

A Conversion of Public Indic Fonts from METAFONT into Type 1 Format with T_EXTRACE

Karel Píška

Institute of Physics, Academy of Sciences

182 21 Prague, Czech Republic

piska@fzu.cz

<http://hp18.fzu.cz/~piska/>

Abstract

The paper presents fonts for Indic languages in the Type 1 format converted from METAFONT sources with the T_EXTRACE program, developed and presented by **Péter Szabó** in 2001. T_EX supports major Indic scripts and the T_EX/L^AT_EX packages together with public font METAFONT sources are available in the T_EX archives (CTAN) in the `tex-archive/language/<lang>`. The fonts in the pfb format, despite their limited quality of approximation and relatively large font file size, may be used as an alternative to corresponding bitmap fonts (represented by pk files), permitting creation of documents in PDF containing only vector outline fonts, and eliminating the use of bitmap fonts.

Introduction

Outline fonts are preferable and more suitable for use in PDF or other final electronic documents than bitmap fonts. Among the outline fonts, the most popular in the T_EX world are Adobe Type 1 PostScript fonts [2]. A single definition of a Type 1 font embedded in a document may cover all the magnifications of the script, is common for all resolutions, and the result can be zoomed in a previewer (like Acrobat Reader), scaled in PostScript printers and devices, etc. It is not necessary again and again to generate bitmap font representations for various sizes and different resolutions of output devices from METAFONT sources. Moreover, the bitmap images look ugly and are displayed slowly.

Today most of the major Indic scripts are available for use with T_EX. Numerous fonts in the METAFONT format together with T_EX support are available from the Comprehensive T_EX Archive Network (CTAN). “An Overview of Indic Fonts for T_EX” was published by Anshuman Pandey in *TUGboat* in 1998 [1]. Following the example of this article we will use the adjective “Indic” (not “Indian”) for languages and scripts of India. Several public Indic fonts in the PostScript Type 1 form exist in CTAN but the exact Type 1 equivalents of the METAFONT originals usually are not freely available. For TUG 2002 in India I decided to prepare a collection of Type 1 Indic fonts corresponding to their METAFONT sources from CTAN.

The T_EXTRACE program developed by Péter Szabó [3] was used for conversion into the Type 1 format from the original METAFONT sources. The font families, converted and used in the contribution, are listed in Table 1 with their versions and author names.

Several public Indic fonts in the Type 1 format (listed in Table 2) can be found in CTAN. Other Type 1 fonts for the Indic languages exist but they are not free or are not available from CTAN. Table 3 presents an overview of the Unicode “common” part of consonants for Indic languages — now in the Type 1 representation.

Conversion with T_EXTRACE and the first stage postprocessing

I intended to test possibilities of the T_EXTRACE program providing a conversion of bitmap images into outlines and I did not want to “reinvent” the methods of analytic conversion developed by Basil K. Malyshev [4] or Richard J. Kinch [5] despite the loss of important information from METAFONT.

The conversion from METAFONT sources into a primary Type 1 format was executed on a SUN workstation with the Solaris 2.6 operating system without problems; no assistance was needed and it took about 10 minutes for a single font.

Unfortunately, after the T_EXTRACE transformation, the Type 1 fonts, without optimization and without hinting, are not perfect; the outline curve approximation contains many nodes, the font files

are relatively large, and postprocessing is necessary. The results of the outlines approximated by T_EXTRACE are worse than I had expected.

The converted font accurately reflects the original *as a whole* and successfully solves a minimalization problem. The primary Type 1 fonts, however, have a number of inconvenient features *in the details*: absence of extrema points and, on the other hand, unexpected bumps may occur; improper starting points; straight lines divided into multiple segments; identical glyph elements may be approximated differently by segments; corners with small angles are not detected and are approximated by arcs or sequences of short segments; node clustering in regions where the curvature is changing; irregularities due to “*simplification*” of some “*details*” in the METAFONT sources; (most notably) short segments have bad tangential angles; slanted straight lines are approximated by curves — and — curves are approximated by straight lines (contrary to other known font families of our experience); single curve segments in the “S” form (tangential vectors from nodes point in contrary directions).

The subsequent multistep postprocessing is and will continue to be executed using T1UTILS [6] for the conversion of pfb files into a readable “raw” text form and back. The AWK programs [7] are used for operations with quasi-automatic or manual marking of the “raw” form by marks — commands, e.g., for inserting extrema points (i.e., a segment is divided into two segments in a point with partial derivation equal 0); merging two curve segments (the previous subpath is changed into a more smooth segment); joining more segments into one straight line (after T_EXTRACE it may be split) or by moving the starting point of a path to the next or the previous node. Hinting information (in the T_EXTRACE output it is missing) may be added by a font editor using an autohinting tool, e.g., using PFAEDIT [8].

