

Phonetic font images for the web

J. Kevin Varden, Tsukuba University

Abstract

This paper introduces a set of web-ready phonetic symbol images created from the free phonetic fonts available from SIL International (formerly the Summer Institute of Linguistics). By accessing the image pages with a web browser, the user can either copy those symbols they need to use or download the entire image file package. The package contains over 150 symbols available in each of the 3 type faces (Doulos, Manuscript and Sophia) for each of 4 font sizes (10pt, 12pt, 14pt and 18pt). The paper also briefly discusses how to embed the image files in html pages, and discusses alternatives for phonetic font display in web pages.

1. Introduction

As a phonologist/phonetician who has worked on several web projects that utilize IPA renderings of sounds, I began searching for a way to display IPA symbols in my web pages. A search of the web led me to the IPA Association's discussion of phonetic fonts for the web,¹ which confirmed my suspicions: there is still no elegant solution to the problem of displaying phonetic fonts in web pages. The IPA page suggests 3 alternatives as viable options: 1) use Adobe Acrobat to save a file containing embedded phonetic font symbols as a pdf file, which can then be displayed as a web page; 2) embed image files in one's html pages, or 3) utilize font style tags in a page's html to define the IPA symbol's font, as well as instruct the web user to install the required IPA font on their machine.

Since time for most researchers and instructions is quite limited, purchasing a new piece of software and learning how to use it (option 1 above) or learning how to author html tags (option 3) is simply not worth the time and effort required to display a limited number of phonetic symbols. Using image files in html pages (option 2 above) would therefore seem to be the best option for most researchers. Unfortunately, I could not connect to either of the sources of image files listed on the IPA's page, nor find a good set at any other site. Hence it appears that the need for a full set of IPA symbol images still exists.

2. The SIL IPA font image files

In order to fill this gap, a set of IPA GIF images have been created from the SIL's freeware phonetic fonts with the permission of the SIL.² In order to facilitate the matching of the image GIFs with the typeface an author is using on a given web page, image files have been created for all 3 typefaces in 10, 12, 14, and 18 points sizes. All GIFs are transparencies; that is, the background of the characters is not specified; the

¹ <http://www2.arts.gla.ac.uk/IPA/ipahtml.html>

² For those not familiar with the SIL IPA fonts, there are three TrueType typefaces available: Doulos and Sophia, roughly equivalent to Times and Helvetica on the Macintosh platform, respectively, and Manuscript, a mono-width font. The freeware fonts are a subset of the SIL Encore Font System, a fuller, richer TrueType font set that is available for US\$100 (individual license) through the SIL. See <http://www.sil.org/computing/fonts/lang/IPA.html> and <http://www.sil.org/computing/fonts/Encore.html>, respectively, for more information.

background of the web page will show around the character itself. In addition, each character appears in the same horizontal position roughly in the center of the image in order to allow concatenation when rendering words or phrases.

The GIF images are currently available for download from the author's web site³. They can be downloaded directly from the index page, or as a single package. In addition to the 131 base symbols listed in the SIL Encore IPA93 Fonts Access Code Chart (hereafter 'SIL guide', included with the package), several common diacritic-sonorants have been included: the umlaut, acute and grave accent versions of the major vowels, voiceless versions of the major vowels and sonorants (n, r, l, w), as well as the nasalized version of the major vowels and nasals, the length mark, and the inclusion of primary and secondary IPA stress symbols.

3. Naming conventions

All images included in html code must have a name, else the image cannot be specified (see §4 for examples). At first blush, my intention was to name the image files follow the naming of the phonetic symbols detailed in Pullman & Ladusaw (1996; hereafter P&L), each of which is "...a convenient, though not necessarily official, working name...". However, this would unnecessarily penalize those who do not have a copy of P&L available. It was therefore decided to utilize the names given in the SIL guide instead even though it contains inconsistencies. The greatest departure of these names from the names used in P&L from those used in the SIL guide is that P&L have taken the time to standardize the names of some of the symbols. For instance, it makes little sense to call an 'i' without a dot an 'undotted i' but a 'j' without a dot a 'dotless j'. The SIL guide still contains such naming inconsistencies.

Simply using the names of the symbols as they appear in the SIL guide, however, would result in some quite long image file names. For example, using the entire given name of the character in the name of the GIF would result in names like "hooktopSmallCapitalG.gif" and "smallCapitalOELigature.gif".

To alleviate the file name length somewhat, the following abbreviation conventions have been adopted for the image files under discussion:

Bar:	barred
Curly:	curly tail
Curs:	cursive
Dbl:	double
HkTp:	hooktop
Inv:	inverted
Lig:	digraph (P&L ligature)
LfTl:	left tail
Rev:	reversed
RtTl:	right-tail
SmCap:	small capital
SupScr:	superscript
Turn:	turned
Undot:	undotted; dotless

³ <http://www.varden.com>, in the Research section.

The general goal in naming the images was to preserve 3-4 characters of the symbol's name, either the first syllable or an initial heavy cluster. These can act as a mnemonic device for remembering the file names when working in the html code. When it was not possible to maintain a transparent portion of the name, an alternative was sought.

