Supplementary material to: De Wet F, et al. S Afr J Sci. 2017;113(1/2), Art. #2016-0038, 9 pages.

How to cite: De Wet F, Kleynhans N, Van Compernolle D, Sahraeian R. Speech recognition for under-resourced languages: Data sharing in hidden Markov model systems [supplementary material]. S Afr J Sci. 2017;113(1/2), Art. #2016-0038, 3 pages. http://dx.doi.org/10.17159/sajs.2017/20160038/suppl

Appendix 1: Mapping between Dutch and Afrikaans phones in XSAMPA format

Dutch phone	Dutch simplified	Dutch example	Afrikaans equivalent
p	p	paal	p
b	b	baal	b
T	t	taal	t
d	d	daal	d
k	k	kaal	k
g	g	goal	g
F	f	fel	f
v	v	vader	v
S	s	samen	s
Z	z	zaad	z
S	S	chauffeur	S
Z	Z	job	Z
x	x	acht	x
G	G	gezien	x
h	h	haat	h
m	m	maand	m
n	n	naad	n
N	N	ring	N
J	n j	vignet	n j
l	l	laat	l
r	r	raat	r
j	j	ja	j
w	w	waal	w
I	I	wil	@
E	E	met	E

Dutch phone	Dutch simplified	Dutch example	Afrikaans equivalent
A	A	acht	A
O	O	of	O
Y	Y	tussen	Y
@	@	het	@
i	i	nieuws	i
e	e	zee	e
a	a	aan	a
o	o	over	o
y	y	uur	y
u	u	goed	u
&	&	deur	&
E^	E^	blijkbaar	E^
O^	O^	vrouw	O^
@^	@^	uitspraak	@^
E:	I	militair (mono)	**
O:	O	kous (mono)	**
@:	@	huis (mono)	**
A~	A n	croissant	**
E~	E n	bulletin	**
O~	O n	conge	**
Y~	Y n	parfum	**
@.	@	een	**

***indicates that a phone does not occur in Afrikaans*

Appendix 2: Nomenclature

Multilayer perceptron	Multiple layer feed-forward artificial neural network. There are typically three layers: input, hidden and output. Each layer is connected to the next by weights which are tuned during training.
Gaussian mixture model (GMM)	A mixture of Gaussian probability density functions. A weighted sum of density functions that are used to model an arbitrary density function.
Deep neural network (DNN)	An artificial neural network that has multiple hidden layers.
Bottleneck features	A DNN in which a single hidden layer has a reduced size compared to the other hidden layers. This is referred to as the bottleneck. The outputs from this layer are used as bottleneck features.
Intrinsic spectral analysis	Intrinsic spectral analysis approximates a set of nonlinear projection maps onto an intrinsic coordinate system using graph Laplacian spectral clustering. For the purpose of acoustic feature extraction from speech signals, it can be regarded as a manifold learning algorithm used as a signal-processing technique.
Hidden Markov model (HMM)	State-based model used to model the acoustic units in speech, i.e. phones. A typical topology is three states, left to right, in which the system can remain in a given state or move to the next. The state parameters are estimated during training and are represented by a GMM. These parameters are latent (hidden) and are inferred from the training data transcriptions and accompanying audio.
Kullback–Leibler (KL) divergence based HMM	Posterior features, derived from phone class conditional posteriors, are modelled using a multinomial emission state. The KL-divergence is measured between the multinomial distribution and the posterior, which is referred to as the emission score.
Mel frequency cepstral coefficients	Biological inspired features that are used in most speech recognition systems. The spectrum is modified from a linear scale to a Mel scale and a set of filters is applied to the modified spectrum. The spectral components within each filter are summed and the discrete cosine transform is applied to the bank of filter energies.
Subspace GMM	Each speech state is represented by a GMM, but the means and weights components are derived from a state vector that is applied to globally shared parameters.
Acoustically weighted Levenshtein distance	The Levenshtein distance is used to calculate the edit distance between strings. In our case, the editing costs are estimated from the acoustic space, i.e. related to the phone space and distance between phones.