Unfortunately attempts to develop automatic postprocessing algorithms have been unsuccessful. Different fonts have different designs and “*produce*” different problems. Many irregularities are “*local*” and a more general approach is impossible to find.

Availability

The Type 1 fonts (pfb files) in α -version in various stages of postprocessing are or will be available from my web site (<http://hp18.fzu.cz/~piska/>) together with the corresponding tfm files (copied

from CTAN) and proofsheets in PDF. The proofsheets are created using my method of generating associated Type 1 fonts that visualizes nodes, control points and hints. This method was used for a comparison of CM/EC fonts and was presented at the EuroBachTeX2002 meeting (April 29–May 3) in Poland [9].

Acknowledgements

I would like to thank all the authors of public METAFONT fonts for Indic languages.

Conclusion – Future work

The outline Type 1 fonts can be relatively simply generated by T_EXTRACE from METAFONT sources and then can successfully be substituted for their bitmap versions. But large sizes of font files and many “*tiny*” details on glyph outline level need complicated postprocessing.

References

- [1] Anshuman Pandey. “An Overview of Indic Fonts for T_EX”, *TUGboat* **19** (2), pp. 115–120, 1998.
- [2] Adobe Systems Inc. *Adobe Type 1 Font Format*. Addison-Wesley Publishing Company, 1990.
- [3] Péter Szabó. “Conversion of T_EX fonts into Type 1 format”, *Proceedings of the EuroT_EX 2001 conference*, pp. 192–206, Kerkrade, the Netherlands, 23–27 September 2001.
- [4] Basil K. Malyshev, “Problems of the conversion of METAFONT fonts to PostScript Type 1”, *TUGboat*, **16** (1), pp. 60–68, 1995.
- [5] Richard J. Kinch, “MetaFog: Converting METAFONT Shapes to Contours”, *TUGboat*, **16** (3), pp. 233–243, 1995.
- [6] T1UTILS package. (Type 1 tools). <http://www.lcdf.org/~eddietwo/type/#t1utils>
- [7] Alfred V. Aho, Brian W. Kernighan, and Peter J. Weinberger, *The AWK Programming Language*, Addison-Wesley, 1988.
- [8] George Williams. PFAedit – A PostScript Font Editor, <http://pfaedit.sourceforge.net/overview.html>.
- [9] Karel Piška. “A comparison of public CM/EC fonts in Type 1 format”, *Proceedings of the XIII European T_EX Conference*, April 29–May 3, 2002; Bachotek, Poland.

Table 1: Indic METAFONT fonts converted into Type 1

	Script	Font Name (Package)	Author(s) of METAFONT CTAN/language subdirectory	Version
Dvg	Devanagari	dvng10 (devnag)	Frans J. Velthuis indian/itrans	1.7 (1998)
San	Sanskrit	skt10	Charles Wikner sanskrit	2.0 (1996)
Ben	Bengali	bnr10	Abhijit Das bengali/pandey	1997
	Gujarati		No METAFONT font in CTAN	
Gmu	Gurmukhi	grmk10 (Gurmukhi)	Amarjit Singh gurmukhi/singh	1.0 (1995)
Pun	Punjabi	pun10	Hardip Singh Pannu gurmukhi/pandey	1999
Ory	Oriya	or10 (Oriya- \TeX)	Jeroen Hellingman oriya	0.93 (1998)
Sin	Sinhalese	sinha10 (sinhala_ \TeX)	Yannis Haralambous, Vasantha Saparamadu sinhala	2.1.1 (1996)
Kan	Kannada	kan10 (Kannada \TeX)	G.S. Jagadeesh indian/itrans	1991
Tel	Telugu	tel10 (Telugu \TeX)	Lakshman Kumar Mukkavilli telugu	1.0 (1991)
Mlm	Malayalam	mm10 (Malayalam- \TeX)	Jeroen Hellingman malayalam	1.6 (1998)
Tam	Tamil	wntml10	Thomas Ridgeway tamil/wntamil	1988–91
Tib	Tibetan	ctib (cTib \TeX)	Sam Sirlin, Oliver Corff tibetan/ctib	0.1 (1999)

