AN ACOUSTIC ANALYSIS OF VOWEL PRODUCTION ACROSS TASKS IN A CASE OF NON-FLUENT PROGRESSIVE APHASIA

Karen Croot

School of Communication Sciences and Disorders, University of Sydney, PO Box 170, LIDCOMBE, NSW, AUSTRALIA 1825, and, Speech, Hearing and Language Research Centre, Macquarie University, NSW, AUSTRALIA 2109

ABSTRACT

This paper presents an acoustic investigation of vowel production in reading, naming and repetition tasks by LM, a man with progressive non-fluent aphasia. Plots in the F1/F2 plane showing the centroids of the acoustic targets of [i: E A V O u:] and the formant trajectories of [ai ei ou] demonstrate that LM achieved greater differentiation of targets in reading than in naming or repetition, and that the vowel space for repetition was distorted relative to that of the other two tasks. An earlier study of LM's speech argued that phonological information available from the stimuli in reading and repetition tasks facilitated the activation of stored phonological representations for speech production (Croot, Patterson & Hodges, 1988); the present study suggests that articulatory processing is also facilitated directly or indirectly by the availability of phonological information.

1. INTRODUCTION

Investigating the breakdown of speech sound production in acquired language disorders (aphasia) can provide valuable insights into the nature of normal speech production. The syndrome of non-fluent progressive aphasia, arising in the context of neurodegenerative disease, has the potential to be particularly informative because the speech sound deficits (together with syntactic impairments) are highly selective in the early years of the disease, while other language and cognitive functions are relatively spared [1].

In an earlier study, Croot, Patterson and Hodges [2] showed that two people with non-fluent progressive aphasia produced single words more successfully in a reading task than in repetition, and were least successful in picture naming. The number of *fully correct* responses was higher in reading than repetition and naming, however the number of *phonologically related responses* (containing at least some part of the correct sound structure of the goal utterance[1]) was higher in both reading and repetition than naming. It was concluded that information about the sound-structure of the goal utterance (phonological information) provided by written and spoken task stimuli facilitated the patients' activation of their own knowledge of the sound structure of words (phonological

representations) in speech production. In naming, by contrast, where picture stimuli do not provide any phonological information, the patients were often unable to give any spoken response whatsoever. The advantage in reading over repetition was attributed to the patients' better receptive processing of written versus spoken language, shown by their performance on other language tasks. The patients' sound production errors were assumed to derive at least in part from impaired activation of phonological representations, however there was also evidence of articulatory disruption. It was not clear, therefore, whether the phonological information provided in reading and repetition may also have facilitated articulatory processing; the present study considers this question.

The error analyses in the earlier study [2] were based on impressionistic phonetic transcriptions which were unable to capture the *graded* quality of many sound errors in the patients' speech. This paper reports instead an acoustic analysis of vowels that one of the patients (LM) produced in that study[2], and demonstrates qualitative differences between tasks which were not apparent from the impressionistic transcription data.

One feature of clearly articulated speech is that the acoustic vowel targets are located more peripherally in the F1/F2 space, allowing a greater acoustic distance and thus differentiation between individual targets [3]. It was therefore hypothesised that if speech production is facilitated by phonological information in task stimuli, then LM's utterances in reading and repetition should demonstrate greater differentiation between individual monophthong targets, and thus a generally larger vowel space, than utterances produced in naming. Similarly, the distance between the first and second targets in diphthongs should be greater when phonological information is available from the stimuli.

2. CASE STUDY

LM, a right-handed retired clerk aged 77 years, participated in this study in early 1994, approximately four years after the onset of a non-fluent progressive aphasic syndrome. Longitudinal neuropsychological testing from 1992 onwards [1]

---

[1]In the earlier study we referred to the words we sought to elicit as "targets", however they are referred to as "goal utterances" here in order to avoid confusion between these and the acoustic vowel targets.

