



STR

Quality that speaks for itself

Phonetic Database (PDB)

Speech Technology Research Ltd., Kay Elemetrics Corporation of Lincoln Park, New Jersey and the University of Victoria in Victoria, B.C., Canada, have collaborated on the design, reproduction and distribution of a phonetic database consisting of digitized speech samples from a variety of languages. This database provides easy access to a wide variety of speech samples. It can be randomly accessed and acoustically analyzed using the Computerized Speech Lab (CSL). CSL includes utilities to export the data to non-CSL formats. *(See the CSL brochure for product/software information).*

Supported by its documentation, the purpose of the **Phonetic Database** is to depict the sounds, symbols and conventions of transcription for a number of languages. In most cases, the items presented are intended to illustrate the extent of phonological contrast in the language, to represent a wide range of phonetic realizations and to give examples of running speech.

The newest version, Version 4 of the **Phonetic Database** contains a total of 5,672 digitized speech files from 57 languages, some of which include time-synchronized IPA transcriptions incorporated directly into the speech files. The supporting documentation attempts to describe the sound system of each language, and provides a phonemic, phonetic and orthographic transcription of each file, along with an English gloss.

We enclose a listing of the languages, their language codes and the number of files available for each language in the Phonetic Database.

**STR***Quality that speaks for itself*

57 Digitized Speech Samples

<u>CODE</u>	<u>LANGUAGE</u>	<u>No. Files</u>
AHO	AHOUSAHT (NOOTKA)	36
ARE	ARABIC (EGYPTIAN)	79
CAN	CHINESE (CANTONESE)	59
MSC	CHINESE (MODERN STANDARD)	158
SZE	(Sichuan)	10
DAN	DANISH	241
DIT	DITIDAHT	66
ENG	ENGLISH (CANADIAN)	214
EST	ESTONIAN	159
EWE	EWE	68
FRS	FARSI (Persian)	402
FIN	FINNISH	73
FRN	FRENCH (Quebec)	106
GAE	GAELIC (SCOTS) (Isle of Lewis)	77
GAR	GARIFUNA	54
GEO	GEORGIAN	97
GYA	GUGU YALANJI (Banambila)	52
HGW	HAN GWICH'IN	70
HNG	HUNGARIAN	124
IGL	INUKTITUT (Iglulik)	71
INU	INUKTITUT (Pangnirtung)	55
JAP	JAPANESE (NS/NNS)	47
JA2	JAPANESE	47
KNJ	KAANTJU	54
KOR	KOREAN	79
LAT	LATVIAN	104
MIR	MIRIAM	124
NYA	NYANGUMARTA	82
PJB	PANJABI	85
RUN	RUNYORO	54
RUT	RUTOORO	52
SAA	INARISAAMI (LAPPISH)	82
SIN	SINHALA	63
SWA	siSWATI	56
SKG	SKAGIT (N. LUSHOOTSEED)	47
SPN	SPANISH (Venezuela)	89
SPK	SPOKANE-KALISPEL-FLATHEAD	85
SWE	SWEDISH	121
THA	THAI (Bangkok)	136



57 Digitized Speech Samples cont'd

<u>CODE</u>	<u>LANGUAGE</u>	<u>No. of Files</u>
TBT	TIBETAN	33
TGR	TIGRINYA	82
TRK	TURKISH (Ankara)	57
UMP	UMPILA	108
VTN	VIETNAMESE (Hanoi)	144
XHS	XHOSA	35
YNS	YI (NOSU)	164
YOR	YORUBA	239
	Subtotal:	4,540

Additional Languages

ARM	ARMENIAN	102
BLG	BULGARIAN	107
GRG	GEORGIAN	114
KZH	KAZAKH	102
KRG	KIRGHIZ	91
RSN	RUSSIAN	108
TDJ	TADJIK	127
TAT	TATAR	141
UKR	UKRAINIAN	142
UZB	UZBEK	98
	Subtotal:	1,132
	TOTAL:	5,672