Table 2: Indic Type 1 fonts from CTAN

Script	Font Name	Directory	Author of Type 1 font
Devanagari	xdvng	indian/itrans	Sandeep Sibal
Bengali	ItxBengali	indian/itrans	Shrikrishna Patil
Gujarati	ItxGujarati	indian/itrans	Shrikrishna Patil
Punjabi	Punjabi	indian/itrans	Hardip Singh Pannu
Perso-Arabic	xnsh14	arabtex	Taco Hoekwater

Table 3: Consonants

UNI	TRA	Dvg/San	Ben	Gmu/Pun	Ory	Sin ^d	Kan ^c	Tel ^c	Mlm	Tam	Tib ^d
KA	k	क क	क	ख ख	क क	ක ක	ಕ ಕ	క క	ക ക	க	ཀ
KHA	kh	ख ख	ख	घ घ	ख ख	ඝ ඝ	ಘ ಘ	ఘ ఘ	ഘ ഘ	஘	ཁ
GA	g	ग ग	ग	ग ग	ग ग	ග ග	ಗ ಗ	గ గ	ഗ	க	ཁ
GHA	gh	घ घ	घ	घ घ	घ घ	ඝ ඝ	ಘ ಘ	ఘ ఘ	ഘ ഘ	஘	ཁ
NGA	ñ	ङ ङ	ङ	ङ ङ	ङ ङ	ඞ ඞ	ಞ ಞ	ఞ ఞ	ഞ	ఞ	ཁ
CA	c	च च	च	च च	च च	ච ච	ಛ ಛ	ఛ ఛ	ച	ச	ཁ
CHA	ch	छ छ	छ	छ छ	छ छ	ඡ ඡ	ಞ ಞ	ఞ ఞ	ച	ச	ཁ
JA	j	ज ज	ज	ज ज	ज ज	ජ ජ	ಞ ಞ	ఞ ఞ	ജ	ச	ཁ
JHA	jh	झ झ	झ	झ झ	झ झ	ඣ ඣ	ಞ ಞ	ఞ ఞ	ജ	ச	ཁ
NYA	ñ	ञ ञ	ञ	ञ ञ	ञ ञ	ඤ ඤ	ಞ ಞ	ఞ ఞ	ഞ	ఞ	ཁ
TTA	t	ट ट	ट	ट ट	ट ट	ඨ ඨ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
TTHA	th	ठ ठ	ठ	ठ ठ	ठ ठ	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
DDA	ḍ	ड ड	ड	ड ड	ड ड	ඳ ඳ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
DDHA	ḍh	ढ ढ	ढ	ढ ढ	ढ ढ	ඬ ඬ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
NNA	ṇ	ण ण	ण	ण ण	ण ण	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
TA	t	त त	त	त त	त त	ඨ ඨ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
THA	th	थ थ	थ	थ थ	थ थ	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
DA	d	द द	द	द द	द द	ඳ ඳ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
DHA	dh	ध ध	ध	ध ध	ध ध	ඬ ඬ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
NA	n	न न	न	न न	न न	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
NNNA	ṅ	ङ ङ	ङ	ङ ङ	ङ ङ	ඞ ඞ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
PA	p	प प	प	प प	प प	ඵ ඵ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
PHA	ph	फ फ	फ	फ फ	फ फ	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
BA	b	ब ब	ब	ब ब	ब ब	භ භ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
BHA	bh	भ भ	भ	भ भ	भ भ	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
MA	m	म म	म	म म	म म	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
YA	y	य य	य	य य	य य	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
RA	r	र र	र	र र	र र	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
RRA	r̥	र̣ र	र̣	र̣ र	र̣ र	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
LA	l	ल ल	ल	ल ल	ल ल	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
LLA	l̥	ल̣ ल	ल̣	ल̣ ल	ल̣ ल	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
LLLA	l̄	ल̄ ल	ल̄	ल̄ ल	ल̄ ल	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
VA	v	व व	व	व व	व व	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
SHA	ś	श श	श	श श	श श	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
SSA	ṣ	ष ष	ष	ष ष	ष ष	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
SA	s	स स	स	स स	स स	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ
HA	h	ह ह	ह	ह ह	ह ह	ඹ ඹ	ಠ ಠ	ఠ ఠ	ത	த	ཁ

UNI – Unicode character names.

TRA – for Romanized transliteration (produced by Dominik Wujastyk)
the Type 1 font CS Bitstream Charter is used.

^c The composite glyphs are not completed here.

^d Sinhalese and Tibetan character sets are significantly different from the Indic repertoire.