

TR-I-0047

Hidden Markov Modelを用いた英単語認識  
English Word Recognition By HMM

丸山 活輝 花沢 利行 川端 豪 鹿野 清宏

Katsuteru Maruyama, Toshiyuki Hanazawa, Takeshi  
Kawabata and Kiyohiro Shikano

1988.10

概要

自動翻訳電話の要素技術である英語音声認識を実現するためにHidden Markov Modelを用いて英単語認識実験を行った。英語の場合は、音声データに対する音韻ラベル付けが日本語に比べてはるかに困難である。そこで、学習時に音韻モデルを結合、分解しながらラベルなしで学習を行う方法を導入した。米国人男性一名の発声により作成した英語データベースの約3000語の単語を用いて認識実験を行った結果、現在のところ約80%の認識率を得ている。

ATR Interpreting Telephony Research Laboratories  
ATR 自動翻訳電話研究所

## 目次

1. はじめに	1
2. 音声データ	1
2.1 分析	1
2.2 学習用データと評価用データ	1
2.3 コードブックの作成	2
3. HMMの連結	2
4. 単語音韻変換テーブル	2
5. HMMの学習	3
5.1 HMMパラメータ初期値の設定	4
5.2 音韻モデルの連結	4
5.3 連結学習	5
5.4 学習回数	5
6. 認識実験	6
7. まとめ	7
参考文献	9
付録-1	10
付録-2	11
付録-3	13
付録-4	15

## 1.はじめに

自動翻訳電話の要素技術である英語音声認識を行うために、日本語の音声認識において既に実績のあるHidden Markov Model<sup>[1]</sup>を用いて英単語の認識実験を行った。自動翻訳電話のように大語彙を音声認識の対象とする場合、音韻を単位としたHMMを用いる方式がその組合せとして単語や文節を扱う事ができ、柔軟性に富んでいる。また、英語の場合は音韻のラベル付けは日本語に比べてはるかに困難である。そこで今回は、ラベル付けのない音声データの認識のために、音韻モデルを基本単位として、それらを連結した単語モデルを用いて英単語の認識実験を行ったので報告する。

## 2.音声データ

特別の発声訓練を受けていない米国人男性1人により発声された英単語2834語(542種類)を使用する。発声内容は、「キーボード会話300」(国際会議についての問合せや、ホテルの予約の電話会話。内容の例を付録-1に示す。)の英文377文を単語に区切って発声したものである。これらのデータにはラベリングは施されていない。

### 2.1分析

上記の音声データを12kHzでサンプリングし窓長21.3msec、周期3msecのハミング窓で切り出し、高域強調をかけた後、14次のLPC分析を行った。

### 2.2学習用データと評価用データ

今後、認識結果に対して統計的手法による単語予測を行って認識率の向上をはかれるように文単位で学習用と評価用に分けた。

そこで、377文を10文毎に区切り、交互に学習用データと評価用データとし、後述(2.3コードブックの作成)するコードブックを用いてコード番号列に変換して実験用データとした。

	学習用	評価用
文番号	001~010	011~020
	021~030	031~040
	:	:
	<u>361~370</u>	<u>371~377</u>
	1487単語	1347単語

## 2.3コードブックの作成

上記の学習用単語中300単語を用いてPWLRを距離尺度としたベクトル量子化を行い256個のベクトルからなるコードブックを作成した。

## 3.HMMの連結

前述した音声データはラベリングされていないので単語内の時間情報がないため、音韻を正確に切り出すことはできない。そこで音韻モデルを連結して単語モデルを構成して学習を行った後、分解して音韻モデルを作る。それらの音韻モデルを連結した単語モデルを用いて単語の認識を行う。

音韻モデルとしては日本語の音韻で検討した結果<sup>[1]</sup>から4状態3ループで最後のアークの組をタイドアークとしたモデルを採用した。また、今回使用するデータは音声区間の前後に無音区間があるのでそのためのモデルとして2状態1ループでタイドアークとしたモデルを採用した。(図1)

