Phonetic phenomena in continuous speech of Putonghua and its relation to speech recognition

Zu Yiqing

汉语普通话连续语流中的语音现象及其同语音识别的关系

祖漪清

[摘要] 汉语语音识别已由全音节识别进入了大词汇的连续语音识别。与此同时,汉语语音学的研究也在连续语流中深入进行。只有掌握了连续语流中的语音现象，才能为语音识别的研究提供实质性的帮助。

为了统计连续语流中的语音现象,我们需要确定基本音位。共有下面 36 个:

Basic phonemes(36):

{ a1, a2, b, c, ch, d, e1, e2, e3, er, f, g, h, i1, i2, i3, j, k, l, m, n, ng, o1, o2, p, q, r, s, sh, t, u, x, yv, z, zh, sil }

其中 sil 为无声段（ silence 的缩写）。在连续语流中，这些基本音位很难再现它们独立存在时的形象,它们通常是以音位变体的形式出现。这就是所谓的协同调音现象。本文在声学层面上，以双音子(diphone)和三音子(triphone)为例讨论音位变体的问题。

在音节内部，也就是说在不跨越音节的情况下，利用音节构成规则，统计出 170 个双音子， 276 个三音子。其中可作为音节开头的有 167 个双音子，称为音节首双音子。可作为音节结尾的双音子有 97 个，称音节尾双音子。一个音节尾音位和一个音节首音位可以构成一个音节间双音子。一个音节尾双音子和一个音节首音位可以构成一个音节间三音子; 同样，一个音节尾音位和一个音节首双音子也可以构成一个音节间三音子; 一个单音位音节和不同的音节首，尾音位可以构成一个跨三个音节的音节间三音子。音位、双音子、三音子等都可作为连续语音中的语音基元。音节内的语音基元和音节间的语音基元共同构成了连续语流中的语音基元。如果考虑双音子，基元的个数是音节首尾音位相乘（ sil-sil 不构成二元音位）加上音节内双音子:

Number of diphones = (32*13-1)+170=580;

如果考虑三音子，音节间的三音子，是音节尾音位乘以音节首双音子加上音节尾双音子乘以音节首音位，以及单音节音位与首尾音位的搭配。

tri_num=32*97+13*167+32*7*13+276=3104+2171+2912+276=8463;

语音基元的选择是语音识别最重要的问题之一。识别采用的语音基元可以是: 音位、多音子、半音节、音节、词，甚至句子。简单构成的基元（如音位），需要较少的训练数据，但结果不能令人满意; 较复杂的语音基元增加了模型的自由度，同时训练数据和模型参数也都随之增加了。为了达到一个合理的折中，许多识别系统都在语音基元的选择上下了一番工夫。三音子较详尽地描写了连续语音的音变，它被许多英语识别系统采用为识别基元。 然而，庞大的三音子系统为语音的训练和识别带来了沉重的负担。而且，根据我们的统计结果，大

多数理论上可能存在的三音子在自然语言中很少出现，因此，如果采用三音子作为识别基元，应当对其进行归类。

建立语音识别数据库的目的是为语音识别系统提供充足的语音数据，以进行训练。当前语音识别的主流方法采用的仍是隐马尔可夫模型（ HMM ）方法，训练每一个模型都需要大量的训练数据。同时，还要考虑到语音识别系统建立模型可能选取不同的语音基元，因此，所有的语音现象都应尽可能地被包括在内，但这在具体的实现过程中，却是十分困难的。

语音识别系统的评测即为评价一个语音识别系统的优劣，评价的方法一般是给出一套测试数据（语音波形），由被测系统进行识别，识别结果（识别率）客观地反映了被测系统的水平。 因此，设计测试数据的原始语音材料也是一项十分重要的工作。

同语音识别数据库一样，语音测试材料的设计要对语音识别的基本方法有所认识。语音识别的测试材料与语音识别数据库的共同点是同样要考虑各种语音现象，所不同的是对每一种语音现象不需要那样大量的数据，也就是说测试集的规模远小于训练集。另一个重要区别是对于汉语连续语音来说，识别系统识别出来的是语音基元串，由于汉语有多音字、同音字、多义字等特点，识别的最终结果不是唯一的，因此需要一个语言模型处理这一问题。为考察这一问题，测试材料还应当包括汉语的不同句型。

