

UNLIMITED

A238 894



RSRE
MEMORANDUM No. 4463

ROYAL SIGNALS & RADAR ESTABLISHMENT

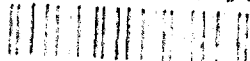
ANALYSIS OF THE PHONEME RECOGNITION
PERFORMANCE OF THE ARM CONTINUOUS SPEECH
RECOGNITION SYSTEM

Author: S R Browning

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
RSRE MALVERN,
WORCS.

Best Available Copy

91-06119



UNLIMITED

MEMORANDUM NO. 4463

01 05 010

0100492

CONDITIONS OF RELEASE

302014

DRIC U

COPYRIGHT (c)
1988
CONTROLLER
HMSO LONDON

DRIC Y

Reports quoted are not necessarily available to members of the public or to commercial organisations.

Royal Signals and Radar Establishment

Memorandum 4463

Analysis of the phoneme recognition performance of the ARM continuous speech recognition system

S R Browning

21st March 1991

Abstract

This memorandum presents the results of a phonetically motivated analysis of the speech recognition system developed as part of the *ARM* (Airborne Reconnaissance Mission) project. The aim of the work described here is to investigate to what extent errors can be explained by phonetic effects; those which cannot may indicate where models may be improved. The background to the investigation, and the problems of evaluating phoneme recognition performance are described, then the remainder of the report is concerned with a detailed analysis of specific types of errors, motivated by a desire to find phonetic explanations of them.

Copyright
(C)
Controller HMSO London
1991



1. PROJECT NO.	
2. PROJECT NAME	
3. PROJECT DESCRIPTION	
4. PROJECT STATUS	
5. PROJECT OFFICER	
6. PROJECT MANAGER	
7. PROJECT CO-ORDINATOR	
8. PROJECT SPECIALIST	
9. PROJECT REVIEWER	
10. PROJECT APPROVER	
11. PROJECT DATE	
12. PROJECT LOCATION	
13. PROJECT CONTACT	
14. PROJECT COMMENTS	
15. PROJECT SIGNATURE	
16. PROJECT INITIALS	
17. PROJECT DATE	
18. PROJECT LOCATION	
19. PROJECT CONTACT	
20. PROJECT COMMENTS	
21. PROJECT SIGNATURE	
22. PROJECT INITIALS	
23. PROJECT DATE	
24. PROJECT LOCATION	
25. PROJECT CONTACT	
26. PROJECT COMMENTS	
27. PROJECT SIGNATURE	
28. PROJECT INITIALS	
29. PROJECT DATE	
30. PROJECT LOCATION	
31. PROJECT CONTACT	
32. PROJECT COMMENTS	
33. PROJECT SIGNATURE	
34. PROJECT INITIALS	
35. PROJECT DATE	
36. PROJECT LOCATION	
37. PROJECT CONTACT	
38. PROJECT COMMENTS	
39. PROJECT SIGNATURE	
40. PROJECT INITIALS	
41. PROJECT DATE	
42. PROJECT LOCATION	
43. PROJECT CONTACT	
44. PROJECT COMMENTS	
45. PROJECT SIGNATURE	
46. PROJECT INITIALS	
47. PROJECT DATE	
48. PROJECT LOCATION	
49. PROJECT CONTACT	
50. PROJECT COMMENTS	
51. PROJECT SIGNATURE	
52. PROJECT INITIALS	
53. PROJECT DATE	
54. PROJECT LOCATION	
55. PROJECT CONTACT	
56. PROJECT COMMENTS	
57. PROJECT SIGNATURE	
58. PROJECT INITIALS	
59. PROJECT DATE	
60. PROJECT LOCATION	
61. PROJECT CONTACT	
62. PROJECT COMMENTS	
63. PROJECT SIGNATURE	
64. PROJECT INITIALS	
65. PROJECT DATE	
66. PROJECT LOCATION	
67. PROJECT CONTACT	
68. PROJECT COMMENTS	
69. PROJECT SIGNATURE	
70. PROJECT INITIALS	
71. PROJECT DATE	
72. PROJECT LOCATION	
73. PROJECT CONTACT	
74. PROJECT COMMENTS	
75. PROJECT SIGNATURE	
76. PROJECT INITIALS	
77. PROJECT DATE	
78. PROJECT LOCATION	
79. PROJECT CONTACT	
80. PROJECT COMMENTS	
81. PROJECT SIGNATURE	
82. PROJECT INITIALS	
83. PROJECT DATE	
84. PROJECT LOCATION	
85. PROJECT CONTACT	
86. PROJECT COMMENTS	
87. PROJECT SIGNATURE	
88. PROJECT INITIALS	
89. PROJECT DATE	
90. PROJECT LOCATION	
91. PROJECT CONTACT	
92. PROJECT COMMENTS	
93. PROJECT SIGNATURE	
94. PROJECT INITIALS	
95. PROJECT DATE	
96. PROJECT LOCATION	
97. PROJECT CONTACT	
98. PROJECT COMMENTS	
99. PROJECT SIGNATURE	
100. PROJECT INITIALS	
101. PROJECT DATE	
102. PROJECT LOCATION	
103. PROJECT CONTACT	
104. PROJECT COMMENTS	
105. PROJECT SIGNATURE	
106. PROJECT INITIALS	
107. PROJECT DATE	
108. PROJECT LOCATION	
109. PROJECT CONTACT	
110. PROJECT COMMENTS	
111. PROJECT SIGNATURE	
112. PROJECT INITIALS	
113. PROJECT DATE	
114. PROJECT LOCATION	
115. PROJECT CONTACT	
116. PROJECT COMMENTS	
117. PROJECT SIGNATURE	
118. PROJECT INITIALS	
119. PROJECT DATE	
120. PROJECT LOCATION	
121. PROJECT CONTACT	
122. PROJECT COMMENTS	
123. PROJECT SIGNATURE	
124. PROJECT INITIALS	
125. PROJECT DATE	
126. PROJECT LOCATION	
127. PROJECT CONTACT	
128. PROJECT COMMENTS	
129. PROJECT SIGNATURE	
130. PROJECT INITIALS	
131. PROJECT DATE	
132. PROJECT LOCATION	
133. PROJECT CONTACT	
134. PROJECT COMMENTS	
135. PROJECT SIGNATURE	
136. PROJECT INITIALS	
137. PROJECT DATE	
138. PROJECT LOCATION	
139. PROJECT CONTACT	
140. PROJECT COMMENTS	
141. PROJECT SIGNATURE	
142. PROJECT INITIALS	
143. PROJECT DATE	
144. PROJECT LOCATION	
145. PROJECT CONTACT	
146. PROJECT COMMENTS	
147. PROJECT SIGNATURE	
148. PROJECT INITIALS	
149. PROJECT DATE	
150. PROJECT LOCATION	
151. PROJECT CONTACT	
152. PROJECT COMMENTS	
153. PROJECT SIGNATURE	
154. PROJECT INITIALS	
155. PROJECT DATE	
156. PROJECT LOCATION	
157. PROJECT CONTACT	
158. PROJECT COMMENTS	
159. PROJECT SIGNATURE	
160. PROJECT INITIALS	
161. PROJECT DATE	
162. PROJECT LOCATION	
163. PROJECT CONTACT	
164. PROJECT COMMENTS	
165. PROJECT SIGNATURE	
166. PROJECT INITIALS	
167. PROJECT DATE	
168. PROJECT LOCATION	
169. PROJECT CONTACT	
170. PROJECT COMMENTS	
171. PROJECT SIGNATURE	
172. PROJECT INITIALS	
173. PROJECT DATE	
174. PROJECT LOCATION	
175. PROJECT CONTACT	
176. PROJECT COMMENTS	
177. PROJECT SIGNATURE	
178. PROJECT INITIALS	
179. PROJECT DATE	
180. PROJECT LOCATION	
181. PROJECT CONTACT	
182. PROJECT COMMENTS	
183. PROJECT SIGNATURE	
184. PROJECT INITIALS	
185. PROJECT DATE	
186. PROJECT LOCATION	
187. PROJECT CONTACT	
188. PROJECT COMMENTS	
189. PROJECT SIGNATURE	
190. PROJECT INITIALS	
191. PROJECT DATE	
192. PROJECT LOCATION	
193. PROJECT CONTACT	
194. PROJECT COMMENTS	
195. PROJECT SIGNATURE	
196. PROJECT INITIALS	
197. PROJECT DATE	
198. PROJECT LOCATION	
199. PROJECT CONTACT	
200. PROJECT COMMENTS	
201. PROJECT SIGNATURE	
202. PROJECT INITIALS	
203. PROJECT DATE	
204. PROJECT LOCATION	
205. PROJECT CONTACT	
206. PROJECT COMMENTS	
207. PROJECT SIGNATURE	
208. PROJECT INITIALS	
209. PROJECT DATE	
210. PROJECT LOCATION	
211. PROJECT CONTACT	
212. PROJECT COMMENTS	
213. PROJECT SIGNATURE	
214. PROJECT INITIALS	
215. PROJECT DATE	
216. PROJECT LOCATION	
217. PROJECT CONTACT	
218. PROJECT COMMENTS	
219. PROJECT SIGNATURE	
220. PROJECT INITIALS	
221. PROJECT DATE	
222. PROJECT LOCATION	
223. PROJECT CONTACT	
224. PROJECT COMMENTS	
225. PROJECT SIGNATURE	
226. PROJECT INITIALS	
227. PROJECT DATE	
228. PROJECT LOCATION	
229. PROJECT CONTACT	
230. PROJECT COMMENTS	
231. PROJECT SIGNATURE	
232. PROJECT INITIALS	
233. PROJECT DATE	
234. PROJECT LOCATION	
235. PROJECT CONTACT	
236. PROJECT COMMENTS	
237. PROJECT SIGNATURE	
238. PROJECT INITIALS	
239. PROJECT DATE	
240. PROJECT LOCATION	
241. PROJECT CONTACT	
242. PROJECT COMMENTS	
243. PROJECT SIGNATURE	
244. PROJECT INITIALS	
245. PROJECT DATE	
246. PROJECT LOCATION	
247. PROJECT CONTACT	
248. PROJECT COMMENTS	
249. PROJECT SIGNATURE	
250. PROJECT INITIALS	
251. PROJECT DATE	
252. PROJECT LOCATION	
253. PROJECT CONTACT	
254. PROJECT COMMENTS	
255. PROJECT SIGNATURE	
256. PROJECT INITIALS	
257. PROJECT DATE	
258. PROJECT LOCATION	
259. PROJECT CONTACT	
260. PROJECT COMMENTS	
261. PROJECT SIGNATURE	
262. PROJECT INITIALS	
263. PROJECT DATE	
264. PROJECT LOCATION	
265. PROJECT CONTACT	
266. PROJECT COMMENTS	
267. PROJECT SIGNATURE	
268. PROJECT INITIALS	
269. PROJECT DATE	
270. PROJECT LOCATION	
271. PROJECT CONTACT	
272. PROJECT COMMENTS	
273. PROJECT SIGNATURE	
274. PROJECT INITIALS	
275. PROJECT DATE	
276. PROJECT LOCATION	
277. PROJECT CONTACT	
278. PROJECT COMMENTS	
279. PROJECT SIGNATURE	
280. PROJECT INITIALS	
281. PROJECT DATE	
282. PROJECT LOCATION	
283. PROJECT CONTACT	
284. PROJECT COMMENTS	
285. PROJECT SIGNATURE	
286. PROJECT INITIALS	
287. PROJECT DATE	
288. PROJECT LOCATION	
289. PROJECT CONTACT	
290. PROJECT COMMENTS	
291. PROJECT SIGNATURE	
292. PROJECT INITIALS	
293. PROJECT DATE	
294. PROJECT LOCATION	
295. PROJECT CONTACT	
296. PROJECT COMMENTS	
297. PROJECT SIGNATURE	
298. PROJECT INITIALS	
299. PROJECT DATE	
300. PROJECT LOCATION	
301. PROJECT CONTACT	
302. PROJECT COMMENTS	
303. PROJECT SIGNATURE	
304. PROJECT INITIALS	
305. PROJECT DATE	
306. PROJECT LOCATION	
307. PROJECT CONTACT	
308. PROJECT COMMENTS	
309. PROJECT SIGNATURE	
310. PROJECT INITIALS	
311. PROJECT DATE	
312. PROJECT LOCATION	
313. PROJECT CONTACT	
314. PROJECT COMMENTS	
315. PROJECT SIGNATURE	
316. PROJECT INITIALS	
317. PROJECT DATE	
318. PROJECT LOCATION	
319. PROJECT CONTACT	
320. PROJECT COMMENTS	
321. PROJECT SIGNATURE	
322. PROJECT INITIALS	
323. PROJECT DATE	
324. PROJECT LOCATION	
325. PROJECT CONTACT	
326. PROJECT COMMENTS	
327. PROJECT SIGNATURE	
328. PROJECT INITIALS	
329. PROJECT DATE	
330. PROJECT LOCATION	
331. PROJECT CONTACT	
332. PROJECT COMMENTS	
333. PROJECT SIGNATURE	
334. PROJECT INITIALS	
335. PROJECT DATE	
336. PROJECT LOCATION	
337. PROJECT CONTACT	
338. PROJECT COMMENTS	
339. PROJECT SIGNATURE	
340. PROJECT INITIALS	
341. PROJECT DATE	
342. PROJECT LOCATION	
343. PROJECT CONTACT	
344. PROJECT COMMENTS	
345. PROJECT SIGNATURE	
346. PROJECT INITIALS	
347. PROJECT DATE	
348. PROJECT LOCATION	
349. PROJECT CONTACT	
350. PROJECT COMMENTS	
351. PROJECT SIGNATURE	
352. PROJECT INITIALS	
353. PROJECT DATE	
354. PROJECT LOCATION	
355. PROJECT CONTACT	
356. PROJECT COMMENTS	
357. PROJECT SIGNATURE	
358. PROJECT INITIALS	
359. PROJECT DATE	
360. PROJECT LOCATION	
361. PROJECT CONTACT	
362. PROJECT COMMENTS	
363. PROJECT SIGNATURE	
364. PROJECT INITIALS	
365. PROJECT DATE	
366. PROJECT LOCATION	
367. PROJECT CONTACT	
368. PROJECT COMMENTS	
369. PROJECT SIGNATURE	
370. PROJECT INITIALS	
371. PROJECT DATE	
372. PROJECT LOCATION	
373. PROJECT CONTACT	
374. PROJECT COMMENTS	
375. PROJECT SIGNATURE	
376. PROJECT INITIALS	
377. PROJECT DATE	
378. PROJECT LOCATION	
379. PROJECT CONTACT	
380. PROJECT COMMENTS	
381. PROJECT SIGNATURE	
382. PROJECT INITIALS	
383. PROJECT DATE	
384. PROJECT LOCATION	
385. PROJECT CONTACT	
386. PROJECT COMMENTS	
387. PROJECT SIGNATURE	
388. PROJECT INITIALS	
389. PROJECT DATE	
390. PROJECT LOCATION	
391. PROJECT CONTACT	
392. PROJECT COMMENTS	
393. PROJECT SIGNATURE	
394. PROJECT INITIALS	
395. PROJECT DATE	
396. PROJECT LOCATION	
397. PROJECT CONTACT	
398. PROJECT COMMENTS	
399. PROJECT SIGNATURE	
400. PROJECT INITIALS	
401. PROJECT DATE	
402. PROJECT LOCATION	
403. PROJECT CONTACT	
404. PROJECT COMMENTS	
405. PROJECT SIGNATURE	
406. PROJECT INITIALS	
407. PROJECT DATE	
408. PROJECT LOCATION	
409. PROJECT CONTACT	
410. PROJECT COMMENTS	
411. PROJECT SIGNATURE	
412. PROJECT INITIALS	
413. PROJECT DATE	
414. PROJECT LOCATION	
415. PROJECT CONTACT	
416. PROJECT COMMENTS	
417. PROJECT SIGNATURE	
418. PROJECT INITIALS	
419. PROJECT DATE	
420. PROJECT LOCATION	
421. PROJECT CONTACT	
422. PROJECT COMMENTS	
423. PROJECT SIGNATURE	
424. PROJECT INITIALS	
425. PROJECT DATE	
426. PROJECT LOCATION	
427. PROJECT CONTACT	
428. PROJECT COMMENTS	
429. PROJECT SIGNATURE	
430. PROJECT INITIALS	
431. PROJECT DATE	
432. PROJECT LOCATION	
433. PROJECT CONTACT	
434. PROJECT COMMENTS	
435. PROJECT SIGNATURE	
436. PROJECT INITIALS	
437. PROJECT DATE	
438. PROJECT LOCATION	
439. PROJECT CONTACT	
440. PROJECT COMMENTS	
441. PROJECT SIGNATURE	
442. PROJECT INITIALS	
443. PROJECT DATE	
444. PROJECT LOCATION	
445. PROJECT CONTACT	
446. PROJECT COMMENTS	
447. PROJECT SIGNATURE	
448. PROJECT INITIALS	
449. PROJECT DATE	
450. PROJECT LOCATION	
451. PROJECT CONTACT	
452. PROJECT COMMENTS	
453. PROJECT SIGNATURE	
454. PROJECT INITIALS	
455. PROJECT DATE	
456. PROJECT LOCATION	
457. PROJECT CONTACT	
458. PROJECT COMMENTS	
459. PROJECT SIGNATURE	
460. PROJECT INITIALS	
461. PROJECT DATE	
462. PROJECT LOCATION	
463. PROJECT CONTACT	
464. PROJECT COMMENTS	
465. PROJECT SIGNATURE	
466. PROJECT INITIALS	
467. PROJECT DATE	
468. PROJECT LOCATION	
469. PROJECT CONTACT	
470. PROJECT COMMENTS	
471. PROJECT SIGNATURE	
472. PROJECT INITIALS	
473. PROJECT DATE	
474. PROJECT LOCATION	
475. PROJECT CONTACT	
476. PROJECT COMMENTS	
477. PROJECT SIGNATURE	
478. PROJECT INITIALS	
479. PROJECT DATE	
480. PROJECT LOCATION	
481. PROJECT CONTACT	
482. PROJECT COMMENTS	
483. PROJECT SIGNATURE	
484. PROJECT INITIALS	
485. PROJECT DATE	
486. PROJECT LOCATION	
487. PROJECT CONTACT	
488. PROJECT COMMENTS	
489. PROJECT SIGNATURE	
490. PROJECT INITIALS	
491. PROJECT DATE	
492. PROJECT LOCATION	
493. PROJECT CONTACT	
494. PROJECT COMMENTS	
495. PROJECT SIGNATURE	
496. PROJECT INITIALS	
497. PROJECT DATE	
498. PROJECT LOCATION	
499. PROJECT CONTACT	
500. PROJECT COMMENTS	
501. PROJECT SIGNATURE	
502. PROJECT INITIALS	
503. PROJECT DATE	
504. PROJECT LOCATION	
505. PROJECT CONTACT	
506. PROJECT COMMENTS	
507. PROJECT SIGNATURE	
508. PROJECT INITIALS	
509. PROJECT DATE	
510. PROJECT LOCATION	
511. PROJECT CONTACT	
512. PROJECT COMMENTS	
513. PROJECT SIGNATURE	
514. PROJECT INITIALS	
515. PROJECT DATE	
516. PROJECT LOCATION	
517. PROJECT CONTACT	
518. PROJECT COMMENTS	
519. PROJECT SIGNATURE	
520. PROJECT INITIALS	
521. PROJECT DATE	
522. PROJECT LOCATION	
523. PROJECT CONTACT	
524. PROJECT COMMENTS	
525. PROJECT SIGNATURE	
526. PROJECT INITIALS	
527. PROJECT DATE	
528. PROJECT LOCATION	
529. PROJECT CONTACT	
530. PROJECT COMMENTS	
531. PROJECT SIGNATURE	
532. PROJECT INITIALS	
533. PROJECT DATE	
534. PROJECT LOCATION	
535. PROJECT CONTACT	
536. PROJECT COMMENTS	
537. PROJECT SIGNATURE	
538. PROJECT INITIALS	
539. PROJECT DATE	
540. PROJECT LOCATION	
541. PROJECT CONTACT	
542. PROJECT COMMENTS	
543. PROJECT SIGNATURE	
544. PROJECT INITIALS	
545. PROJECT DATE	
546. PROJECT LOCATION	
547. PROJECT CONTACT	
548. PROJECT COMMENTS	
549. PROJECT SIGNATURE	
550. PROJECT INITIALS	
551. PROJECT DATE	
552. PROJECT LOCATION	
553. PROJECT CONTACT	
554. PROJECT COMMENTS	
555. PROJECT SIGNATURE	
556. PROJECT INITIALS	
557. PROJECT DATE	
558. PROJECT LOCATION	
559. PROJECT CONTACT	
560. PROJECT COMMENTS	
561. PROJECT SIGNATURE	
562. PROJECT INITIALS	
563. PROJECT DATE	
564. PROJECT LOCATION	
565. PROJECT CONTACT	
566. PROJECT COMMENTS	
567. PROJECT SIGNATURE	
568. PROJECT INITIALS	
569. PROJECT DATE	
570. PROJECT LOCATION	
571. PROJECT CONTACT	
572. PROJECT COMMENTS	
573. PROJECT SIGNATURE	
574. PROJECT INITIALS	
575. PROJECT DATE	
576. PROJECT LOCATION	
577. PROJECT CONTACT	
578. PROJECT COMMENTS	
579. PROJECT SIGNATURE	
580. PROJECT INITIALS	
581. PROJECT DATE	
582. PROJECT LOCATION	
583. PROJECT CONTACT	
584. PROJECT COMMENTS	
585. PROJECT SIGNATURE	
586. PROJECT INITIALS	
587. PROJECT DATE	
588. PROJECT LOCATION	
589. PROJECT CONTACT	
590. PROJECT COMMENTS	
591. PROJECT SIGNATURE	
592. PROJECT INITIALS	
593. PROJECT DATE	
594. PROJECT LOCATION	
595. PROJECT CONTACT	
596. PROJECT COMMENTS	
597. PROJECT SIGNATURE	
598. PROJECT INITIALS	
599. PROJECT DATE	
600. PROJECT LOCATION	
601. PROJECT CONTACT	
602. PROJECT COMMENTS	
603. PROJECT SIGNATURE	
604. PROJECT INITIALS	
605. PROJECT DATE	
606. PROJECT LOCATION	
607. PROJECT CONTACT	
608. PROJECT COMMENTS	
609. PROJECT SIGNATURE	
610. PROJECT INITIALS	
611. PROJECT DATE	
612. PROJECT LOCATION	
613. PROJECT CONTACT	
614. PROJECT COMMENTS	
615. PROJECT SIGNATURE	
616. PROJECT INITIALS	
617. PROJECT DATE	
618. PROJECT LOCATION	
619. PROJECT CONTACT	
620. PROJECT COMMENTS	
621. PROJECT SIGNATURE	
622. PROJECT INITIALS	
623. PROJECT DATE	
624. PROJECT LOCATION	
625. PROJECT CONTACT	
626. PROJECT COMMENTS	
627. PROJECT SIGNATURE	
628. PROJECT INITIALS	
629. PROJECT DATE	
630. PROJECT LOCATION	
631. PROJECT CONTACT	
632. PROJECT COMMENTS	
633. PROJECT SIGNATURE	
634. PROJECT INITIALS	
635. PROJECT DATE	
636. PROJECT LOCATION	
637. PROJECT CONTACT	
638. PROJECT COMMENTS	
639. PROJECT SIGNATURE	
640. PROJECT INITIALS	
641. PROJECT DATE	
642. PROJECT LOCATION	
643. PROJECT CONTACT	
644. PROJECT COMMENTS	
645. PROJECT SIGNATURE	
646. PROJECT INITIALS	
647. PROJECT DATE	
648. PROJECT LOCATION	
649. PROJECT CONTACT	
650. PROJECT COMMENTS	
651. PROJECT SIGNATURE	
652. PROJECT INITIALS	

