Conditional Probabilities for IBM Voice Browser 2.0 Alpha and Alphanumeric Recognition

TR 29.3498
April 25, 2002

Lenora E. Wright
Matthew W. Hartley
James R. Lewis

IBM Voice Systems

West Palm Beach, Florida

Abstract

The ability to recognize spoken numbers and letters of the English alphabet is an important property of speech engines used in telephony applications. Due to the acoustic similarity of this type of input, though, it is a difficult and potentially error-prone process. This report provides preliminary recognition accuracy data for developers who work with Version 2 of the IBM WebSphere Speech Browser so they can design more effective recovery mechanisms for misrecognitions of spoken letters and numbers.

ITIRC Keywords

Voice spelling
Alphabetic entry
Alphanumeric entry
Error recovery
Correction of misrecognitions
Speech recognition
Speech browser
Voice browser
Telephony applications

Contents

Introduction

In a previous report (Hartley & Lewis, 2001), we described recognition accuracy data for spoken letters using the IBM WebSphere[1] Speech Browser Version 1.0. The motivation for this work was that telephony applications cannot not provide visual feedback when users need the system to recognize spoken letters for the purpose of spelling names (or other words) that are out-of-vocabulary or hard to recognize. There are no standard mechanisms for spelling with a telephone keypad, and those that exist all have usability problems of one type or another (Lewis, Potosnak, and Magyar, 1997), especially when the keypad is part of the telephone handset. Speech recognition systems that provide *n*-best lists (a list that contains the most likely word or letter for a given spoken input plus the *n* best alternatives -- see Balentine & Morgan, 1999 for more details) require greater resources than systems that do not. For this reason, some systems provide *n*-best lists for spoken input (including letters), but others do not.

When a system does not provide an *n*-best list, an alternative approach is to determine empirically the distribution of misrecognitions among the letters of the alphabet and to use the data from that distribution to guide error recovery schemes.

The purpose of this report is to describe preliminary misrecognition data for spoken U.S. English letters and numbers using the IBM WebSphere Speech Browser (Version 2.0). The work in this report builds on our previous work, but extends it to Version 2.0 of the browser and includes an assessment of both alpha (letters only) and alphanumeric (letters and numbers) patterns of recognition.

[1] WebSphere is a trademark or registered trademark of International Business Machines Corp. IBM is a registered trademark of International Business Machines Corp.

Method

Participants

The participants in this study were twenty IBM employees. The sample included ten males and ten females. The mean age of the sample was 36 with a standard deviation of 8.28 (ranging from 26 to 55). All but three participants spoke with standard American English accents. One male participant was from Thailand and one male and one female participant were from China.

Materials

Participants used their phones at work to place calls to Cisco 2600 gateway, connecting to a voice browser (GA version of the IBM Voice Server 2.0) running a VXML program created for the purpose of collecting user speech. The speakers' audio was captured exactly as it came from the gateway for maximum validity. After capture, the files were edited with Cool Edit 2000 to create a separate file for each speaker's pronunciation of each letter of the English alphabet and each digit.

Procedure

The recordings were played into an accuracy-testing program (using the GA version of the IBM Voice Server SDK 2.0 running on a Dell Dimension XPS T550). This procedure was replicated three times to check for outliers or other indications of unstable data. The output of the program provided the information required for estimating misrecognition rates among the spoken letters and digits.

Results

Misrecognitions

Tables 1 and 2 show the raw count data and counts converted to rates (correct recognition rates along the diagonal, misrecognition rates in off-diagonal cells) for the evaluation of alpha recognition. With twenty speakers and three trials per recording, there are 60 opportunities for error for each letter. The number of times that the system returned a silence timeout, low confidence recognition, or a substituted command appear in their respective rows. Tables 3 and 4 show the same analyses for the alphanumeric data.