作为一个为语音识别服务的例子，我们从《人民日报》文库中挑选出 500 个句子作为测试候选句。其中我们考虑了汉语的句型分布，还包括了不含声调区别的 403 个音节以及 300 多个音节间双音子。

研究连续语音中的音位变体现象对于语音识别是有重要意义的，我们需要进一步探讨它们在语言学及声学语音学上的交。

Abstract

This paper intends to systematically describe the phonetic phenomena in continuous speech of Putonghua. 36 basic phonemes are introduced at first, and then statistics are provided for both intra- and/or inter-syllabic data. It is calculated that within syllables there are 170 diphones and 276 triphones, and 415 diphones and 8187 triphones between syllables. In totality, there are 585 diphones and 8463 triphones. The phonetic phenomena in continuous speech are closely related to speech recognition, and they supply a frame of reference in selecting proper speech units for speech recognition and designing materials of speech database. Finally, the design of material for testing speech recognition systems is given as an example of showing how the phonetic knowledge is used in speech recognition.

1. Introduction

Beginning with isolated-syllable recognition, the speech recognition in China has now entered a new stage of development in large-vocabulary and continuous speech (Ji, Yang, Wang & Lu, 1995; Sun, Wang, Wang & Li, 1995; Ma, Xu, Huang & Zhang, 1995). At the same time, the phonetic research on Chinese has also gone deeply in progress. To meet the needs of continuous speech recognition, phoneticians should describe the phonetic variations in different contexts systematically. Only by understanding the phonetic phenomena in continuous speech,

can speech materials be provided for training data and testing data and can speech units and models of speech recognizer be created. The Chinese speech recognition systems which have taken the transitions between syllables into account have got obviously higher score on recognition rate (Ji, et. al.,1995). Therefore the phonetic phenomena in continuous speech is also the key problem in speech recognition. By statistics, this paper come from basic phonemes, introduces all possible allophones in different phonetic contexts.

2. The basic phonetic phenomena in Putonghua

2.1. *Basic phonemes*

To obtain statistic data in continuous speech, basic phonemes should be defined. The basic phonemes are those that occupy independently some domain in acoustic plane. In totality, there are 36 phonemes:

Basic phonemes (36): {a1,a2,b,c,ch,d,e1,e2,e3,er,f,g,h,i1,i2,i3,j,k,l,m,n,ng,o1,o2,p,q,r,s,sh,t, u,x,yv,z,zh,sil}

where sil is "silence". The introduction of silence may provide convenience for continuous speech recognition. There may be disputes on vowels. Figure 1 is a vowel chart showing phonemes in basic phonemes. And Table I lists those phonemes in juxtaposition with IPA and some examples.

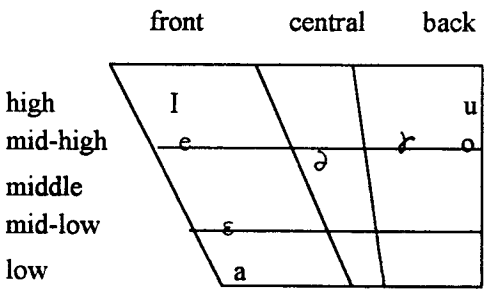

Figure 1 A vowel chart showing some main basic phonemes

Among those 36 basic phonemes, 7 can construct a syllable independently:

mono- phonemic syllables (8): {a1,e1,er,o1,u,i1,yv,sil}

32 phonemes can be taken as head in syllables:

heads(32):{sil,a1,a2,b,c,ch,d,e1,e2,e3,er,f,g,h,i1,j,k,l,m,n,o,p,q,r,s,sh,t,u,x,yv,z,zh}

13 phonemes can be used as tail in syllables.

tails(13):{sil,a1,i1,i2,i3,n,ng,u,e2,o,e,er,yv}

In continuous speech, it is difficult for basic phonemes to maintain their forms in isolation. They always appear in the form of allophones, which results in the co-articulatary effects in continuous speech. Diphones and triphones will be introduced as examples of explaining

allophones in the following discussion.