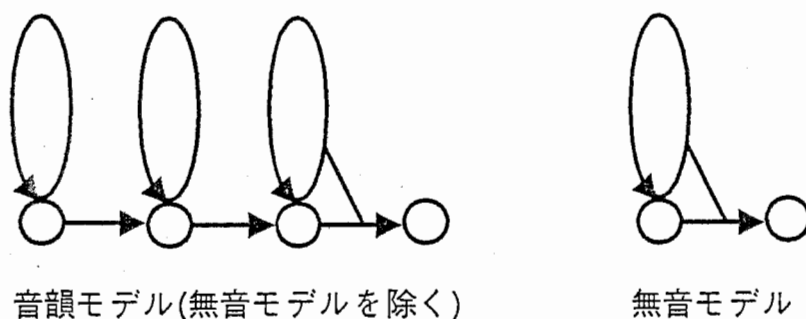


図1 音韻モデル

音韻モデルから単語モデルを作るには、図2のように、音韻モデルでの最終状態を次の音韻モデルの初期状態とすることにより音韻モデルを連結する。従って、単語の音韻数がN個のとき、状態数 $(3N+1)+2$ 、ループ数 $3N+2$ のモデルが単語モデルとしてできる。

## 4.単語音韻変換テーブル

今回使用する音声データはラベリングが施されていないので単語に応じた音韻列を与えて単語モデルを作る。ここでは音韻としては英和辞典<sup>[2]</sup>で使用されている発音記号(付録-2)を用いた。図3に示すような単語のスペルとその単語の発音記号を記載した単語-音韻変換テーブルを用意して単語のスペルに応じた音韻列を与える。このテーブルには発声リスト「キーボード会話300」に



HMMの学習方法は、前述のように音韻モデルを連結した単語モデルを用いて音韻モデルの学習を行う。音韻モデルは、発音記号58個のうち実験用単語に現れる56個と無音モデル1個の計57個である。

## 5.1 HMMパラメータ初期値の設定

音韻モデルのHMMパラメータ(遷移確率、出力確率)の初期値は以下のように設定した。

・遷移確率 $a_{ij}$ の初期値

状態 $i$ から出るアークの数を $n$ としたとき

$$\begin{aligned} a_{i1} + a_{i2} + \dots + a_{in} &= 1 \\ a_{i1} = a_{i2} = \dots = a_{in} &= 1/n \end{aligned}$$

と設定した。すなわち同じ状態 $i$ から出るアークの遷移確率には一様な値を与えた。

・出力確率 $b_{ij,k}$ の初期値

単語の音韻数で学習データを均等分割した区間を音韻区間と仮定して、その中の各コード番号の出現頻度によって各音韻モデルの出力確率を設定した。すなわち、

$$b_{ij,k} = \frac{\text{学習データ中に含まれるコード番号}k\text{の数}}{\text{学習データ中に含まれるデータの総和}}$$

とし、各々のアークに等しい初期値を与えた。

## 5.2 音韻モデルの連結

前述の単語-音韻変換テーブルを用いて得た学習用単語の音韻列に従い単語モデルをつくる。この際、今回使用する音声データは有音区間(単語発声区間)の前後に約300msecの無音区間があるので音韻モデルを連結したモデルの両端に無音モデルをつけた。以下にその例を示す。

単語(スペル)                      音韻モデル列  
able                      →            # ei b l #  
(#は無音モデル記号)

### 5.3 連結学習

前述した学習用単語についてそれぞれ単語モデルを作り、forward-backward アルゴリズム<sup>[3]</sup>によりHMMパラメータ(遷移確率、出力確率)の学習を行う。単語モデルで学習した後、音韻モデルに分解する。

### 5.4 学習回数

図4に、学習過程での学習データの平均対数生起確率、すなわち各々の学習データがHMMから出力される確率の対数値の和を全学習データ長の和で正規化した値の変化の様子を示す。

