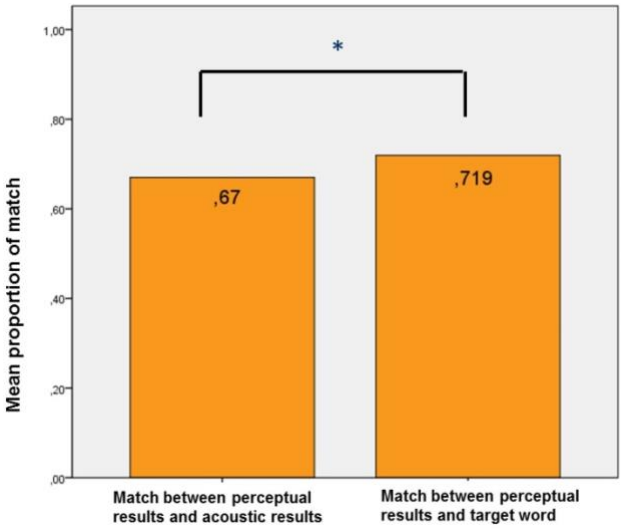

Figure 1. Mean proportion of match between perceptual results and acoustic results, and mean proportion of match between perceptual results and target word for all stimuli.

Additional analysis was conducted for items with clusters in initial position and final position. For clusters in initial position, the mean proportion of match between listeners' responses and the acoustic analysis was 0.7489 (SD=0.04683) while the mean proportion of match between listeners' responses and the target word was 0.7787 (SD=0.10996). These results are visible in Figure 2. While the difference between means indicates that listeners were slightly more successful at identifying the target word than acoustic analysis would suggest, paired t-test results did not find a significant difference in this case (t(9)=-1.121, p=0.291).

For items with clusters in final position, the mean proportion of match between listeners' responses and the acoustic analysis was 0.5667 (SD=0.05885) while the mean proportion of match between listeners' responses and the target word was 0.6417 (SD=0.08013). These results are visible in Figure 3. Paired t-test confirmed that this difference was significant (t(9)=-3.948, p=0.003). Therefore, we can conclude that listeners were more successful at identifying the CV structure of the target word than the acoustic analysis would suggest, even though the overall proportion of match between both listeners' responses and the acoustic analysis and listeners' responses and the target word were lower for final clusters than for initial clusters.

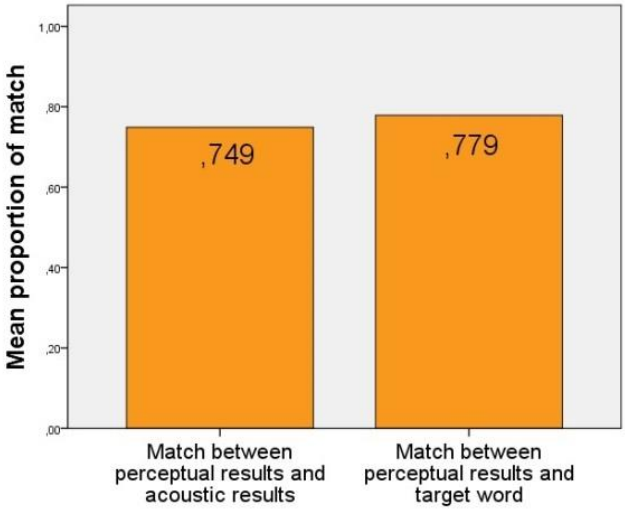

Figure 2. Mean proportion of match between perceptual results and acoustic results, and mean proportion of match between perceptual results and target word for initial clusters.

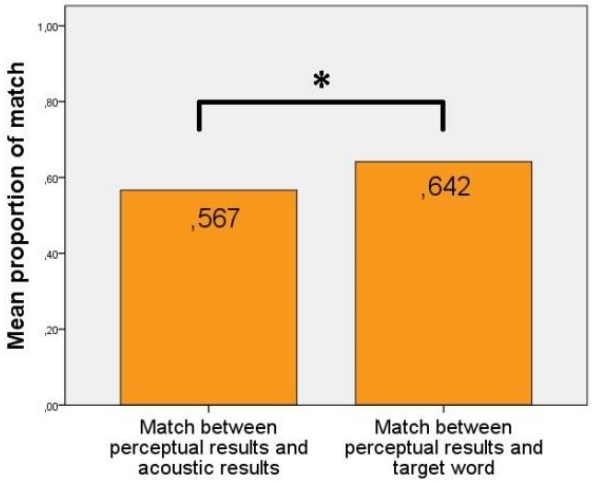

Figure 3. Mean proportion of match between perceptual results and acoustic results, and mean proportion of match between perceptual results and target word for final clusters.

Table 1

Itemized list of reasons for mismatch between acoustic information and perceptual information, separated into acoustic reasons for mismatch and perceptual reasons for mismatch

Acoustic reasons for mismatch		Perceptual reasons for mismatch	
Reason	Number of instances	Reason	Number of instances
Deletion of consonant from cluster	87	Deletion of consonant from cluster	80
Extra initial vowel	31	Extra initial vowel	38
Extra final consonant	9	No response	22
More than 2 extra initial sounds	1	More than 2 extra final sounds	16
		More than 2 extra initial sounds	8
		Extra final consonant	3
		Extra initial consonant	3
		Other differences	2
		Extra final vowel	2
		Extra vowel within consonant cluster	1

Table 2

Itemized list of reasons for mismatch between target word and perceptual information

Perceptual reasons for mismatch	
Reason	Number of instances
Deletion of a consonant from the cluster	127
Extra initial vowel	69
No response	24
2 or more extra final sounds	17
2 or more extra initial sounds	15
Extra final consonant	5
Other differences	3
Extra final vowel	3
Extra initial consonant	2
Extra vowel within consonant cluster	1

A number of different reasons for mismatch between listeners' responses and the acoustic information were identified. For some responses the acoustic analysis revealed deleted or added sounds, while in some cases the reason for the mismatch seemed to lie within the listeners' responses. Reasons for mismatch between the listeners' responses and the target word were also identified. Itemized results of this coding are presented in Tables 1 and 2. In all cases, the lack of

a consonant cluster, i.e. deletion of one of the consonants from the consonant cluster, was the top reason for mismatch, followed by addition of an initial vowel to initial consonant clusters.