---

indicated that he had initially poor phonological and syntactic processing, a reduced digit span and impaired naming and reading aloud; these abilities further declined over time. Visuospatial, perceptual, mnestic and semantic abilities were initially preserved, however, and were still within two standard deviations of normal at the time of this study. By early 1994 LM's spontaneous speech was effortful, and characterised by word-finding difficulties, incomplete syntactic constructions, and errors in sound production which on occasions rendered his speech unintelligible. His nonverbal communication was still effective, and apart from his aphasia he was able to carry out the activities of everyday life without difficulty.

Brain imaging in early 1994 indicated atrophy in the left parietal and temporal regions. LM spoke with a regional British accent typical of the area in East Anglia where he grew up. He wore a hearing aid, and audiometric assessment one year after this study showed unimpaired hearing at and below 1000 Hz (10-20 dB loss), moderately impaired hearing at 1500 Hz (60 dB loss) and a severe hearing loss (> 80 dB at higher frequencies.

3. METHOD

3.1 Speaking Tasks and Materials

LM produced single words in citation form in picture naming, reading aloud and repetition tasks, in response to stimuli consisting of line drawings, written words and spoken words respectively. There were 180 stimuli in each task, of 1, 2 or 3 syllables to manipulate phonological difficulty. All goal utterances were concrete nouns with typical spelling-to-sound correspondence. Words were divided into three blocks, with equal numbers of 1-, 2- and 3-syllable words matched for frequency in each block. One block was presented in each task on each of three testing sessions; block and task order were counterbalanced across sessions.

The earlier study [2] was not designed with acoustic analyses in view, therefore did not control for the number of instances of each vowel elicited, nor for the phonetic environment in which the vowels occurred. In the present study the monophthongs [i: E A V O u:][3] were selected for analysis because there were sufficient goal utterances containing these to assume in most cases a reliable estimate of the centre formant values of the acoustic targets in LM's responses, and because these are typically distributed quite widely in the vowel space and would thus demonstrate the extent and shape of LM's vowel space in each task. The diphthongs [ou ai ei] were also selected for analysis on the basis of sufficiently large Ns. Although the aim was to elicit the same words in each task, it was impossible to obtain identical phonetic environments for vowels across tasks because LM's response did not always match the goal utterance. For example, not all picture stimuli elicited the goal utterance

---

[3]List of vowel symbols used in text, with examples of stimulus words used to elicit them: /i:/ sheep, /E/ bed, /A/ cat, /V/ duck, /O/ sock, /u:/ spoon, /ou/ goat, /ai/ five, /ei/ rake, /au/ house

(or any response at all) in naming, some items elicited more than one response, and variable speech sound errors involving neighbouring consonants occurred in all tasks.

3.2 Recording, Transcription, Scoring

Testing was carried out in LM's home and recorded on VHS-C videotape. Broad phonetic transcriptions were carried out in situ by the author (KC) and later checked against the video for accuracy. Because of accent differences between KC (an Australian English speaker) and LM, a lenient criterion for transcribing LM's /I/, /A/ and /au/ vowels was adopted. /I/ vowels were transcribed as correct if produced in KC's perceptual [I] or [E] space, /A/ vowels as correct in KC's [A] or [V] space and /au/ as correct if perceived as [au] or [ou]. Responses were scored as correct/incorrect, and categorised as phonologically related/unrelated to the goal utterance. Related utterances were presumed to reflect successful production of at least some part of the phonological form of the goal utterance, and shared at least one stressed vowel or consonant phoneme with the goal utterance. Although this was a lenient criterion for assuming phonological relatedness, the earlier study [2] demonstrated that LM's responses classified under this criterion showed more phonological overlap with the goal utterance than would be expected on the basis of chance alone. Only stressed vowels in related responses were acoustically analysed; unrelated responses (no phonological overlap with the goal utterance) were excluded because there was no evidence that LM's intended production was the goal utterance.