Table 1. Raw Count Data for Alpha

Said:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**Total**
Returned A:	43				1																					2	46
B:		45		1	2																						48
C:			57		3																	4	1			4	69
D:		2		41	4																	4					51
E:		6		10	48		2															1					67
F:						51													4					3			58
G:				3			53													4							60
H:								57																			57
I:									52									3	1								56
J:							3			55																3	61
K:	11	3								2	55					3				3		3					80
L:												54															54
M:												2	57	3													62
N:	1												3	55													59
O:												3			57												60
P:		4									2					48				1		1					56
Q:											3						58										61
R:									8									55									63
S:	2					4		3											54	1							64
T:				5			2			3						9				48	1						68
U:																					59	1					60
V:																						30				6	36
W:																							54	1			55
X:																								56	1		57
Y:																									59	1	60
Z:																						16				44	60
Timeout:						2									3		1	2	1								9
Low Confidence	3		3		2	3						1		2			1						5				20
Command:																				3							3
Total:	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	1560

Table 2. Rate Data for Alpha

Said:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
Returned A:	0.72				0.02																					0.03
B:		0.75		0.02	0.03																					
C:			0.95		0.05																	0.07	0.02			0.07
D:		0.03		0.68	0.07																	0.07				
E:		0.10		0.17	0.80		0.03															0.02				
F:						0.85													0.07					0.05		
G:				0.05			0.88													0.07						
H:								0.95																		
I:									0.87									0.05	0.02							
J:							0.05			0.92																0.05
K:	0.18	0.05								0.03	0.92					0.05				0.05		0.05				
L:												0.90														
M:												0.03	0.95	0.05												
N:	0.02												0.05	0.92												
O:												0.05			0.95											
P:		0.07									0.03					0.80				0.02		0.02				
Q:											0.05						0.97									
R:									0.13									0.92								
S:	0.03					0.07		0.05											0.90	0.02						
T:				0.08			0.03			0.05						0.15				0.80	0.02					
U:																					0.98	0.02				
V:																						0.50				0.10
W:																							0.90	0.02		
X:																								0.93	0.02	
Y:																									0.98	0.02
Z:																						0.27				0.73
Timeout						0.03									0.05		0.02	0.03	0.02							
Low Confidence	0.05		0.05		0.03	0.05						0.02		0.03			0.02						0.08			
Commands																				0.05						
Total:	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3. Raw Count Data for Alphanumeric

Said:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
Returned A:	41				2																					
B:		45		2																						
C:			55		3																	3				3
D:				40	3																	4				1
E:		6	2	8	49																	2				
F:						51													6					3		
G:				3			51													3						
H:								57																		
I:									51									3								
J:							3			54																3
K:	6	3								2	52					3				3		3				
L:												54														
M:												1	57	3												
N:	2												3	57												
O:												2			18											
P:		6									3					49						3				
Q:											3						60									
R:									6									55								
S:	1					3		3											53							
T:				5			3			3						8				48						
U:																					59					
V:																						30				9
W:																							56			
X:																								54		
Y:																									60	
Z:																						15				38
0															39											
1												1														
2							3				2									3						
3	3			2																						3
4																										
5									3																	
6			2																					3		
7																										
8	4																									3
9																										
Timeout:						2				1					3			2	1							
Low Confidence	3		1		3	4						2									1		4			
Command:																				3						
Total:	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60	60

Table 3. Raw Count Data for Alphanumeric (cont.)

Said:	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
Returned A:											43
B:											47
C:											64
D:											48
E:											67
F:											60
G:											57
H:											57
I:						3					57
J:											60
K:											72
L:											54
M:											61
N:									3		65
O:											20
P:											61
Q:			1								64
R:											61
S:											60
T:									2		69
U:											59
V:											39
W:											56
X:											54
Y:											60
Z:											53
0	60										99
1		60									61
2			56								64
3				60							68
4			3		59						62
5						57					60
6							60				65
7								58			58
8									53		60
9										60	60
Timeout:					1						10
Low Confidence								2	2		22
Command:											3
Total:	60	60	60	60	60	60	60	60	60	60	2160