この図において縦軸が平均対数生起確率、横軸が学習の繰り返し回数である。この図から確率の収束に要する学習回数は10回程度の学習でほぼ収束していることがわかる。そこで、学習回数は10回とした。

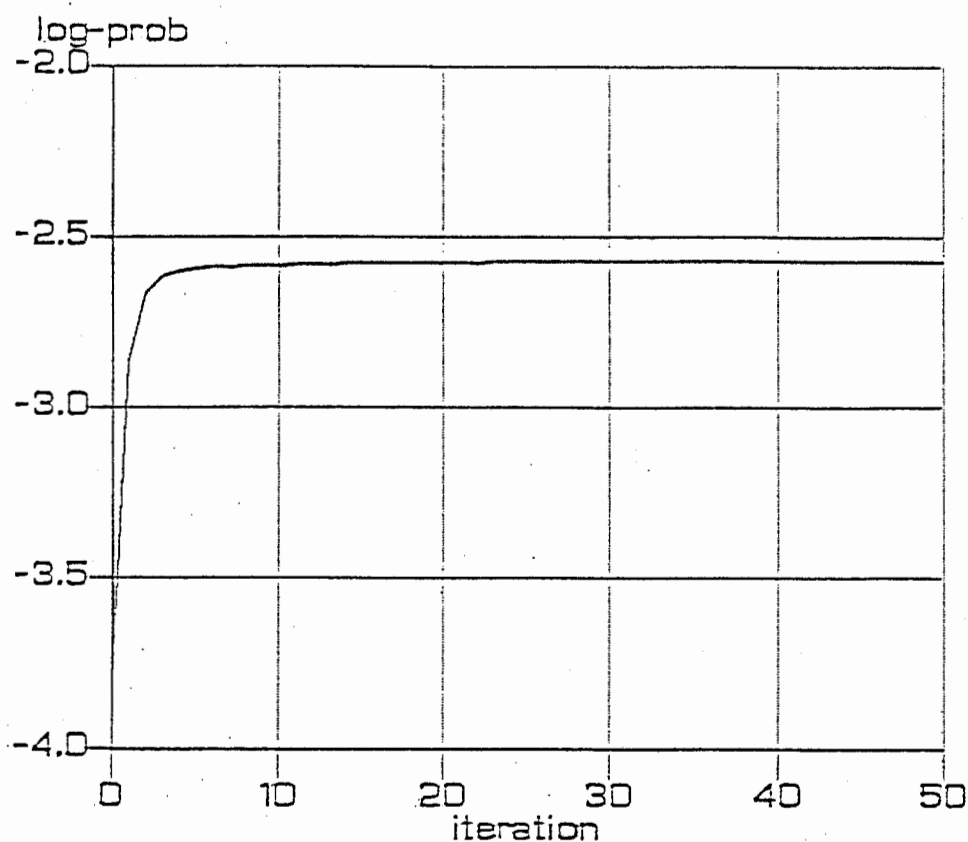


図4 HMMの学習過程における平均対数生起確率の変化

## 6. 認識実験

評価用単語データ(1347語)について、それぞれ、学習された音韻モデルで構成された辞書単語(発声リストに出現する542語。付録-4に示す。)の単語モデルの生起確率を求め、認識実験を行った結果、認識率は80.85%であった。ここで、正しく認識された単語1089語中107語は、for-four,to-too-twoのように発音の等しい同音異義語であった。

また、クローズ・データでの認識率は92.13%であった。

表1に評価単語の生起確率が辞書単語の中で何番目であるかをまとめた累積正答率を示す。この表から4-5位までに約95%,10位までには約97%の正解単語が含まれている事がわかる。

順位	単語数	累積単語数	累積正答率
1	1089	1089	80.85
2	109	1198	88.94
3	38	1236	91.76
4	36	1272	94.43
5	20	1292	95.92
6	4	1296	96.21
7	2	1298	96.36
8	2	1300	96.51
9	4	1304	96.81
10	6	1310	97.25
11位以下	37	1347	

