

[jɑ:mes] or [dʒɛɪmz] or Perhaps Something In-between? Recapping Three Years of Xenophone Studies

Anders Lindström & Robert Eklund
Telia Research AB, S-123 86 Farsta, Sweden

Abstract

This paper summarises work on ‘xenophones’ (foreign sounds) carried out at Telia Research. The inclusion of “foreign” sounds in Swedish is described, as well as their implications on speech recognition and speech synthesis. Results from two earlier studies are summarised and described: the nature of the expansion of what is normally regarded as the Swedish phone set, and the nature of some possible underlying factors.

1. Introduction

Current automatic speech recognition (ASR) and text-to-speech (TTS) systems have attained a level that allow inclusion in every-day applications. It goes without saying, though, that this does not mean that all problems within these two fields are solved. The problem discussed in this paper concerns the inclusion of “foreign” sounds into what is normally regarded as the Swedish phone set. In order to create Swedish ASR/TTS systems, this expansion of the Swedish phone inventory needs to be covered. The problem can be given the following “nut-shell” description: A name or word of foreign origin can be pronounced in Swedish with some of the phonetic/phonemic traits of the source language still preserved, or only slightly altered towards the target (Swedish) phone system. Although full rephonematization *does* occur, it is in effect most often the case that Swedish speakers to at least a certain degree adjust their production system in the pronunciation of some foreign, notably English, items. The phenomenon per se has been acknowledged since days of yore, but although several researchers have discussed the inclusion of “foreign” items into the Swedish phone set, no formal study, to the best of our knowledge, has been carried out that describes the nature of this extension of the Swedish phone set. The status of these sounds of foreign origin is somewhat tricky. On the one hand, they can clearly not be regarded as phonemes in Swedish, nor can they be described as allophones, lacking underlying phonemes in Swedish. On the other hand, they are so established that their use in certain contexts is more or less taken for granted, at least by the bulk of the Swedish population. The phenomenon as such is described by Maddieson (1984), who calls these sounds ‘anomalous’. Since we find that this term is not very satisfactory, we have coined the term **xenophone** to denote such sounds. The problem of accomodating foreign words and names into a given language is far from new, and we have come across Swedish and German works that date back to the 16th century. Slightly more recently, Abelin (1985) discusses the problem of representing pronunciation of words and names of predominantly English origin in *Svensk Ordbok*. Möbius et al. (1997) mention that the German version of the Bell Labs multi-lingual TTS system has been augmented with phones outside the German phone inventory in order to cover French and English items. In Eklund & Lindström (1996) we investigated what English phones Swedish speakers actually used in their speech. This study was based on 35 speakers from Stockholm and 35 speakers from Scania (Skåne). As a result of the study, the Telia Research concatenative speech synthesizer was modified to encompass xenophone polyphones and a large pronunciation dictionary was created. In later work (Eklund & Lindström 1998), we studied the production of 460 subjects from 40 different locations in Sweden, and also presented examples of synthesized speech.

A preliminary discussion of some of the underlying factors behind the expanded phone set was presented in Lindström & Eklund (1999). The aim of the present paper is to summarise our previous research, and also to present it from slightly different angles. We also point at work planned and underway.

2. Method

All our work so far has been carried out based on a production study. By studying production, insight may be gained in several dimensions, beyond a mere description of what English phones affect Swedish speakers' production. Also, even speakers who do not produce English or English-similar phones might indicate an awareness of the difference between the English, "original", pronunciation and a fully rephonematised Swedish rendering of the word/name in question. This provides important information in the "attitude dimension", since it points at the expectations that Swedish speakers (or rather listeners) might have on a Swedish TTS system. A set of twelve sentences was constructed containing the 15 English speech sounds [tʃ, dʒ, ʃ, ʒ, θ, ð, z, ʒ, ɹ, w, aɪ, eɪ, əʊ, ju:, æ]. These sounds were chosen because they were judged to be possible candidates for the processes described, and differ phonetically from Swedish speech sounds to varying degrees. The twelve sentences included names and words of English origin that were deemed to be commonly known, embedded in a Swedish sentence in a natural way. An example is shown below:

