

REKENAAR SPRAAKSINTESE: Die Omskakeling van Teks na Klank

M.J. Wagener

Departement Rekenaarwetenskap
Universiteit van Port Elizabeth

OPSOMMING

Verslag word gelewer oor die ontwikkeling en implementering van 'n stel letter-na-klank reëls vir die omskakeling van Afrikaanse teks na foneme. 'n Inleidende oorsig van rekenaar spraaksintese en 'n kort beskrywing van die verskillende metodes van teks na klank transkripsie, word ook gegee.

1. INLEIDING

Die afgelope vyf jaar het 'n nuwe afvoertoestel op die kommersiële mark verskyn, naamlik die elektroniese spraaksintiseerder. Daar is hoofsaaklik twee tipes spraaksintiseerders: die digitale spraaksintiseerder en die fonetiese spraaksintiseerder.

Die digitale spraaksintiseerder het die vermoë om 'n digitale voorstelling van stempatrone om te skakel in klank. Die gehalte van die spraak wat gelewer word, is direk afhanklik van die aantal waarnemings wat van die oorspronklike stempatrone gemaak word. Die digitale spraaksintiseerder kan dus spraak van baie hoë kwaliteit produseer indien groot hoeveelhede geheue beskikbaar is. Byvoorbeeld, as een greep benodig word vir 'n enkele waarneming en ag duisend waarnemings per sekonde gemaak word, sal ag kilogrepe geheue benodig word om een sekonde spraak te stoor.

Die tweede tipe spraaksintiseerder het die vermoë om foneme om te skakel in klank. Indien die fonetiese transkripsie van 'n woord in die fonetiese spraaksintiseerder ingevoer word, sal dit die betrokke woord "sê". Alhoewel hierdie tipe sintiseerder nie dieselfde kwaliteit spraak as die digitale sintiseerder lewer nie, vereis dit baie minder geheue. Die fonetiese transkripsie van die teks benodig ongeveer dieselfde hoeveelheid geheue as die teks self.

2. TEKS NA KLANK OMSKAKELING

Die maklikste manier om die fonetiese transkripsie van woorde te bepaal, is deur die woorde en hul fonetiese transkripsies in 'n woordeboek te stoor. Woorde kan getranskribeer word deur dit op te soek en die ooreenstemmende foneme te produseer. Hierdie metode verseker absoluut akkurate en korrekte transkripsies en die gebruiker kan sy voorkeur in uitspraak van woorde bepaal. 'n Woordeboekstelsel het egter verskeie beperkende nadele.

- (1) Dit vereis genoegsame geheue om die woorde te stoor, asook effektiewe en vinnige middele om die woorde te onttrek. 'n Toename in die grootte van die woordeboek het ook 'n gepaardgaande afname in die spoed van woordherwinning tot gevolg.
- (2) Slegs woorde bevat in die woordeboek kan getranskribeer word.
- (3) 'n Minder beduidende nadeel is dat woorde met dieselfde spelling en verskillende uitspraak nie suksesvol hanteer kan word nie.

So 'n woordeboekstelsel is geïmplementeer deur die woorde en hul ooreenstemmende foneemstringe in 'n relasiedatabasis te stoor. 'n Volledige stelsel vir die instandhouding van die databasis asook die gebruik daarvan vir transkripsie, is ook voltooi. Die spesifieke implementasie het waardevolle hulp verleen met die beoordeling van die sukses van die letter-na-klank algoritme vir Afrikaans. (sien paragraaf 3)

'n Prakties aanvaarbare metode van teks na klank omskakeling moet "slim" genoeg wees om te bepaal wat die uitspraak van 'n gegewe woord is. Die NRL stelsel van Elovitz, Johnson en andere ontwikkel vir Engels, beantwoord aan hierdie vereiste. Hierdie stelsel gebruik 'n stel van letter-na-klank reëls wat die uitspraak van 'n letter of groep letters bepaal deur te kyk na die konteks waarin die letter of groep letters in 'n woord voorkom. 'n Soortgelyke stelsel vir Afrikaans is ontwikkel, geïmplementeer en getoets. Die prestasie van so 'n stelsel verswak as die

groepe letters wat getranskribeer word te groot is of selfs volledige woorde gebruik word vir 'n spesifieke reël en kan in dié geval oorgaan tot 'n tipiese woordeboekstelsel. Dit is egter ook 'n voordeel van die metode dat ongewone woorde as reëls ingesluit kan word sonder om die logika en struktuur van die algoritme te benadeel. Die sukses van hierdie metode word bepaal deur die spesifieke taal waarvoor dit ontwikkel is: as daar 'n sterk verband tussen die spelling van woorde en die fonetiese transkripsie daarvan bestaan, kan beter resultate verkry word.

