Proceedings of the Institute of Acoustics

HUMAN FACTORS AND AUDITORY PROCESSING IN THE RECOGNITION OF PROPER NAMES

James Monaghan
University of Hertfordshire, UK

1. INTRODUCTION

The research behind this paper arose from a pilot project to investigate the best way of developing an automated captioning system for use by TV editors working on the broadcasting of proceedings in the British Houses of Parliament. There are several difficulties in this, unknown to other assemblies.

Because of the rather *ad hoc* development of the broadcasting of the proceedings of Parliament and the suspicion of TV coverage on the part of the members, the broadcasting facilities that have been provided can only be described as inadequate. Considering the high levels of ambient noise, the sound capture is very patchy and limited to only a few microphones with poor coverage, and the visual cues are haphazard as the cameras are few and can only be operated under severe restrictions. Unlike some assemblies, there is no fixed seating plan and there are no individual microphones for each member.

In fact, Members of Parliament, when they rise to speak in the House, must "catch the Speaker's eye" and then may only speak when nominated by the Speaker. The Speaker does not have a microphone in front of him or her and the sound of the name called is often difficult to pick up. The MPs take their cue from making eye contact, but at present, the editor in the studio is dependent on hearing the Speaker call the name of the next Member of Parliament to speak, and, having identified this correctly, he must access a database to produce a printed caption for the TV screen, showing the member's name, party and constituency. An important factor in this process is the speed at which this can be done accurately, and only too often there is an unacceptable delay due to the initial problem of recognition of the name.

2. PILOT STUDY

Before the present experiment various techniques were tried on various groups, in order to get some idea of how bad the misrecognition problem was likely to be. A small pilot study was carried out on a group of 20 people. The names, in random order (but each occurring only once), were read out 'live' (i.e. not using a tape recorder) and the subjects were asked to transcribe what they heard. Although the reading was done live, with the reader in the same room as the subjects, the possibility of visible clues (from lip-reading) was excluded, as the reader was out of sight of the subjects.

The subjects carried out the experiment in three groups. Some members were familiar with the reader's voice, while others had never heard it before. In all cases, the subjects were told that the list was of MPs' names. The error rate across all subjects (those familiar with the speaker as well as those unfamiliar) was 14.3% - that is, the subjects misrecognised one name in seven.

Following on from this pilot study, we expected that a more formal set of experiments, over a wider range of subjects would result in a very similar rate of misrecognition. However, this was not the case, as the following sections will illustrate. The error rate was greatly increased in the more rigorous experiments. Some possible explanations for this are suggested in section 2.2. below.

3. METHOD

Our initial strategy was to see if we could identify the names most likely to cause confusion with a view to setting up an intelligent knowledge-based system to assist the database querying process. We then collected a random subset of 50 names of members of the House of Commons as they were recorded from a particular day's business in Hansard, scrambled them and taped them twice in a random order so that any influences from the linear sequence in which the words were used could be minimised. No particularly elaborate sound capture techniques were thought to be needed as the real situation does not have optimal sound either. We then proceeded on three separate lines of investigation.

Firstly, we tried to establish an *a priori* working hypothesis of the likelihood of their confusability by transcribing and grouping the names with a view to isolating sequences that *a priori* seemed likely to be confused on the basis of their phonology. Secondly we searched the telephone directory for all the names and listed them in terms of their frequency as words. Thirdly, the tapes of the names in random order were played to a total of 78 listeners who were asked to write down what they heard.

The informants were not told that the items were names and there was a limited improvement in the recognition of second appearances of names for those candidates who knew MPs by name. For the present experiment, we restricted ourselves to human subjects but at a later date we plan to test the list on an automatic speech recognition system and compare the results. Our corpus of test data consists of the responses from 78 subjects. They were graded correct or incorrect on the basis of whether they produced a spelling which could represent the word actually pronounced rather than insisting that total accuracy be achieved. This meant that spelling errors or variation were ignored. For example, we accepted Birmingham for Bermingham, Hywel for Howell and Louis for Lewis etc. As we shall see in the case of Blunkett (where we accepted Blunkit(t)) we rejected all variations that could not represent the pronunciation as on the tape.

3.1. The Names and their Sounds
Of the fifty names in the original selection there were

- 10 monosyllables
- 33 disyllables
- 5 trisyllables
- 1 disyllable-disyllable combination (Campbell- Savours)
- 1 disyllable-monosyllable (Garel-Jones)

Including the monosyllables but treating the double-barrelled names as two names each, there were

- 48 cases of initial stress
- 3 with stress on the second syllable (Boscawen, McMaster and Tredinnick)
- 1 with stress on the final syllable (Molyneaux.).

In initial position, which is crucial for recognition of names, there were

- 22 stops,
- 14 fricatives,
- 5 nasals,
- 3 laterals,
- 4 affricates,
- 2 vowels, and
- 2 semivowels.

Frequencies of the names were found by consulting the London telephone directory and are listed in Appendix 3. Frequency - ranged from Jones (4372), Davies (3185), Harris (2237), Lewis (1913) and Murphy (1440) to Blunkett, Caborn, Garel-Jones and Thurnham (1 each) and Campbell-Savours, Chope and Janman (0). Clearly frequency does not tell all as, for instance, Bermingham, with only 36 mentions is obviously a lot more familiar than this from the place name. Names which were incorrectly recognised were by and large those of lowest frequency and so presumably of lowest familiarity. More interesting were the substitutions which were made, which revealed several complex processing characteristics over and above the phonological features.