Having defined the basic phonemes, we assume that:

(a). On phonetic level, there are 36 basic phonemes which resemble the symbols used in narrow transcription. When this problem is discussed, we should understand the fact that the number of segments on phonetic level is different from that of segmentd on acoustic level. Figure 2 may illustrate this point clearly.

(b). Every basic phoneme has different allophones when it is in different left and right context between and/or within syllables.

The following discussions are supported by those two assumptions.

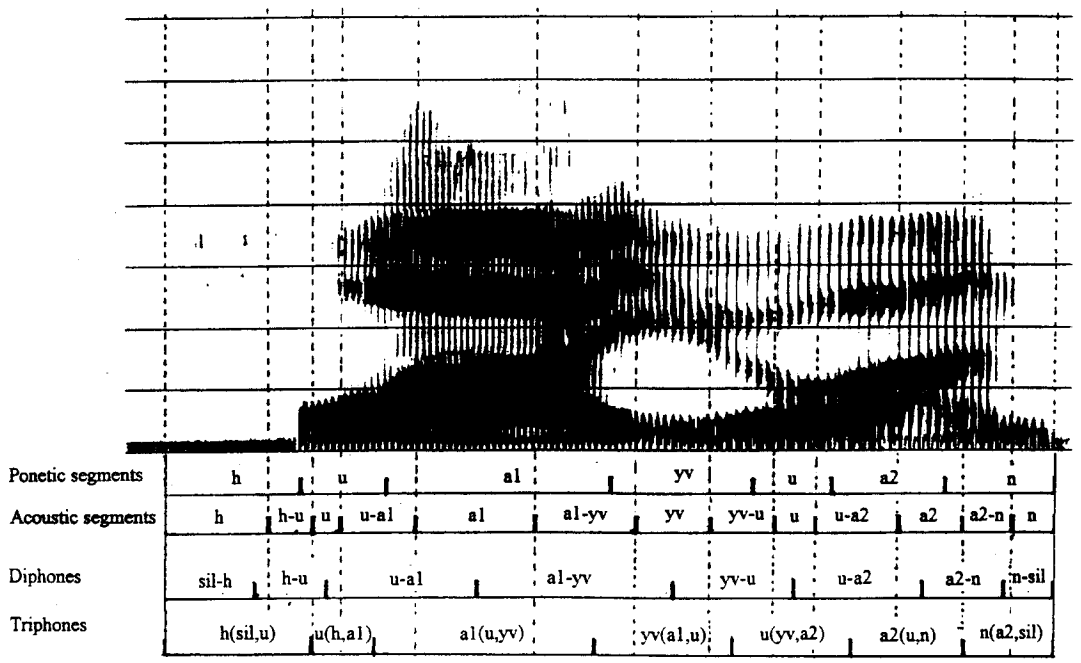

Figure 2 Segmentations of "花园 hua yuan /xua yuan/" on phonetic level, acoustic level, diphones and triphones

Table 1 The contrast symbols between basic phonemes and IPA

Basic phonemes	IPA	Examples
a1	a	ba,bang,bao
a2	ɛ	an,ai
b	p	ba,bu
c	ts^h	ca,cu
ch	$tʂ^h$	cha,chu
d	t	da,di
e1	ɤ	he,ge
e2	e	ei,ye,yue
e3	ə	en,eng
er	ər	er
f	f	fa,fu
g	k	gei,ge
h	x	he,hu
i1	I	bi,yi
i2	ɿ	zi,ci,si
i3	ʅ	zhi,chi,shi
j	tɕ	ji,jin
k	k^h	ka,ke
l	l	la,li
m	m	ma,mu
n	n	na,ni
ng	ŋ	ong,ang,eng
o1	ɔ	wo
o2	o	ou
p	p^h	pa,pi
q	$tɕ^h$	qa,qi
r	ɻ	ri,re
s	s	sa,si
sh	ʂ	shi,shu
t	t^h	ti,ta
u	u	u
x	ɕ	xi,xia
yv	y	yu,yue
z	ts	zi,za
zh	tʂ	zi,zhu
sil(silence)		

2.2. *The intra-syllabic allophones*

Within syllables, 170 diphones and 276 triphones are calculated by a rule which condtructs all of syllables by basic phonemes:

Intra-syllabic diphones (170):