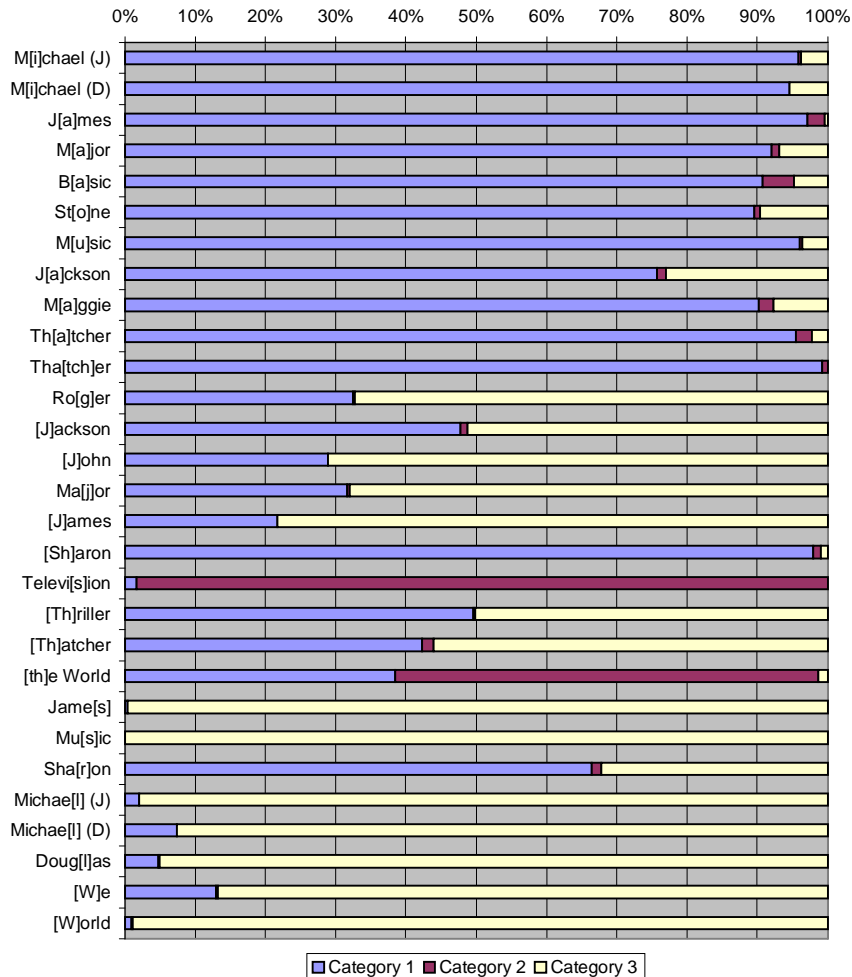
Många har Roger Moore som favorit i rollen som James Bond.
("Many prefer Roger Moore's interpretation of James Bond.")

The sentences were included in a much larger session of linguistic material recorded to train the Telia/SRI Swedish speech recognizer (Becket et al. 1997), and were presented under the heading 'Kändisar' (Celebrities), so it can be assumed that the subjects were unaware of the fact that the object of study was in fact their pronunciation. 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. In this way, all major dialects were covered and a total of 13,343 tokens have been studied. 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 from Two Studies

In this section, the results from two previous studies are summarised. We have shown (Eklund & Lindström 1996) that very few subjects resort to full rephonematization (cf. Table 1), but instead expand their phone repertory considerably when pronouncing items of English origin. From the data, one can also make observations along an "awareness" (that something non-Swedish should be produced) and a "fidelity" (how well it is produced) dimension. In Lindström & Eklund (1999) we took a deeper look at nine target xenophones and tried to correlate the results with regard to gender, age and region (cf. Table 2). Our observations lead to the conclusion that **gender** is of little or no importance, whereas **age** is an important factor, which in turn implies that a socio-cultural dimension is at play here, since it can safely be assumed that certain age groups are more exposed to Anglo-American cultural influx than other age groups. When it comes to **regional differences**, things are not so simple. Dialect regions are not as clear-cut, and the dialect groups given in Elert (1994), that we based our studies on, are defined partly for prosodic reasons, a factor we have yet to look at. Thus, for the nonce, we find that it is premature to draw any conclusions concerning any potential regional effects on the use of xenophones of English origin.