Several other considerations came into play. Since the images themselves need to be placed on a server, all hyphens have been removed from the names; i.e. the names have been normalized so as not to prevent problems with those servers with strict file naming conventions. Also, in order to shorten names, the label 'lower-case' has not been included in image file names ('lower-case' is the underspecified default form of a character, if you will). Unrelated to servers, but important for readability, each 'morpheme' of the name begins with a capital letter, and all other letters are in small type; e.g. "A" for "lower-case a", "HkTpBarUndotJ.gif" for "hooktop barred dotless (undotted) j". While the latter example is not particularly short in and of itself, it does seem an improvement over "HooktopBarredUndottedJ.gif". A readable convention like this will hopefully facilitate character identification and minimize errors when naming image files in html code.

Another point to be noted is that there are minor differences in the ordering of the characters in the SIL guide and P&L. For example, whereas in the SIL guide the handwritten lower-case Roman v is called 'cursive v' and listed between 'barred u' and 'upsilon', in Pullum & Ladusaw (1996) is called 'script v' and listed after 'lower-case v'. It is hoped that these differences should be readily understandable to users of P&L.

One final note on diacritics: when entering text using the SIL fonts, secondary characteristics of the sounds in question are generated by adding diacritics to the primary character; e.g. to type an nasalized 'a', one types 'a' followed by 'shift-0'. This is not possible with image files, but would require extra html encoding to properly superimpose the diacritic image over the primary vowel image. Since many diacritics are commonly used, the following diacritic version of the primary vowels and sonorant consonants have been included, despite the fact that the approved IPA usage of the acute and grave accents are for marking tones. They are provided to allow the user to choose between the IPA method of marking tone and the common use of accent marks. Superscript symbols marking secondary articulations, e.g. the labial marker ^w, denoted by the prefix 'SupScr', have been included since they do not require overlap to display properly.

- Acute:** acute accent (primary stress) mark
- Dev:** devoiced
- Grave:** grave accent (secondary or tertiary stress) mark
- Nas:** nasalized (a tilde over the vowel)
- Um:** umlaut
- Wedge:** wedge symbol (palatalization in American use)

These specifications have been included at the beginning of the file name; i.e. "NasA" indicates a nasalized 'a', "GraveO" indicates an 'o' with a grave accent.

Finally, it will be noted that the images are available in all 3 font faces and 4 point sizes, and so are arranged in 12 folders/directories. Because of this, the name of the image file "HkTpHeng.gif" can actually refer to any one of 12 image files: 10pt

Doulos, 14pt Manuscript, etc. The specification of the full file name will be covered in the next section.

4. Using the image files

4.1 Downloading the full package

The simplest way to obtain the SIL IPA image files is to download the entire package. It is available via the author's web site (see footnote 3). Once this file is downloaded, it will need to be expanded with Stuffit Expander™, a free utility available from Aladdin⁴ for both Macintosh and Windows. If one's browser settings are not set to expand downloaded archives automatically, one can opt to open Stuffit Expander, select the IPA image archive, and then select the save destination.

This should result in a parent folder/directory entitled 'IPA images'. In it, one should find an html page entitled 'ipa.html', another page entitled 'sampler.html', and three folders/subdirectories: Doulos, Manuscript and Sophia. In each of the 3 folders/subdirectories, there should be a folder/directory for each of the 4 point sizes of each image file: 10pt, 12pt, 14pt and 18pt. Inside each of these should be the image files for all symbols currently included.

4.2 Copying the files

For those who only need a small number of symbols, the images available at the URL cited above should be readily accessible via a user's web browser. Once at the specified URL, a link entitled "IPA images" will present itself. Navigation through the links in that and subsequent pages will take the browser to the desired font face and point size.

Almost all current browsers will allow the user to drag an image file to a window or icon in order to copy it. Failing that ability, they will let the user save an image file by clicking on it and selecting the "Save image as" or equivalent command from the menu that pops up/down. Select the desired location in the dialog, and save the file. The selected image file will then be available for use without the need for unstuffing with Stuffit Expander discussed above.

4.3 Sample html code

The following sample html code is for the benefit of those who are not using a WYSIWYG (what you see is what you get) html software package such as GoLive™ from Adobe, or who simply prefer to handle the html coding themselves. Although this is well-known by those who have worked with web page authoring before, a brief explanation seems in order.

Html web pages display images by loading the image from a location that you specify in your html page. For example, to display a file called 'ImageFile.gif',

⁴ On the Macintosh platform: http://www.aladdinsys.com/expander/expander_mac_login.html; on Windows: http://www.aladdinsys.com/expander/expander_win_login.html. Registration is required.

located in the Macintosh directory called 'images', you would reference it with the following html code:

```

```

The 'img' portion of the tag tells the browser it is an image file; the 'src' specifies the source of the image file. Including this tag in a page's html code would result in the image file "ImageFile.gif" being displayed in the web browser's window, along with the surrounding text, etc. that was specified in the page's code.

