THE IMPORTANCE OF ASPIRATED INITIAL STOPS
IN ENGLISH AS A LINGUA FRANCA

Mara Haslam, Stockholm University
Elisabeth Zetterholm, Stockholm University

A significant proportion of the population of the world is made up of users of English as a Lingua Franca (ELF). Jenkins (2000) published the Lingua Franca Core (LFC), a syllabus for ELF pronunciation, including the claim that the fortis/lenis distinction must be preserved on English stop consonants for successful ELF intelligibility. The present study evaluates the relationship between Voice Onset Time (VOT) and how the sounds are perceived by ELF listeners. 101 tokens produced during ELF interaction which contained the stops /b/, /p/, /d/, /t/, /g/, or /k/ were played for 9 Swedish listeners, who could indicate that they heard either the word or its minimal-pair counterpart, e.g. *bees* or *peas*. The relationship between VOT and perceived stop was analyzed, with the expectation that longer VOTs would be associated with fortis consonants and shorter VOTs would be associated with lenis consonants. Results followed the predicted pattern for /d/ and /g/ but not for /t/ and /k/. In addition, the pattern observed for /p/ and /b/ is the reverse of the pattern found for the other consonants. These results suggest that further research into the LFC's claim about the fortis/lenis distinction and other LFC claims are warranted.

INTRODUCTION

An estimated one quarter to one third of the world's population speaks English, making it a very important language in the world (Crystal, 2008). Of these speakers, an estimated 75 percent are non-native speakers. Many of these people will engage in communication with other non-native speakers of English, making English as a Lingua Franca (ELF) an important form of communication, worthy of research.

ELF has only recently emerged as a field of study and much remains to be learned about ELF phonology. In 2000, Jenkins published the seminal book "The Phonology of English as an International Language" in which she investigated phonological factors that contribute or do not contribute to misunderstanding in ELF communication. Her conclusions are presented as the Lingua Franca Core (LFC), a list of phonological characteristics which Jenkins claims ELF speech must have in order to be intelligible. The LFC is the only current standard available for ELF pronunciation and has even been included in popular pronunciation textbooks (e.g. Celce-Murcia, Brinton, Goodwin, and Griner, et al., 2010; Hancock, 2012).

However, there is reason to believe that the LFC warrants further testing. First, the data upon which the LFC was based was recorded for other purposes, not specifically for the purpose of investigating intelligibility during ELF communication. The participants in the study upon which the LFC is based were principally Japanese users of English communicating with Swiss German-speaking users of English and therefore represent a somewhat narrow cross-section of ELF users. In addition, Jenkins discovered the phonological points outlined in the LFC by listening for communication breakdown and then attempting to interpret which aspect of the pronunciation caused the breakdown rather than testing the cause of unintelligibility, perception, more directly. While the LFC represents a good step forward in understanding ELF intelligibility and intelligibility of non-native speech in general, it is perhaps prudent to view the LFC as a hypothesis which can be tested, a starting point from which further research can depart. Deterding (2013) uses a methodology similar to Jenkins' (2000), based on external observation of communication breakdown. The present study represents an attempt to apply laboratory methods to one of the aspects of the LFC with the purpose of testing and refining the LFC to further our understanding of ELF phonology and second-language phonology in general. We have turned our attention first to the fortis-lenis distinction on stop consonants as a starting point for investigating the LFC.

One of the claims of the LFC is that ELF intelligibility will break down unless the fortis-lenis distinction on stop consonants is preserved, such as the distinction between /b/ (lenis) and /p/ (fortis). The fortis-lenis distinction is a well-established but somewhat difficult to quantify characteristic of native speakers' speech in English. Lisker and Abramson (1967) identified that for native speakers and for word-initial stops, the fortis-lenis distinction is mostly reflected in Voice Onset Time (VOT), with lenis stops showing shorter VOTs than their fortis counterparts. Therefore the present study takes as its research questions the following:

- Do ELF listeners perceive stops with short VOT as lenis and stops with longer VOTs as fortis?
- Is the fortis-lenis distinction necessary for ELF intelligibility?

METHODS

Stimuli

Word tokens from the Wildcat corpus' (VanEngen, Baese-Berk, Baker, Choi & Bradlow, 2010) Diapix task were used. In this task two participants looked at similar images and talked with each other to find the differences between the images. In order to capture ELF pronunciation, only Diapix conversations in English between speakers with different L1s were used. Table 1 shows the language backgrounds of the talkers and how many talkers of each native language created these conversations.