INTENTIONALLY BLANK

Contents

1	Introduction	1
2	Background	1
2.1	The ARM task	1
2.2	The Speakers	1
2.3	The System	2
2.4	Evaluating the performance	2
3	The Analysis	3
3.1	Analysis of correctness/accuracy	3
3.1.1	Individual phonemes	3
3.1.2	Manner of articulation	4
3.1.3	Place of articulation	6
3.2	Substitutions	6
3.2.1	Individual phonemes	6
3.2.2	Manner of articulation	7
3.2.3	Place of articulation	8
3.2.4	Contextual effects	8
3.3	Deletions	9
3.3.1	Individual phonemes	9
3.3.2	Manner of articulation	9
3.3.3	Place of articulation	10
3.3.4	Contextual effects	10
3.4	Insertions	11
3.4.1	Individual phonemes	11
3.4.2	Manner of articulation	12
3.4.3	Place of articulation	13
3.4.4	Contextual effects	13
4	Discussion and conclusions	14
	References	15
	Appendix A Phoneme frequency	16
	Appendix B The ARM Vocabulary	17
	Appendix C Results in full	21

List of tables

Table 1	Summary of phoneme recognition results	3
Table 2	Individual phoneme correct/accuracy for each and all speakers	4
Table 3	Phoneme substitutions (%) for each and all speakers	7
Table 4	Confusion matrix for manner of articulation – all speakers	8
Table 5	Confusion matrix for place of articulation – all speakers	9
Table 6	Phoneme deletions (%) for each and all speakers	10
Table 7	Phoneme insertions for each and all speakers	12

List of figures

Figure 1	Key of Manner and place class membership	5
Figure 2	Graph of manner class correct/accuracy	5
Figure 3	Graph of place class correct/accuracy	6
Figure 4	Graph of manner class substitutions	7
Figure 5	Graph of place class substitutions	8
Figure 6	Graph of manner class deletions	9
Figure 7	Graph of place class deletions	10
Figure 8	Graph of manner class insertions	12
Figure 9	Graph of place class insertions	13

1. Introduction

This memorandum presents the results of a phonetically motivated analysis of the speech recognition system developed as part of the *ARM* (Airborne Reconnaissance Mission) project. The aim of the *ARM* project is accurate recognition of continuously spoken airborne reconnaissance reports using sub-word (phoneme) hidden Markov modelling techniques. The version of the system on which this study is based is speaker-dependent and has a vocabulary of 497 words. The *ARM* system is described in [5]. The version of the system on which this investigation was based scores an average of 86.8% word accuracy with word level syntax (i.e. perplexity = 497).

The aim of the work described here is to investigate to what extent errors can be explained by phonetic effects; those which cannot may indicate where models may be improved. For instance, if /p/ is misrecognised as /b/, this is understandable from the phonetic point of view as the two are acoustically rather similar; however, if /p/ were to be consistently misrecognised as /@U/ or /z/ this error would be difficult to explain in acoustic-phonetic terms, and would probably indicate that there is something wrong with the model(s).

The following section describes the background to the investigation, and the problems of evaluating phoneme recognition performance. The remainder of the report is then concerned with a detailed analysis of specific types of errors, motivated by a desire to find phonetic explanations of them. The phonemic transcriptions in this report are in the SAM-PA notation [2, 8], and see Appendix A for the list of phonemes and examples.

2. Background

2.1 The *ARM* task

The airborne reconnaissance mission reports which the *ARM* system recognises follow a standard format, beginning with some highly structured sentences recounting the mission details, such as time and place of observation. Then follows a slightly more free-format section where the reconnaissance pilot describes what he sees and assesses its condition. The report concludes with a brief description of the weather and visibility conditions. The vocabulary of the system with its citation-form phoneme transcription can be found in Appendix B. An example of an *ARM* report is given below.

Recce report two stroke charlie stroke six eight one. Military activity at map co-ordinates india hotel eight four three four. Time over target eleven oh seven GMT. New target cat zero one; operational airstrip. Roughly fifteen light aircraft of type possibly foxbat. Main runways heading southwest wholly unusable. SAM defences intact. TARWI five eighths at niner hundred; end of report.

It is important to note that this is not a natural use of language, and this may influence the generality of the results of this study, in that the relative frequency of phonemes in the *ARM* vocabulary will not necessarily match that in natural language. In particular, the phoneme /D/, which ranks eighth in normal use (due to the high frequency of the word "the" in natural speech), does not occur at all in the data of two of the speakers examined here, and only occurs once for Speaker 2, who pronounces the word "with" as /wID/, rather than /wIT/. A comparison of phoneme frequencies in normal speech [3] with those in the *ARM* data can be found in Appendix A.