DISCUSSION

To summarize the results, overall, listeners seemed to perform better at identifying the target word than the acoustic analysis would suggest. This pattern holds for final clusters but not for initial clusters. These results seem to be in line with previous results (Haslam & Zetterholm, 2016) which demonstrated that ELF listeners were often successful at identifying target words regardless of VOT of the initial consonant. In combination with previous results on VOT, the present results may indicate that the particular characteristics of the consonant/s (e.g. VOT) are not especially important for recognition of a word as long as an initial consonant/s is present; that is, listeners may be using top-down skills to recognize words rather than bottom-up processing. This supports Field's (2004) analysis that L2 listeners rely more on top-down processing than bottom-up.

Another area of interest in these results has to do with the relative status of initial clusters vs. final clusters. Listeners performed more accurately in general on recognizing the structure of initial clusters than final clusters. These results can possibly be explained by the fact that some sort of phonetic reduction often happens at the ends of words, such as word-final devoicing produced by some non-native speakers of English (Edge, 1991). However, there was significantly more match between the result of the perceptual test and the target word than there was match between the perceptual result and the acoustic information for final clusters, but there was no corresponding significant difference for initial clusters. These results together indicate that listeners do find final clusters important, but that they may be depending on other information, such as the vowel, to identify words with initial clusters.

The present study was an attempt to support or refute the LFC's points about consonant clusters. These results can be interpreted as partial support of the LFC. The LFC's claim that deletion of consonants from consonant clusters would result in intelligibility seems to be supported: deletion was the top reason for mismatch in all cases. When these results are broken down into initial and final clusters, however, only the final clusters support the LFC's claim.

In addition, the LFC's claim that epenthesized consonant clusters would be intelligible seems to not be supported: the presence of an extra initial vowel (i.e., a vowel inserted before the initial consonant cluster) was the second most common reason for mismatch in these results. Further research can investigate in more detail whether the location of epenthesis is important. In the present study, epenthesis before the initial consonant cluster was found to be a major reason for mismatch, but epenthesis between the consonants of the consonant cluster was a very uncommon reason for mismatch.

Since the present study used only monosyllabic words, further research can also investigate the effect of strategies such as epenthesis and deletion when the consonant cluster is part of a two- or multi-syllable word. As both Haslam and Zetterholm (2016) and the present study suggest, using perceptual methodology to investigate ELF perception can be a valuable line of inquiry. Further research can also focus on other aspects of the LFC such as its claims about suprasegmental pronunciation in addition to segmental aspects.

ABOUT THE AUTHORS

Mara Haslam is a Senior Lecturer in English with language education orientation at the Department of Language Education at Stockholm University. Her research interests include how aspects of pronunciation affect intelligibility in different languages and combinations of speakers and listeners. Her research is inspired by her experiences as a teacher of pronunciation and she hopes that it will help pronunciation teachers to prioritize which aspects of pronunciation they should teach. Mara's teaching efforts focus on helping pre-service teachers prepare to teach English in the Swedish school system. Mara completed a Ph.D. at the University of Utah in 2011 with a dissertation titled *The effect of perceptual training including required lexical access and meaningful linguistic context on second language phonology*.

Elizabeth Zetterholm is a Senior Lecturer and Associate Professor in Swedish as a second language at the Department of Language Education at Stockholm University. Her main research interest is second language pronunciation with a special focus on Swedish. She has been studying how pronunciation is taught in classroom with adult second language learners of Swedish. Another research interest is how orthography can have an impact on pronunciation for second language learners.

REFERENCES

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-345.

Bradlow, A. R., Baker, R. E., Choi, A., Kim, M., & Van Engen, K. J. (2007). The Wildcat corpus of native-and foreign-accented English. *Journal of the Acoustical Society of America*, *121*(5), 3072.

Dauer, R. M. (2005). The Lingua Franca Core: A new model for pronunciation instruction? *TESOL Quarterly*, *39*(3), 543-550.

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

Edge, B. (1991). The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition, 13*(3), 377-393. doi:10.1017/S0272263100010032

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, *32*(3), 363-377.

Hahn. L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201–223.

Haslam, M., & Zetterholm, E. (2016). The importance of aspirated initial stops in English as a lingua franca. In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference.* ISSN 2380-9566, Dallas, TX, October 2015 (pp.66-75). Ames, IA: Iowa State University.

Kang, O., & Pickering, L. (2011). The role of objective measures of suprasegmental features in judgments of comprehensibility and oral proficiency in L2 spoken discourse. *Speak Out, 44*, 4–8.

Lewis, M.P., Simons, G.F., & Fennig, C.D. (Eds.). (2015). *Ethnologue: Languages of the world*, (18th edition). Dallas, Texas: SIL International.

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

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83-103.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(4), 510-540.