{a1-i1,a1-ng,a1-u,a2-n,a2-i1,b-a1,b-a2,b-e2,b-e3,b-i1,b-o1,b-u,c-a1,c-a2,c-e1,c-e3,c-i2,c-u,c-o2,
ch-a1,ch-a2,ch-e1,ch-e3,ch-i3,ch-u,ch-o2,d-a1,d-a2,d-e1,d-e2,d-e3,d-i1,d-u,d-o2,e2-i1,e2-n,
e3-n,e3-ng,f-a1,f-a2,f-e2,f-e3,f-o1,f-o2,f-u,g-a1,g-a2,g-e1,g-e2,g-e3,g-u,g-o2,h-a1,h-a2,h-e1,
h-e2,h-e3,h-u,h-o2,i1-e2,i1-a1,i1-n,i1-ng,i1-o2,i1-u,i1-o1,j-i1,j-yv,k-a1,k-a2,k-e1,k-e2,k-e3,k-u,
k-o2,l-a1,l-a2,l-e1,l-e2,l-e3,l-i1,l-o1,l-u,l-o2,l-yv,m-a1,m-a2,m-e1,m-e2,m-e3,m-i1,m-o1,m-o2,
m-u,n-a1,n-a2,n-e1,n-e2,n-e3,n-i1,n-u,n-o2,n-yv,o2-u,p-a1,p-a2,p-e2,p-e3,p-i1,p-o1,p-o2,p-u,
q-i1,q-yv,r-a2,r-a1,r-e1,r-e3,r-i3,r-u,r-o2,s-a1,s-a2,s-e1,s-e3,s-i2,s-u,s-o2,sh-a1,sh-a2,sh-e1,
sh-e2,sh-e3,sh-i3,sh-o2,sh-u,t-a1,t-a2,t-e1,t-e2,t-e3,t-i1,t-u,t-o2,u-ng,u-a2,u-e2,u-e3,u-o1,u-a1,
x-i1,x-yv,yv-e2,yv-n,z-a1,z-a2,z-e1,z-e2,z-e3,z-i2,z-u,z-o2,zh-a1,zh-a2,zh-e1,zh-e2,zh-e3,zh-i3,
zh-u,zh-o2}

Where the form x-y defines a diphone in the context of x-y.

Intra-syllabic triphones (276):