2.2 The Speakers

The system currently recognises the speech of three speakers, and is trained separately for each, using approximately fifteen minutes of speech (airborne reconnaissance mission reports) from each of the three. Speakers 1 and 2 are male; Speaker 1 is basically RP, while Speaker 2 has Midlands overtones. Speaker 3 is female and has north-eastern colours in her accent. Each speaker has their own dictionary to take account of dialectal variations. In this report I will be trying to draw some general conclusions about error types which apply to all three speakers, but the more important speaker differences will also be pointed out.

2.3 The System

The *ARM* system is described in detail in [7]. Sub-word (phoneme-like) hidden Markov models are used, but it is well known that the acoustic realisation of phonemes varies in different contexts. In order to take account of this context-sensitivity, approximately 1500 triphones are used. Triphone modelling assumes that it is the immediately surrounding context which exerts the most influence on the acoustic realisation of a particular phoneme, so a triphone is a model of a phoneme in its left and right context. In the current system this is restricted to word-internal contexts. See [6] for a full description of the triphone methods used in the *ARM* system.

In addition to the triphones for each context-sensitive phoneme, a number of short words are modelled explicitly at the word level. Non-speech sounds, such as breath noise or lip smacks are also modelled explicitly with a set of single state models. Both the word models and the noise models are treated in exactly the same way as the triphones.

For the purposes of the analysis described here the system was configured as a phoneme recogniser, with no dictionary and no syntax. There is, however, some measure of constraint, in that the right context of each triphone must match the left context of the next. This is no small constraint; as the triphones are word-internal, and the vocabulary so limited, the number of different triphones that actually occur is very small. (There are 1456 different triphones in the *ARM* set, while a 68000-word dictionary has 14378.)

2.4 Evaluating the performance

The system has so far been tested on ten *ARM* reports (that were not in the training set) from each speaker, containing a total of approximately 2290 phonemes per speaker, 6873 in all. The arrangement described above produced an average phoneme error rate of 26.2% for the three speakers. Phoneme recognition performance is measured by aligning the output of the system with a phonemic transcription of the test material. The latter is obtained by replacing each word in the orthographic transcription of the data with its phonemic transcription from the (speaker-dependent) dictionary. Errors are classified as substitutions, deletions or insertions. Substitutions occur when a phoneme is misrecognised as another phoneme, deletions when a phoneme has been missed by the system, and insertions when the system has recognised an extra phoneme. Recognition performance is stated in terms of correctness and accuracy. The first is simply a measure of how many times the system produced the same label as the dictionary transcription, while the second is a more stringent measure, which is calculated by subtracting the number of insertions from the number of correctly recognised phonemes, and as such is a more satisfactory indicator of the recognition performance.

The alignment of recognition results and transcription is automatic, and a summary of individual phoneme performance is also produced, along with a confusion matrix. However, this process is not accurate, in that sometimes errors in the alignment obscure correct matches, and insertions are counted as substitutions, etc. If alignment errors are taken into account the overall results are not significantly different (on average slightly over 1% either way), but as it is the distribution and details of the error types that are most important for this investigation, it is necessary to hand correct the alignment and scoring. This is quite a lengthy process as it involves listening to the speech at the same time as observing the labelling produced by the system on the spectrograms, and then re-compiling the phoneme statistics and confusion matrix. All the analyses in this paper are based on hand-corrected alignments.

It is in practice extremely difficult to assess performance, as in many cases the speaker may not actually produce the somewhat idealised pronunciation represented in the dictionary. For example, in the sequence "six six" the speaker is likely to produce only one /s/ (though it may be somewhat lengthened) for the two which phonemically occur over the word boundary. In this example, if the system recognises only one /s/ it is penalised for having deleted a phoneme. There are numerous examples of this nature, and these will be discussed under the appropriate categories below. In spite of these shortcomings, results are scored strictly against the dictionary transcription in order to ensure that the evaluation system is both consistent and automatic. We are, however, currently investigating the inclusion of alternative transcriptions in the dictionary, which will allow us to take account of many of these so-called errors.

3. The Analysis

In this section the phoneme recognition results are analysed in some detail. A summary of the phoneme recognition results for each speaker and for all speakers combined is shown in Table 1. From this it can be seen that the results for all three speakers are in the same range, although Speaker 2 and Speaker 3 have slightly better performance than Speaker 1. This general trend is evident in most of the more detailed analyses of phoneme performance; particular differences between speakers will be pointed out below.

Speaker	% correct	% substitution	% deletion	% accuracy	total no. of phonemes
1	75.5	13.2	11.3	68.8	2290
2	82.0	10.9	7.1	76.8	2290
3	80.8	11.6	7.6	76.0	2293
All speakers	79.4	11.9	8.7	73.8	6873

Table 1 Summary of phoneme recognition results

The rest of Section 3 is devoted to a discussion of the different types of error. A complete set of tables showing the individual phoneme performance for each speaker and all speakers combined can be found in Appendix C (tables C1-4). For convenience and clarity in the following sections only the information about the factors under discussion will be presented.

3.1 Analysis of Correctness/accuracy

3.1.1 Individual phonemes:

Table 2 shows the phoneme correct and accuracy scores for each speaker and all speakers. A number of phonemes (/D, ol, 3, e@ and U/) occur so rarely in the *ARM* reports that their results are unreliable indicators of performance, so these will be ignored in this analysis. The models with the poorest performance were those for whole words, which tended to be confused with one or more phonemes. Although there are slight differences between speakers at the top end of performance, it can be seen that in general /A/ was recognised most reliably, closely followed by /eI/, /S/ and /O/.

It is at the lower end, however, that more obvious differences emerge. None of the speakers has good performance for /N/, for instance, and only Speaker 2 has a reasonable score for /v/. In Speaker 1's data, /p/, although well recognised, suffered from some insertions, and was one of the least accurately recognised phonemes. For this speaker too /m/ scored particularly badly, but /d/ was the least accurate due to an unusually large number of insertions (these will be discussed later). It was /V/ that was least correctly recognised in Speaker 2's case, and this phoneme was relatively often inserted too, making it the least accurate phoneme for this speaker. This may be due to there being relatively few occurrences of this phoneme in Speaker 2's data (18 as opposed 35 in the other speakers' reports). In Speaker 3's case the least correct (apart from /v/ and /N/ which was common to the other speakers) was /b/.

In trying to find general trends in phoneme recognition performance the phonemes were classified into phonetically motivated groups, namely 'manner' and 'place' of articulation. (I have disregarded the word-level models in this classification.) Under 'manner' there is a broad classification into vowels and consonants, which should be self-evident, and a finer one where consonants are split into more specific classes. A list of the members of each of these classes is given in Figure 1, along with the total number of phonemes in each class. The fineness of the place classification was chosen in an attempt to make sure that there were enough members

of a given class to give a reasonable sample. In the case of the centring diphthongs and palatal-alveolars there are probably too few, but it would not have been reasonable to include them in any of the other classes. It would therefore be unwise to draw any conclusions about these two classes. It may be useful to note that in general these classes of phonemes occur comparatively rarely, either in normal speech or in the *ARM* test data (see Appendix A). So /e@/ and /I@/ rank 40th and 41st in normal speech and 40th and 30th in *ARM*. And for the palatal-alveolars, /S/ ranks 31st in normal, 27th in *ARM*; /tS/ 38th (37th); /tZ/ 36th in both; /j/ ranks 32nd in normal, 33rd in *ARM*, and /Z/ does not occur at all in *ARM*, and ranks 43rd in normal.

Phon	Speaker 1		Speaker 2		Speaker 3		All		Total
	% Cor	% Acc	% Cor	% Acc	% Cor	% Acc	% Cor	% Acc	
s	83.8	75.7	86.8	80.9	86.8	83.8	85.8	80.2	408
z	86.0	70.2	71.9	50.8	82.5	77.2	80.1	65.5	171
S	83.9	83.9	100.0	100.0	96.8	96.8	93.5	93.5	93
f	89.8	85.9	89.8	83.3	83.4	80.8	87.6	83.3	234
v	36.7	20.0	83.3	70.0	50.0	36.7	56.7	42.3	90
T	83.3	72.2	65.7	54.3	75.0	69.4	74.8	65.5	107
D	--	--	0.0	0.0	--	--	0.0	0.0	1
h	61.5	61.5	76.9	76.9	61.5	46.2	66.7	61.6	39
uS	66.7	66.7	88.9	88.9	88.9	88.9	81.5	81.5	27
dZ	72.7	72.7	81.8	63.6	63.6	63.6	72.7	66.6	33
p	92.7	58.5	90.2	61.0	82.9	58.5	88.6	59.3	123
b	60.0	46.7	60.0	60.0	53.4	26.7	57.9	44.4	45
t	85.3	81.8	85.3	83.1	87.0	82.7	85.9	82.6	693
d	59.8	19.5	63.4	52.4	70.0	59.8	64.6	43.9	246
k	89.7	84.5	92.7	88.7	90.7	89.7	91.1	87.6	291
g	75.6	75.6	78.0	78.0	87.8	87.8	80.5	80.5	123
m	38.8	34.7	85.7	71.4	91.8	85.7	72.2	64.0	147
n	55.0	54.4	80.1	73.7	77.2	76.0	70.8	66.7	513
N	55.6	55.6	50.0	44.4	50.0	50.0	51.9	50.0	54
l	78.7	70.7	86.7	81.3	76.0	66.7	80.5	72.9	225
r	88.4	84.5	96.1	96.1	88.4	85.4	90.9	88.7	309
w	79.6	79.6	86.4	81.8	90.9	86.4	85.6	82.6	132
j	100.0	85.7	71.4	71.4	85.7	85.7	85.7	80.9	42
i	84.1	77.3	84.1	80.6	93.7	93.7	86.1	82.1	224
l	65.5	53.8	70.9	70.1	76.3	66.9	72.1	64.0	394
E	74.7	72.3	84.3	83.1	85.6	80.7	81.5	78.7	249
(	90.2	88.2	89.5	84.2	86.0	86.0	88.5	86.1	165
A	97.6	97.6	100.0	97.1	94.3	91.4	97.3	95.5	111
Q	76.9	76.9	96.7	96.7	84.6	84.6	89.3	89.3	56
O	81.1	78.4	94.6	94.6	100.0	100.0	91.9	91.0	111
U	100.0	100.0	66.7	33.3	66.7	66.7	77.8	66.7	9
u	72.7	67.3	83.6	81.8	69.1	67.3	75.2	72.2	165
3	100.0	66.7	100.0	100.0	100.0	66.7	100.0	77.2	9
@	56.0	50.0	66.0	58.7	64.7	51.3	62.2	53.3	450
V	82.8	77.1	50.0	27.7	74.3	68.6	72.7	63.6	88
el	89.8	89.8	100.0	100.0	91.8	91.8	93.9	93.9	147
al	86.3	86.3	94.1	90.2	94.1	94.1	91.5	89.2	153
ol	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	3
aU	87.5	87.5	87.5	87.5	93.8	87.5	89.6	87.5	48
@U	71.4	66.1	82.1	80.4	76.8	76.8	76.8	74.4	168
I@	94.1	94.1	100.0	94.1	82.4	82.4	92.2	90.2	51
e@	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	6
<at>	19.0	19.0	38.1	38.1	42.9	42.9	33.3	33.3	63
<oh>	50.0	50.0	33.3	33.3	50.0	50.0	44.4	44.4	18
<of>	45.5	45.5	27.3	27.3	45.5	45.5	39.4	39.4	33
<or>	50.0	50.0	50.0	50.0	0.0	0.0	33.3	33.3	6
Overall	75.5	68.8	82.0	76.8	80.8	76.0	79.4	73.8	6873

Table 2 Individual phoneme correct/accuracy for each and all speakers

3.1.2 Manner of articulation

There is no significant difference in the recognition performance between vowels and consonants, with vowel correctness 79.5% (n=2607) and consonants 80.6% (n=4146). However, consonants are more than twice as likely to be inserted as vowels; 267 insertions compared with 119, making the accuracy for the consonants slightly lower; consonants 74.2%, vowels 75.0%.

Again the full set of results for all speakers can be found in Appendix C (tables C5-8); only the information relevant to the current discussion will be presented here. The results analysed in terms of phoneme correctness/accuracy by manner of articulation are therefore shown in Figure 2. The overall manner class accuracy was 87.1%.

MANNER			PLACE		
Plosive	p b t d k g	1521	Labial	p b m f v T D w	879
Affricate	tS dZ	60	Alveolar	t d n s z l r	2565
Strong fricative	s z S	672	Palatal-alveolar	S tS dZ j	195
Weak fricative	f v T D h	471	Velar	k g N h	507
Liquid/Glide	l r w j	708	Front	i I E {	1032
Nasal	n m N	714	Central	V @ 3	547
Vowel	i I E { A Q O	2607	Back	A O Q U u	452
	u U V @ 3 e l		Fronting	al el oi	303
	al ol aU @ U		Centring	l@ e@	57
	l@ e@		Backing	aU @ U	216

Figure 1 Key of Manner and place class membership

Liquids/glides and strong fricatives were recognised most correctly and accurately for all speakers. Nasals were quite clearly the worst, especially for Speaker 1, though the accuracy of weak fricatives was also poor because of the high number of insertions. Both of these classes may be acoustically weak, and /v/ especially is easily missed, which might explain their poor performance. It is not surprising that strong fricatives should be well modelled, as they are generally acoustically prominent (compared to weak fricatives, especially). More unexpected was the good performance of liquids and glides which are often thought to be problematic for systems with limited ability to model temporal dynamics. The explanation for this may be provided by the variable frame rate analysis which is used [4]; areas which are acoustically stable are compressed into a smaller number of frames/states, while those that vary rapidly, such as /r/, /j/ and /w/ are modelled using comparatively more states, giving the improved time resolution needed to identify these sounds.

