Deliberate mispronunciation in EFL e-dictionaries: integrating PDI with TTS

Włodzimierz Sobkowiak

Abstract

The Phonetic Difficulty Index (PDI) is a quantitative/qualitative measure of word pronouncing difficulty to L1 learners of a given L2. Specifically, in its current implementation (see http://ifa.amu.edu.pl/~swlodek/public.htm for bibliography), it assigns numerical (0-10 range) and difficulty (57 pronouncing problems) Polglish-sensitive tags to an English word-list or text. The range of applications of the current version of PDI extends from evaluation of pedagogical materials, such as texts, word-lists, dictionaries, etc., in terms of phonetic difficulty, to generation of word-lists meeting user-specified phonetic criteria for teaching, learning, testing and materials preparation.

One application of PDI which has not so far been considered is in modeling learners' pronunciation of English lexical items through deliberately mispronouncing e-dictionary entries in ways characteristic of the given L1, in this case – Polish, or, more accurately, Polglish, i.e. the Polish-English interlanguage of Polish learners of English as a foreign language (EFL). PDI identifies for each lexical entry in an e-dictionary expected Polglish mispronunciations, generates a mispronounced phonetic representation in the orthographic or transcriptional form, and passes it on to the TTS module for conversion into audio. The model and the mispronunciation can now be audially produced on the fly, with no need for prior recording with human speakers. The exact mispronunciation can be controlled down to minute phonetic detail to suit the proficiency level and phonetic idiosyncracies of the user (as constructed by the user-modeling component of the dictionary) or the pedagogical agenda of the learner/teacher (for example, the amount of final obstruent voicing in English can be exaggerated).

1. TTS for EFL?

As I wrote at the beginning of this decade, "not only is (top-quality) synthesized speech intelligible and natural, but it can also actually function as a model of pronunciation. For example, *Filoglossia*, a CALL package with (modern) Greek as a foreign language, already employs TTS synthesis: http://www.ilsp.gr/filoglossia_plus_eng.html, and *WordPilot* from http://www.compulang.com, also has this feature" (Sobkowiak 2003). To take another example of TTS in CALL, "ScanSoft's *RealSpeak*™ *Word* uses a ground-breaking approach to text-to-speech to achieve superb quality speech output from a dictionary of words and idioms, allowing language learners to hear how words should be accurately pronounced" (http://www.nuance.com/realspeak/word/). Finally, TTS was used in the highly acclaimed electronic dictionary *Grote Van Dale* as early as 1999 in the form of "diphone-based generated speech, taking as its input the phonetic transcription" of the ¼ million headwords (Geeraerts 2000:77).

There is little research to demonstrate the actual pedagogical superiority of synthesized speech over human recordings (or at least no inferiority), but favourable opinions are easy to find: "TTS applications can render many benefits to EFL students while making teachers job easier. I have found that my students have improved their pronunciation since I started using

them in my classes, not to mention that they have become more autonomous" (González 2007). All in all, it seems that TTS is here to stay for all kinds of EFL applications, whether CALL in the strict sense or e-dictionaries for learners. Considering the commercial (smaller cost) and technological (little storage) factors which additionally speak in its favour, it is a safe bet that the scope of TTS will grow dynamically. Some examples of English and Polish synthesized speech by ScanSoft and IvoSoftware TTS systems (web demo downloads) can be listened to here: http://ifa.amu.edu.pl/~swlodek/PLM07.htm.

In the 2003 paper mentioned above I also noticed that "simulating a foreign accent of English by computer for didactic purposes is not a new idea. In 1997 Hyouk-Keun Kim created his *Korean Accented English Pronunciation Simulator* (http://english-korean.com/projects/kaeps/index.html), rightly noticing that "Most adult ESL/EFL learners [...] do not recognize the problems of their English pronunciation", and that it might be a good idea to demonstrate these under computer control. Eventually a rule-based KAEPS system was set up, simulating "three types of English pronunciations in the IPA symbols: 1) a phoneme-based English pronunciation, 2) a desirable allophone-based American English pronunciation, and 3) a possible Korean accented English pronunciation". While Kim's system has never advanced beyond accented graphemic (i.e. IPA) representation, it would be easy enough to attach the IPA-to-speech engine to it. After all, most TTS systems use phonetic transcription at some stage of the synthesis process [...] An L1-sensitive TTS system would be able to dynamically adjust its parameters to realistically simulate spoken Polglish at these various stages of proficiency" (Sobkowiak 2003). It is this last idea which will be tested in the following sections.

