

# XENOPHONES REVISITED: LINGUISTIC AND OTHER UNDERLYING FACTORS AFFECTING THE PRONUNCIATION OF FOREIGN ITEMS IN SWEDISH

Anders Lindström & Robert Eklund  
*Telia Research AB, Farsta, Sweden*

## ABSTRACT

In this paper, the distribution of Swedish subjects' productions of foreign speech sounds, here termed xenophones, is studied, and tabulated across gender, age, and region. The results are grouped in three categories along the "awareness" and "fidelity" dimensions. Results indicate that age is by far the most decisive underlying factor, which can be explained in terms of cultural exposure.

## 1. INTRODUCTION

Increasing international contact and cross-cultural influences are becoming more and more palpable in our modern society. From a language point of view, this affects the language user on all levels, e.g. syntax, morphology, lexicon, phonology. The particular problem discussed in this paper is the inclusion of "foreign" speech sounds in Swedish.

### 1.1. The Pronunciation Problem

A foreign item (name or word) produced by a Swedish speaker in an otherwise Swedish context can be pronounced with varying degrees of adjustment to the Swedish phonological system, spanning the whole range from virtually no adjustment, via some degree of rephonematization, to total adjustment. This adjustment may well change over time. From a more synchronic perspective, the sounds that are 'foreign' to the phone inventory of a language have a special status in the phonological system of that language. On the one hand, it is hard to claim that they be a part proper of that system, but on the other hand, they might still have such a status that most people would expect them to be used in certain linguistic contexts. We have introduced the term **xenophones** (meaning 'foreign sounds') to denote such sounds.

### 1.2. Previous Work

In a 1996 study [2], we first discussed the problem, and made a first attempt to determine the nature of this extended phone set by presenting results from a small production study of speakers from Stockholm and Skåne (Scania). The foreign words examined were all of English origin. In a follow-up study [3], we examined speakers from all Swedish dialectal areas, and showed that a majority of Swedish speakers do indeed add several foreign speech sounds to their phone repertoire when reading Swedish sentences with embedded foreign names and words. Based on these results, we recorded a set of xenophone polyphones and demonstrated how they could be used to enhance the quality of our Swedish text-to-speech system.

**1.3. The Present Study** As discussed previously [2, 3], a number of underlying factors can be assumed to be involved in governing the degree of adjustment, including the speaker's

competence and performance capabilities with respect to the source language, the speaker's expectations of the listener's competence, the relative social status of speaker and listener, the socio-cultural distance to the country of origin, recency and frequency of the lexical item in question, and similarities/dissimilarities between the two phonological systems in question. Moreover, since the dialects of a given language might differ in terms of phonology, it would not be surprising to find dialect-specific variation in the treatment of foreign items. In the present study, we are focussing on the three parameters age, gender and regional (dialectal) variation in order to see whether the aforementioned results can be explained in terms of these factors. In the previous studies, it was apparent that the foreign sounds produced can be divided into three groups: (1) A group of sounds produced by more or less all the subjects, e.g. [tʃ, aɪ, eɪ, ju:, æʃ]; (2) A group of sounds produced by almost no one, e.g. [z, ʒ, ɫ, wʃ]; and (3) A group of sounds with more dispersed distribution, e.g. [dʒ, θ, ð, æ, ɔ]. In this paper, we have selected primarily sounds from the third group.

## 2. METHOD

### 2.1. The Linguistic Material

A set of twelve sentences containing names and words of English origin was created. These names and words contained 15 English speech sounds that were chosen so that they would differ phonetically from Swedish speech sounds to varying degrees, and would not be included in any traditional description of the Swedish phonological system. An example is shown below:

*Många har Roger Moore som favorit i rollen som James Bond.*  
("Many prefer Roger Moore's interpretation of James Bond")

### 2.2. The Recordings

The sentences were included in a much larger session of linguistic material recorded to train the Telia/SRI Swedish speech recognizer [1]. The sentences were presented under the heading 'Kändisar' (Celebrities), and it can be assumed that subjects were unaware of the fact that their pronunciation was the object of study. The subjects were all Telia employees or relatives of Telia employees. The age span was 15 to 75. The sentences were recorded hi-fi on disk by more than 460 subjects on 40 different locations covering the whole of Sweden. Thus, all major dialects were covered. In this way a total of 13,343 tokens were collected.