Table 4. Rate Data for Alphanumeric

Said:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
Returned A:	0.68				0.03																					
B:		0.75		0.03																						
C:			0.92		0.05																	0.05				0.05
D:				0.67	0.05																	0.07				0.02
E:		0.10	0.03	0.13	0.82																	0.03				
F:						0.85													0.10					0.05		
G:				0.05			0.85													0.05						
H:								0.95																		
I:									0.85									0.05								
J:							0.05			0.90																0.05
K:	0.10	0.05								0.03	0.87					0.05				0.05		0.05				
L:												0.90														
M:												0.02	0.95	0.05												
N:	0.03												0.05	0.95												
O:												0.03			0.30											
P:		0.10									0.05					0.82						0.05				
Q:											0.05						1.00									
R:									0.10									0.92								
S:	0.02					0.05		0.05											0.88							
T:				0.08			0.05			0.05						0.13				0.80						
U:																					0.98					
V:																						0.50				0.15
W:																							0.93			
X:																								0.90		
Y:																									1.00	
Z:																						0.25				0.63
0															0.65											
1												0.02														
2							0.05				0.03									0.05						
3	0.05			0.03																						0.05
4																										
5									0.05																	
6			0.03																					0.05		
7																										
8	0.07																									0.05
9																										
Timeout:						0.03				0.02					0.05			0.03	0.02							
Low Confidence	0.05		0.02		0.05	0.07						0.03									0.02		0.07			
Command:																				0.05						
Total:	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. Rate Data for Alphanumeric (cont.)

Said:	0	1	2	3	4	5	6	7	8	9
Returned A:										
B:										
C:										
D:										
E:										
F:										
G:										
H:										
I:						0.05				
J:										
K:										
L:										
M:										
N:									0.05	
O:										
P:										
Q:			0.02							
R:										
S:										
T:									0.03	
U:										
V:										
W:										
X:										
Y:										
Z:										
0	1.00									
1		1.00								
2			0.93							
3				1.00						
4			0.05		0.98					
5						0.95				
6							1.00			
7								0.97		
8									0.88	
9										1.00
Timeout:					0.02					
Low Confidence								0.03	0.03	
Command:										
Total:	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Conditional Probabilities

For the data to be useful for the purpose of intelligent error recovery, it is important to compute them as conditional probabilities ($P(A)|A$). The reason for this is that if a user rejects a returned letter or set of letters as being incorrect, the developer (and by extension, the system) can only know the returned letter(s) -- not the spoken letter(s). What the developer needs to know about the returned letter is the probability distribution of all the other letters given that returned letter -- their conditional probabilities -- the ratio of the number of times the system returned a given letter given that the speaker said that letter divided by the total number of times the system returned that letter. For the purpose of error recovery, this is much more important information than the standard measurement of recognition accuracy (the ratio of number of times a recognizer returns a letter correctly divided by the number of times speakers said that letter).

For example, the standard recognition accuracy of "A" (from the A-A cell of Table 2) was a fairly low 72%. Often, when a speaker said "A" the system returned "K" (18% of the time). On the other hand, the conditional probability that the speaker had said "A" when the system returned an "A" (from the A-A cell of Table 5, shown below) was a fairly high 93%. In other words, when a speaker said "A" the system often misrecognized it, but if the system returned an "A" there was a very good chance that the speaker actually had said "A".

On the other hand, the standard recognition accuracy of "K" was a fairly high 92%. When a speaker said "K", the system was very likely to return "K". However, the conditional probability that a returned "K" occurred when the speaker actually said "K" was a fairly low 69% because the system was likely to produce a "K" when the speaker actually said "A". This means that a voice-spelling application should have fairly low confidence that a returned "K" is evidence of a spoken "K".

Tables 5 and 6 show the conditional probabilities for alpha and alphanumeric recognition computed using the data from the previous tables.

Table 5. Conditional Probabilities for Alpha

Said:	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
Returned A:	0.93				0.02																					0.04
B:		0.94		0.02	0.04																					
C:			0.83		0.04																	0.06	0.01			0.06
D:		0.04		0.80	0.08																	0.08				
E:		0.09		0.15	0.72		0.03															0.01				
F:						0.88													0.07					0.05		
G:				0.05			0.88													0.07						
H:								1.00																		
I:									0.93									0.05	0.02							
J:							0.05			0.90																0.05
K:	0.14	0.04								0.03	0.69					0.04				0.04		0.04				
L:												1.00														
M:												0.03	0.92	0.05												
N:	0.02												0.05	0.93												
O:												0.05			0.95											
P:		0.07									0.04					0.86				0.02		0.02				
Q:											0.05						0.95									
R:									0.13									0.87								
S:	0.03					0.06		0.05											0.84	0.02						
T:				0.07			0.03			0.04						0.13				0.71	0.01					
U:																					0.98	0.02				
V:																						0.83				0.17
W:																							0.98	0.02		
X:																								0.98	0.02	
Y:																									0.98	0.02
Z:																						0.27				0.73
Timeout:						0.22									0.33		0.11	0.22	0.11							
Low Confidence	0.15		0.15		0.10	0.15						0.05		0.10			0.05						0.25			
Command:																				1.00						