Table 1

First languages and number of talkers from that language background who produced the tokens used in the present study

Native Language	Number of Talkers
Korean	8
Chinese (Mandarin)	6
Spanish	2
Italian	1
Japanese	1
Russian	1
Persian	1
Thai	1
Marathi/Hindi	1
TOTAL	**22**

From these recordings, the monosyllabic words starting with /p/, /t/, /k/, /b/, /d/, and /g/ which were each said by the most talkers were identified and selected as target words. This was done by compiling all words in the corpus from the corpus transcription and then manually selecting the words which fit the requirements. Table 2 shows the target words and the number of talkers who said that target word during their Diapix conversation. The first occurrence of the word in the Diapix conversation for each talker was selected as the token used for the present research. The total number of tokens was 101. No non-target words were used; all items in the experiment were recordings from the corpus. Recordings of the minimal-pair counterparts of the target words (see **Procedure**, below) were not used because such recordings did not always exist in the corpus.

Table 2

Words and number of tokens of each word used in the present study

Word	Number of tokens used
bees	15
cat	20
door	21
got	6
pet	19
two	20
TOTAL TOKENS	**101**

Listeners

Nine listeners were recruited from the Stockholm University community. All identified as native speakers of Swedish. Each listener received a coupon good for a beverage and a pastry at a local café for his/her participation. Swedish-speaking listeners were used in order to simulate an ELF listening situation as much as possible, considering that none of the talkers were native speakers of Swedish.

Procedure

The experiment took place in a quiet office environment. The main task consisted of an identification task where two words in orthographic representation appeared on a computer screen while the listener heard one of the token recordings. Listeners were instructed to press a key labeled "RIGHT" if the word they heard was the word displayed on the right side of the screen and a key labeled "LEFT" if the word they heard was the word displayed on the left side of the screen. The words displayed were always the target word associated with the token and the other member of the minimal-pair containing the voiced counterpart if the word started with a voiceless consonant, or vice versa. For example, if the item involved a token where the target word was "bees", the words "bees" and "peas" were displayed on the computer screen. Each of the 101 tokens was tested in this fashion. The experimental task took approximately 10 minutes to complete.

Table 3 shows the target words and their minimal-pair counterparts. Display of the words was randomized so that the target word did not always appear on the left or right side of the screen. A response was counted as correct if the participant pressed the key corresponding to the side of the screen where the target word was displayed. A response was counted incorrect if the participant did not respond within 4 seconds or if the person pressed the key corresponding to the side of the screen where the minimal-pair counterpart of the target word was displayed.

Table 3

Words used in the study (see Table 2) and their minimal-pair counterparts

Target Word	Minimal-Pair Counterpart
bees	peas
cat	gat
door	tore
got	cot
pet	bet
two	do

After completion of the identification task, the participants filled in a computerized survey with information about their age and background information. Participants were 8 females and 1 male, ranging in age from 21 to 47. All were native speakers of Swedish.

VOT for each of the tokens was measured by one of the authors using the analysis program Praat (Boersma, 2001) and calculations comparing the perception results and the VOT measurements were performed.

RESULTS

Figure 1 depicts, for each target phoneme, the mean proportion of tokens correctly identified as the target phoneme by the listeners.
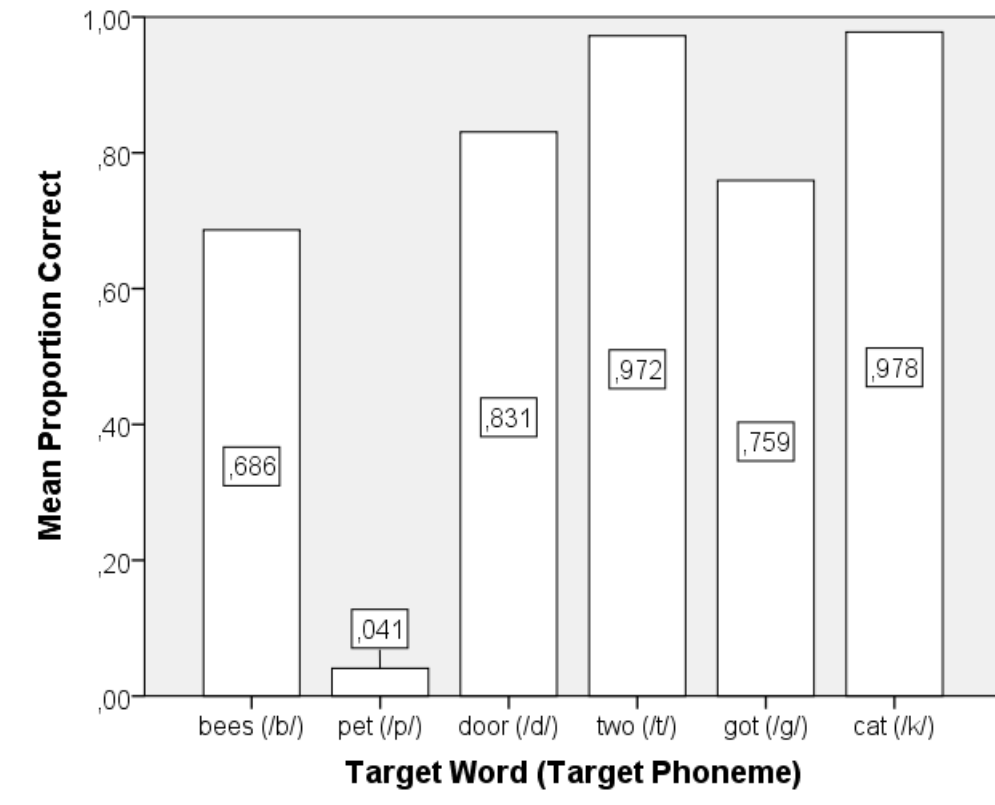
Figure 1. Proportion correct responses (i.e. tokens identified as the target) for the six target phonemes /b/, /p/, /d/, /t/, /g/, and /k/ as measured in 101 total tokens.