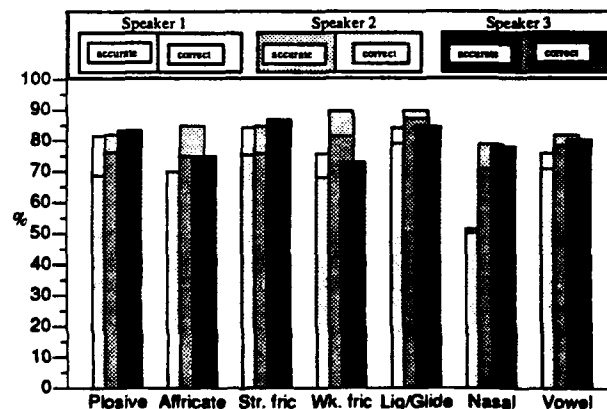


Figure 2 Graph of manner class correct/accuracy

3.1.3 Place of articulation

Figure 3 gives the analysis of the results grouped by place of articulation (and see Appendix C, tables C9-12). The overall place class accuracy was 84.4%. From the graph it can be seen that diphthongs which move towards a front position are most accurately recognised; while among the consonants, palatal-alveolars are the best recognised. Perhaps not surprisingly, central vowels were poorly dealt with. The /@/ vowel represents a large proportion (over 80%) of the central vowels and as this vowel is unstressed and notoriously variable, it is not surprising it is rather loosely modelled, and is not only easily confusable, but frequently inserted too. Labial consonants are only moderately well modelled, perhaps because most of the weak fricatives are in this group, and these are often acoustically indistinct. These results are strikingly consistent across speakers.

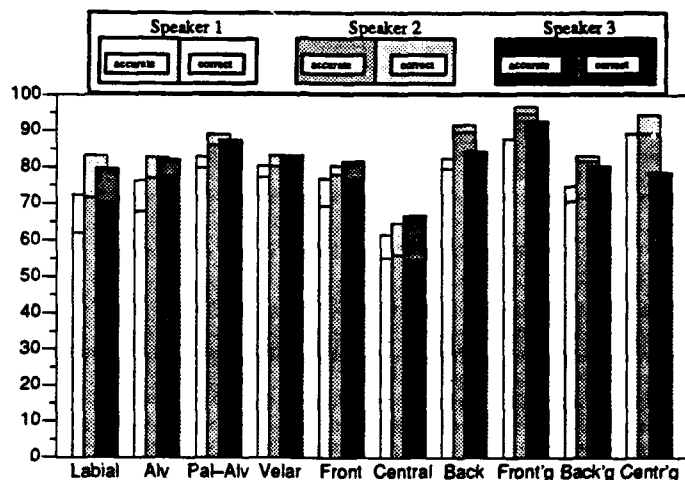


Figure 3 Graph of place class correct/accuracy

3.2 Substitutions

3.2.1 Individual phonemes

The substitution rates for individual phonemes for each and all speakers is presented in Table 3. The function words quite clearly were much substituted, and of the phonemes /N/, /V/ and /@U/ are most likely to be substituted. /S/, /A/ and /h/ are least confused (ignoring those phonemes mentioned earlier that occur only a few times).

When the system misrecognises one phoneme as another it is important to be able to explain why this has happened. If the two phonemes involved differ minimally, in one phonetic feature (/p/ and /b/, for instance) then it may be difficult to improve either model to separate them. If, however, larger differences are involved, there may be more scope for better modelling. In order to investigate what proportion of the substitution errors were phonetically predictable, phoneme confusion matrices were constructed, and these can be found in Appendix C (tables C13-16). In general confusions are with phonetically similar sounds, though there are some exceptions to this, which are difficult to explain, even when there appears to be some pattern to them. For example, all 4 of Speaker 1's /v/-/@/ confusions occurred in the word "seven", but there were as many occasions when the /v/ in this word was correctly recognised, so it is not possible to make any generalisations about the cause of this error.

	Spkr 1	Spkr 2	Spkr 3	All	Total		Spkr 1	Spkr 2	Spkr 3	All	Total
s	11.8	6.6	7.4	8.6	408	i	12.5	13.6	4.2	11.2	224
z	10.5	21.1	7.0	12.9	171	l	18.8	21.4	15.6	18.3	394
S	3.2	0.0	0.0	1.1	93	E	18.1	9.7	4.8	10.8	249
f	5.1	5.1	12.8	7.7	234	l	7.8	10.5	12.3	10.3	165
v	23.3	10.0	6.7	13.3	90	A	2.4	0.0	5.7	2.7	111
T	2.8	14.3	16.7	11.2	107	Q	7.7	3.3	0.0	3.6	56
D	--	100.0	--	100.0	1	O	8.1	2.7	0.0	3.6	111
h	23.1	7.7	23.1	17.9	39	U	0.0	0.0	0.0	0.0	9
IS	33.3	11.1	11.1	15.4	27	u	14.6	14.6	27.3	18.8	165
dZ	18.2	18.2	18.2	18.2	33	3	0.0	0.0	0.0	0.0	9
p	4.9	4.9	12.2	7.3	123	@	14.7	18.0	18.0	16.9	450
b	13.3	6.7	33.3	17.8	45	V	14.3	33.3	17.1	21.6	88
t	5.6	6.9	10.4	7.6	693	el	10.2	0.0	8.2	6.1	147
d	14.6	17.1	15.9	15.9	246	al	5.9	5.9	5.9	5.9	153
k	3.1	2.1	5.2	3.4	291	ol	0.0	0.0	0.0	0.0	3
g	17.1	17.1	9.8	14.6	123	aU	12.5	12.5	6.2	10.4	48
m	32.6	10.2	4.1	15.6	147	@U	23.2	14.3	21.4	19.6	168
n	25.1	8.8	12.3	15.4	513	k@	5.9	0.0	17.6	7.8	51
N	27.8	22.2	33.3	27.8	54	ed@	50.0	50.0	50.0	50.0	6
l	5.3	9.3	10.7	8.4	225	<st>	81.0	57.1	57.1	65.1	63
r	1.9	1.0	5.8	2.9	309	<oh>	50.0	66.7	50.0	55.6	18
w	15.9	9.1	6.8	10.6	132	<of>	54.5	72.7	36.4	54.5	33
j	0.0	14.3	14.3	9.5	42	<or>	50.0	50.0	100.0	66.7	6

Table 3 Phoneme substitutions (%) for each and all speakers

3.2.2 Manner of articulation

There is no evidence that either vowels or consonants are more subject to substitution. Figure 4 shows the rate of substitutions for manner of articulation.

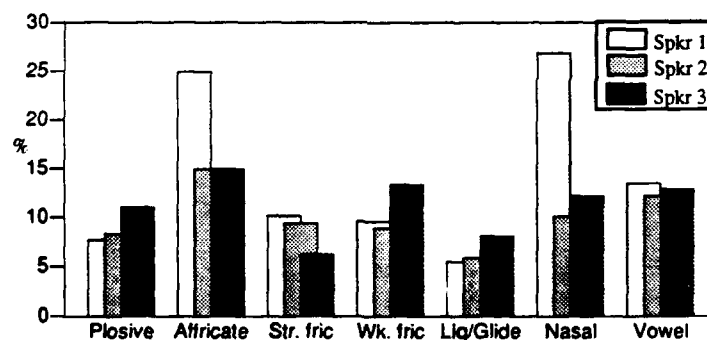


Figure 4 Graph of manner class substitutions

Confusions with phonemes from the same class would be more explicable than those with a different one, although there is a hierarchy of class similarities. For instance, plosives are more like affricates than vowels; nasals are more like vowels than they are strong fricatives. And in general this is what we find in the ARM results. Consonants are recognised as consonants 93%, and vowels as vowels nearly 90% of the time. The results of the finer manner class analysis of confusions for all speakers are shown in Table 4 (those for individual speakers can be found, as usual in Appendix C, tables C17–20). This matrix shows how often phonemes from one class were recognised as phonemes from other classes. The matrix diagonal shows within-class recognitions.

Nasals were the most confused, though most of the confusions are predictable; nasals share stop-like characteristics with plosives, and a vowel-like structure with liquids and vowels. It is interesting that almost all (95%) of the nasal/plosive confusions were for Speaker 1, where /n/ was mostly misrecognised as /b/ or /d/.

Plosives were misrecognised most often as vowels. Nearly half of these unexpected confusions are with central vowels, indicating that /ə/ is a major culprit in misrecognition (as well as being misrecognised itself). In general plosives are the most often substituted class.

		%Recognised							Total
		Plo	Aff	SF	WF	L/G	Nas	Vow	
S	Plosive	87.4	0.3	0.8	0.7	0.6	0.3	1.0	137
	Affricate	3.3	81.7	6.7			1.7	1.7	11
p	Str Fric	1.5	0.4	90.8	0.3		0.7	0.3	574
	Wk Fric	5.3	0.2	0.4	80.0			1.3	362
k	Liq/Glide	0.7	0.1		0.6		87.3	0.4	611
e	Nasal	2.2			0.3	1.8	77.2	3.8	497
n	Vowel	0.6		0.1	0.2	0.6	0.5	89.9	335

Table 4 Confusion matrix for manner of articulation – all speakers

The rest of the matrix is very much as one would expect. In general in-class recognition is good. Affricates are confused with plosives and strong fricatives with which they share many features. Weak fricatives are also confused with plosives, particular confusion being /f/ with /p/, and as these share place of articulation, being broadly speaking labial, this is not unexpected.

3.2.3 Place of articulation

The substitution rates for place of articulation are shown in Figure 5. As might be expected, central vowels are the weakest; they are confused with a wide range of different classes, and are the most widely substituted class too. Backing, and for Speaker 3, centring diphthongs are also frequently confused. It can be seen from the place confusion matrix in Table 5 that much of the poor recognition of labials is likely to be due to them being confused with each other, with alveolars being the most likely substitute.

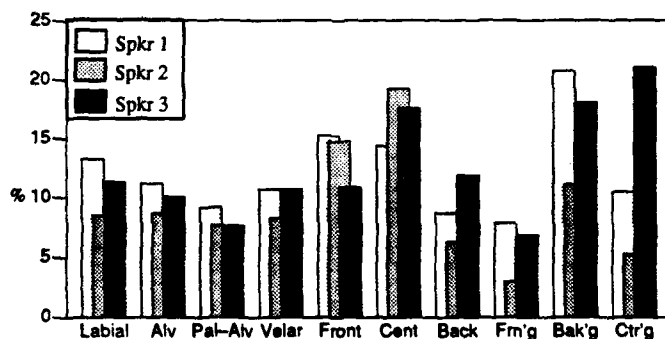


Figure 5 Graph of place class substitutions

3.2.4 Contextual Effects.

Substitution errors can sometimes be explained by the normal co-articulatory processes. Examples such as /n/ recognised as /N/ in "machine gun", /m/ as /n/ before an alveolar in "platforms", /g/ as /d/ and /d/ as /p/ in the sequence "target grid ref", /s/ as /z/ in voiced environment "zero seven", and the sequence /st/ recognised as /zd/ in the voiced environment "fuel station" are not hard to find. A more detailed description of these errors is contained in [1]. These examples account for 10% of substitutions.

In addition, as has already been mentioned above, some substitution errors are due to the quite legitimate variations which occur in fluent speech, and these nearly always involve minimal difference between target and recognised phoneme, such as place of articulation or voicing. The alternation of /i/ with /I/ in final unstressed syllables, such as in "facility", and "twenty", and /ə/ with practically any unstressed vowel is well known, and was the source of on average 15% of the substitution errors. Such errors may serve to bear out the hypothesis that a major part of the substitution errors made by the system have a phonetic explanation.

		%Recognised										Total
		Lab	Alv	P-A	Vel	Fm	Bck	Cen	F's	B's	C's	
S p o k e r	Labial	86.1	2.6	0.1	0.5	.	0.2	1.0	0.1	0.1	.	879
	Alveolar	1.9	85.4	0.3	0.2	0.3	0.2	1.2	0.2	0.1	.	2565
	Pal-Alv	.	1.0	91.3	.	1.5	.	0.5	0.5	.	.	195
	Velar	1.6	4.3	0.6	84.4	0.4	.	0.2	0.6	.	0.2	5.7
	Front	0.1	0.8	0.4	.	85.5	1.5	2.9	1.0	0.6	0.5	1032
	Back	0.6	1.1	.	.	2.7	86.5	1.1	0.2	1.3	.	547
	Central	2.2	2.0	0.2	0.2	5.7	1.5	67.1	0.5	1.1	0.4	452
	Fronting	.	.	.	.	3.0	0.7	1.0	93.1	0.7	0.3	303
	Backing	.	0.5	.	.	2.8	0.9	7.9	2.8	79.6	0.5	216
	Centring	.	.	.	.	8.8	.	1.8	.	.	89.5	57

Table 5 Confusion matrix for place of articulation - all speakers

3.3 Deletions

3.3.1 Individual phonemes

Deletions account for 42% of the recognition errors, so it would be useful to find out why they occur. Table 6 shows the deletion rates for individual phonemes.

Among the consonants /v/ scores poorly, as does /b/. We have already discussed the possible reasons for the poor performance of /v/, and of weak fricatives and nasals in general, but it is not so clear why a sound such as /b/ should be missed, but since this is consistent across speakers, it is possible that the models are defective in some way. There also appears to be a problem with /m/ specific to Speaker 1; 28.6% of this speaker's /m/s were deleted, as compared to 4.1% for both Speakers 2 and 3. There does not seem to be any particular pattern to these deletions, and there is at the moment no explanation for them, except that the models may be unreliable.

3.3.2 Manner of articulation

The deletions according to manner of articulation are shown in Figure 6. There is no real patterning to manner class deletions, although strong fricatives appear more robust for all speakers than other classes. The large percentage of nasal deletions for Speaker 1 is due in part to the predominance of /m/ deletions already mentioned, but this speaker also has nearly twice as many /n/ deletions as the others.