3.3 Digitising and Labelling

Recordings were digitised at 20 000 Hz, and the first three vowel formants and their bandwidths were tracked using a 12th order LPC model in the speech signal processing package ESPS/Waves. The automatically-calculated formant tracks were later checked and hand-corrected where necessary. Labelling was carried out in EMU, a hierarchical speech database management system [4], using criteria derived from those in the Australian National Database of Spoken Language [5]. Acoustic vowel targets (a single time point for monophthongs, two for diphthongs) were marked at the point where there was minimum movement of the formant tracks. In high vowels, this was where F2 reached a peak; in open vowels where F1 reached a peak, and in back vowels where F2 reached a trough. Where there was no formant movement between the vowel onset and offset, the vowel midpoint was marked as the target [see further, 6].

3.4 Formant Analyses

Ellipse plots in the F1/F2 plane were calculated for the targets of the monophthongs and diphthongs, and included at least 95% of tokens in each vowel category. For the monophthongs, plots of the centroids of these ellipses were examined to determine the extent of the vowel space in each speaking task and to identify any overlap between vowels. For each diphthong, centroids of both targets were plotted to indicate the extent of the formant trajectories in each task. Vowels preceded by [w] or [r] or followed by [l] were excluded because the lip-rounding

of the former two consonants systematically lowers F2 and F3 and the retraction of dark [l] lowers F2: 20% of vowels in reading, and 42% of vowels in each of naming and repetition were excluded.

The first analysis included all monophthongs produced by LM in related responses, thus including tokens which had been transcribed as other vowels as well as those transcribed as correct. The vowel space for each task in this analysis therefore reflects acoustic variability arising from both phonological and articulatory sources of error. The second analysis, carried out both for monophthongs and diphthongs, only included vowels transcribed as correct productions of the vowel in the goal utterance. This analysis therefore excluded overt phonological substitutions and articulatory errors resulting in perceptually distant vowels, and was therefore informative about the degree of articulatory variability across tasks. Table 1 shows the number of tokens per analysis.

Table 1. Number of tokens in each analysis

	Monophthongs						Diphthongs		
	i:	E	A	V	O	u:	ai	ei	ou
All tokens									
Read	16	16	31	13	12	13	-	-	-
Name	8	17	18	7	13	7	-	-	-
Repeat	6	10	22	15	11	3	-	-	-
Vowel Correct									
Read	16	13	28	13	12	12	13	7	9
Name	7	13	18	4	12	6	12	7	15
Repeat	6	4	12	15	11	3	7	4	15

4. RESULTS AND DISCUSSION

Figures 1 and 2 clearly reveal variation in the monophthong targets due to task. The vowel space was largest for reading, and showed the greatest differentiation of targets, consistent with the hypothesis that the orthographic stimuli in reading provide phonological information which facilitates speech production in non-fluent progressive aphasia. Although the previous study found that the spoken stimuli in repetition assisted LM to produce responses which were phonologically related to the goal utterances [2], there was no evidence of a repetition advantage compared with naming for these vowels. While all six monophthongs were well differentiated in naming, there was no reliable distinction between /E/ and /A/ in repetition. Further, the vowel space for naming, although smaller, preserved approximately the same shape as that for reading except for a more fronted /i:/, whereas the repetition space was distorted relative to that of the other tasks. In particular, /i:/ was centralised, and /V/ and /O/ were considerably lower. It therefore appears that for LM, spoken information in task stimuli helped the activation of at least part of the response, but did not contribute to any "fine-tuning" of the vowels. This is unsurprising, as the information available to him from that source about the required vowel was probably of poor quality due to a combination of impaired hearing, impaired receptive
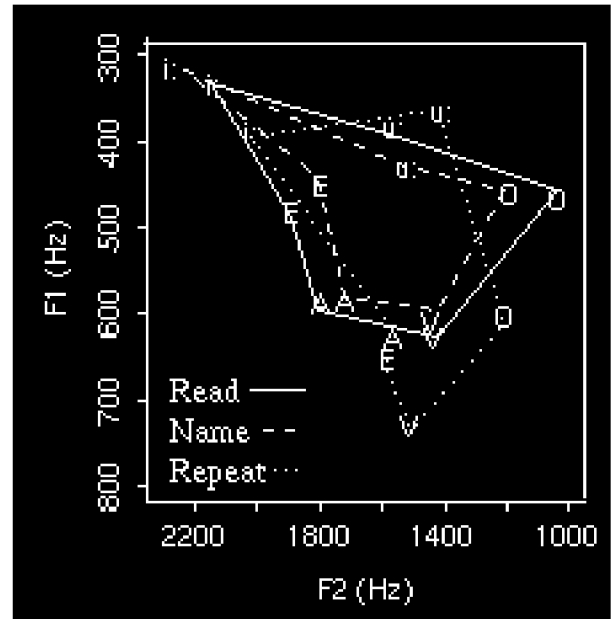