{a1(b,ng),a1(b,u),a1(i1,u),a1(c,ng),a1(c,u),a1(ch,ng),a1(ch,u),a1(u,ng),a1(d,ng),a1(d,u),a1(f,ng),
a1(g,ng),a1(g,u),a1(h,ng),a1(h,u),a1(i1,ng),a1(k,ng),a1(k,u),a1(l,ng),a1(l,u),a1(m,ng),a1(m,u),
a1(n,i1),a1(n,ng),a1(n,u),a1(p,i1),a1(p,ng),a1(p,u),a1(r,ng),a1(r,u),a1(s,ng),a1(s,u),a1(sh,ng),
a1(sh,u),a1(t,ng),a1(t,u),a1(z,ng),a1(z,u),a1(zh,ng),a1(zh,u),a2(b,i1),a2(b,n),a2(c,i1),a2(c,n),
a2(u,n),a2(ch,i1),a2(ch,n),a2(u,i1),a2(d,i1),a2(d,n),a2(f,n),a2(g,i1),a2(g,n),a2(h,i1),a2(h,n),
a2(k,i1),a2(k,n),a2(l,i1),a2(l,n),a2(m,i1),a2(m,n),a2(n,n),a2(p,n),a2(r,n),a2(s,i1),a2(s,n),a2(sh,i1),
a2(sh,n),a2(t,i1),a2(t,n),a2(z,i1),a2(z,n),a2(zh,i1),a2(zh,n),e2(b,i1),e2(i1,n),e2(u,i1),e2(d,i1),
e2(f,i1),e2(g,i1),e2(h,i1),e2(yv,n),e2(k,i1),e2(l,i1),e2(m,i1),e2(n,i1),e2(p,i1),e2(sh,i1),e2(t,i1),
e2(z,i1),e2(zh,i1),e3(b,n),e3(b,ng),e3(c,n),e3(c,ng),e3(u,n),e3(ch,n),e3(ch,ng),e3(d,n),e3(d,ng),e
3(f,n),e3(f,ng),e3(g,n),e3(g,ng),e3(h,n),e3(h,ng),e3(k,n),e3(k,ng),e3(l,ng),e3(m,n),e3(m,ng),
e3(n,n),e3(n,ng),e3(p,n),e3(p,ng),e3(r,n),e3(r,ng),e3(s,n),e3(s,ng),e3(sh,n),e3(sh,ng),e3(t,ng),
e3(u,ng),e3(z,n),e3(z,ng),e3(zh,n),e3(zh,ng),i1(b,e2),i1(b,a1),i1(b,n),i1(b,ng),i1(d,a1),i1(d,e2),
i1(d,ng),i1(d,o2),i1(j,a1),i1(j,e2),i1(j,n),i1(j,ng),i1(j,u),i1(j,o2),i1(l,a1),i1(l,e2),i1(l,n),i1(l,ng),
i1(l,o2),i1(m,e2),i1(m,a1),i1(m,n),i1(m,ng),i1(m,o2),i1(n,e2),i1(n,a1),i1(n,n),i1(n,ng),i1(n,o2),
i1(p,e2),i1(p,a1),i1(p,n),i1(p,ng),i1(q,a1);i1(q,e2),i1(q,n),i1(q,ng),i1(q,u),i1(q,o2),i1(t,e2),i1(t,a1),
i1(t,ng),i1(x,a1),i1(x,e2),i1(x,n),i1(x,ng),i1(x,u),i1(x,o2),o2(c,u),o2(ch,u),o2(i1,u),o2(d,u),o2(f,u),
o2(g,u),o2(h,u),o2(k,u),o2(l,u),o2(m,u),o2(n,u),o2(p,u),o2(r,u),o2(s,u),o2(sh,u),o2(t,u),o2(z,u),
o2(zh,u),u(c,ng),u(c,a2),u(c,e2),u(c,e3),u(c,o1),u(ch,ng),u(ch,a2),u(ch,a1),u(ch,e2),u(ch,e3),
u(ch,o1),u(d,ng),u(d,a2),u(d,e2),u(d,e3),u(d,o1),u(g,ng),u(g,a1),u(g,a2),u(g,e2),u(g,e3),u(g,o1),
u(h,ng),u(h,a1),u(h,a2),u(h,e2),u(h,e3),u(h,o1),u(i1,ng),u(k,ng),u(k,a1),u(k,a2),u(k,e2),u(k,e3),
u(k,o1),u(l,ng),u(l,a2),u(l,e3),u(l,o1),u(n,ng),u(n,a2),u(n,o1),u(r,ng),u(r,a1),u(r,a2),u(r,e2),u(r,e3),
u(r,o1),u(s,ng),u(s,a2),u(s,e2),u(s,e3),u(s,o1),u(sh,a1),u(sh,a2),u(sh,e2),u(sh,e3),u(sh,o1),
u(t,ng),u(t,a2),u(t,e2),u(t,e3),u(t,o1),u(z,ng),u(z,a2),u(z,e2),u(z,e3),u(z,o1),u(zh,ng),u(zh,a1),

u(zh,a2),u(zh,e2),u(zh,e3),u(zh,o1),yv(j,e2),yv(j,n),yv(l,e2),yv(n,e2),yv(q,e2),yv(q,n),yv(x,e2), yv(x,n)}

The form of x(y,z) means phoneme x in the left context y and right context z. Table 2 illustrates some intra-syllabic diphones, triphones with some examples.

Table 2 Some examples to illustate diphone & triphone within syllables

example	phone	diphone	triphone
chuang	ch,u,a1,ng	ch-u,u-a1,a1-ng	u(ch,a1),a1(u,ng)
sang	s,a1,ng	s-a1,a1-ng	a1(s,ng)
ca	c,a	c-a	/
wu	u	/	/

2.3. *Inter-syllabic allophones*

Apart from mono-phonemic syllables(eg. a,i,u), 167 diphones can be used as heads of syllables (i.e. the first and second phonemes in syllables). We call them head diphones.