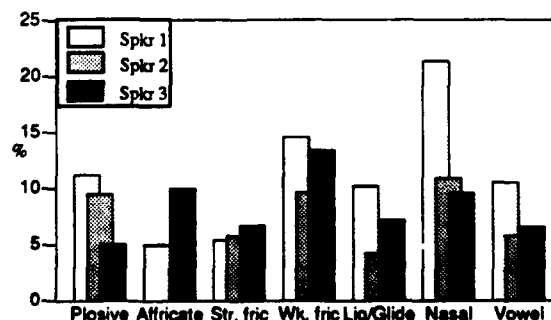


Figure 6 Graph of manner class deletions

	Spkr 1	Spkr 2	Spkr 3	All	Total		Spkr 1	Spkr 2	Spkr 3	All	Total
s	4.4	6.6	5.8	5.6	408	i	3.4	2.3	2.1	2.7	224
z	3.5	7.0	10.5	7.0	171	l	13.7	7.7	8.1	9.6	394
S	12.9	0.0	3.2	5.4	93	E	7.2	6.0	9.6	7.6	249
f	5.1	35.1	3.8	4.7	234	ɪ	2.0	0.0	1.7	1.2	165
v	40.0	6.7	43.3	30.0	90	A	0.0	0.0	0.0	0.0	111
T	13.9	20.0	8.3	14.0	107	Q	15.4	0.0	15.4	7.1	56
D	--	0.0	--	0.0	1	O	10.8	2.7	0.0	4.5	111
h	15.4	15.4	15.4	15.4	39	U	0.0	33.3	33.3	22.2	9
tS	0.0	0.0	0.0	0.0	27	u	12.7	1.8	3.6	6.0	165
dZ	9.1	0.0	18.2	9.1	33	3	0.0	0.0	0.0	0.0	9
p	2.4	4.9	4.9	4.1	123	@	29.3	16.0	17.3	20.9	450
b	26.7	33.3	13.3	24.2	45	V	2.9	16.7	6.6	8.0	88
t	9.1	7.8	2.6	6.5	693	eɪ	0.0	0.0	0.0	0.0	147
d	25.6	19.5	13.4	19.5	246	aɪ	7.8	0.0	0.0	2.6	153
k	7.2	5.2	4.1	5.5	291	oɪ	0.0	0.0	0.0	0.0	3
g	7.3	4.9	2.4	4.9	123	aU	0.0	0.0	0.0	4.1	48
m	28.6	4.1	4.1	12.2	147	@U	5.4	3.6	1.8	73.6	168
n	19.9	11.1	10.5	13.8	513	ɪ@	0.0	0.0	0.0	0.0	51
N	16.6	27.8	16.7	20.3	54	e@	0.0	0.0	0.0	0.0	6
l	16.0	4.0	13.3	11.1	225	<ae>	0.0	4.8	0.0	1.6	63
r	9.7	2.9	5.8	6.2	309	<oh>	0.0	0.0	0.0	0.0	18
w	4.5	4.5	2.3	3.8	132	<of>	0.0	0.0	18.2	6.1	33
j	0.0	14.3	0.0	4.8	42	<or>	0.0	0.0	0.0	0.0	6

Table 6 Phoneme deletions (%) for each and all speakers

3.3.3 Place of articulation

Figure 7 shows the deletions analysed by place of articulation. By far the most deleted class is that of the central vowels, and this is mainly due to the phoneme /@/, which accounts for 93% of all central vowel deletions. The reasons for this are often contextual, as is discussed in 3.3.4 below. Diphthongs are not often deleted, and this could be due to the fact that they are relatively long, and usually have quite a clear structure.

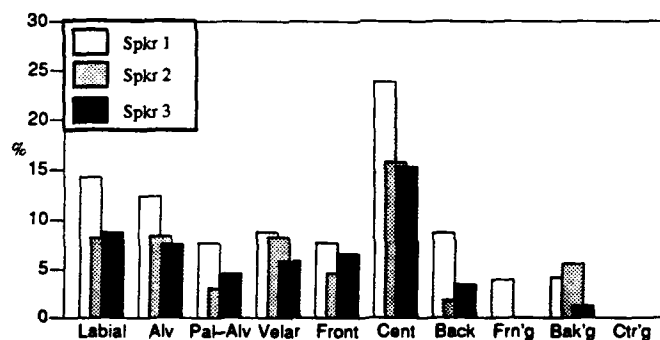


Figure 7 Graph of place class deletions

3.3.4 Contextual effects

A scored deletion is often the result of the system labelling two phonemes as one. The most typical examples of this occur when the same sound occurs at the end of one word and the beginning of the next, as in "five five" or "six six" (five is pronounced "fife" to help avoid confusion with "nine", which is pronounced "niner"). When pronounced in fluent speech the phonemes tend to run into each other, and the system recognises only one, so the above examples will be recognised as /falfalf/ and /siksiks/. This is another example of how the system is

penalised for an error which is due to the normal phonological processes of fluent speech. A different error of this type may be attributed to the fact that we are working with a very limited, and rather specialised vocabulary. Part of the second /n/ in "niner" is often labelled as part of the /aI/. This may be due to the fact that "niner" occurs frequently in the database, so it will have a significant influence on the (aI:n_n) triphone model (/aI/ with /n/ as its left and right context).

Many of the deletion errors are caused by genuine elisions by the speaker. For example, the unstressed /ə/ vowel is often elided, particularly in unstressed syllables before a nasal or liquid. The speaker-specific dictionaries account for a number of such cases, for example in "seven" (/sEvn/) and "hidden" (/hIdn/), but in the present analysis if an /ə/ appears in the dictionary transcription it will be scored as a deletion if it isn't recognised, even if in reality it wasn't produced. Words like "correction" are transcribed /k@rEkS@n/, but either (or both) of these schwas may be deleted in fast speech – /krEkSn/. It is probably for this reason that /ə/ is the most deleted phoneme and is twice as likely to be deleted as any other vowel. Speaker 1 has almost twice as many /ə/ deletions as the other two speakers (44 as opposed to 24 for Speaker 2 and 26 for Speaker 3), but 48% of these deletions may be attributed to legitimate variation in the way in which some words are pronounced. A slightly larger proportion of Speaker 2's schwa deletions can be so explained (58%), but less for Speaker 3 (38%).

Another case where deletion is predictable is in word-final stops, which are frequently omitted, particularly in fast speech before a word initial stop (e.g. "target category" is realised as /tAgI k(ɪ@gri/). There are many of this type of error, and they are analysed in more detail in [1].

Deletions of /h/ are yet another example of phonologically predictable errors. This phoneme can be very variable, as it tends to take on the spectral structure of the following vowel, and is often indistinct from it. In addition one third of /h/ deletions happened after a voiceless plosive e.g. "stroke hotel" which was recognised as /stɹ@Uk @UɪEl/, where it is likely that the /h/ has been merged with the aspiration of the /k/, causing it to be missed by the system. On average 15% of all deletions can be explained by phonological effects.

3.4 Insertions

3.4.1 Individual phonemes

Insertions occur when the system has put in an extra phoneme label, and the numbers of insertions for each and all speakers is presented in Table 7. Some of the highly inserted phonemes are speaker specific – /d/ for Speaker 1 is an interesting example for which there is no explanation; others are common to all speakers, and here /ə/ is the clearest example. As has already been mentioned the speaker-specific dictionaries account for a number of predictable cases of elision of this phoneme, but there are occasions when the speaker does pronounce /ə/, for instance produces not a syllabic /n/, but /n@n/. On average 45% of /ə/ insertions could be accounted for in this way.

As an interesting aside; of the /l/ insertions, all six of those in Speaker 1's case were following the phoneme /O/. This contrasts with four out of seven, for Speaker 3 and one out of four for Speaker 2. This is interesting as for many speakers the so-called "dark /l/" resembles an /O/ vowel in spectral structure (but with a slightly lower intensity), and it could be that an off-glide of /O/ would be confused with /l/.

The phoneme /h/ was inserted only twice, in Speaker 3's data, and both after voiceless plosive (e.g. "time" recognised as /thalm/), and we can hypothesise that the aspiration of the /t/ was what caused the insertion of /h/.

	Spkr 1	Spkr 2	Spkr 3	Total		Spkr 1	Spkr 2	Spkr 3	Total
s	11	8	4	23	i	6	3	0	9
z	9	12	3	24	l	16	1	15	32
S	0	0	0	0	E	2	1	4	7
f	3	5	2	10	(	1	3	0	4
v	5	4	4	13	A	0	1	1	2
T	4	4	2	10	Q	0	0	0	0
D	0	0	0	0	O	1	0	0	1
h	0	0	2	2	U	0	1	0	1
IS	0	0	0	0	u	3	1	1	5
4Z	0	2	0	2	3	1	0	0	1
p	14	12	10	36	@	9	11	20	40
b	2	0	4	6	V	2	4	2	8
t	8	5	10	23	eI	0	0	0	0
d	33	9	9	51	aI	0	2	0	2
k	5	4	1	10	oI	0	0	0	0
g	0	0	0	0	aU	0	0	1	1
m	2	7	3	12	@U	3	1	0	4
n	1	11	2	14	I@	0	1	0	1
N	0	1	0	1	@@	0	0	0	0
l	6	4	7	17	<at>	0	0	0	0
r	4	0	3	7	<oh>	0	0	0	0
w	0	2	2	4	<of>	0	0	0	0
j	2	0	0	2	<or>	0	0	0	0

Table 7 Phoneme insertions for each and all speakers

3.4.2 Manner of articulation

We have already mentioned that consonants are more than twice as likely to get inserted as vowels. The comparatively high level of consonant insertion was common to Speaker 2 (90 consonants compared with 30 vowels) and Speaker 1 (109 consonants and 44 vowels), but not so conspicuous in Speaker 3's results. The distribution of insertions with respect to where they occur is interesting. As many as 65% of consonant insertions were between words, perhaps being confused with breath noise or lip smacks; while 69% of vowel insertions were within words. Figure 8 shows the distribution of the rather finer manner class insertions.

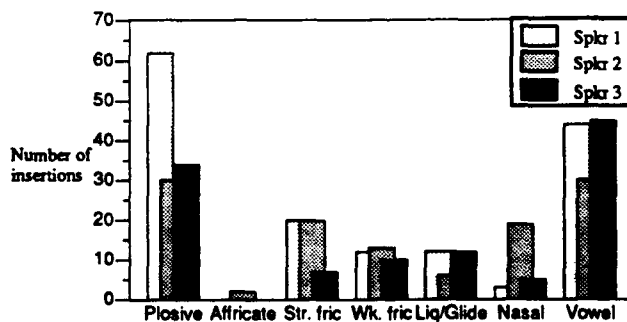


Figure 8 Graph of manner class insertions

Plosives are most frequently inserted, and 83% of these were between words, possibly bearing out the hypothesis that although there are explicit models for breath noise, lip smacks and other glitches, these sounds are nevertheless being recognised as plosives. The only class of consonants that are more often inserted (77%) within words than between them are the strong fricatives, and this may be for the same reason as vowel insertion, for which see 3.4.4 below.

3.4.3 Place of articulation

The distribution of insertions according to place of articulation is given in Figure 9. The alveolars were most likely to be inserted (though the prevalence of /d/ insertions may account for this). Diphthongs were very rarely inserted, as were the velars and palatal-alveolars, though there are too few of the latter to allow any conclusion to be drawn.

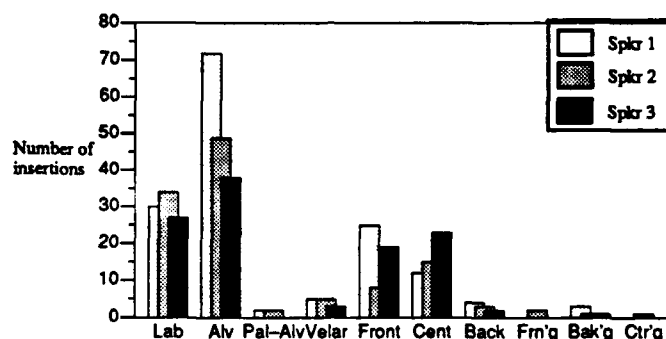


Figure 9 Graph of place class insertions

3.4.4 Contextual effects

A summary of the contexts of insertions can be found in [1]. It often happens that a long phoneme has been recognised as two separate phonemes. Sometimes these phonemes will be identical, as when /@U/ (in "zero", for instance) is transcribed as /@U @U/; in other cases the insertion is phonetically related – "many" is recognised as /mEnil/; or diphthongs may be recognised as two vowels, so "eight" gets recognised as /elit/. Off glides from vowels are often recognised as vowel+/@/, e.g. /O/ in "four" as /O@/, and /@U/ in "zero" as /@U@/. Examples such as repetition of identical phoneme labels, split diphthongs, and offglide schwa account for 80% of the vowel insertions (26%, 21% and 33% respectively).

With the consonants the reasons for insertions are not so clear. Some of the insertions, like the vowels, are due to two identical labels being assigned to one phoneme (13%); others (9%) are phonetically related, as when /s/ following a voiced sound (and usually word initial) is transcribed as /z s/. For example, "four six" was recognised /fO z siks/.

In addition, around 4% of the total consonant insertions are due to the speaker's insertion of certain sounds (mainly glides) as linkers to ease the transition between sounds. Examples of this are insertion of /t/ in "4/8" /fOrelTs/, and between "niner" and "oh". A linking /w/ is inserted in 2/8 – /tuweitTs/, and between "tango" and "eight" /tNg@Uwelt/; and /j/ in "virtually unusable" /v3tS@U j @n.../. The numbers of such insertions are small in our test data, but this probably reflects the limited occasions when such links could occur. For instance, /w/ was inserted only twice each by Speaker 2 and Speaker 3, and not at all by Speaker 1, but on each of those occasions it could be classed as a linking sound. Similarly, /j/ was only inserted twice by Speaker 1 and not at all by the others, and one of those insertions was a linking one. Three out of four of Speaker 1's, and two out of three of Speaker 3's /r/ insertions were linking cases. The fact that the s; stem inserts the appropriate labels in these cases may indicate that we need to model triphones across word boundaries, rather than just word-internally as we do at the moment.

4. Discussion and conclusions

There are many interesting observations to be made from this data. What has been presented here has been an attempt to pull these together and point out general trends which might indicate what the phoneme models are doing right, as well as what they are doing wrong.

From this short discussion there have emerged two types of error: those which are genuine misrecognitions, and those which are due to the normal co-articulatory effects in fluent speech, and are thus to be expected. As far as the former are concerned, phonological effects appear to be involved in around 30% of such errors.