表1 累積正答率

表2には、音韻数別の誤認識を示した。音韻数の多い単語は誤認識率が低くなっている。音韻数の多い単語はサンプル数が少ないのではつきりとは言えないが、音韻数の多いほどそれだけ音韻の組合せが複雑になり他の単語と識別しやすくなると思われる。

誤認識された単語の中で数の多い単語(ここでは3個以上)について、誤認識の割合、学習単語中に含まれるその単語の数、その単語に対して認識した単語について表3にまとめた。bank,I'm,tour,very,wayはすべて、on,any,do,attendはほとんど誤認識されている。これらの単語はほとんどが音韻数2~3で構成されており、また特にbank,tour,way,any等は学習単語にふくまれる数が少な



音韻数	評価単語	誤認識単語	誤認識率
1	159	33	20.75
2	422	90	21.33
3	339	63	18.58
4	144	36	25.00
5	93	19	20.43
6	65	11	16.92
7	32	3	9.38
8	26	2	7.69
9	38	1	2.6
10	27	0	0.0
11	2	0	0.0

表2 音韻数別誤認識率

いことから誤認識率が高くなっていると思われる。その他の単語についても比較的音韻数の少ないものが表に現れている。

#### 無音区間の削除

今回使用した音声データは単語ファイルに切り出す際、約300msecの無音区間を有声区間の前後に付加している。音韻数の少ない単語については有声区間に対する無音区間の割合が大きくなっている。そこで無音区間を音声データの始点、終点それぞれから削除した音声データを使って学習、認識実験を行った。無音区間の削除時間を150msec,240msec,300msecとした場合の結果を表4に示す。ここで削除時間240msecの場合には無音モデルは付加していない。

削除時間300msecの場合だけ他に比べて認識率が低くなっていて、他はほぼ同じ認識率であり無音区間の長さが認識率に関係しているとは言えない。また、削除区間300msecの場合、音声データが単語切り出しプログラムによって正確に切り出されているとは限らないので、単語の語頭、語尾が削られて認識率を悪くしているとも考えられる。

#### 7.まとめ

HMMの音韻モデルを連結した単語モデルを用いた手法で、ラベル付けのない音声データの認識実験を行って約80%の認識率を得た。この結果と言語情報が

単語	誤認識数 全数	学習単語中 の単語数	認識した単語 ( )内は度数
bank	5/5	0	thank(3),think(2)
I'm	3/3	7	from(2),time(1)
tour	6/6	1	for(2),record(2),before(1),cover(1)
very	7/7	7	vary(6),interpreter(1)
way	3/3	2	away(2),while(1)
on	7/8	7	foreign(2),join(2),one(2),from(1)
any	6/7	2	ATR(4),japanese(1),sending(1)
do	15/18	7	to(12),june(3)
attend	4/5	3	intend(2),content(1),understand(1)
pay	5/8	6	take(4),paid(1)
no	3/5	1	double(2),number(1)
this	3/5	2	Ms.(1),office(1),tours(1)
the*	6/11	11	fee(2),me(2),be(1),Miki(1)
if	4/9	5	it(1),is(1),live(1),six(1)
and	10/23	32	end(4),attend(2),send(2),chance(1),sent(1)
so	3/8	3	social(2),Tomoko(1)
will	7/19	25	well(7)
I	15/47	37	I'd(4),five(3),like(3),I've(2),apply(1),by(1),I'm(1)
have	4/13	12	however(2),happy(1)
in	5/17	20	been(2),begin(1),Heian(1),seen(1)
be	3/17	25	before(1),fee(1),need(1)
you	9/65	50	Yukiko(4),future(2),Yamamura(2),used(1)

表3 誤認識の多い単語  
(\* [ði:]のthe)