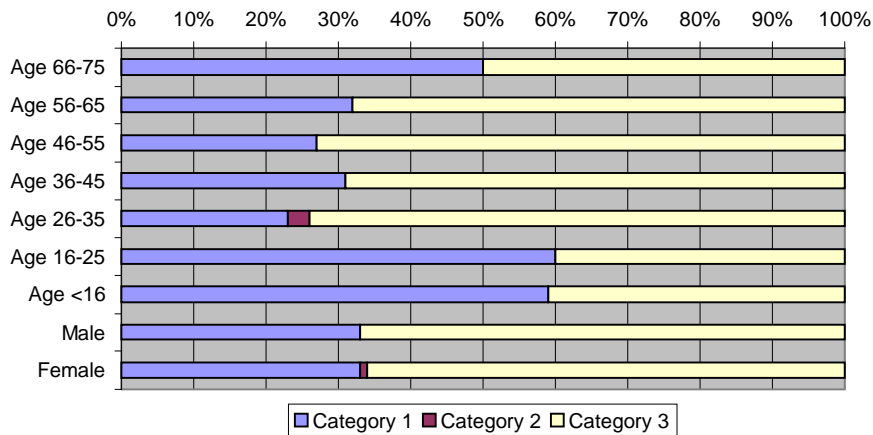
Table 1. Distribution of Swedish subjects' productions of xenophones assigned to three categories along the "awareness" and "fidelity" (i.e. closeness to English) dimensions. Category 1 implies high awareness/high fidelity. Category 2 implies high awareness/low fidelity. Category 3 implies low awareness/low fidelity. Results are shown for 29 target xenophones.



4. Summary And Future Work

Summing up, we have so far looked at the problem per se, carried out a large-scale investigation to establish the nature and distribution of xenophones of Anglo-American origin. We have also recorded a set of xenophone polyphones and used them in the implementation of a concatenative synthesizer as well as carried out a preliminary study with regard some of the possible underlying factors at play. What remains to be done? Clearly, a much more thorough study of the underlying factors is needed, in particular with regard to what rôle the underlying phonologies of different regions may have in the realisation of xenophones. Another thing of interest would be to study the inclusion of xenophones from languages other than English, but in order to do that we would need to record new data, since our present corpus only contains but a few non-English sounds. Planned work includes an evaluation of our TTS system, to see what kind of effect the inclusion of xenophones would have on listeners.

Table 2. Distribution of Swedish subjects' productions of the target xenophone [dʒ] in the name "Roger" assigned to three categories along the "awareness" and "fidelity" dimensions (cf. Table 1 caption). Results are shown for the parameters age and gender.



Acknowledgements

We would like to thank Olle Engstrand for pointing out Maddieson's work to us, Per Sautermeister for help with the listening task and Mats Wirén for fabulous proof-reading.

References

- Abelin, Åsa. 1985. Om uttalsmarkering och uttalsregler i svensk ordbok. *Rapporter från Språkdata (21)*, Gothenburg University, Dept. of Computational Linguistics, Gothenburg.
- 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. 1997. *Spoken Language Translator: Phase Two Report*. Telia Research AB and SRI International.
- Eklund, Robert & Anders Lindström. 1998. 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, 2831–2835, Sydney.
- Eklund, Robert & Anders Lindström. 1996. 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.
- Elert, Claes-Christian. 1994. 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.
- Lindström, Anders & Robert Eklund. 1999. Xenophones Revisted: Linguistic and other underlying factors affecting the pronunciation of foreign items in Swedish. *Proc. International Conference of Phonetic Sciences 1999, Berkeley, California*. [Forthcoming.]
- Maddieson, Ian. 1964. *Patterns of sounds*, Cambridge University Press.
- Möbius, Berndt, Richard Sproat, Jan P.H.van Santen & Joseph P. Olive. 1997. The Bell Labs German Text-to-Speech System: An Overview. *Proc. ESCA. Eurospeech97, Rhodes, Greece*, ISSN 1018–4071, 2443–2446.