The vast majority of genuine errors are not unexpected, involving as they do, confusions with rather similar phonemes, or deletions of acoustically weak segments. Weak sounds such as nasals or weak fricatives predictably cause problems, as does the neutral /@/. Equally, strong and long sounds such as strong fricatives and diphthongs are well handled. The surprisingly good recognition of liquids and glides may provide an independent vindication of the use of variable frame rate analysis. A large number of the insertions and deletions could probably be prevented if our duration modelling were more sophisticated.

Although the majority of the errors appear to have a phonetic basis, there are cases where the errors are as yet inexplicable from a phonetic point of view – the unusually large number of /d/ insertions in Speaker 1's data, and the poor recognition of the same speaker's /m/, Speaker 3's /b/ and Speaker 2's /V/ for example. A small number of phonemes (and Speaker 2's /V/ may belong to this group) may simply not occur frequently enough for a reliable indication of performance to be made. Where there isn't a phonetic explanation of an error, it would be interesting to find out if the system's own measure of its goodness of match is consistent with our judgement of its performance.

It is important to remember that this study was based on a system which used no dictionary, although the triphones are forced to match at the edges. When lexical and syntactic constraints are available, as they are when the system is run in its usual mode, as a word recogniser, then many of the problems discussed above no longer occur. However, a general improvement in the sub-word level modelling would provide a sound basis for better word recognition and this study has enabled us to pinpoint a few areas where our models might be improved, and may indicate that we need to give some consideration to phonological effects across word boundaries. The level of performance depends ultimately depends on the task and vocabulary, and a next step might be to assess the extent to which the somewhat specialised vocabulary of the ARM task has influenced these results, by looking at other tasks, and bigger vocabularies, as well as at a wider range of speakers.

References

- [1] J D Crowe, "Contextual influences on phoneme recognition in the *ARMADA* speech recognition system", RSRE SP4 Research Note No. 129, 1990.
- [2] A J Fourcin, G Harland, W Barry, and V Hazan, "Speech Input and Output Assessment. Multilingual Methods and Standards", Ellis Horwood Ltd., Chichester, 1989.
- [3] A C Gimson, An introduction to the pronunciation of English. (Second Edition) Edward Arnold (Publishers) Ltd. 1970.
- [4] S M Peeling and K M Ponting, "Further experiments in variable frame rate analysis for speech recognition", RSRE Memorandum No. 4336, 1990.
- [5] K M Ponting and M J Russell, "The ARM project: automatic recognition of spoken airborne reconnaissance reports", Military and Government Speech Tech '89, Arlington Va., 1989.
- [6] M J Russell, K M Ponting, S R Browning, S Downey and P Howell, "Triphone clustering in the *ARM* system", RSRE Memorandum No. 4357, 1990.
- [7] M J Russell, K M Ponting, S M Peeling, S R Browning, J S Bridle, R K Moore, I Galiano and P Howell, "The *ARM* Continuous Speech Recognition System", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, 1990.
- [8] J Wells et al., "Specification of the SAM Phonetic Alphabet (SAM-PA)", included in P Winski, W J Barry and A Fourcin (Eds.) "Support available from the SAM Project for other ESPRIT speech and language work", the SAM Project, University College London, 1990.

INTENTIONALLY BLANK

Appendix A. Phoneme frequency

Rank	Frequency [3]	Phoneme	Example	ARM frequency	ARM rank
1	10.74	@	alpha	6.67	3
2	8.33	l	civil	5.80	5
3	7.58	n	new	7.60	2
4	6.42	t	target	10.27	1
5	5.14	d	damaged	3.64	10
6	4.81	s	six	6.04	4
7	3.66	l	jima	3.33	12
8	3.56	D	then	0.01	42
9	3.51	r	rail	4.58	6
10	3.22	m	map	2.18	=18
11	3.09	k	correction	4.31	7
12	2.97	E	enemy	3.67	9
13	2.81	w	well	1.96	20
14	2.46	z	comprising	2.53	13
15	2.00	v	over	1.33	27
16	1.97	b	about	0.67	33
17	1.83	aI	dimension	2.27	17
18	1.79	f	five	3.46	11
19	1.78	p	papa	1.82	=21
20	1.75	V	up	1.30	28
21	1.71	eI	containing	2.18	=18
22	1.65	i	beacon	3.91	8
23	1.51	@U	close	2.49	14
24	1.46	h	hotel	0.58	=35
25	1.45	{	damaged	2.44	=15
26	1.37	Q	approx	0.83	29
27	1.24	O	four	1.64	=23
28	1.15	N	bearing	0.80	30
29	1.13	u	two	2.44	=15
30	1.05	g	grass	1.82	=21
31	0.96	S	ambush	1.38	26
32	0.88	j	yards	0.62	34
33	0.86	U	woods	0.13	=38
34	0.79	A	charlie	1.64	=23
35	0.61	aU	south	0.71	32
36	0.60	dZ	damaged	0.49	36
37	0.52	3	heard	0.13	=38
38	0.41	tS	charlie	0.40	37
39	0.37	T	three	1.58	25
40	0.34	e@	air	0.08	40
41	0.21	l@	clear	0.76	31
42	0.14	oI	destroyed	0.04	41
43	0.10	Z	pleasure	0.00	=43
44	0.06	U@	poor	0.00	=43

INTENTIONALLY BLANK

Appendix B. The ARM vocabulary

about	@baUt	above	@bVv
access	/ksEs	ack-ack	{k{k
acquisition	/kwIzIS@n	action	{kS@n
active	{kIv	activities	{kIvItiz
activity	{kIvIti	aerial	e@ri@I
acrials	e@ri@Iz	AEW	elidVb@Iju
aircraft	e@krAft	airfield	e@fi@Id
air-raid	e@reId	airstrip	e@striP
alpha	{li@	ambush	{mbUS
ammo	{m@U	ammunition	{mjuniS@n
and	{nd	antenna	{ntEn@
antennae	{ntEn@	anti-aircraft	{ntie@krAft
anti-armour	{ntiAm@	anti-tank	{ntit{nk
approx	@prQks	approximately	@prQksIm@di
APVs	elPiviz	armour	Am@
armoured	Am@d	arms	Amz
array	@reI	artillery	AUl@ri
as	{z	assembled	@sEmblD
assembly	@sEmbli	associated	@s@Usielu@d
badger	b{dZ@	be	bi
beacon	bik@n	bear	be@
bearing	be@riN	being	biIN
below	blI@U	blocked	blQkI
bomber	bQm@	bowyers	baUz@z
bravo	brAv@U	bridge	brIdZ
brow	braU	C2	situ
camouflaged	k{m@fIAZd	camp	k{mp
canal	k@n{I	canilever	k{nIliv@
carriage	k{rIdZ	carriages	k{rIdZlZ
capability	kelp@bIlIti	capacity	k@p{slu
casevac	k{z@v{k	cat	k{t
category	k{u@grI	centre	sEnt@
charlie	tSAli	circular	s3kjul@
civil	slv@I	civilian	s@vIlj@n
clear	kIl@	close	kl@Us
comms	kQmz	collection	k@IEkS@n
column	kQI@m	communications	k@mjunIkElS@nz
complete	k@mplit	comprehensive	kQmprIhEnslv
comprising	k@mpratZIN	concealed	k@nsild
concrete	kQnkrit	conical	kQnlk@I
considerable	k@nsIdr@bl	consisting	k@nsIsItIN
construction	k@nstrVks@n	containing	k@nteInIN
convoy	kQnvOl	co-ordinates	k@UOdIn@ts
co-ords	k@UOdZ	correction	k@rEkS@n
covert	k@Uv3t	crossing	krQsIN
damaged	d{mIdZd	dash	d{S
data	delt@	deck	dEk
defence	dIfEns	defences	dIfEnslz
defended	dIfEndId	degrees	d@griz
delta	dElI@	depot	dEp@U
diameter	dal{mIt@	difficult	dIfIk@It
dimension	dalmEnS@n	dipole	dalp@Ul
dipoles	dalp@Ulz	direct	dIrEkI
direction	dIrEkS@n	dish	dIS
dispersal	dIsp3s@I	dispersals	dIsp3s@Iz
destroyed	dIstroid	dug	dVg
dump	dVmp	each	itS
easy	izi	east	ist
echo	Ek@U	eight	elt
eighteen	eltin	eighty	elti
eleven	IlEv@n	emplacement	Emplei:m@nt
end	End	enemy	En@mi

Appendix B. The ARM vocabulary

engineering	EndZInI@rIN	erect	IrEkI
ESM	iEsEm	estimated	EstImeId
evidence	EvId@ns	evident	EvId@ni
EW	idVblju	facility	f@stlItu
feet	fit	a_few	@fju
field	fi@Id	fifteen	flfun
fifty	flfti	fighter	falt@
fire	fal@	firing	fal@rIN
fishbed	fISbEd	five	falf
fixed	flkst	flanker	fl(nk@
flogger	fIQg@	forger	fOdZ@
fortified	fOulfalD	forty	fOti
four	fO	fourteen	fOtin
foxbat	fQksb(t	foxtrot	fQkstrQt
freight	freIt	frenzied	frEnzId
fuel	fju@l	fulcrum	fUlkr@m
GMT	dZiEmu	golf	gQlf
going	g@UIN	goods	gUdz
grass	grAs	grid	grId
ground	graUnd	guidance	galD@ns
gun	gVn	guns	gVnz
hangar	h(N@	hardened	hAd@nd
having	h(vIN	havok	h(v@k
heading	hEdIN	heavy	hEvi
height	halt	helicopters	hElIkOpt@z
helos	hEl@Uz	hidden	hIdn
hind	halnd	hip	hIp
hocum	h@Uk@m	holding	h@UIdIN
horizontal	hQrlzQnt@l	hom	hOn
hospital	hQspli@l	hotel	h@UeI
hour	aU@	howitzer	haUwltz@
hundred	hVndrEd	hurried	hVrid
hyphen	halI@n	including	InkludIN
incomplete	Ink@mplIt	incorporating	InkOp@reIdIN
india	Indj@	infantry	Inf@ntri
inoperative	InQpr@dV	installaion	Inst@leIS@n
intact	Int(kI	intelligence	IntElIdZ@ns
joint	dZolnt	juliet	dZuliEt
junction	dZVnkS@n	kilo	kiI@U
kilometres	kiIQm@t@z	kilometres/hour	kiIQm@t@z3aU@
knots	nQIs	lanes	leInz
launch	IONtS	launcher	IONtS@
launchers	IONtS@z	length	IENT
less	IEs	level	IEv@l
lift	lIt	light	lalt
like	laIk	lima	lim@
limited	lIm@lId	lines	laInz
little	lIl	loading	l@UdIN
located	l@UkelId	location	l@UkelIS@n
logistics	lQdZIsuks	loop	lup
lorry	lQri	machine	m@Sin
machine-gun	m@SingVn	main	meIn
maintenance	meInt@n@ns	major	meIdZ@
many	mEni	map	m(p
marshalling	mAS@IIN	mast	mAst
material	m@U@rI@l	MCVs	Emsiviz
mechanised	mEk@nalzd	medium	midj@m
message	mEsIdZ	metres	mi@z
mike	maIk	miles	maIiz
miles/hour	maIlzp3aU@	military	mIk@tri
minor	maIn@	missile	mIsall
missiles	mIsallz	mixed	mIkst

Appendix B. The ARM vocabulary

mobile	m@Uball	modified	mQdIfald
more	mO	most	m@Ust
motorised	m@U@ralzd	motorway	m@U@wel
movement	muvm@nt	much	mVtS
navaid	n{veld	near	nl@
new	nju	nine	naIn@
nineteen	naIntin	ninety	naInu
no	n@U	normal	nOm@l
north	nOT	northeast	nOTist
northwest	nOTwEst	not	nQt
noticed	n@Udst	november	n@UvEmb@
number	nVmb@	a-number	@nVmb@
numerous	njum@r@s	observed	Qbz3vd
obstructed	QbstrVktld	occupied	QkjupaId
on	Qn	one	wVn
operational	Qp@relS@n@l	oscar	Qsk@
out	aUt	over	@Uv@
pack	p{k	papa	p{p{
parabolic	p{r@bQllk	partially	pAS@li
partly	pAui	passenger	p{s@ndZ@
peak	pik	per	p3
perhaps	p@h{ps	permanent	p3m@n@nt
personnel	p3s@nEl	pipeline	palplaln
platforms	pl{ifOmz	platoon	pl@tun
plus	plVs	police	p@lis
pontoon	pQntun	position	p@zIS@n
possibly	pQsIbli	practically	pr{ktldi
prepared	pripe@d	principal	prInslp@l
proceeding	pr@sidIN	projeciles	pr@dZEktallz
protected	pr@tEktd	quebec	kw@bEk
radar	reIdA	radio	reIdj@U
rail	rell	railway	rellwel
ready	rEdi	re-arming	riAmIN
recce	rEki	receiver	r@siv@
reconnaissance	r@kQnls@ns	red-cross	rEdkrQs
ref	rEf	reference	rEfr@ns
re-fuelling	rifju@lIN	refurbishing	rif3blSIN
repair	r@pe@	repaired	r@pe@d
repeat	r@pit	report	r@pOt
rhombic	rQmbIk	river	rlv@
road	r@Ud	rocket	rQklt
rockets	rQkltS	romeo	r@Umi@U
rotary	r@Ut@ri	roughly	rVfli
rounds	raUndz	runway	rVnweI
runways	rVnweIz	SAM	s{m
SAM-7	s{msEvn	scientific	sal@ntIfIk
scout	skaUt	section	sEkS@n
sections	sEkS@nz	self-propelled	sElfpr@pEld
semi	sEmi	serviceable	s3vls@bl
sets	sEts	seven	sEvn
seventeen	sEvntin	seventy	sEv@nti
several	sEvr@l	siding	saldIN
sidings	saldINz	sierra	siEr@
sighted	saltId	sighting	saltIN
similar	slmIk@	single	sINgl
site	salt	six	sIks
sixteen	slksun	sixty	slksu
size	salz	skip	sklp
slash	sl{S	small	smOl
so	s@U	some	sVm
something	sVmTIN	south	saUT
southeast	saUTist	southwest	saUTwEst