'n Derde bekende metode van teks na klank omskakeling doen 'n morfologiese ontleding van woorde. Enige woord kan opgebreek word in morfeme, wat basies ooreenstem met die voorvoegsels, stamwoorde en agtervoegsels van die woord. So 'n stelsel hou 'n woordeboek van morfeme. Dit is 'n beter uitweg as 'n woordeboek van woorde aangesien daar baie meer woorde in 'n taal bestaan as morfeme en die aantal morfeme in so 'n woordeboek 'n versadigingspunt bereik. Verder sal die weglating van 'n morfeem uit die woordeboek nie die transkripsieproses beëindig nie, maar slegs 'n minder korrekte uitspraak tot gevolg hê. Die nadeel van morfologiese ontleding is dat dit nie altyd moontlik is om die morfeme van 'n betrokke woord te bepaal nie. Die proses van leksikale samestelling het gewoonlik tot gevolg dat die individuele morfeme van 'n woord verander word. Om die proses weer om te keer en die oorspronklike morfeme te verkry, word 'n stel samestellingsreëls gebruik. Die kompleksiteit van so 'n stel reëls is afhanklik van die grootte van die woordeboek van morfeme. Morfeme word getranskribeer deur weereens 'n stel van letter-na-klank reëls te gebruik of deur die foneemstringe van morfeme ook in die woordeboek van morfeme te stoor.

3. TEKS-NA-KLANK ALGORITME VIR AFRIKAANS

Die doel van 'n teks-na-klank algoritme is om die gaping tussen die geskrewe taal en die uitspraak daarvan te probeer oorbrug. Mense skryf nie 'n taal soos hulle dit praat nie. Dit is egter ook verkeerd om te sê dat daar geen verband tussen ortografie en uitspraak bestaan nie, want ons is tog in staat om te lees. 'n Meganiese omskakelingsalgoritme kan dus poog om die menslike leesproses na te boots. Die probleem is egter dat wanneer 'n persoon lees, sien hy nie net 'n string losstaande letters nie, maar wel woorde en die woorde het sekere betekenis in die konteks waarin dit voorkom. Die semantiek van 'n woord bepaal tot 'n mindere mate die uitspraak daarvan.

Die algoritme wat ontwikkel is vir Afrikaans, beskou 'n letter of 'n groep letters in die konteks van die woord waarin dit voorkom, en maak daarvolgens 'n afleiding oor die uitspraak daarvan. Dit neem geensins die semantiek van die woord in ag nie — dit sal dus byvoorbeeld nie die verskil in uitspraak tussen “ge'klik” en “gek'lik” kan bepaal nie. Daar word ook geensins gepoog om intonasie en ritme toe te pas nie. Die algoritme, op sigself, is dus 'n suiwer vertaling van letters na foneme. Die algoritme bestaan uit die toepassing van 'n stel reëls, en word aangegee in aanhangsel B. Elke reël het die volgende formaat:

$$A[B]C = D$$

B is die letter of letters wat omskep word in D — een of meer foneme soos aangegee in aanhangsel A. A en C is weereens 'n letter of groep letters wat B onderskeidelik moet voorafgaan en volg in die woord. In die reëls self bevat A en C ook karakters wat nie as sulks in die konteks voorkom nie, maar wel ander groepe karakters voorstel. Bv. 'n \$ in 'n reël dui enige medeklinker in die konteks aan en 'n & 'n klinker. Hierdie karakters word aangegee in aanhangsel B. Die reëls is gesorteer in groepe volgens die eerste letter in B. Vir die transkripsie van 'n lettergreep in die teks, word daar slegs na die groep reëls wat ooreenstem met die eerste letter in die lettergreep, gekyk. Die reëls word een vir een, beginnende by die eerste reël in die groep, getoets en die eerste korrekte een word dan gebruik. (Die laaste reël in 'n groep sal altyd pas en die soekproses beëindig.) Die algoritme vir die transkripsie van 'n woord is soos in figuur 1.