Note that in Figure 1 the pattern of responses to the six target phonemes is quite different depending on which phoneme pair the target phoneme belongs to. While the proportion of correct responses to /d/, /t/, /g/, and /k/ were all quite high, the correct responses to /b/ were 69% and the proportion of correct responses to /p/ were 4%, much lower than chance. One–way ANOVA calculation with target phoneme as a within-subjects factor and proportion correct as the dependent variable showed a significant effect of target phoneme (F=106.634, p=0.00) and subsequent comparison of phoneme pairs found significant effects of target phoneme for all three pairs /b/-/p/, /d/-/t/, and /g/-/k/, indicating that participants responded significantly differently to different members of the pairs (/b/-/p/ F=152.407, p=0.00; /d/-/t F=22.708, p=0.001; /g/-/k/ F=17.966, p=0.003).

Figures 2, 3, and 4 depict the relationship between VOT and which phoneme in each pair was selected for /b/-/p/, /d/-/t/, and /g/-/k/ respectively.
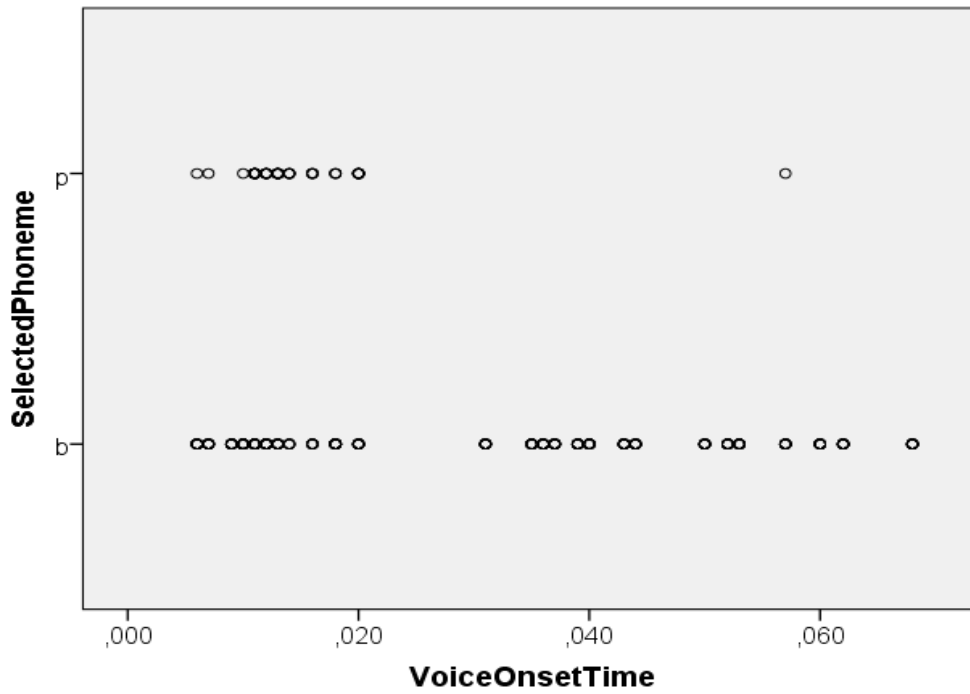
Figure 2. Plot for /b/-/p/ showing VOT and which phoneme in the pair was selected.
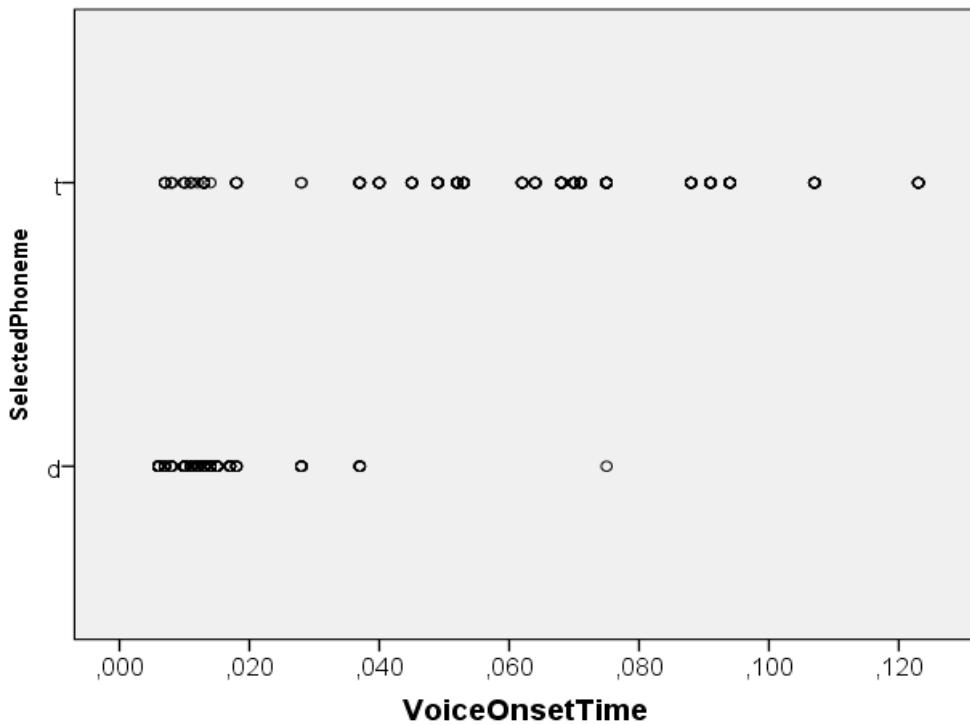