head diphones (167):
{(a1-i1),(a1-ng),(a1-u),(a2-n),(b-a1),(b-a2),(b-e2),(b-e3),(b-i1),(b-o1),(b-u),(c-a1),(c-a2),(c-e1), (c-e3),(c-i2),(c-u),(c-o2),(ch-a1),(ch-a2),(ch-e1),(ch-e3),(ch-i3),(ch-u),(ch-o2),(d-a1),(d-a2), (d-e1),(d-e2),(d-e3),(d-i1),(d-u),(d-o2),(e2-i1),(e3-n),(e3-ng),(f-a1),(f-a2),(f-e2),(f-e3),(f-o1), (f-o2),(f-u),(g-a1),(g-a2),(g-e1),(g-e2),(g-e3),(g-u),(g-o2),(h-a1),(h-a2),(h-e1),(h-e2),(h-e3),(h-u), (h-o2),(i1-a1),(i1-e2),(i1-n),(i1-ng),(i1-o1),(i1-u),(i1-o2),(j-i1),(j-yv),(k-a1),(k-a2),(k-e1),(k-e2), (k-e3),(k-u),(k-o2),(l-a1),(l-a2),(l-e1),(l-e2),(l-e3),(l-i1),(l-o1),(l-u),(l-o2),(l-yv),(m-a1),(m-a2), (m-e1),(m-e2),(m-e3),(m-i1),(m-o1),(m-o2),(m-u),(n-a1),(n-a2),(n-e1),(n-e2),(n-e3),(n-i1),(n-u), (n-o2),(n-yv),(o2-u),(p-a1),(p-a2),(p-e2),(p-e3),(p-i1),(p-o1),(p-o2),(p-u),(q-i1),(q-yv),(r-a2),(r-a1), (r-e1),(r-e3),(r-i3),(r-u),(r-o2),(s-a1),(s-a2),(s-e1),(s-e3),(s-i2),(s-u),(s-o2),(sh-a1),(sh-a2),(sh-e1), (sh-e2),(sh-e3),(sh-i3),(sh-o2),(sh-u),(t-a1),(t-a2),(t-e1),(t-e2),(t-e3),(t-i1),(t-u),(t-o2),(u-a1),(u-a2), (u-e2),(u-e3),(u-o1),(x-i1),(x-yv),(yv-e2),(yv-n),(z-a1),(z-a2),(z-e1),(z-e2),(z-e3),(z-i2),(z-u),(z-o2), (zh-a1),(zh-a2),(zh-e1),(zh-e2),(zh-e3),(zh-i3),(zh-u),(zh-o2)}

97 diphones can be used as tail of syllables (i.e. the last phonemes and the one before last in syllables). They are called tail diphones.

tail diphones (97):
{(a1-i1),(a1-ng),(a1-u),(a2-n),(a2-i1),(b-a1),(b-i1),(b-o1),(b-u),(c-a1),(c-e1),(c-i2),(c-u),(ch-a1),

(ch-e1),(ch-i3),(ch-u),(d-a1),(d-e1),(d-i1),(d-u),(e2-i1),(e2-n),(e3-n),(e3-ng),(f-a1),(f-o1),(f-u),
(g-a1),(g-e1),(g-u),(h-a1),(h-e1),(h-u),(i1-e2),(i1-n),(i1-ng),(i1-a1),(i1-o1),(j-i1),(j-yv),(k-a1),(k-e1),
(k-u),(l-a1),(l-e1),(l-i1),(l-o1),(l-u),(l-yv),(m-a1),(m-e1),(m-i1),(m-o1),(m-u),(n-a1),(n-e1),(n-i1),
(n-u),(n-yv),(o2-u),(p-a1),(p-i1),(p-o1),(p-u),(q-i1),(q-yv),(r-e1),(r-i3),(r-u),(s-a1),(s-e1),(s-i2),(s-u),
(sh-a1),(sh-e1),(sh-i3),(sh-u),(t-a1),(t-e1),(t-i1),(t-u),(u-ng),(u-o1),(u-a1),(x-i1),(x-yv),(yv-e2),
(yv-n),(z-a1),(z-e1),(z-i2),(z-u),(zh-a1),(zh-e1),(zh-i3),(zh-u)}

A tail phoneme and a head diphone can construct a inter-syllabic triphone; and so do a tail diphone and a head phoneme. A mono-phonemic syllable with different left and right conexts will construct a triphone across three syllables. Diphones and triphones are both speech units in continuous speech and can be used as units in recognition system.