Appendix B. The ARM vocabulary

span	sp(n	spans	sp(nz
speed	spid	SPGs	EspidZiz
spotted	spQld	squad	skwQd
squadron	skwQdr@n	static	st(tlk
station	stlS@n	steel	sti@l
stone	st@Un	stop	stOp
storage	stOrdZ	stores	stOz
strip	strip	stroke	str@Uk
structural	strVktS@r@l	summit	sVmlt
supply	s@plal	surface	s3fls
support	s@pOt	suspension	s@spEnS@n
swing	swIN	tactical	t(ktlk@l
tango	t(Ng@U	tanker	t(Nk@
tankers	t(Nk@z	tanks	t(Nks
target	tAgl	TARWI	tAwi
task	tAsk	taxiway	t(ksIwel
taxiways	t(ksIwelz	temporarily	tEmp@re@rlli
temporary	tEmp@ri	ten	tEn
than	D@n	thirteen	T3tin
thirty	T3ti	thousand	TaUz@nd
three	Tri	time	taln
to	tu	total	t@U@l
tracked	tr(kt	tracks	tr(ks
train	trIn	trains	trInz
transmitter	trAnzmIt@	transport	trAnspOt
trees	triz	troop	trup
troops	trups	twelve	twElv
twenty	twEnu	twenty-one	twEntwVn
twenty-three	twEnuTri	twenty-two	rwEntu
twin	twIn	two	tu
type	talp	undamaged	Vnd(mIdZd
undefended	VndIfEndId	under	Vnd@
unidentified	VnaldEnUfaId	unipole	junIp@Ul
unknown	Vn@Un	unloading	Vnk@UdIN
unobstructed	Vn@bstrVktId	unoccupied	VnQkjupald
unoperational	VnQp@reIS@n@l	unrepaired	Vnr@pe@d
unserviceable	Vns3vIs@bl	unusable	Vnjuz@bl
uniform	junIfOm	up	Vp
U/S	juEs	usable	juz@bl
use	jus	vehicle	vi@kl
vehicles	vi@klz	vertical	v3tlkl
victor	vJkt@	virtually	v3tS@li
visible	vIz@bl	VSTOL	vIstQl
wagon	w(g@n	wagons	w(g@nz
water	wOt@	weapon	wEp@n
weapons	wEp@nz	well	wEl
west	wEst	whiskey	wIski
wholly	h@Uli	wing	wIN
wire	wal@	width	wIdT
with	wIT	wood	wUd
wooden	wUd@n	woods	wUdz
work	w3k	worked	w3kt
x-ray	Eksrel	YAGI	jAgi
yankee	j(Nki	yard	jAd
yards	jAdz	zero	zl@r@U
zulu	zulu		

Appendix C. Results in full

Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc	Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc
s	136	83.8	11.8	4.4	11	75.7	i	88	84.1	12.5	3.4	6	77.3
z	57	86.0	10.5	3.5	9	70.2	I	117	65.5	18.8	13.7	16	53.8
S	31	83.9	3.2	12.9	0	83.9	E	83	74.7	18.1	7.2	2	72.3
f	78	89.8	5.1	5.1	3	85.9	(	51	90.2	7.8	2.0	1	88.2
v	30	36.7	23.3	40.0	5	20.0	A	41	97.6	2.4	0.0	0	97.6
T	36	83.3	2.8	13.9	4	72.2	Q	13	76.9	7.7	15.4	0	76.9
D	0	--	--	--	--	--	O	37	81.1	8.1	10.8	1	78.4
h	13	61.5	23.1	15.4	0	61.5	U	3	100.0	0.0	0.0	0	100.0
uS	9	66.7	33.3	0.0	0	66.7	u	55	72.7	14.6	12.7	3	67.3
dZ	11	72.7	18.2	9.1	0	72.7	3	3	100.0	0.0	0.0	1	66.7
p	41	92.7	4.9	2.4	14	58.5	@	150	56.0	14.7	29.3	9	50.0
b	15	60.0	13.3	26.7	2	46.7	V	35	82.8	14.3	2.9	2	77.1
t	231	85.3	5.6	9.1	8	81.8	el	49	89.8	10.2	0.0	0	89.8
d	82	59.8	14.6	25.6	33	19.5	al	51	86.3	5.9	7.8	0	86.3
k	97	89.7	3.1	7.2	5	84.5	ol	1	100.0	0.0	0.0	0	100.0
g	41	75.6	17.1	7.3	0	75.6	aU	16	87.5	12.5	0.0	0	87.5
m	49	38.8	32.6	28.6	2	34.7	@U	56	71.4	23.2	5.4	3	66.1
n	171	55.0	25.1	19.9	1	54.4	l@	17	94.1	5.9	0.0	0	94.1
N	18	55.6	27.8	16.6	0	55.6	e@	2	50.0	50.0	0.0	0	50.0
l	75	78.7	5.3	16.0	6	70.7	<ad>	21	19.0	81.0	0.0	0	19.0
r	103	88.4	1.9	9.7	4	84.5	<oh>	6	50.0	50.0	0.0	0	50.0
w	44	79.6	15.9	4.5	0	79.6	<of>	11	45.5	54.5	0.0	0	45.5
j	14	100.0	0.0	0.0	2	85.7	<or>	2	50.0	50.0	0.0	0	50.0

Table C1. Phoneme recognition results for Speaker 1.

Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc	Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc
s	136	86.8	6.6	6.6	8	80.9	i	88	84.1	13.6	2.3	3	80.6
z	57	71.9	21.1	7.0	12	50.8	I	117	70.9	21.4	7.7	1	70.1
S	31	100.0	0.0	0.0	0	100.0	E	83	84.3	9.7	6.0	1	83.1
f	78	89.8	5.1	5.1	5	83.3	(	57	89.5	10.5	0.0	3	84.2
v	30	83.3	10.0	6.7	4	70.0	A	35	100.0	0.0	0.0	1	97.1
T	35	65.7	14.3	20.0	4	54.3	Q	30	96.7	3.3	0.0	0	96.7
D	1	0.0	100.0	0.0	0	0.0	O	37	94.6	2.7	2.7	0	94.6
h	13	76.9	7.7	15.4	0	76.9	U	3	66.7	0.0	33.3	1	33.3
uS	9	88.9	11.1	0.0	0	88.9	u	55	83.6	14.6	1.8	1	81.8
dZ	11	81.8	18.2	0.0	2	63.6	3	3	100.0	0.0	0.0	0	100.0
p	41	90.2	4.9	4.9	12	61.0	@	150	66.0	18.0	16.0	11	58.7
b	15	60.0	6.7	33.3	0	60.0	V	18	50.0	3.3	16.7	4	27.7
t	231	85.3	6.9	7.8	5	83.1	el	49	100.0	0.0	0.0	0	100.0
d	82	63.4	17.1	19.5	9	52.4	al	51	94.1	5.9	0.0	2	90.2
k	97	92.7	2.1	5.2	4	88.7	ol	1	100.0	0.0	0.0	0	100.0
g	41	78.0	17.1	4.9	0	78.0	aU	16	87.5	12.5	0.0	0	87.5
m	49	85.7	10.2	4.1	7	71.4	@U	56	82.1	14.3	3.6	1	80.4
n	171	80.1	8.8	11.1	11	73.7	l@	17	100.0	0.0	0.0	1	94.1
N	18	50.0	22.2	27.8	1	44.4	e@	2	50.0	50.0	0.0	0	50.0
l	75	86.7	9.3	4.0	4	81.3	<ad>	21	38.1	57.1	4.8	0	38.1
r	103	96.1	1.0	2.9	0	96.1	<oh>	6	33.3	66.7	0.0	0	33.3
w	44	86.4	9.1	4.5	2	81.8	<of>	11	27.3	72.7	0.0	0	27.3
j	14	71.4	14.3	14.3	0	71.4	<or>	2	50.0	50.0	0.0	0	50.0

Table C2. Phoneme recognition results for Speaker 2.

Appendix C. Results in full

Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc	Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc
s	136	86.8	7.4	5.8	4	83.8	i	48	93.7	4.2	2.1	0	93.7
z	57	82.5	7.0	10.5	3	77.2	I	160	76.3	15.6	8.1	15	66.9
S	31	96.8	0.0	3.2	0	96.8	E	83	85.6	4.8	9.6	4	80.7
f	78	83.4	12.8	3.8	2	80.8	I	57	86.0	12.3	1.7	0	86.0
v	30	50.0	6.7	43.3	4	36.7	A	35	94.3	5.7	0.0	1	91.4
T	36	75.0	16.7	8.3	2	69.4	Q	13	84.6	0.0	15.4	0	84.6
D	0	--	--	--	--	--	O	37	100.0	0.0	0.0	0	100.0
h	13	61.5	23.1	15.4	2	46.2	U	3	66.7	0.0	33.3	0	66.7
uS	9	88.9	11.1	0.0	0	88.9	u	55	69.1	27.3	3.6	1	67.3
dZ	11	63.6	18.2	18.2	0	63.6	3	3	100.0	0.0	0.0	0	66.7
p	41	82.9	12.2	4.9	10	58.5	@	150	64.7	18.0	17.3	20	51.3
b	15	53.4	33.3	13.3	4	26.7	V	35	74.3	17.1	8.6	2	68.6
t	231	87.0	10.4	2.6	10	82.7	eI	49	91.8	8.2	0.0	0	91.8
d	82	70.7	15.9	13.4	9	59.8	al	51	94.1	5.9	0.0	0	94.1
k	97	90.7	5.2	4.1	1	89.7	oI	1	100.0	0.0	0.0	0	100.0
g	41	87.8	9.8	2.4	0	87.8	aU	16	93.8	6.2	0.0	1	87.5
m	49	91.8	4.1	4.1	3	85.7	@U	56	76.8	21.4	1.8	0	76.8
n	171	77.2	12.3	10.5	2	76.0	I@	17	82.4	17.6	0.0	0	82.4
N	18	50.0	33.3	16.7	0	50.0	e@	2	50.0	50.0	0.0	0	50.0
!	75	76.0	10.7	13.3	7	66.7	<ad>	21	42.9	57.1	0.0	0	42.9
r	103	88.4	5.8	5.8	3	85.4	<oh>	6	50.0	50.0	0.0	0	50.0
w	44	90.9	6.8	2.3	2	86.4	<of>	11	45.4	36.4	18.2	0	45.4
j	14	85.7	14.3	0.0	0	85.7	<or>	2	0.0	100.0	0.0	0	0.0

Table C3. Phoneme recognition performance for Speaker 3.

Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc	Phoneme	Total	% Cor	% Sub	% Del	No. of Ins	% Acc
s	408	85.8	8.6	5.6	23	80.2	i	224	86.1	11.2	2.7	9	82.1
z	171	80.1	12.9	7.0	24	65.5	I	394	72.1	18.3	9.6	32	64.0
S	93	93.5	1.1	5.4	0	93.5	E	249	81.5	10.8	7.6	7	78.7
f	234	87.6	7.7	4.7	10	83.3	I	165	88.5	10.3	1.2	4	86.1
v	90	56.7	13.3	30.0	13	42.3	A	111	97.3	2.7	0.0	2	95.5
T	107	74.8	11.2	14.0	10	65.5	Q	56	89.3	3.6	7.1	0	89.3
D	1	0.0	100.0	0.0	0	0.0	O	111	91.9	3.6	4.5	1	91.0
h	39	66.7	17.9	15.4	2	61.6	U	9	77.8	0.0	22.2	1	66.7
uS	27	81.5	18.5	0.0	0	81.5	u	165	75.2	18.8	6.0	5	72.2
dZ	33	72.7	18.2	9.1	2	66.6	3	9	100.0	0.0	0.0	2	77.2
p	123	88.6	7.3	4.1	36	59.3	@	450	62.2	16.9	20.9	40	53.3
b	45	57.9	17.8	24.2	6	44.4	V	88	72.7	19.3	8.0	8	63.6
t	693	85.9	7.6	6.5	23	82.6	eI	147	93.9	6.1	0.0	0	93.9
d	246	64.6	15.9	19.5	51	43.9	al	153	91.5	5.9	2.6	2	89.2
k	291	91.1	3.4	5.5	10	87.0	oI	3	100.0	0.0	0.0	0	100.0
g	123	80.5	14.6	4.9	0	80.5	aU	48	89.6	6.3	4.1	1	87.5
m	147	72.2	15.6	12.2	12	64.0	@U	168	76.8	19.6	3.6	4	74.4
n	513	70.8	15.4	13.8	21	66.7	I@	51	92.2	7.8	0.0	1	90.2
N	54	51.9	27.8	20.3	1	50.0	e@	6	50.0	50.0	0.0	0	50.0
!	225	80.5	8.4	11.1	17	72.9	<ad>	63	33.3	65.1	1.6	0	33.3
r	309	90.9	2.9	6.2	7	88.7	<oh>	18	44.4	55.6	0.0	0	44.4
w	132	85.6	10.6	3.8	4	82.6	<of>	33	39.4	54.5	6.1	0	39.4
j	42	85.7	9.5	4.8	2	80.9	<or>	6	33.3	66.7	0.0	0	33.3

Table C4. Phoneme recognition performance for all speakers combined.

Appendix C. Results in full

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Plosive	81.1	7.7	11.2	62	65.8	507
Affricate	70.0	25.0	5.0	0	70.0	20
Str Fric	84.4	10.2	5.4	20	75.4	224
Wk Fric	75.8	9.6	14.6	12	68.2	157
Liq/Glide	84.3	5.5	10.2	12	79.2	236
Nasal	51.7	26.9	21.4	3	50.4	238
Vowel	76.0	13.5	10.5	44	71.0	868
Average	74.8	14.1	9.8	21.9	68.6	2250

Table C5. Manner results for Speaker 1.

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Plosive	82.2	8.3	9.5	30	76.3	507
Affricate	85.0	15.0	0.0	2	75.0	20
Str Fric	84.8	9.4	5.8	20	75.9	224
Wk Fric	81.5	8.9	9.6	13	73.2	157
Liq/Glide	89.9	5.9	4.2	6	87.3	236
Nasal	79.0	10.1	10.9	19	71.0	238
Vowel	82.0	12.2	5.8	30	78.6	868
Average	83.5	10.0	6.5	17.1	76.7	2250

Table C6. Manner results for Speaker 2.