らの単語予測結果を合わせることにより認識率はかなり向上すると思われる。また、出現頻度の高い単語は音韻モデルの連結でなく初めから単語モデルを作る等さらに認識結果を解析してアルゴリズムの改良を行うことや、セパレート・ベクトル量子化を用いた動的特徴量の導入、ファジー・ベクトル量子

削除区間	0msec	150msec	240msec	300msec
認識率	80.85%	80.92%	77.06%	81.66%

表4 無音区間を削除した時の認識率

化の導入、デュレーション・コントロール等により認識率の向上がはかれるはずである。

#### 参考文献

[1]花沢、川端、鹿野：[Hidden Markov Modelを用いた日本語有声破裂音の識別] ,ATR Technical Report TR-I-0018(December,1987)

[2]「新英和中辞典第5版」, 研究社(1985)

[3]Levinson,S.E, Rabiner,L.R. and Sondhi,M.M. : “An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition“ , The Bell System Technical Journal,vol.62,No.4(April,1983)

付録-1 「キーボード会話300」の英文例

話者A,Bによる電話会話文

A Hello.

B Hello.

A I would like to attend the conference.

B Would you like us to send you the details?

A Yes, please. Could you tell me how I make an application?

B We will send you an application form.

A Thank you. I don't intend to give a paper, so is it all right if I just attend as a listener?

B Yes, it is. You can register and pay at the conference in August.

A How much is the registration fee?

B \$100 U.S.

A Is it possible to pay in yen?

B I'm sorry, but we would like you to pay in dollars.

A Is it \$100 for three days?

B That's right, hotel and food expenses are extra.

A I see. If I can only attend on the first day, is it possible to make a cancellation?

B Do you mean be entitled to a refund?

A Yes, is it possible to get 2/3 of the fee back?

B I'm sorry, but it is not possible.

A I see. What language will the conference be in?

B Both English and Japanese.

A I do not have an English capability. Will simultaneous interpretation be provided?

B Yes.

A I would like to converse with some of the participants. Will I be able to have an interpreter at my disposal?

B Yes.

A I see. Thank you very much.

B Do you have any other queries?

A No, thank you.

B O.K. Goodbye.

付録-2 音韻表(発音記号)

発音記号	実験用記号	単語例
i:	i:	be <u>ɪ</u>
i	i	bee <u>n</u>
e	e	ge <u>t</u>
æ	a1	a <u>n</u>
a:	a:	fa <u>th</u> er
a	a	no <u>t</u>
ɔ:	o1:	a <u>ll</u>
u:	u:	do <u>o</u>
ʊ	u	cou <u>ld</u>
ʌ	a2	bu <u>t</u>
ɝ:	e2:	hea <u>rd</u>
ə	e1	he <u>ll</u> o
ɝ	e2	af <u>te</u> r
o	o	om <u>i</u> t
ei	ei	sta <u>y</u>
ai	ai	by <u>ɪ</u>
ɔi	o1i	jo <u>in</u>
au	au	ho <u>w</u>
ou	ou	ho <u>p</u> e
ju:	ju:	yo <u>u</u>
ju	ju	popu <u>l</u> ar
iə	ie2	hea <u>r</u>
ɛə	Ee2	the <u>i</u> r
aə	ae2	ca <u>r</u>
ɔə	o1e2	fo <u>r</u>
uə	ue2	to <u>u</u> r
juə	jue2	yo <u>u</u> r
aiə	aie2	fi <u>r</u> e
auə	ae2	ho <u>u</u> r
p	p	pl <u>a</u> n
b	b	ba <u>n</u> k
t	t	ta <u>k</u> e
d	d	da <u>y</u>
k	k	ca <u>n</u>
g	g	go
f	f	fi <u>v</u> e
v	v	gi <u>v</u> e
θ	th1	tha <u>n</u> k