After a brief introduction of the Phonetic Difficulty Index and the overview of my attempts to validate and calibrate it empirically over a selection of English words, I will demostrate samples of synthesized Polglish renditions of these words in two variants, compared with a model native-like pronunciation. Due to the nature of the current medium, the actual sound samples are stored on a server to be retrieved as part of the multimedia presentation of this paper held at the PLM2007 conference in Gniezno on September 14th 2007: http://ifa.amu.edu.pl/~swlodek/PLM07.htm. As far as it was feasible, the phonetic transcription below matches exactly the synthesized pronunciation, for readers with no (immediate) internet access.

2. PDI – predicted phonetic difficulty

The Phonetic Difficulty Index (PDI) is a quantitative/qualitative measure of word pronouncing difficulty to L1 learners of a given L2. Specifically, in its current implementation (see http://ifa.amu.edu.pl/~swlodek/public.htm for bibliography), it assigns numerical (0-10 range) and difficulty (57 pronouncing problems) Polglish-sensitive tags to an English word-list or text. Table 1. holds a presentation of some of the tags with their associated symbols, likely phonetic problems and Polglish errors.

Table 1. Example PDI codes with likely Polglish errors

phonetic difficulty code (PDI code)	likely Polglish error
b – <ur> in word	schwa, r?
s – <age_> in stem and not eɪdʒ_	eɪdʒ
w – <ey_> in stem and not eɪ_	eɪ
B – eə	j breaking, smoothing, schwa
E – ʌ	Polish a

J – short schwa	schwa quality
L – voiced apico-dental	d, z, v
N – final voiced obstruent	devoicing
O – pre-voiced dɪs- or mɪs-	z
Q – vowel overnasalization	Polish-like fully nasal vowels
V – glottal fricative h	Polish velar fricative x
X – word-final syllabic sonorants	schwa insertion
2 – more than 5 syllables	stress and articulation problems
7 – <ary_>/<ory_>/<ery_> in bisyllabic-plus stems	stress, vowel quality
9 – proper noun	graphophonemically irregular

PDI was originally conceived on an entirely intuitive basis, but backed up by many years of research and teaching experience in the area of Polglish pronunciation. The prima facie feasibility of indicated problems and errors had to be checked empirically, of course, on a sample of Polish learners of English. This was done in two stages. First, a large group of advanced Polglish learners were asked about the perceived phonetic difficulty of a selection of English words. Second, a smaller group of beginners/intermediates were actually recorded while saying the same words in carrier sentences. The following sections briefly report on these experiments. Interested readers are referred to the indicated papers for extended analyses and discussions.

3. The 20 words – perceived phonetic difficulty (Sobkowiak 2000)

In the study originally written in 2000, and published on the web (http://ifa.amu.edu.pl/~swlodek/diffind2.pdf), 208 Polish students of English philology filled in a questionnaire concerning the perceived phonetic difficulty of twenty English words stratified on two dimensions: (a) a-priori rule-based assessment of phonetic difficulty and (b) word frequency rank. The words were, alphabetically: *almost, appear, author, awkward, belief, carry, coloured, debate, defect, dissolve, kingdom, mother, oblige, relax, server, southern, survive, taxi, tired, youngster*. A two-way ANOVA confirmed the significance of both main effects and their synergetic interaction, i.e. the perceived difficulty rating was affected by both the word's rule-based difficulty index and its frequency independently, as well as by their product. Thus, high-PDI and low-freq words obtained higher difficulty ratings than low-PDI and high-freq ones.