**Table 1:** Distribution of Swedish subjects' productions of nine target xenophones (a-i) across **gender** and category [per cent].

Gender	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
		a) Target $\zeta$ occurring in the word Roger			b) Target $\zeta$ occurring in the word James			c) Target $\zeta$ occurring in the word Jackson				
M	245	33 $\zeta$		67 $\alpha$	245	20 $\zeta$		80 $i$	240	42 $\zeta$	1 -	57 $i$
F	230	33 $\zeta$	1 -	66 $g$	230	24 $\zeta$		76 $j$	226	52 $\zeta$	2 -	46 $j$
		d) Target $\theta$ occurring in the word Thriller			e) Target $\theta$ occurring in the word Thatcher			f) Target $\delta$ occurring in the World				
M	240	48 $\theta$	1 $\uparrow$	51 $t$	243	43 $\theta$	1 -	56 $t$	243	37 $\delta$	1 $r$	58 $d$
											1 $d$	2 $t$
											1 $v$	
F	226	51 $\theta$	3 $\uparrow$	46 $t$	230	42 $\theta$	2 -	56 $t$	230	40 $\delta$	1 $v$	56 $d$
											1 $r$	1 $t$
											1 -	
		g) Target $\alpha$ occurring in the word Stone			h) Target $\alpha$ occurring in the word Maggie			i) Target $\alpha$ occurring in the word Jackson				
M	242	88 $\alpha$	2 -	4 $\uparrow$	243	88 $\alpha$	2 -	10 $a$	240	74 $\alpha$		26 $a$
				6 $\alpha$								
F	228	91 $\alpha$	1 -	3 $\uparrow$	230	93 $\alpha$	2 -	5 $a$	226	76 $\alpha$	2 $\epsilon$	22 $a$
				5 $\alpha$								

**Table 2:** Distribution of Swedish subjects' productions of nine target xenophones (a-i) across **age** and category [per cent].