Figure 3. Plot for /d/-/t/ showing VOT and which phoneme in the pair was selected.
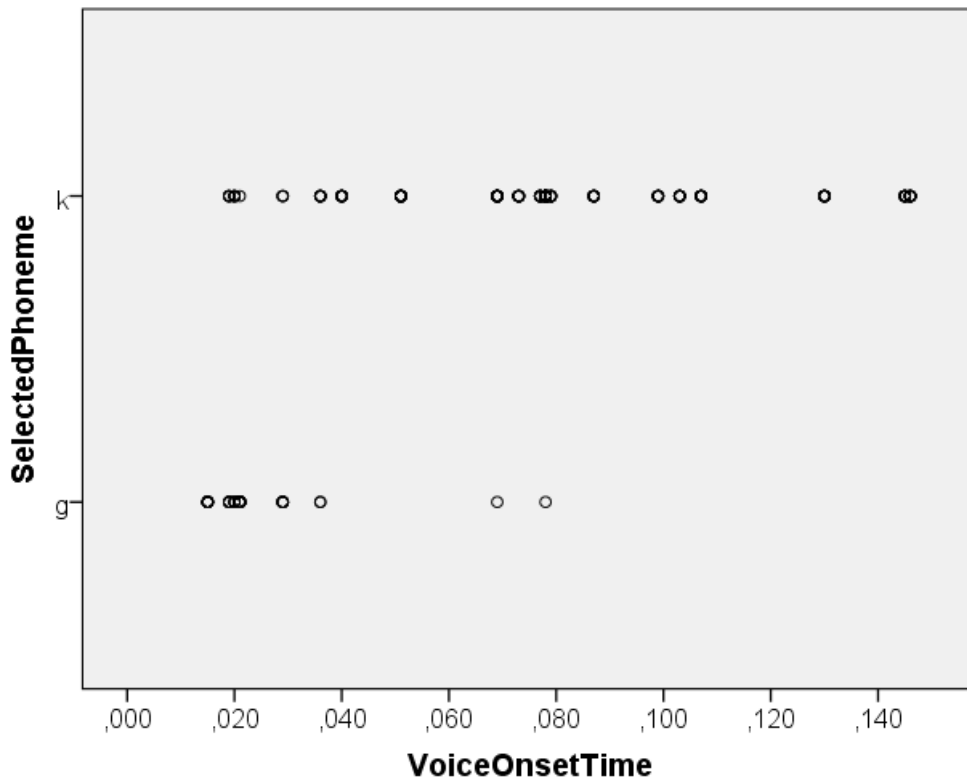
Figure 4. Plot for /g/-/k/ showing VOT and which phoneme in the pair was selected.