ð	th2	<u>th</u> an
s	s	<u>s</u> ix
z	z	<u>z</u> ero
ʃ	sh	<u>d</u> ash
ʒ	zy	pleas <u>u</u> re
h	h	<u>h</u> er
ts	ts	<u>c</u> ats
dz	dz	<u>r</u> eads
tr	tr	<u>t</u> ry
dr	dr	<u>d</u> rink
tʃ	tsh	teach <u>e</u> r
dʒ	dzy	<u>j</u> uly
m	m	<u>m</u> ake
n	n	<u>n</u> ow
ŋ	ng	morning <u>g</u>
l	l	<u>l</u> ike
r	r	pres <u>e</u> nt
j	j	<u>y</u> es
w	w	<u>w</u> eek

付録-3 実験データに含まれる音韻数

A. 学習用データ

音韻	音韻数	音韻	音韻数	音韻	音韻数	音韻	音韻数
i:	155	i	408	e	195	a1	151
a:	1	a	93	o1:	32	u:	72
u	33	a2	78	e2:	27	e1	357
e2	62	E	3	o	26	ei	104
ai	129	o1i	5	au	31	ou	79
ju:	58	ju	3	ie2	5	Ee2	16
ae2	20	o1e2	39	ue2	1	jue2	10
aie2	1	ae2	0	p	99	b	92
t	345	d	176	k	221	g	35
f	126	v	95	th1	56	th2	128
s	268	z	127	sh	48	zy	0
h	74	ts	13	dz	22	tr	9
dr	17	tsh	27	dzy	30	m	147
n	411	ng	67	l	196	r	159
j	36	w	103	#	2974		

B. 評価用データ

音韻	音韻数	音韻	音韻数	音韻	音韻数	音韻	音韻数
i:	133	i	315	e	140	a1	185
a:	0	a	60	o1:	39	u:	82
u	29	a2	90	e2:	23	e1	278
e2	56	E	0	o	13	ei	97
ai	117	o1i	1	au	15	ou	70
ju:	71	ju	1	ie2	2	Ee2	7
ae2	22	o1e2	32	ue2	7	jue2	11
aie2	0	ae2	0	p	75	b	84
t	313	d	177	k	192	g	33
f	130	v	52	th1	36	th2	113
s	240	z	94	sh	53	zy	0
h	59	ts	12	dz	1	tr	17
dr	9	tsh	28	dzy	27	m	99
n	340	ng	61	l	156	r	106
j	22	w	92	#	2694		



a  
able  
about  
absorb  
abstract  
accept  
accepting  
according  
account  
acquire  
acquisition  
actually  
address  
addressing  
advice  
afraid  
after  
afternoon  
agenda  
ahead  
all  
allowed  
along  
also  
am  
american  
an  
and  
another  
any  
anyone  
anything  
application  
apply  
applying  
appreciate  
arata  
are  
as  
ask  
aspects  
assist  
assistance  
at  
atr  
attend  
attendance  
attending  
august  
average  
awaiting  
away  
back  
bank  
be  
been  
before

begin  
being  
between  
book  
booking  
both  
branch  
breakdown  
bus  
but  
by  
call  
calling  
can  
cancel  
cancellation  
cannot  
capability  
car  
card  
cases  
castle  
catch  
certainly  
chance  
charge  
chaya-machi  
check  
cheque  
children  
chome  
city  
coming  
company  
completed  
conference  
confidence  
contact  
contain  
content  
converse  
correct  
cost  
could  
couple  
cover  
credit  
currently  
dash  
daughter  
daughter's  
day  
days  
deadline  
deal  
decide  
depending  
depends  
detailed  
details

development  
discounts  
discuss  
discussions  
disposal  
dissemination  
distance  
do  
documents  
does  
doing  
dojima  
dollars  
don't  
double  
each  
ehime  
eight  
eighth  
eighty  
either  
eleven  
else  
enclose  
end  
english  
entitled  
etsuko  
evening  
everybody  
exactly  
expense  
expenses  
experience  
express  
extra  
facet  
faster  
fastest  
fee  
fees  
field  
fifteen  
fifth  
fill  
findings  
fine  
finished  
first  
five  
follow  
food  
for  
foreign  
form  
forty  
four  
fourth  
free  
freely