Age	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
		a) Target $\zeta$ occurring in the word Roger			b) Target $\zeta$ occurring in the word James			c) Target $\zeta$ occurring in the word Jackson				
66-75	4	50 $\zeta$		50 $\alpha$	4			100 $i$	4			100 $i$
56-65	31	32 $\zeta$		68 $g$	31	13 $\zeta$		87 $j$	27	37 $\zeta$		63 $j$
46-55	146	27 $\zeta$		73 $\alpha$	146	16 $\zeta$		84 $i$	143	39 $\zeta$	1 -	60 $i$
36-45	133	31 $\zeta$		69 $g$	133	20 $\zeta$		80 $j$	130	55 $\zeta$	1 $\epsilon$	44 $j$
26-35	86	23 $\zeta$	3 -	74 $\alpha$	86	30 $\zeta$		70 $i$	88	53 $\zeta$	3 $d_{\uparrow}$	43 $i$
16-25	43	60 $\zeta$		40 $g$	43	35 $\zeta$		65 $j$	43	46 $\zeta$		54 $j$
<16	22	59 $\zeta$		41 $\alpha$	22	27 $\zeta$		73 $i$	21	62 $\zeta$		38 $i$
		d) Target $\theta$ occurring in the word Thriller			e) Target $\theta$ occurring in the word Thatcher			f) Target $\delta$ occurring in the World				
66-75	4			100 $t$	4			100 $t$	2		50 $r$	50 $d$
56-65	27	59 $\theta$	8 $\uparrow$	33 $t$	29	42 $\theta$	3 $s$	52 $t$	30	27 $\delta$	3 $d_{\downarrow}$	57 $d$
							3 $tc$				3 $r$	10 $t$
46-55	143	34 $\theta$	4 $\uparrow$	62 $t$	147	36 $\theta$	1 $tc$	63 $t$	146	40 $\delta$	1 $r$	54 $d$
											4 -	1 $t$
36-45	130	54 $\theta$	1 $\uparrow$	45 $t$	133	46 $\theta$	1 $tc$	53 $t$	134	42 $\delta$	1 $d_{\downarrow}$	54 $d$
											3 -	
26-35	88	62 $\theta$		38 $t$	86	50 $\theta$	1 $tc$	50 $t$	86	59 $\delta$		41 $d$
16-25	43	56 $\theta$	2 $\uparrow$	42 $t$	43	44 $\theta$		56 $t$	42	38 $\delta$	2 $h$	60 $d$
<16	21	57 $\theta$		43 $t$	21	43 $\theta$	5 -	52 $t$	21	18 $\delta$		82 $d$
		g) Target $\alpha$ occurring in the word Stone			h) Target $\alpha$ occurring in the word Maggie			i) Target $\alpha$ occurring in the word Jackson				
66-75	4	50 $\alpha$		50 $\uparrow$	4	50 $\alpha$		50 $a$	4	25 $\alpha$		75 $a$
56-65	30	83 $\alpha$		10 $\alpha$	29	72 $\alpha$	4 -	24 $a$	27	63 $\alpha$		37 $a$
				7 $\uparrow$								
46-55	143	83 $\alpha$	4 -	9 $\alpha$	147	88 $\alpha$		12 $a$	143	68 $\alpha$	4 $\epsilon$	28 $a$
				4 $\uparrow$								
36-45	133	90 $\alpha$	1 -	3 $\alpha$	133	95 $\alpha$	1 $\alpha$	4 $a$	130	82 $\alpha$		18 $a$
				6 $\uparrow$								
26-35	84	92 $\alpha$		4 $\alpha$	86	94 $\alpha$	2 $\alpha$	2 $a$	88	78 $\alpha$		22 $a$
				4 $\uparrow$			2 -					
16-25	43	100 $\alpha$			43	95 $\alpha$	5 -		43	91 $\alpha$		9 $a$
<16	23	100 $\alpha$			21	95 $\alpha$	5 -		21	71 $\alpha$		29 $a$

**Table 3: Regional Groups.**

Region	Recordings made in	Region	Recordings made in
A	Malmö, Trelleborg, Helsingborg, Kristianstad	F	Visby
B	Varberg, Halmstad, Falkenberg	G	Eskilstuna, Stockholm, Västerås, Örebro
C	Göteborg, Lidköping, Uddevalla, Trollhättan, Skövde	H	Borlänge, Särna, Mora, Falun, Idre, Orsa, Ludvika, Saxdalen, Nyhammar
D	Karlskrona, Jönköping, Växjö, Kalmar	I	Hudiksvall, Sundsvall, Örnsköldsvik, Östersund, Skellefteå, Umeå, Boden, Luleå, Piteå
E	Linköping, Norrköping, Västervik		

**Table 4: Distribution of Swedish subjects' productions of nine target xenophones (a-i) across region and category [per cent].**