Figure 1: The centroids of the monophthong vowel targets in all utterances.
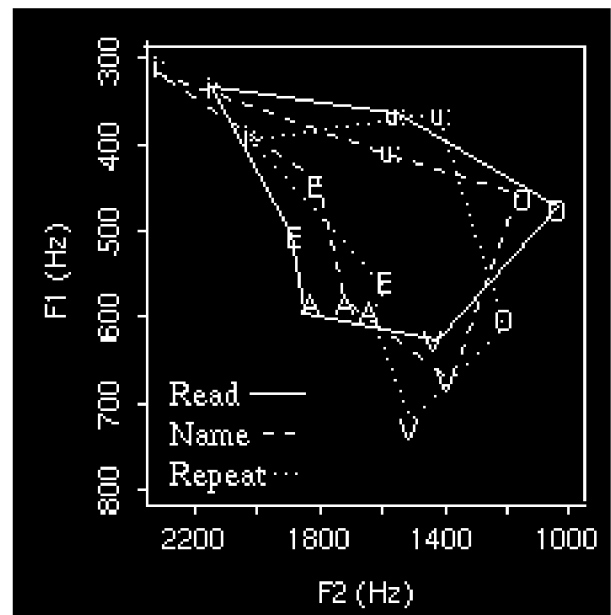

Figure 2: The centroids of the monophthong vowel targets in utterances where the vowel was transcribed as correct.

processing of language, and accent differences between LM and KC (who provided the model word for repetition).

The differences in monophthongs across tasks remained even when the analysis was restricted to vowels transcribed as correct (Figure 2). Figures 3 and 4 also reveal a qualitative difference between the formant trajectories of /ai/ and /ei/ in reading versus naming and repetition. Consistent with the larger monophthong space in this task, these trajectories are longer for reading, although there is little difference in the trajectory lengths of /ou/ (Figure 5). There is no evidence for
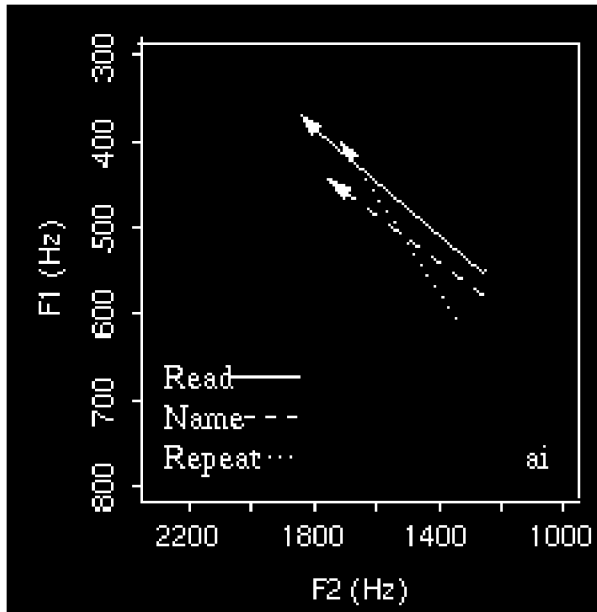
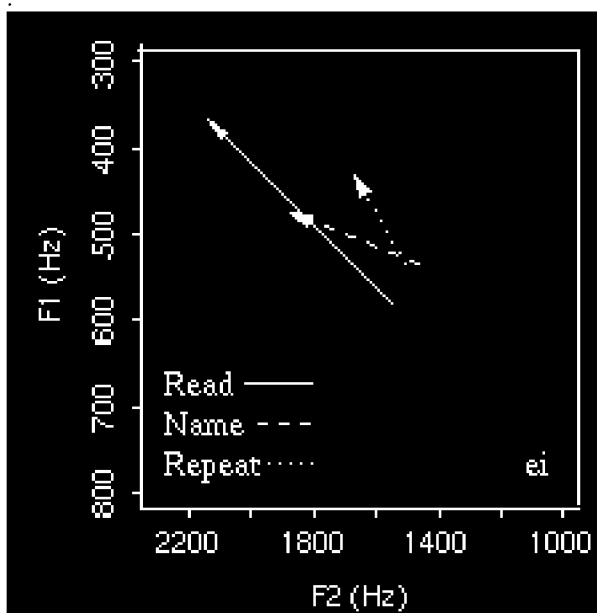**Figure 3:** The formant trajectories of [ai] in the three tasks