While providing some empirical support for my intuitive assessment of what is pronunciation-wise difficult to Polglish learners, this study did not actually retrieve any readings, with their unavoidably associated errors, from the respondents, to compare them with the PDI predictions. An experiment like this was conducted six years later.

4. The 20 words – attested phonetic difficulty (Sobkowiak and Ferlacka 2006)

In a recent study, Sobkowiak and Ferlacka (2006) tried to "calibrate the Phonetic Difficulty Index" empirically. The twenty English words of Sobkowiak 2000 were read in carrier sentences by 38 Polish learners of English aged 17-18. The sentences were definitions taken from the *Macmillan English Dictionary for Advanced Learners'*, first edition (MEDAL1). A total of 617 word-readings yielded 1211 errors, for the grand mean of 1.96 phonetic errors per reader per word. Predictably, the PDI phonolapsological intuitions turned out to be taken from the academic EFL context, and as such showed little correlation with the actual errors made by Polish schoolchildren. It is obvious that, to be a useful tool over a variety of Polglish

contexts, PDI would have to be made sensitive to many variables, such as – first of all – learner proficiency, age, preferred learning strategies and styles, and many others. The creation of a flexible and parameter-adjustable PDI like this would require a major project, and at this stage remains a plan for the future.

The recordings did yield a lot of material for further work with PDI, however. In particular, it was now possible to confront the intuitive phonolapsological predictions of the present author and the student sample of 2000 with the actually commited pronunciation errors. Some of the latter are demonstrated in a sample of 5 sentences/definitions from one learner (keywords bolded) can be retrieved in audio form from http://ifa.amu.edu.pl/~swlodek/PLM07.htm:

- *Pallbearer - someone who helps to **carry** a coffin at a funeral..*
- *Melt - if you melt into or against someone you **relax** as they hold you close in a romantic way.*
- *Hail - to signal a **taxi** or bus so that it stops for you.*
- ***Defect** - a fault in someone or something.*

The most commonly attested keyword errors out of the total 1211 were used to program the TTS engine to generate deliberately mispronounced 'accented' Polglish speech to be presented in the next section.

5. PDI-based TTS in EFL e-dictionaries

EFL learners often have problems perceiving the phonetic difference between their 'accented' pronunciation of a given lexical item and the native speaker model (see e.g. Baker & Trofimovich 2006 and the references therein). The modern techniques offered by contemporary e-dictionaries of allowing the learner to record his/her pronunciation to compare audially or visually with the recorded native model may not work in this situation. Demonstrating an actual Polglish mispronunciation of the word alongside the correct native version, spoken in the same voice and keeping all the other phonetic variables constant, might be more useful.

This has not been feasible so far in e-dictionaries: no professional native English speaker could be expected to persuasively mimic Polglish mispronunciation, not to mention the cost of such a procedure. With PDI and Text-to-Speech synthesis (TTS) we have the two key technologies to make such believable mispronunciations possible.

Speech synthesis mechanisms can be tweaked to produce human-sounding audio output of an arbitrary phonemic/allophonic string, including deliberate mispronunciations illustrating selected interlanguage features. These can then be offered to the EFL e-dictionary user, suitably adjusted to their needs and wants. In Table 2. two mispronunciation versions are given, one containing the PDI-predicted error(s), the other showing the most common of the actually attested errors in Sobkowiak & Ferlacka study of 2006 (the actual transcription coding was made by Ferlacka). I am grateful to Mr. Dawid Pietrala for tweaking the Festival speech synthesis system to obtain phonolapsologically accented 'Polglish' speech. All sounds in the 'error' columns are to be interpreted as having basically Polish qualities, e.g. /a/ is Polish /a/, similarly for other vowels and consonants, e.g. /dʒ/ or /tʃ/. The audio files can be listened to at: http://ifa.amu.edu.pl/~swlodek/PLM07.htm.