Region	Tokens	Category			Tokens	Category			Tokens	Category		
		1	2	3		1	2	3		1	2	3
<i>a) Target ɕ occurring in the word Roger</i>				<i>b) Target ɕ occurring in the word James</i>				<i>c) Target ɕ occurring in the word Jackson</i>				
A	36	67 ɕ		33 a	36	22 ɕ		78 i	27	30 ɕ		70 i
B	40	45 ɕ		55 g	40	25 ɕ		75 j	40	45 ɕ		55 j
C	77	23 ɕ		77 a	77	13 ɕ		87 i	74	45 ɕ	1 -	54 i
D	45	41 ɕ		59 g	45	24 ɕ		76 j	45	49 ɕ	2 d	47 j
E	34	41 ɕ		59 a	34	12 ɕ		88 i	34	50 ɕ	3 d	47 i
F	31	29 ɕ		71 g	31	23 ɕ		77 j	32	44 ɕ		56 j
G	75	35 ɕ		65 a	75	29 ɕ		71 i	76	64 ɕ		36 i
H	29	21 ɕ	3 -	76 g	29	21 ɕ		79 j	29	55 ɕ		45 j
I	104	22 ɕ	1 ad	77 a	104	22 ɕ		78 i	105	42 ɕ		58 i
<i>d) Target θ occurring in the word Thriller</i>				<i>e) Target θ occurring in the word Thatcher</i>				<i>f) Target ð occurring in the World</i>				
A	27	37 θ		63 t	37	51 θ		49 t	36	44 ð		56 d
B	40	45 θ		55 t	40	55 θ	3 -	42 t	40	20 ð	2 v	78 d
C	74	53 θ	5 ʃ	40 t	75	49 θ	2 s	49 t	76	41 ð	3 t	54 d
D	45	53 θ	2 -	38 t	44	52 θ		48 t	45	31 ð		69 d
E	34	47 θ	9 ʃ	47 t	33	27 θ	3 c	70 t	31	45 ð	3 v	45 d
F	32	41 θ		59 t	32	41 θ		59 t	30	37 ð	3 h	4 t
G	76	59 θ		41 t	73	56 θ	3 c	41 t	75	40 ð	3 -	57 d
H	29	59 θ		41 t	29	31 θ		69 t	29	48 ð	4 t	48 d
I	105	44 θ		56 t	106	24 θ	2 c	74 t	106	41 ð	3 r	53 d
<i>g) Target ɔ occurring in the word Stone</i>				<i>h) Target æ occurring in the word Maggie</i>				<i>i) Target æ occurring in the word Jackson</i>				
A	36	97 ɔ		3 u	37	86 æ	11 -	3 a	27	74 æ		26 a
B	40	85 ɔ	2 u	5 u	40	92 æ	3 -	5 a	40	68 æ		32 a
C	73	82 ɔ	3 o	5 o;	75	92 æ	3 a	5 a	74	77 æ	4 -	19 a
D	44	89 ɔ	2 o	7 o;	44	96 æ	2 a	2 a	45	78 æ		22 a
E	34	91 ɔ		2 u	33	94 æ	3 -	3 a	34	68 æ		32 a
F	31	84 ɔ	3 -	10 o;	32	81 æ		19 a	32	62 æ		38 a
G	75	92 ɔ		3 o;	73	88 æ		12 a	76	78 æ	1 e	21 a
H	28	94 ɔ		5 u	29	97 æ	3 -		29	86 æ		14 a
I	105	90 ɔ	1 o	7 o;	106	88 æ	1 a	8 a	105	80 æ	2 e	18 a
			1 o	1 u;			1 e					

**2.3. Evaluation**

Three phonetically trained native speakers of Swedish, with an above-average knowledge of English, transcribed the target phones, using a fairly narrow allophonic transcription scheme.

**3. RESULTS**

The results of the production study regarding nine target xenophones, labelled a) through i), are tabulated across gender

(Table 1), age (Table 2) and region (Tables 3 and 4). The different productions are assigned to one of three categories along two dimensions, the “awareness” dimension (to what extent people are aware of the difference between Swedish and English pronunciation), and the “fidelity” dimension (how well they succeed in the production of the foreign sounds). Category 1 corresponds to a high awareness among the subjects coupled with a high capability in rendering a sound close to the one in the source language. Category 2 corresponds to the case where the subjects were apparently aware that something “non-Swedish” would be appropriate, but failed to produce a good approximation. Finally, Category 3 corresponds to full adjustment to Swedish.

### 3.1. Gender Differences

The results in Table 1 indicate that gender does not play an important role in governing the production distribution. The percentage of category 1 productions is slightly larger for female subjects for eight targets out of nine, but this result has not been tested for significance.