The hypothesized pattern was that the lenis member of each pair (i.e., /b/, /d/, or /g/) would be selected for items shorter VOTs and the fortis member of each pair (i.e., /p/, /t/ or /k/) would be selected for items with longer VOTs. For the /d/-/t/ pair and the /g/-/k/ pair this hypothesis is partially followed in that the lenis member of the pair was generally chosen only for short VOTs, but participants selected the fortis member of the pair for both shorter and longer VOTs. However, this pattern is reversed in the /b/-/p/ pair, where /p/ was generally selected only in combination with short VOTs while /b/ was selected for both longer and shorter VOTs.

DISCUSSION

This study was conducted primarily to examine the LFC's claim that the fortis-lenis contrast is important to preserve in ELF speech and that not preserving this contrast would cause problems in intelligibility. The results from this study provide a mixed view on the question, partially supporting the LFC's requirement for the fortis-lenis contrast and partially providing evidence against the LFC's claim. As shown above, listeners were not always successful at identifying whether they were listening to the target word or whether they were listening to its minimal-pair counterpart, especially in the case of /b/-

/p/ words. Listeners' relatively poor performance on this pair of sounds would cause them to need to depend on something other than segmental phonology for understanding.

Taking the results into account in general, we can say that participants responded to the /b/-/p/ pair quite differently than they did to the other pairs, and this itself provides important clarification of the LFC. The LFC treats stop consonants as a unified group, arguing that the fortis-lenis distinction is important for all stops, while this study demonstrates that the listeners responded very differently to the /b/-/p/ pair of stops than they did to the other two pairs. This suggests that a more granular approach to the LFC is warranted: instead of treating sounds in large groups such as "stop consonants" or "vowels" perhaps more specific examination will reveal other unexpected information about certain sounds or contrasts that act as intelligibility pitfalls for ELF listeners.

Since the VOT vs. selected phoneme results are noticeably different from the expected pattern of selecting the fortis member of the phoneme pair for long VOTs and the lenis member of the pair for short VOTs, further research will hopefully be able to pin down what listeners are responding to when making their selections. In addition to VOT, a probable candidate is transitions into the following vowel, length of the following vowel, or perhaps some combination of these factors with VOT.

Another issue for further research is the issue of how accurate at perception ELF listeners really are. ELF research in general shows that not much conversational repair is needed and this is taken as evidence that ELF communication is relatively successful. However, the results of this study demonstrate that perhaps ELF listeners are not as successful at perception of consonants as they seem to be. Further research with more controlled methods is warranted to investigate just how successful bottom-up phonological processing in ELF communication is. Perhaps ELF listeners, because of their poor stop consonant perception, are required to depend heavily on top-down processing in ELF communication situations. This has potential implications for both future ELF research and English pronunciation and listening instruction.

ABOUT THE AUTHORS

Mara Haslam is a Senior Lecturer at the Department of Language Education, Stockholm University. Her primary research interest is in second-language phonology with a view to how pronunciation instruction can be improved. This includes interest in second-language intelligibility.

Elisabeth Zetterholm is a Senior Lecturer and Associate Professor in Swedish as a Second Language at Stockholm University. Her current research interests focus on pronunciation in second language learning, comprising segmental and prosodic issues. Zetterholm is also interested in background languages' influence on third language learning, with a focus on phonology transfer and pronunciation. When teaching, most of her students are prospective teachers of Swedish as a second language.

REFERENCES

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

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide*. Cambridge University Press.

Crystal, D. (2008). Two thousand million? *English Today*, *24*, 3-6. doi:10.1017/S0266078408000023.

Deterding, D. (2013). *Misunderstandings in English as a lingua franca: An analysis of ELF interactions in South-East Asia* (Vol. 1). Walter de Gruyter.Hancock, M. (2012). English pronunciation in use intermediate with answers. Cambridge University Press.

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

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). (2015). *Ethnologue: Languagesof the World, Eighteenth edition*. Dallas, Texas: SIL International. Online version:http://www.ethnologue.com.

Lisker, L., & Abramson, A.S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, 10, 1-28.

Seidlhofer, B. (2004). 10. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, *24*, 209-239. doi:10.1017/S0267190504000145.

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*(Pt 4), 510–540.