3.2. Misrecognitions
Over the whole group of testers the average of errors was 28%, but this included some rogue results as high as 48% wrong, presumably based on deafness, boredom or severe unfamiliarity with British surnames. Even for those who do know British names Thurnham, Caborn and Garel-Jones, which appeared only once, caused some trouble. These were overwhelmingly recognised as Vernon, Claybourne and Darryl (or even Dow!) Jones respectively. Of the names that did not appear in the telephone directory - Campbell-Savours, Chope and Janman - guessing by many respondents gave Cable-Savers, Choke, frequently, and Jarman (or German).

Blunkett also appeared only once in the telephone directory and the spread of guesses is listed in Appendix 2. Since it was the first name on the list many users guessed blanket since they did not realise that the list was a list of names. Even when they did catch on many were still unable to match the name with the sound and the majority of responses were for blanket. Only 5 responded with Blunkett the first time itwas heard, but this increased to 26 on the second mention. Of the 32 who chose blanket on the first occasion, 17 got Blunkett correct the second time (although one switched in the opposite direction!). Overall, nobody got the syllable count or the stress assignment wrong. There was significant consonant cluster simplification in several directions with blunkett being replaced by bunkett, berney etc. (see appendix 2). There was also widespread consonant confusion.

This trend is to be found especially with single initial stop consonants. Powell and Howell are always difficult to tell apart, even when spoken in isolated word mode, although they are both common enough. We were rather more surprised at Harris being mistaken for Paris, which happened less frequently. Liquids and nasals frequently got confused as in Namby for Lambie, Hallam for Hannam and Tredillick for Tredinnick. There was also frequent loss of initial /h/ in Healey. Although there are clearly phonological reasons involved in misrecognitions, it seems clear from the much higher success rate in our pilot study that pragmatically derived information from the respondent's previous knowledge of the words and the instructions given before the tests has major significance.

4. CONCLUSION

Enough has been learned so far to suggest that this topic should be pursued. As well as trying the same tape on an automatic speech recognition system we would wish to avoid much of the difficulty we experienced with illegible responses - which explains some of the lists that don't add up - by using the present tapes again and giving the respondents multiple choice checklists on the basis of the responses we got this time. As our results were so much worse than the pilot study, we would like to try the present tape on another large group who had been told that the items were MPs' names.

APPENDIX 1 NAMES AS PRESENTED TO THE SUBJECTS

1. Blunkett
2. Ground
3. Boscawen
4. Davies
5. Ashdowne
6. Currie
7. Banks
8. Caborn
9. Dobson
10. Garel-Jones
11. Colvin
12. Hannam
13. Chope
14. Hayhoe
15. Campbell-Savours
16. Howell
17. Flynn
18. Bradley
19. Grist
20. Armstrong
21. Healey
22. Powell
23. Gill
24. Riddick
25. Gale
26. Shepherd
27. Hunt
28. Sedgemore
29. Janman
30. Wiggin
31. Lewis
32. Stokes
33. Kirkhope
34. Tredinnick
35. Molyneaux
36. Thurnham
37. McMaster
38. Maxton
39. Murphy
40. Mowlam
41. Shelton
42. Jones
43. Randall
44. Lloyd
45. Foulkes
46. Lambie
47. Adley
48. Gorst
49. Bermingham
50. Harris
51. Bradley
52. Blunkett
53. Armstrong
54. Boscawen
55. Powell
56. Ashdowne
57. Riddick
58. Banks
59. Shepherd
60. Dobson
61. Sedgemore
62. Colvin
63. Wiggin
64. Chope
65. Stokes
66. Campbell-Savours
67. Tredinnick
68. Flynn
69. Thurnham
70. Grist
71. Maxton
72. Healey
73. Mowlam
74. Gill
75. Jones
76. Gale
77. Lloyd
78. Hunt
79. Lambie
80. Janman
81. Gorst
82. Lewis
83. Harris
84. Kirkhope
85. Ground
86. Molyneaux
87. Davies
88. McMaster
89. Currie
90. Murphy
91. Caborn
92. Shelton
93. Garel-Jones
94. Randall
95. Hannam
96. Foulkes
97. Hayhoe
98. Adley
99. Howell
100. Bermingham

APPENDIX 2 VARIATIONS ON BLUNKETT

39	blanket	1	bucket
31	Blunkett	1	glarkin
22	(no response)	1	lambert
6	plunkit	1	Lampitt
4	locket	1	landfit
4	lunket	1	lankit
3	larkin	1	lonkage
3	market	1	luckett
2	borkit	1	lugitt
2	bunkit	1	Lumpit
2	marker	1	lunford
2	monket	1	lunkey
1	b__ket	1	monkey
1	barking	1	Nodcat
1	bernly	1	packet
1	blonkett	1	bunion
1	blousket		

APPENDIX 3 OVERALL NAME FREQUENCIES FROM THE LONDON TELEPHONE DIRECTORY

4372	Jones	219	Stokes	10	Mowlam
3185	Davies	217	Randall	9	Riddick
2237	Harris	174	Dobson	8	Gorst
1913	Lewis	165	Currie	4	Adley
1440	Murphy	64	Shelton	4	Boscawen
933	Hunt	36	Bermingham	4	Maxton
838	Powell	36	Colvin	4	Tredinnick
696	Shepherd	33	Ashdowne	3	Kirkhope
575	Lloyd	32	Molyneaux	2	Sedgemore
403	Bradley	31	Hannam	1	Blunkett
383	Howell	25	Foulkes	1	Caborn
374	Gill	16	Grist	1	Garel-Jones
369	Armstrong	16	McMaster	1	Thurnham
335	Flynn	16	Wiggin	0	Campbell-Savours
309	Banks	14	Hayhoe	0	Chope
308	Healey	12	Lambie	0	Janman
221	Gale	10	Ground		