### 3.2. Age Differences

The results in Table 2 indicate that age is an important factor across the entire material. This implies that the socio-cultural dimension is of primary importance. Across most of Table 2, the share of productions in category 3 tends to increase with the age of the subjects. In addition to this tendency, Tables 2b, 2d, 2e and 2f also show a tendency where category 3 productions have a minimum around ages 26–35, and then increase also for younger subjects. One particularly interesting result is apparent from Table 2a, where the data is drawn from the example sentence in 2.1 regarding the actor Roger Moore. For ages above 35, around 70 % of the productions fall in category 3, with a steep fall to around 40 % for younger subjects. This distribution is quite distinct from the others in Table 2.

### 3.3. Regional Differences

In Table 4, the same data is tabulated across regions, roughly corresponding to dialect groups suggested by Elert [4]. Dialect regions are not as clear-cut and easily defined as are gender and age, and the dialect groups given by Elert are defined partly for prosodic reasons, something we have yet to look at. Thus, the groups given in Table 3 differ slightly from those of Elert in order to correspond more to the task at hand. The results in Table 4 show quite large but unsystematic differences between the groups.

## 4. DISCUSSION

Gender does not seem to be a primary factor, which is not surprising, given the small, if any, differences between the genders with regard to factors such as educational level.

That age seemingly is of importance is not surprising. The cultural influx probably influences certain age groups more than other age groups. The fact that the youngest subjects exhibit a larger share of category 3 productions in some cases can be explained in the following way: Either their educational level is systematically lower, because of lower age, or the amount of cultural exposure is lower, for partly the same reason, or a combination of these two factors. The particular distribution in

Table 2a (“Roger Moore”) could also have at least two possible tentative explanations. One is simply that “Roger” is also a fairly common Swedish name, at least compared with “James” etc. The other explanation would be centered around the fact that Roger Moore’s career as James Bond during the 70’s would have influenced those age groups frequenting the cinemas during that period. Yet another, almost anecdotal, explanation, is related to a nationally very famous Swedish TV sketch, where two comedians (“Hasse å Tage”) mention “Roger Moore” several times with marked, almost exaggerated, Swedish pronunciation. This points to the fact that the specific items you decide to use in a study such as this, can seriously affect and skew the results.

The fact that there are no apparent systematic regional differences between the groups, as defined in the regional part of this study, can be due to many factors. First, as already mentioned, although there are undoubtedly differences between the phonological systems of different Swedish dialects, there still remains the non-trivial problem of “defining” these dialects. Second, even if it were possible to define such groups, the phonological differences may be of secondary importance, compared to other factors, such as educational level and age, two parameters that are not controlled for in the regional tabulation.

### ACKNOWLEDGMENTS

The authors thank Per Sautermeister for help with the listening task and Johan Boye and Mats Wirén for comments on the (almost) final paper.

### REFERENCES

- [1] Becket, Ralph, Pierrette Boullion, Harry Bratt, Ivan Bretan, Dave Carter, Vassilios Digalakis, Robert Eklund, Horacio Franco, Jaan Kaja, Martin Keegan, Ian Lewin, Bertil Lyberg, David Milward, Leo Neumeyer, Patti Price, Manny Rayner, Patti Price, Per Sautermeister, Fuliang Weng & Mats Wirén. *Spoken Language Translator: Phase Two Report*. Telia Research AB and SRI International, 1997.
- [2] Eklund, Robert & Anders Lindström: Pronunciation in an internationalized society: a multi-dimensional problem considered. Proc. from “Fonetik 96” (Swedish Phonetics Conference), published in *TMH-QPSR* 2, Dept. of Speech, Music and Hearing, KTH, Stockholm, 1996.
- [3] Eklund, Robert & Anders Lindström: How to Handle “Foreign” Sounds in Swedish Text-to-Speech Conversion: Approaching the ‘Xenophone’ problem. *Proc. from The 5th International Conference on Spoken Language Processing*, Vol. 7, pp. 2831–2835, Sydney, 1998.
- [4] Elert, Claes-Christian: Indelning och gränser inom området för den nu talade svenskan — En aktuell dialektografi. Lars-Erik Edlund (ed.) *Kulturgränser — Myt eller verklighet?*, DIABAS 4, skrifter från den dialektgeografiska institutionen för nordiska språk vid Umeå universitet, Umeå 1994.