Table 6. Conditional Probabilities for Alphanumeric

Said:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Returned A:	0.95				0.05																					
B:		0.96		0.04																						
C:			0.86		0.05																	0.05				0.05
D:				0.83	0.06																	0.08				0.02
E:		0.09	0.03	0.12	0.73																	0.03				
F:						0.85													0.10					0.05		
G:				0.05			0.89													0.05						
H:								1.00																		
I:									0.89									0.05								
J:							0.05			0.90																0.05
K:	0.08	0.04								0.03	0.72					0.04				0.04		0.04				
L:												1.00														
M:												0.02	0.93	0.05												
N:	0.03												0.05	0.88												
O:												0.10			0.90											
P:		0.10									0.05					0.80						0.05				
Q:											0.05						0.94									
R:									0.10									0.90								
S:	0.02					0.05		0.05											0.88							
T:				0.07			0.04			0.04						0.12				0.70						
U:																					1.00					
V:																						0.77				0.23
W:																							1.00			
X:																								1.00		
Y:																									1.00	
Z:																						0.28				0.72
0															0.39											
1												0.02														
2							0.05				0.03									0.05						
3	0.04			0.03																						0.04
4																										
5									0.05																	
6			0.03																					0.05		
7																										
8	0.07																									0.05
9																										
Timeout:						0.20				0.10					0.30			0.20	0.10							
Low Confidence	0.14		0.05		0.14	0.18						0.09									0.05		0.18			
Command:																				1.00						

Table 6. Conditional Probabilities for Alphanumeric (cont.)

Said:	0	1	2	3	4	5	6	7	8	9
Returned A:										
B:										
C:										
D:										
E:										
F:										
G:										
H:										
I:						0.05				
J:										
K:										
L:										
M:										
N:									0.05	
O:										
P:										
Q:			0.02							
R:										
S:										
T:									0.03	
U:										
V:										
W:										
X:										
Y:										
Z:										
0	0.61									
1		0.98								
2			0.88							
3				0.88						
4			0.05		0.95					
5						0.95				
6							0.92			
7								1.00		
8									0.88	
9										1.00
Timeout:					0.10					
Low Confidence								0.09	0.09	
Command:										

Most Likely Substitutions

From the table of conditional probabilities (Tables 5 and 6), it is possible to develop a list of the most likely substitutions for a given letter. This list appears in Tables 7 and 8. In these tables, capitalized bold letters indicate substitution probabilities that exceeded .10. Standard face letters indicate substitutions that occurred during the study, but had substitution probabilities less than .10. For example, if a user rejects a returned "V" in a voice-spelling application, then the letter most likely to have actually been spoken is "Z".

Six of the letters in Table 7 (alpha) have only one substitution for which the probability of substitution exceeded 10%. Twenty of the letters don't have any substitutes for which the probability of substitution exceeded 10%, and eight of those didn't have any substitutions at all. This means that whenever the system returned these letters (A, B, H, L, U, W, X, and Y), the developer could have very high confidence that the speaker actually said that letter.

Table 7. Most Likely Substitutions for Alpha

Returned		**Likely Substituted For**	
A			
B			
C		v	z
D		e	v
E		**D**	b
F		s	x
G		t	d
H			
I		r	
J		g	z
K		**A**	
L			
M		n	
N		m	
O		l	
P		b	
Q		k	
R		**I**	
S		f	h
T		**P**	d
U			
V		**Z**	
W			
X			
Y			
Z		**V**	

In Table 8 (alphanumeric), eight letters and one number have one substitution for which the probability of substitution exceeded 10%. Eighteen of the letters and nine numbers don't have any substitutes for which the probability of substitution exceeded 10%. Seven letters and four numbers didn't have any substitutions at all. This means that whenever the system returned these letters or numbers (B, H, L, U, W, X, Y, 1, 3, 7, or 9), the developer could have very high confidence that this is what the speaker actually said. The introduction of digits in this evaluation did not lead to any cases in which a digit substantially interfered with the recognition of a letter. The primary cause of digit misrecognition was the substitution of the letter 'O' for the digit '0'.