Table 2. 20 English words: correct and in two Polglish versions

	word	correct (Festival)	PDI-predicted error	most common attested error	comment
1.	*almost*	'oːlməʊst	'ælməʊst	'alməst	
2.	*appear*	ə'pɪə	ə'pijə	a'pir	dipthong breaking predicted; spelling pronunciation attested
3.	*author*	'oːθə	'aʊtə	'aʊtor	spelling pronunciation attested
4.	*awkward*	'oːkwəd	'aʊkwət	'afkwart	spelling pronunciation attested
5.	*belief*	bɪ'lif	---	be'lif	no error predicted; spelling pronunciation attested
6.	*carry*	'kærɪ	---	'keri	no error predicted; spelling pronunciation attested
7.	*coloured*	'kʌləd	'kaləʊrt	'koloret	spelling pronunciation attested
8.	*debate*	dɪ'beɪt	---	de'beɪt	no error predicted; spelling pronunciation attested
9.	*defect*	'difekt	---	'defekt	no error predicted; spelling pronunciation attested
10.	*dissolve*	dɪ'zɔlv	dɪ'zolf	dɪ'solf	final devoicing predicted; spelling pronunciation attested
11.	*kingdom*	'kɪŋdəm	'kɪŋgdom	'kiŋgdom	spelling pronunciation attested
12.	*mother*	'mʌðə	'madə	'mader	
13.	*oblige*	ə'blaɪdʒ	o'blaɪtʃ	o'blik	spelling pronunciation attested
14.	*relax*	rɪ'læks	---	'relaks	no error predicted; spelling pronunciation and interference attested
15.	*server*	'səːvə	'sevə	'səːrver	long schwa attested: error of coding or familiarity effect?
16.	*southern*	'sʌðən	'saʊzən	'saʊtern	two of the common three Polglish substitutions
17.	*survive*	sə'vaɪv	sə'vaɪf	sur'vaɪf	spelling pronunciation attested
18.	*taxi*	'tæksɪ	---	'taksi	no error predicted; spelling pronunciation and interference attested
19.	*tired*	'taɪəd	'taɪət	'taɪret	spelling pronunciation attested
20.	*youngster*	'jʌŋstə	'jaŋkstə	'joŋkster	spelling pronunciation attested

A number of observations are in order at this point. Notice first that there are words for which PDI in its present implementation makes no error predictions, such as *belief* or *relax*. This means that at the advanced proficiency (academic) level these words are relatively easy pronunciation-wise. This was confirmed in Sobkowiak 2000, where these words were student-rated as the easiest of the whole group. As it turned out, however, the less phonetically proficient sample of learners had pronouncing problems with these words as well. These mispronunciations were duly coded into the TTS representation.

Second, this time completely predictably, many of the attested mispronunciations are clear cases of interference from Polish, combined with spelling pronunciation. Some actually bordered on malapropisms, such as /'belfel/ for *belief*, but these were not the most common renditions, so they do not appear in the table.

Third, there is only one word where the predicted and (most commonly) attested errors match exactly, namely *kingdom*. For all other words the predictions, based as they were on highly advanced Polglish, differ from the most common attested erroneous productions in various ways. This does not mean, of course, that there was mismatch in every case. Yet, the actual mispronunciations matching the predictions were simply not the most frequent. Obviously, with proper tweaking, the TTS system can produce any arbitrary mispronunciation, and it would be easy enough to, say, take one word out of the sample and demonstrate its full range of predicted/attested mispronunciations in audio form.

Fourth, some of the synthesized strings sound more authentic/believable/natural/human than others, of course, as can be heard in http://ifa.amu.edu.pl/~swlodek/PLM07.htm. This is to be expected, considering that the TTS system used for this project was the freely available Festival (http://www.cstr.ed.ac.uk/projects/festival/). Even the best TTS engines now available slip occasionally, so that in some rare cases it is actually possible to tell synthesized speech from human recording. This, however, does not undermine the usefulness of TTS in the pedagogical EFL context.