Die ontwikkeling van die stel reëls is begin deur 'n basiese stel reëls saam te stel. Hierin is bevat die reëls waarvoor daar 'n sterk verband tussen letters en foneme bestaan. Byvoorbeeld waar 'n medeklinker herhaal word, word slegs een klank gehoor, en 'n medeklinker alleen het gewoonlik 'n standaard uitspraak wat maklik afleibaar is van die letter. Dubbelklinkers soos “aa” en dubbelklanke soos “ooi” en “eu” kan ook sonder probleme getranskribeer word. Dit is egter enkel klinkers en veral die “e” waarvan die uitspraak in 'n spesifieke konteks moeilik bepaalbaar is. Om 'n nuwe reël op te stel moet daar gesoek word na 'n spesifieke konteks waarin 'n letter baie voorkom en dan ook meerendeels 'n spesieke transkripsie het. Om dit met die hand te doen, is haas onmoontlik en dit was in die opsig waar die woordeboekstelsel van onskatbare waarde was. Die woordeboek is gelaai met 'n groot steekproef van Afrikaanse woorde. Transkripsies is

gedoen met beide metodes en sodoende kon foute gemaak deur die stel reëls maklik uitgelig word en analiseer word om te kyk of spesifieke foute nie herhaal word nie. Verder is die suksespersentasie van die stel reëls ook bepaal en kon die byvoeging of weglating van 'n reël onmiddellik gemotiveer word. Op so 'n wyse is die stel reëls uitgebrei en verfyn om maksimale sukses te verkry.

```

I = 1
J = lengte(WOORD)
terwyl I <= J
{
  beskou eerste reël met B[1] = WOORD[I]
  reël gevind = vals
  terwyl die korrekte reël nie gevind is nie
  (
    LEN = lengte(B)
    as WOORD[I,I+LEN-1] = B
      as (linker konteks pas) en (regter konteks pas)
        produseer D
        reël gevind = waar
      beskou volgende reël met B[1] = WOORD[I]
    )
  I = I + LEN
}

```

figuur 1

Die alfabet foneme wat gebruik word in die stel reëls is saamgestel deur die Internasionale Fonetiese Vereniging van Parys. Die spesifieke karakters wat gebruik word om elke foneem voor te stel, is gewysig om saam te val met die karakterstel van 'n standaard sleutelbord. Die stel foneme wat hanteer kan word deur 'n spesifieke spraaksintiseerder sal egter nie presies met hierdie alfabet ooreenstem nie. Spraaksintiseerders het 'n kleiner stel foneme en sluit slegs die foneme in wat gebruik word in die taal waarvoor dit ontwikkel is. 'n Verdere omskakelingsalgoritme is dus nodig om 'n afbeelding tussen "logiese" en "fisiese" foneme te verkry. 'n Algoritme soortgelyk aan die letter-na-klank algoritme is ontwikkel om die "logiese" foneme, soos aangegee in aanhangsel A, af te beeld op die "fisiese" foneme van die VOTRAX SC-OI spraaksintesevlokkie. Alhoewel daarin geslaag is om elke "logiese" foneem af te beeld, het die uitspraak van Afrikaanse woorde 'n duidelike Amerikaanse aksent.

4. GEVOLGTREKKINGS

Die stel van 161 letter-na-klank reëls vir Afrikaans het bewys dat alhoewel daar nie 'n direkte ooreenkoms tussen ortografie en uitspraak bestaan nie, dit tog moontlik is om vir enige gegewe teks byna korrekte transkripsies te produseer sonder die gebruik van 'n lywige en lomp woordeboek. (Resultate van 'n prestasiemeting van die stel reëls sal in 'n volgende verslag bespreek word.) Die grootte van die stel reëls maak dit ook moontlik om 'n spraaksintese stelsel vir 'n mikrorekenaar te ontwikkel en die transkripsieproses intyds te bedryf.

Verdere byvoegings en veranderinge van die reëls mag moontlik beter transkripsies tot gevolg hê, maar sal geen noemenswaardige verbetering in die finale uitspraak veroorsaak nie. Die byvoeging van algoritmes vir intonasie en ritme asook die gebruik van 'n spraaksintiseerder wat spesifiek vir Afrikaans ontwikkel is, sal die kwaliteit van uitspraak sodanig verbeter dat dit die gebruik daarvan in die praktyk regverdig.