Having said that, to use the IPA image files in a web page need to be specify both the image file name and location. Assuming that the entire IPA image package has been downloaded and its folder/directory structure left in tact, the following html code would display the IPA image files for the English word 'bang' in the SIL IPA Doulos font face:

```
  
  
5
```

These 3 tags would result in the following image series being displayed in the browser's window:

bæŋ

To change the font face of the image files, one would only need change the name of the uppermost folder/directory; e.g. to utilize the Manuscript font face, replace 'Doulos' with 'Manuscript' in all 3 tags above. Likewise, to change the font size, one would only need change the point size in the tags above; e.g. replace '12' with '18'.

In this way, copying the image files to the html page being authored (alternatively, utilizing the names and locations of the image files when writing the tags) would result in the proper IPA strings being displayed, without resorting to any further html coding.

5. Alternatives

As noted in the introduction, there are two alternatives to using image files currently available. These are both discussed briefly below.

5.1 pdf files

The first alternative discussed on the IPA page is to use Adobe Acrobat⁶ to save the file you are working on as a pdf file. As long as the author of the page is using a phonetic font installed on their machine, Acrobat can embed the required characters into the pdf document when it saves it as a pdf file. The pdf file can then be made available for download or viewing in a browser via the web, as long as the user has downloaded the freeware Adobe Reader and/or has installed the Reader plug-in for their browser. The latest version of Acrobat vastly improves for web authoring and management; see Adobe's site for more information.

⁵ A browser will not care if these three references, or html "tags", are on the same line of text or not.

⁶ <http://www.adobe.com/products/acrobat/>

A less expensive alternative to Adobe Acrobat is PrintToPDF⁷, a utility by James Walker that will save a file from almost any application to a (somewhat simple) pdf file. While results vary, it produces a high-quality pdf file from within many applications, and so may be worth the time to investigate.

5.2 Specifying fonts in the html code

The other alternative discussed on the IPA page and elsewhere is specify the font to use when displaying a string of text. This can be done in two ways: 1) use html tags, such as those used in §4 above to reference the IPA image files; or 2) use style sheets to specify a font family, and then use a font tag to specify that a string of text be displayed in that font family. Both of these can be done quite easily in most current html authoring software packages (see your documentation for details).

Of the two options, specifying the font face to be used for a string of text is the most straightforward, and will probably suffice for most IPA symbol display. A sample html coding is desirable for those working in text editors. To set the font face that text will be presented in, one uses a font tag () containing the 'face' attribute. For example, to specify that some text be presented in the Sophia font face, one would use the following tags:

```
<font face="SILSophia IPA93">this text is in Sophia</font>
```

This would result in the text being displayed in the appropriate font, e.g.:

this text is in fophia

Note that in the case of a font tag such as this, the tag is actually one of a pair (termed "opening tag" and "closing tag"). If the closing tag is not included in the html code for this and most font and paragraph tags, the style specified by the opening tag will be applied from that point on in the document; the browser will not know when to stop using that style.

In addition to specify the display font for the text, it is necessary to inform the person browsing the page that they will need to download and install the specified font on their machine. If the specified font is not available, the web browser will default to the font specified in the browser preferences; in the case of a phonetic font, this will invariably result in garbage being displayed. As a courtesy to the reader, it behooves a web page author to use a font such as the SIL IPA fonts that are both free and available for all popular platforms. As long as the web author sticks with the SIL offerings, there will be no compatibility or copyright issues.

5.3 Unicode

An elegant solution to the problem of displaying any complex text, including IPA symbols, is Unicode. Unicode is a browser encoding that has the ability to display thousands of individual characters rather than the 255 characters that are standard internationally (plus the next 255 that vary by platform and font). Each language is assigned a section of the Unicode character values; e.g. Japanese kanji characters are located in one section of codes, Bulgarian Cyrillic script in another. Instead of

⁷ <http://www.jwwalker.com/pages/pdf.html>

specifying a font face tag as done in the sample above, the html author includes character codes in the html, and directs the user to utilize the font encoding “Unicode” when viewing the document. The browser will then display the Unicode character code using the Unicode font that the user has specified in their browser settings.

To view a sample of a Unicode rendering of IPA symbols, the reader is directed to the SAMPA (Speech Assessment Methods Phonetic Alphabet) home page (see the References for the URL), maintained by John Wells at the University College London. It contains both examples and a good deal of discussion of both IPA and Unicode display, including Unicode values for most of the phonetic symbols.

As a final note, for those who are interested in Unicode and its possibilities, an excellent source of information for both Unicode fonts and utilities is available from Alan Wood’s Unicode site; the image is in the References section.

In this day and age, when so many are making their work available on the web, it would serve us all to be able to include accurate phonetic renderings of our data and conclusions. Hopefully the IPA image file set discussed here will play a small part in that process.

References

Pullman, Geoffrey K. and William A. Ladusaw. 1996. *Phonetic Symbol Guide 2nd Ed.* Chicago: University of Chicago Press.

Alan Wood’s Unicode page: <http://www.hclrss.demon.co.uk/unicode/index.html>

IPA Symbols, Unicode and the Web: <http://www2.arts.gla.ac.uk/IPA/ipahtml.html>

SAMPA: <http://www.phon.ucl.ac.uk/home/sampa/home.htm>