from  
full  
further  
future  
gathering  
get  
give  
glad  
go  
golden  
good  
goodbye  
great  
guest  
guests  
hall  
handle  
handling  
happy  
has  
hasegawa  
have  
haven't  
hear  
heard  
heian  
held  
hello  
help  
helpful  
her  
higashi-ku  
hiroyuki  
hotel  
hotels  
how  
however  
hundred  
hundreds  
i  
i'd  
i'll  
i'm  
i've  
if  
in  
include  
included  
indicate  
information  
intend  
interested  
international  
interpretation  
interpreter  
interpreting  
involved  
is  
isn't  
it

it's  
its  
itself  
japanese  
join  
july  
june  
just  
kida  
kind  
kita-ku  
kitaoji  
know  
kyoto  
kyoto's  
laboratory  
language  
last  
late  
least  
like  
linguistics  
listener  
live  
lodging  
long  
longer  
lunch  
major  
make  
mariko  
materials  
matsumoto  
matsuyama  
may  
me  
mean  
methods  
miki  
mikiko  
minamiyogo-cho  
minato-ku  
minutes  
money  
month  
more  
morning  
most  
mr  
ms.  
much  
must  
my  
myself  
name  
names  
near  
nearest  
necessary  
need

new  
night  
nijo  
nine  
nineteen  
no  
not  
now  
number  
o  
objective  
observe  
of  
office  
ok  
on  
one  
only  
open  
opportunities  
or  
order  
organized  
osaka  
other  
our  
out  
own  
package  
paid  
paper  
papers  
part  
participants  
participate  
pay  
paying  
payment  
percent  
percents  
photos  
pity  
plan  
planned  
please  
plenty  
possible  
prefer  
present  
presentations  
presented  
presenting  
price  
prices  
prince  
probably  
procedures  
proceedings  
process  
program

provided  
publish  
queries  
question  
questions  
quicker  
range  
rather  
read  
ready  
received  
receiving  
record  
refund  
refunded  
register  
registering  
registration  
research  
researching  
reservation  
reservations  
reserve  
reveals  
ride  
right  
room  
royal  
rule  
same  
scheduled  
seat  
see  
seems  
seen  
send  
sending  
sent  
september  
service  
session  
seven  
shall  
she  
shimbashi  
shiromi  
should  
shrine  
sightseeing  
simultaneous  
since  
single  
six  
sixty  
small  
so  
social  
soliciting  
some  
something

soon  
sorry  
speak  
speaker  
speakers  
speakers'  
specializing  
speeches  
stage  
start  
station  
stay  
still  
stranger  
student  
students  
studying  
submission  
submissions  
submit  
submitting  
subway  
suggest  
sumitomo  
sure  
takarazuka  
take  
takenaka  
takes  
taking  
talk  
talking  
taro  
taxi  
technical  
telephone  
telephonic  
telephony  
tell  
temple  
ten  
than  
thank  
thanks  
that  
that's  
the  
their  
then  
theory  
there  
therefore  
these  
they  
thing  
think  
thirds  
thirty  
this  
those



thousands  
three  
time  
to  
today  
tokyo  
tomoko  
too  
topics  
tour  
tours  
traffic  
transfer  
tsurumidai  
twenty  
twin  
two  
type  
u.s.  
unable  
understand  
university  
us  
use  
used  
various  
vary  
venue  
very  
vicinity  
waiting  
was  
watanabe  
way  
ways  
we  
we'd  
we'll  
weeks  
welcome  
well  
what  
when  
where  
which  
while  
who  
who's  
will  
wish  
with  
won't  
would  
yamamura  
year  
yen  
yes  
yet  
you  
you'll

you're  
your  
yukiko  
zero