BIBLIOGRAFIE

1. Witten, Ian H., [1982], *Principles of Computer Speech*, Academic Press, New York.
2. Heath Company, [1981], *Voice Synthesis*, Michigan.
3. Elovitz, H.S., Johnson, R., Mchugh, A., Shore, J.E., [1976], Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics, *IEEE Transactions on Acoustics, Speech and Signal processing*, ASSP-24, 6, 446-459, December.
4. Le Roux, T.H., Pienaar, P. de V., [1976], *Uitspraak-Woordeboek van Afrikaans*, J L van Schaik, Pretoria .

AANHANGSEL A : Lys van fonetiese karakters

	Engels	Afrikaans		Engels	Afrikaans
klinkers					
a	father	vat)i	toy	
a:		vate)i	nôï	
A	sat		ui	groei	
e		melaats	medeklinkers		
e:		mening	b	back	bak
E	get	les	c		paadjie
E:		sê	d	dime	doen
]	about	wig	f	fault	vis
);		wie	g	goat	ghoen
i	beet	vlieg	H		hand
i:		vlieë	h	hand	
I	bit		j	young	ja
o		orals	k	coat	kat
o:		bode	l	laugh	lam
O		deur	m	sum	mond
Q		brug	n	sun	nag
Q:		brûe	J		kantjie
u	full	ploeg	N	sung	ding
u:	fool	ploeë	p	pack	pad
y		rusie	r	rate	rooi
y:		uur	s	sue	son
)	lawn	os	F	leash	sjarmant
);		môre	z	leisure	genre
&	not		t	time	tien
^	but		T	char	tjek
R	murder		D	just	
~		kans	v	vault	water
			V		twee
			x		goed
			Z	zoo	Zoeloe
			\$	father	
			@	think	
			w	wall	
			wh	where	
			%	silent	stil klank
				sound	
dubbelklanke					
au	how				
ai	hide	aitsa			
a:i		raai			
e:u		leeu			
]i	gate	lei,ly			
o:i		mooi			
ou	lone	koud			
Qy		muis			

AANHANGSEL B : Letter-na-klank reëls vir Afrikaans

Die spesiale karakters wat in kolomme A en C gebruik word, het die volgende betekenisse:

- * - die begin of einde van 'n woord
- \$ - 'n medeklinker
- & - 'n klinker
- : - twee soortgelyke klinkers
- ; - 'n B, G, M of T (slegs in kolom A)
- - 'n frikatief (slegs in kolom C)
- . - twee karakters waarmee 'n woord nie kan begin nie (slegs in kolom C)

A	B	C	D	A	B	C	D
	AAI		a:i	; V	E] E
*	AAN	; R.	a:	V	ER]r
	AA		a:		Eë		e:]
	AI		ai		E	\$&&]]
*	A	* \$IE	a:	&	E]]
	A	\$&&	a	:	E	\$E	e:]
	A	\$&	a:	&\$	E]]
	AE		a:]		E]]
&\$	A	* \$&	a		E	\$&	e:]
	A	*	a:		E	.	E
	A		a		E	\$	E
					E	\$	E
					E]]
	BB		b		FF		f
*	B	*	be:		FV		f
	BESIG		be:s]x	*	F	*	Ef
	B		b		F		f
					G	G	
*	C	*	se:	*	GH		g
	C		k	*	G	*	xe:
				\$	GE	*	g]
					G		x
	D	D					
A	DJIE		ici				
O	DJIE		ici	*	H	*	Ha:
	DJIE		ci		H		H
*	D	*	de:				
	DT		t		IE		i
	DR		dr		Ië		i]
	DW		dv	*	IN	;]:~]]
	DERS		d]rs		I	*	i
	DERE		d]r]		I	\$IE	i
	DES		d]s		I	&	i
	DELIK		d]l]k		I]]
	DEL		d]l				
	D	&	d				
	D		t	*	J	*	je:
					J		j
	EEU		e:u				
	EE		e:		KK		k
	EI]i	*	K	*	ka
	EU		O		KERE		k]r]
	E	\$IE*	e:		KEL		k]l
	E	\$E*	e:		K		k
*	E	*	e:				
	E	*]]		LL		l
	E	R*]]	*	L	*	E]l
; ;	E	\$*	E		L		l
	E		E				