2.4. *Phonetic phonmena in continuous speech*

The speech units of continuous speech is the sum of inter- and intra-syllabic units. If diphones are considerd, the number of units is:

di_num=(32*13-1)+170=585;

where the first item is the number of head phonemes that multiply the number of tail phonemes (sil-sil is not a diphone), and the second item is the intra-syllabic diphone. If triphones are used as units,the number of triphones is:

tri_num=32*97+13*167+32*(8-1)*13+276=8463;

where the first item is the number of head phonemes together with tail diphones; the second item is the one of tail phonemes together with head diphones, the third item is mono-phonemic syllables in different context; the forth item is the intra-syllabic triphone (except sil).

To note that: The numbers of diphones and triphones are only reference figures because they are calculated on the basis of 36 basic phonemes.

2.5. *Choice of speech units for speech recognition*

In the process of continuous speech recgonition, the speech units should be selected at first, and then all models are constructed by training. The result of recognition is a sequence of speech units. The selection of speech units is one of most important problems in speech recognition. Followings are some candidates of speech units:

phoneme multi-phone semi-syllable syllable

small ----------> large

The simpler the units(eg.phoneme), the fewer the training data needed, but it is difficult to get delighted results. The complex units allow more freedom for models. But at the same time, the training data and the coefficiences of model increase. To obtain a rational stage, a lot of recognizer have worked on speech units (Lee, 1989; Hwang, 1993; Deng, 1994). Triphone describes allophones in detail and is widely used by English speech recognizers. But tremendous number of triphones give rise to more difficulties in recognition system. According to our

statistics, the most triphones do not occur in continuous speech materials. Therefore triphones should be classified into smaller number and the problem of un-seen triphone should be taken into account.

In general, the speech units have great influence on recognition results. Table 3 displays some statistical relation within and/or between syllables.

Table 3 The statistic numbers of different speech units between & within syllables

	intra-syllable	inter-syllable
phoneme	36	32*13-1
diphone	170	415
triphone	276	8187
initial & final	22+38	38*22
semi-syllable	100+38	38*100
syllable	410	410*410

3. Speech database for recognition and recognition assesment

Providing sufficient data for training a recognizer is the prime aim of database. The activity of constructing large speech database are more dependent on the scale and quality worldwide.The main method in speech recognition is still Hidden Markov Model(HMM). The speech database is crucial to the success of HMM (Huang, 1990), phonetic phonemena in continuous speech should be included in speech matriel to the full although it is very difficult to achieve.

Speech recognition assessment is evaluating performence of a given recognition system. The general way is providing a set of testing data and then results of recognition system(recognition rate) show the performence of the relatiye recognizer.

Both training data and testing data should include the phonetic phonemena. But testing data should not be as large as the training one. Another significent difference between them is that the recognition result is at first a sequence of speech units. Final speech result is not unique due to the presence of polysements and synonyms in Putonghua. A language model should come in at the end of recognition system, which can not only help recognizer to deal with such problems but also correct errors. For this resaon, the distribution of sentence patterns should be considered in devising testing matriels.

4. The design of material for testing a recognition system — an example of phonetics knowlage used for speech recognition

500 sentences were collected from database of 1993 "People's Daily" as candidates of testing sentences.

In those sentences, both sentence patterns and phonetic phenomena are included. According to the results of department of Chinese, Tsing Hua University, 209 patterns in Putonghua sentences are identified (Department of Chinese Language and Literature, Tsing Hua Univerisity, 1993). We reclassified them into 17 patterns. To simplify the problem, only syllable and inter-syllabic diphones were included.

A kind of entropy (S) is introduced to describ phonetic balances:

$$S = -\sum_{1}^{N} (Pn * \ln(Pn))$$

where Pn is the occurrence probability of the n-th phenomena and N is the number of different units. S takes a maximum value when all Pn (n=2,...,N) are equal.

Figure 3 shows the process of collecting testing sentences. These 500 sentences include 403 syllables("den,ei,eng,kei,lo,nou,rua" are absent) if tonal distincton is suppressed. It covers 95% sllables and 80% diphones.