Figure 5: The formant trajectories of [ou] in the three tasks.

processing, or an indirect effect, for example, LM may have spoken more clearly when he was more confident of the required phonological form.

This study represents a preliminary acoustic investigation of task differences in the speech of a single person with non-fluent progressive aphasia. Future investigations will explore the role of phonological information in articulatory processing, and the relationship between the receptive processing of speech and its production.

Figure 4: The formant trajectories of [ei] in the three tasks.

greater differentiation of diphthong targets in repetition than naming (trajectories are of similar lengths in both tasks), again suggesting that the spoken stimulus in repetition does not facilitate clearer speech production for LM.

The presence of task differences in the perceptually correct vowels indicates that these differences did not simply reflect a higher rate of phonological substitutions in one task than another. Instead, clarity of articulation, as well as the activation of phonological representations, was influenced by the nature of the speaking task. This may reflect a direct influence of phonological information on articulatory
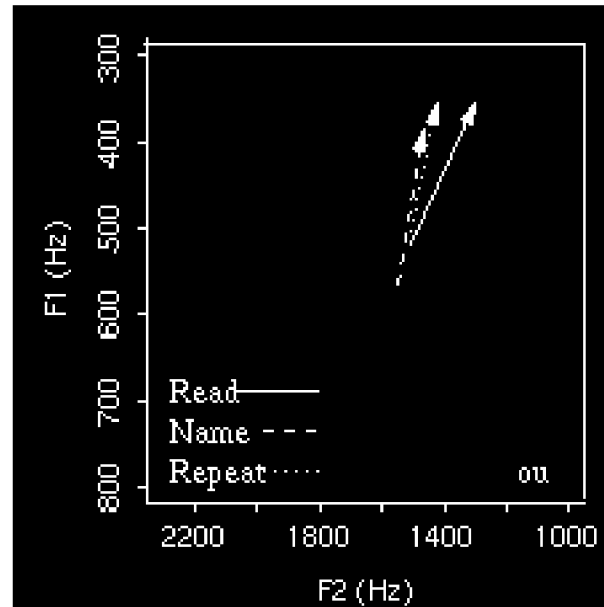
6. REFERENCES

1. Hodges JR, Patterson K "Non-fluent progressive aphasia and semantic dementia: A comparative neuropsychological study", *Journal of the Int. Neuropsychological Society*, 2, 511-524, 1996.

2. Croot K, Patterson K, Hodges JR "Single word production in non-fluent progressive aphasia", *Brain and Language*, 61, 226-273, 1998.

3. Palethorpe, S "Speech intelligibility in communicat-ive difficulty", *Proc. 4th Int. Conf. on Speech Science and Technology (Brisbane)*, 86-90, 1992.

4. Cassidy S, Harrington J "EMU: an enhanced hierarchical speech database management system", *Proc. 6th Int. Conf. on Speech Science and Technology (Adelaide)*, 361-366, 1996.

5. Croot K, Fletcher J, Harrington J "Levels of segment-ation and labelling in the Australian National Data-base of Spoken Language," *Proc. 4th Int. Conf. on Speech Science & Technology (Brisbane)*, 86-90, 1992.

6. Watson CI, Harrington J, Evans Z "An acoustic comparison between New Zealand and Australian English vowels", submitted for publication.