So, once it is agreed that deliberate mispronunciation by a TTS system can and should be implemented in CALL and electronic dictionaries for learners, the time may not be too distant when such functionalities will be routinely built into the user interface, somewhat like in Figure 1., illustrating the pronunciation drill component in e-Longman.

Figure 1. LDOCE3 audio playback screen: "Play Polglish pronunciation"?

6. Synthesizing examples/definitions?

There is no technical reason why TTS should stop at isolated keywords, of course. As I pointed out in my 2006 book, "the first ideas to audiolize EFL e-dictionary examples (but not definitions!), for instance, appeared long ago. In an overview of electronic learners' dictionaries, published in 1997, Perry dreamed: <<Not only could the pronunciation of headwords and derivatives be given, but also the use of sound could be extended to cover some of the usage examples >>. With the recent introduction of recorded audio example sentences in LDOCE4 (http://www.longman.com/ldoce/about.html) there may be a distant glimmer of hope" (Sobkowiak 2006:81). But, rather than actually recording the sentences, TTS could be used, thus potentially rendering all of the material contained in the dictionary in an acoustic form for the benefit of the learner. Some examples of ScanSoft-synthesized MEDAL1 definitions are provided at http://ifa.amu.edu.pl/~swlodek/PLM07.htm. It will be appreciated that the quality of the synthesized voice is quite adequate for pedagogical purposes.

- ***Youngster*** *- a child or a young person.*
- *Run-down - so* ***tired*** *that you do not feel well.*
- ***Oblige*** *– to help someone by doing something that they have asked you to do.*

- ***Debate*** *- if people debate a subject, they discuss it formally before making a decision, usually by voting.*
- *Double cream - thick cream that becomes* ***almost*** *solid when you mix it quickly.*
- *Hail - to signal a* ***taxi*** *or bus so that it stops for you.*
- *Pallbearer - someone who helps to* ***carry*** *a coffin at a funeral.*

7. Conclusion

The field is now ready for testing the many hypotheses concerning the technological and didactic feasibility of applying TTS in ways envisaged in this contribution. The PDI is by no means the central issue here. It is of course possible to simply synthesize any attested or thinkable mispronunciation, whether it be rule-governed or not. The attraction and challenge of PDI in the phonolexicographic context, however, is rather obvious. Automatic coding of keywords, examples and/or definitions for pronouncing errors transforms the e-dictionary into a very powerful pronunciation teaching/learning resource, i.e. something that it has so far failed to be, despite all its rich phonetic content. Some form of a phonetic difficulty index will be unavoidable in the process.

Bibliography

Baker,W. & P.Trochimovich. 2006. "Perceptual paths to accurate production of L2 vowels: The role of individual differences". *IRAL 44*.231-50.

Geeraerts,D. 2000. "Adding electronic value. The electronic version of the Grote Van Dale". In U.Heid et al. (eds). 2000. *Euralex 2000 proceedings*. Stuttgart: IMS. 75-84.

González,D. 2007. "Text-to-speech applications used in EFL contexts to enhance pronunciation". *TESL-EJ 11.2*. [access 11th Sep 2007]

Sobkowiak,W. 2000. "Rule-based and empirical rating of perceived phonetic difficulty of English words according to Polish learners: does frequency matter?" [published electronically: http://ifa.amu.edu.pl/~swlodek/diffind2.pdf]

Sobkowiak,W. 2003. "TTS in EFL CALL - some pedagogical considerations". *Teaching English with Technology 3.4*. [http://www.iatefl.org.pl/call/j_article15.htm]

Sobkowiak,W. 2006. *Phonetics of EFL dictionary definitions*. Poznań: Wydawnictwo Poznańskie.

Sobkowiak,W. & W.Ferlacka. 2006. "Calibrating the Phonetic Difficulty Index". In W.Sobkowiak & E.Waniek-Klimczak (eds). 2006. *Dydaktyka fonetyki języka obcego w Polsce*. Konin: PWSZ w Koninie. 173-187. [Proceedings of the *Phonetics in FLT 6 Conference* in Mikorzyn, 8-10.5.2006]