Table 8. Most Likely Substitutions for Alphanumeric

Returned		**Likely Substituted For**		
A		e		
B				
C		e	v	z
D		e	v	
E		**D**	b	
F		**S**	x	
G		t	d	
H				
I		r	5	
J		g	z	
K		a		
L				
M		n		
N		m		
O		**L**		
P		**B**	k	v
Q		k	v	
R		**I**		
S		f	h	
T		**P**	d	
U				
V		**Z**		
W				
X				
Y				
Z		**V**		
0		**O**		
1				
2		g	t	
3				
4		2		
5		i		
6		x		
7				
8		a	z	
9				

Comparison of Results

Table 9 summarizes the results for highly likely substitutions ($p > 10\%$) for our previous study and the two current studies, and indicates the consensus for likely substitutions. The recognition accuracy for the current studies is clearly better than for the previous study, as indicated by fewer entries in the cells of Table 9 for those studies. A pair of *t*-tests on the difference scores for the letters of the alphabet across the alpha studies confirmed this observation (Alpha1 vs. Alpha2: $t(25) = 3.82$, $p = .001$), and indicated that the observed decline in overall accuracy for letters of the alphabet caused by adding digits to the grammar (86% for Alpha2, 83% for Alphanumeric) was not statistically significant (Alpha2 vs. Alphanumeric: $t(25) = 1.64$, $p = .11$).

Table 9. Comparison of Highly Likely Substitutions

	Alpha 1	Alpha 2	Alphanumeric	Consensus
A	I	-	-	-
B	-	-	-	-
C	V	-	-	-
D	E	-	-	-
E	V	D	D	D
F	S	-	S	S
G	P, T	-	-	-
H	-	-	-	-
I	-	-	-	-
J	-	-	-	-
K	A	A	-	A
L	-	-	-	-
M	N	-	-	-
N	-	-	-	-
O	L	-	L	L
P	-	-	B	B
Q	U	-	-	-
R	I	I	I	I
S	F	-	-	-
T	D, E	P	P	P
U	-	-	-	-
V	Z	Z	Z	Z
W	-	-	-	-
X	-	-	-	-
Y	-	-	-	-
Z	V	V	V	V
0	na	na	O	O
1	na	na	-	-
2	na	na	-	-
3	na	na	-	-
4	na	na	-	-
5	na	na	-	-
6	na	na	-	-
7	na	na	-	-
8	na	na	-	-
9	na	na	-	-

Discussion

Overall, the data indicate significant improvements in voice spelling accuracy for WebSphere Speech Browser Version 2.0 relative to Version 1.0. The conditional probabilities provided in this report show that the patterns of substitution are somewhat different for Version 2.0, so it will be important to update any programs using our previously reported conditional probabilities.

Adding numbers to the alpha grammar did not significantly reduce the recognition accuracy of letter recognition, but did lead to a potentially troublesome substitution of the letter ‘O’ for ‘0’. Developers should be aware of this possibility, and should take advantage of any pattern in the data they are trying to acquire with an alphanumeric grammar that can help them substitute ‘0’ for a returned ‘O’ when they know the character in that position must be numeric.

It is important to note that the conditional probabilities and most likely substitutions should not be the only data that developers bring to bear on the problem of developing their error recovery schemes. If users spell English words, then it would be possible to use published tables of unigram and digram frequencies (Card, Moran, & Newell, 1983) to supplement the conditional probabilities presented in this report. If users "spell" strings that are not English words (for example, codes using alphabetic characters, such as part numbers or membership numbers), then developers should use whatever they know about the characteristics of these strings to guide efficient error recovery.

References

Balentine, B., & Morgan, D. P. (1999). *How to build a speech recognition application: A style guide for telephony dialogues* . San Ramon, CA: Enterprise Integration Group.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hartley, M. W., & Lewis, J. R. (2001). *Conditional probabilities for IBM Voice Browser recognition of letters of the alphabet* (Tech. Report 29.3421). West Palm Beach, FL: International Business Machines Corp.

Lewis, J. R., Potosnak, K. M., & Magyar, R. (1997). Keys and keyboards. In M. Helander, T. K. Landauer, and P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 1285-1315). Amsterdam: North-Holland.