Table 4 reports the phonetic phenomena in testing sentences.

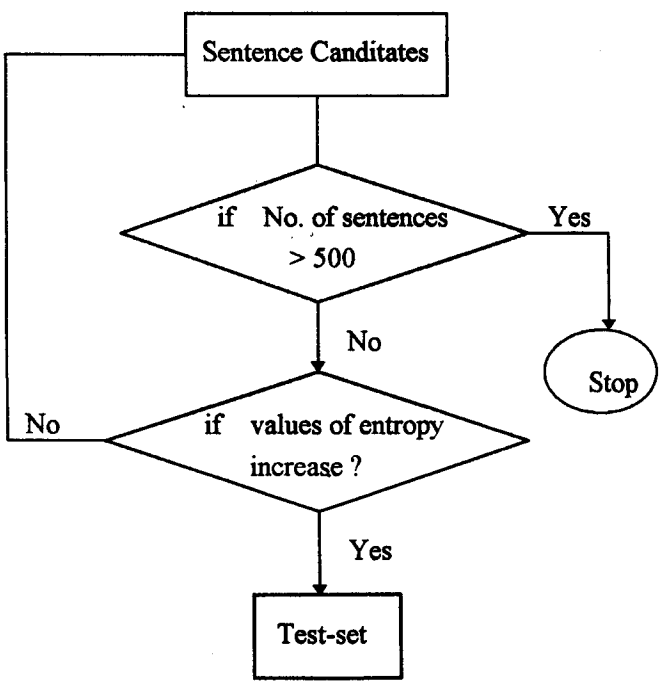

Figure 3 The process of collecting test sentenses

Table 4 The phonetic phenomena in 500 test sentences

	syllable	inter-syllable
Total number	410	415
appearance	403	323
sum of appearance	6181	6640
Max. value of S	6.016	5.955
S	5.177	5.169

5. Discussion

Among very large number of allophones in continuous speech, only a small part often appears. What the further work should be directed to give a set of rational allophones of Putonghua from two aspects. The first one is on the linguistic level, and the second is on the acoustic-phonetic level. Statistics on large speech materials will give an evaluation to all possible allophones on the linguistic level. Labeling continuous speech and studying transitions between connected syllables will give allophone classification on the acoustic-phonetic level.

Acknowledgment

I would like to thank Prof. Cao Jianfen and Ms. Chen Xiaoxia for their valuable advice on basic phonemes and Mr. Li Zhiqiang for his work on clarifying sentence patterns of sentences.

References

Ji, T. Y., Yang, H. R., Wang, Z. Y. & Lu, D. J. (1995) Connected Chinese syllables recognition based on semi-syllabic HMM. In *Proceedings of the 2th National Academic Conference on the Latest Development of Computer Intelligent Interface and Intelligent Application.* pp. 122-127. Beijing: Tsinghua University Press.

Sun, J. S., Wang, Z. Y., Wang, X. & Li, B. (1995) Constructing a word tabel for training of continuous speech. In *Proceedings of the 2th National Academic Conference on the Latest Development of Computer Intelligent Interface and Intelligent Application.* pp. 116-121. Beijing: Tsinghua University Press.

Ma, B., Xu, B., Huang, T. Y., Zhang, X. J. (1995) An acoustic model of context-dependent based on initial/final. *Proceedings of the 7th National Academic Conference on Signal Processing of*

Speech, Image and Communication and Communication (Yi Kechu & Xie Weixin, editors), pp.81-84.

Lee, K.F. (1989) *Automatic Speech Recognition: The Development of the SPHINX System*, Kluwer : Academic Publishers.

Hwang,M. (1993) Phonetic acoustic modeling for speaker-independent continuous speech recognition. In *Ph.D. Thesis*, Canegie Mellon University, School of Computer Science.

Deng, L. (1994) A statistical approach to automatic speech recognition using the atomic speech units constructed from overlapping articulatory features, *J. Acoust. Soc. Am.* 95(5), Pt.1, May, 2703-2718.

Huang, X.D. (1990) *Hidden markov models for speech recognition.* Edinburgh University Press.

Department of Chinese Language & Literature , Tsinghua University (1993) The probability distribution of sentences pattern in Chinese, *manuscript*.