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Plosive	83.8	11.1	5.1	34	77.1	507
Affricate	75.0	15.0	10.0	0	75.0	20
Str Fric	87.0	6.3	6.7	7	83.9	224
Wk Fric	73.2	13.4	13.4	10	66.9	157
Liq/Glide	84.7	8.1	7.2	12	79.9	236
Nasal	78.2	12.2	9.6	5	76.1	238
Vowel	80.5	12.9	6.6	45	75.3	871
Average	80.3	11.3	8.4	16.1	76.3	2253

Table C7. Manner results for Speaker 3.

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Plosive	82.4	9.0	8.6	126	74.1	1521
Affricate	76.7	18.3	5.0	2	73.3	60
Str Fric	85.4	8.6	6.0	47	78.4	672
Wk Fric	76.9	10.6	12.5	35	69.4	471
Liq/Glide	86.3	6.5	7.2	30	82.1	708
Nasal	69.6	16.4	14.0	27	65.8	714
Vowel	79.5	12.9	7.6	119	75.0	2607
Average	79.5	10.6	8.7	55.1	74.0	6753

Table C8. Manner results for all speakers.

Appendix C. Results in full

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Labial	72.4	13.3	14.3	30	62.1	293
Alveolar	76.4	11.2	12.4	72	68.0	855
Pal-Al	83.1	9.2	7.7	2	80.0	65
Velar	80.5	10.7	8.8	5	77.5	169
Front	77.0	15.3	7.7	25	69.6	339
Central	61.7	14.4	23.9	12	55.3	188
Back	82.6	8.7	8.7	4	79.9	149
Fronting	88.1	7.9	4.0	0	88.1	101
Backing	75.0	20.8	4.2	3	70.8	72
Centring	89.5	10.5	0.0	0	89.5	19
Average	78.6	12.2	9.1	15.3	74.1	2250

Table C9. Place results for Speaker 1.

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Labial	83.3	8.5	8.2	34	71.7	293
Alveolar	82.9	8.7	8.4	49	77.2	855
Pal-Al	89.2	7.7	3.1	2	86.2	65
Velar	83.4	8.3	8.3	5	80.5	169
Front	80.6	14.8	4.6	8	78.3	345
Central	64.9	19.3	15.8	15	56.1	171
Back	91.9	6.3	1.8	3	90.0	160
Fronting	97.0	3.0	0.0	2	95.0	101
Backing	83.3	11.1	5.6	1	81.9	72
Centring	94.7	5.3	0.0	1	89.5	19
Average	85.1	9.3	5.6	12.0	80.6	2250

Table C10. Place Results for Speaker 2.

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Labial	79.9	11.3	8.8	27	70.6	293
Alveolar	82.3	10.1	7.6	38	77.9	855
Pal-Al	87.7	7.7	4.6	0	87.7	65
Velar	83.4	10.7	5.9	3	81.7	169
Front	82.5	10.9	6.6	19	77.0	348
Central	67.0	17.6	15.4	23	54.8	188
Back	84.6	11.9	3.5	2	83.2	143
Fronting	93.1	6.9	0.0	0	93.1	101
Backing	80.6	18.1	1.3	1	79.2	72
Centring	78.9	21.1	0.0	0	78.9	19
Average	82.0	12.6	5.4	11.3	78.4	2253

Table C11. Place Results for Speaker 3.

Class	% Cor	% Sub	% Del	No. of Ins	% Acc	Total
Labial	78.5	11.0	10.5	91	68.1	879
Alveolar	80.5	10.0	9.5	159	74.3	2565
Pal-Al	86.7	8.2	5.1	4	84.6	195
Velar	82.4	9.9	7.7	13	79.9	507
Front	80.0	13.7	6.3	52	75.0	1032
Central	64.5	17.1	18.4	50	55.4	547
Back	86.6	8.8	4.6	9	84.5	452
Fronting	92.8	5.9	1.3	2	92.1	303
Backing	79.6	16.7	3.7	5	77.3	216
Centring	87.7	12.3	0.0	1	86.0	57
Average	73.9	11.5	6.7	38.6	77.7	6753

Table C12. Place results for all speakers.

Appendix C. Results in full

[illegible]

Table C13. Confusion matrix for Speaker 1

Appendix C. Results in full

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

Table C14. Confusion matrix for Speaker 2

Appendix C. Results in full

[illegible]

Table C15. Confusion matrix for Speaker 3.

Appendix C. Results in full

	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	Other	Total
A	350	25	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	408
B	13	137	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	171
C	1	1	202	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
D	1	1	1	51	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	234
E	1	1	1	1	80	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	90
F	1	1	1	1	1	26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	107
G	1	1	1	1	1	1	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	39
H	1	1	1	1	1	1	1	24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	27
I	1	1	1	1	1	1	1	1	109	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	33
J	1	1	1	1	1	1	1	1	1	26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
K	1	1	1	1	1	1	1	1	1	1	99	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	45
L	1	1	1	1	1	1	1	1	1	1	1	23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	863
M	1	1	1	1	1	1	1	1	1	1	1	1	159	1	1	1	1	1	1	1	1	1	1	1	1	1	1	346
N	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	291
O	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
P	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
Q	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
R	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
S	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
U	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
V	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
W	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
X	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
Y	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
Z	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123
Other	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	123

Table C16. Confusion matrix for all speakers.

Appendix C. Results in full

		%Recognised							Total
		Plo	Aff	SF	WF	L/G	Nas	Vow	
S p o k e n	Plosive	84.8	0.4	1.4	0.4	0.2	.	1.4	507
	Affricate	5.0	70.0	15.0	.	.	5.0	.	20
	Str Fric	2.7	0.9	90.6	0.4	.	.	.	224
	Wk Fric	3.8	0.6	0.6	77.1	.	.	2.5	157
	Liq/Glide	1.3	.	.	0.4	84.3	.	3.8	236
	Nasal	8.0	.	.	0.8	4.6	60.5	4.2	238
n Vowel		0.9	.	0.1	0.2	0.6	0.2	87.2	868

Table C17. Confusion matrix for manner of articulation – Speaker 1.

		%Recognised							Total
		Plo	Aff	SF	WF	L/G	Nas	Vow	
S p o k e n	Plosive	87.8	.	0.6	0.8	0.6	0.4	0.4	507
	Affricate	.	95.0	.	.	.	.	5.0	20
	Str Fric	.	.	91.5	.	.	2.2	0.4	224
	Wk Fric	6.4	.	.	83.4	.	.	0.6	157
	Liq/Glide	0.4	.	.	1.3	90.3	.	3.8	236
	Nasal	0.4	.	.	.	0.4	85.3	2.9	238
n Vowel		0.7	.	0.1	0.1	0.5	0.8	91.0	868

Table C18. Confusion matrix for manner of articulation – Speaker 2.

		%Recognised							Total
		Plo	Aff	SF	WF	L/G	Nas	Vow	
S p o k e n	Plosive	89.7	0.6	0.6	1.0	1.0	0.6	1.2	507
	Affricate	5.0	80.0	5.0	.	.	.	.	20
	Str Fric	1.8	0.4	90.2	0.4	.	.	0.4	224
	Wk Fric	5.7	.	0.6	79.6	.	.	0.6	157
	Liq/Glide	0.4	0.4	.	.	87.3	1.3	3.4	236
	Nasal	.	.	.	.	0.4	85.7	4.2	238
n Vowel		0.2	.	.	0.2	0.8	0.5	91.8	868

Table C19. Confusion matrix for manner of articulation – Speaker 3.

		%Recognised							Total
		Plo	Aff	SF	WF	L/G	Nas	Vow	
S p o k e n	Plosive	87.4	0.3	0.8	0.7	0.6	0.3	1.0	1521
	Affricate	3.3	81.7	6.7	.	.	1.7	1.7	60
	Str Fric	1.5	0.4	90.8	0.3	.	0.7	0.3	672
	Wk Fric	5.3	0.2	0.4	80.0	.	.	1.3	471
	Liq/Glide	0.7	0.1	.	0.6	87.3	0.4	3.7	708
	Nasal	2.2	.	.	0.3	1.8	77.2	3.8	714
n Vowel		0.6	.	0.1	0.2	0.6	0.5	89.9	2607

Table C20. Confusion matrix for manner of articulation – all speakers

Appendix C. Results in full

		%Recognised										Total
		Lab	Alv	P-A	Vel	Fm	Bck	Cen	F'g	B'g	C'g	
S p o k e r 1	Labial	82.3	0.3	0.3	0.3	.	0.3	1.7	0.3	0.3	.	293
	Alveolar	0.9	82.2	0.2	0.4	0.5	0.2	1.1	0.1	0.1	0.1	855
	Pal-Alv	.	3.1	90.8	.	.	.	.	.	.	.	65
	Velar	1.8	2.4	1.8	82.8	0.6	.	0.6	1.2	.	.	169
	Front	.	.	.	.	81.7	2.1	4.1	1.5	0.9	0.9	339
	Back	.	0.7	.	.	4.0	82.6	1.3	.	1.3	.	149
	Central	3.2	1.6	0.5	0.5	3.2	1.1	63.3	0.5	1.1	0.5	188
	Fronting	.	.	.	.	4.0	.	2.0	89.1	.	1.0	101
	Backing	.	.	.	.	4.2	1.4	6.9	5.6	75.0	1.4	72
	Centring	.	.	.	.	5.3	.	.	.	.	94.7	19

Table C21. Confusion matrix for place of articulation – Speaker 1.

		%Recognised										Total
		Lab	Alv	P-A	Vel	Fm	Bck	Cen	F'g	B'g	C'g	
S p o k e r 2	Labial	87.4	3.4	.	0.7	.	.	0.3	.	.	.	293
	Alveolar	2.6	86.7	.	0.1	0.4	0.2	0.8	0.1	0.1	.	855
	Pal-Alv	.	.	92.3	.	3.1	.	1.5	.	.	.	65
	Velar	0.6	3.6	.	85.7	0.6	.	.	0.6	.	0.6	169
	Front	0.3	2.3	1.2	.	84.3	2.3	3.2	0.9	0.3	.	345
	Back	0.6	1.3	.	.	.	91.9	.	0.6	0.6	.	160
	Central	.	2.9	.	.	5.8	2.9	69.0	.	1.2	0.6	171
	Fronting	.	.	.	.	.	2.0	1.0	97.0	.	.	101
	Backing	.	1.4	.	.	.	1.4	6.9	1.4	83.3	.	72
	Centring	.	.	.	.	5.3	.	.	.	.	94.7	19

Table C22. Confusion matrix for place of articulation – Speaker 2.

		%Recognised										Total
		Lab	Alv	P-A	Vel	Fm	Bck	Cen	F'g	B'g	C'g	
S p o k e r 3	Labial	88.7	4.1	.	0.3	.	0.3	1.0	.	.	.	293
	Alveolar	2.1	87.1	0.6	0.1	0.1	0.1	1.9	0.2	.	.	855
	Pal-Alv	.	.	90.8	.	1.5	.	.	1.5	.	.	65
	Velar	2.4	7.1	.	84.6	.	.	.	.	.	.	169
	Front	.	.	.	.	90.2	.	1.4	0.6	0.6	0.6	348
	Back	1.4	1.4	.	.	4.2	84.6	2.1	.	2.1	.	143
	Central	3.2	1.6	.	.	7.8	0.5	69.1	1.1	1.1	.	188
	Fronting	.	.	.	.	5.0	.	.	93.1	2.0	.	101
	Backing	.	.	.	.	4.2	.	9.7	1.4	80.6	.	72
	Centring	.	.	.	.	15.8	.	5.3	.	.	78.9	19

Table C23. Confusion matrix for place of articulation – Speaker 3.

		%Recognised										Total
		Lab	Alv	P-A	Vel	Fm	Bck	Cen	F'g	B'g	C'g	
S p o k e r n	Labial	86.1	2.6	0.1	0.5	.	0.2	1.0	0.1	0.1	.	879
	Alveolar	1.9	85.4	0.3	0.2	0.3	0.2	1.2	0.2	0.1	.	2565
	Pal-Alv	.	1.0	91.3	.	1.5	.	0.5	0.5	.	.	195
	Velar	1.6	4.3	0.6	84.4	0.4	.	0.2	0.6	.	0.2	5.7
	Front	0.1	0.8	0.4	.	85.5	1.5	2.9	1.0	0.6	0.5	1032
	Back	0.6	1.1	.	.	2.7	86.5	1.1	0.2	1.3	.	547
	Central	2.2	2.0	0.2	0.2	5.7	1.5	67.1	0.5	1.1	0.4	452
	Fronting	.	.	.	.	3.0	0.7	1.0	93.1	0.7	0.3	303
	Backing	.	0.5	.	.	2.8	0.9	7.9	7.8	79.6	0.5	216
	Centring	.	.	.	.	8.8	.	1.8	.	.	89.5	57

Table C24. Confusion matrix for place of articulation – all speakers.

REPORT DOCUMENTATION PAGE

DRIC Reference Number (If known)

Overall security classification of sheetUNCLASSIFIED.....
 (As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the field concerned must be marked to indicate the classification eg (R), (C) or (S).)

Originators Reference/Report No. MEMO 4463		Month MARCH	Year 1991
Originators Name and Location RSRE, St Andrews Road Malvern, Worcs WR14 3PS			
Monitoring Agency Name and Location			
Title ANALYSIS OF THE PHONEME RECOGNITION PERFORMANCE OF THE ARM CONTINUOUS SPEECH RECOGNITION SYSTEM			
Report Security Classification UNCLASSIFIED		Title Classification (U, R, C or S) U	
Foreign Language Title (in the case of translations)			
Conference Details			
Agency Reference		Contract Number and Period	
Project Number		Other References	
Authors BROWNING, S R			Pagination and Ref 30
Abstract This memorandum presents the results of a phonetically motivated analysis of the speech recognition system developed as part of the ARM (Airborne Reconnaissance Mission) project. The aim of the work described here is to investigate to what extent errors can be explained by phonetic effects; those which cannot may indicate where models may be improved. The background to the investigation, and the problems of evaluating phoneme recognition performance are described, then the remainder of the report is concerned with a detailed analysis of specific types of errors, motivated by a desire to find phonetic explanations of them.			
			Abstract Classification (U,R,C or S) U
Descriptors			
Distribution Statement (Enter any limitations on the distribution of the document) UNLIMITED			

1

INTENTIONALLY BLANK