EVIDENCE OF DUAL-ROUTE PHONETIC ENCODING FROM APRAXIA OF SPEECH: IMPLICATIONS FOR PHONETIC ENCODING MODELS

R. A. Varley and S. P. Whiteside

Department of Human Communication Sciences, University of Sheffield, UK.

ABSTRACT

Contemporary psycholinguistic models suggest that there may be dual routes operating in phonetic encoding: a direct route which uses stored syllabic units, and an indirect route which relies on the on-line assembly of sub-syllabic units. The more computationally efficient direct route is more likely to be used for high frequency words, while the indirect route is most likely to be used for novel or low frequency words. We suggest that the acquired neurological disorder of apraxia of speech (AOS), provides a window to speech encoding mechanisms and that the disorder represents an impairment of direct route encoding mechanisms and, therefore, a reliance on indirect mechanisms.

We report an investigation of the production of high and low frequency words across three subject groups: non-brain damaged control (NBDC, N=3); brain damaged control (BDC, N=3) and speakers with AOS (N=4). The results are presented and discussed within the dual-route phonetic encoding hypothesis.

1. INTRODUCTION

Acquired apraxia of speech (AOS) is a motor speech disorder that results from damage in the area of the pre-motor cortex in the language-dominant hemisphere. The characteristics of this motor speech disorder include: a slow rate of speaking [1]; prosodic abnormalities [2]; inconsistent and variable articulatory movements [3], reduced coarticulation [4] and disruptions in all production systems involved in the phonetic encoding of the linguistic message. These disruptions result in multiple phonetic and perceived phonemic errors in speech output, and therefore, lowered speech intelligibility [5].

The traditional conceptualisation of AOS is influenced by psycholinguistic models that suggest that the speech segment plays a critical role in the encoding of speech [6, 7]. In these models, speech encoding is viewed as a process of segment-by-segment access and the subsequent assembly of the syllable/word. Much of the empirical evidence for such theories rests on speech error data from normal speakers (e.g. segmental switches which convert 'car park' to 'par cark'), and AOS is viewed as a failure of these segmental access and syllable assembly processes.

Contemporary psycholinguistic models have, however, begun to challenge the role of the segment in speech production [8]. Levelt and Wheeldon [8] suggest that there are two possible routes in speech encoding: a direct[1] route that operates via stored syllabic schemas, and an indirect[2] assembly route that utilises sub-syllabic units. Direct route encoding is more likely to be used for higher frequency syllables and the indirect route for lower frequency or novel syllables. Direct route encoding permits an efficient and relatively error free output as the multiple degrees of freedom of the speech motor system are constrained into schemas or gestural gestalts [9].

Whiteside and Varley [10] suggest that AOS can usefully be reconceptualised within the direct/indirect or dual-route route hypothesis. They argue that in AOS, either the access to and/or the storage of syllable or word level verbo-motor patterns are disrupted, and that much of the abnormal speech behaviours of speakers with AOS represents an attempt to compensate for this fundamental processing impairment by a reliance on indirect mechanisms. They also suggest that indices of encoding route include parameters such as response latency, utterance and word durations and the degree of coarticulation, with increasing latencies and durations and decreasing coarticulation with indirect route encoding.

This paper reports an investigation of the production of high and low frequency words across three subject groups: non-brain damaged control (NBDC, N=3); brain damaged control (BDC, N=3) and speakers with AOS (N=4). The results are presented and discussed within the dual-route phonetic encoding hypothesis.

2. METHOD

2.1. Subjects

Three groups participated in the study:

1) non-brain damaged controls (NBDC, N=3, mean age=57 years);

2) brain damaged controls (BDC, N=3, mean age=64 years); and

3) speakers with acquired apraxia of speech (AOS, N=4, mean age=61 years).

All participants in the study were female and from South Yorkshire except for one brain-damaged control. The brain-damaged subjects all:

[1] This is referred to as the 'indirect' route by Levelt and Wheeldon [8]

[2] This is referred to as the 'direct' route by Levelt and Wheeldon [8]

- had left-hemisphere lesions;
- were pre-morbidly right-handed;
- were neurologically stable (a minimum of 1 year post onset);
- had varying degrees of aphasic impairment; and
- had no significant dysarthria, which was determined by a test measuring diadokenetic rates.

The brain-damaged controls had no apraxic difficulties. Two were fluent conduction aphasics and one an agrammatic aphasic.

The criteria used in the diagnosis of AOS included:

- the agreement of two independent clinicians;
- prosodic abnormality which was characterised by marked syllable segregation and effort, accompanied by sudden bursts of energy with rapid decay; and
- intact gross oral movement.

2.2. Speech Data

The speech material used in the study consisted of three repetitions of a set of phonetically matched 10 high frequency words (occurrence of more than 100 per million) and 10 low frequency words (occurrence of less than 10 per million). These words were selected from Thorndike and Lorge [11].

The high frequency words were: 'mile', 'base', 'pound', 'bag', 'school', 'cold', 'foot', 'group', 'cup' and 'car'. The low frequency words were: 'bile', 'mace', 'mound', 'bog', 'stool', 'colt', 'soot', 'croup', 'cub', and 'tar'. This gave a total of thirty tokens for each of the high and low frequency word groups, which were then randomised into a single list. Subjects were instructed to repeat each word after the experimenter. Each word was preceded either by 'a' or 'the'. All sessions were recorded in a quiet room using a DAT recorder. The data reported here includes only the first repetition of the data sets.

2.3. Acoustic analysis

All speech data were digitized (SR 10kHz) and analysed using a KAY Computerized Lab (CSL) Model 4300. Speech pressure waveforms, wideband FFT spectrograms and LPC analyses were used to obtain acoustic measures. Only stimuli that were perceived as being 'on target' were analysed.

The measures that were taken included:

- response (or repetition) latencies (in milliseconds) - these were measured from the end of the experimenter's utterance to the start of the participant's utterance;
- utterance durations (in milliseconds) - these were measured from the beginning of the utterance to the end of the entire utterance (i.e. 'The Mile');
- word durations (in milliseconds) - these were measured from the start of the acoustic activity of the target word to the end of the word;
- consonant-vowel coarticulation measures were taken for the bilabial high/low frequency pairs (mile/bile, base/mace, pound/mound), which for these contexts, were diphthongs.

The coarticulation measures for the bilabial subset included measurements of the second formant frequency (F2 in Hz) at the onset of the vowels and the temporal midpoint of the vowels. For the plosive contexts, the vowel onset was taken at the point of the release of the plosive and for the nasal contexts, this was taken at the onset of the vowel following cessation of the low frequency nasal murmur. The difference between the F2 values of the vowel midpoint and the vowel onset were then calculated and measured as the 'F2 change'. Here, a greater F2 change could be interpreted as being indicative of less coarticulation, whereas conversely, a smaller F2 change could be indicative of greater coarticulation.

Further to this, the time interval between the vowel onset and temporal midpoints was measured as the 'CV Timelag' duration (in milliseconds).

3. RESULTS

3.1. Temporal Parameters

The mean and standard deviation values for response latency, utterance duration and word duration are given by word frequency for all three groups in Tables 1, 2 and 3 respectively.

Group	Word Frequency	
	High	Low
Non-brain damaged control	177.03 (84.4)	248.4 (89.1)
Brain-damaged control	348.5 (249.0)	526.5 (381.1)
AOS	1125. 1 (1260.0)	659.8 (447.2)

Table 1: Mean and standard deviation values for all three groups by frequency: response latency.

Group	Word Frequency	
	High	Low
Non-brain damaged control	679.2 (106.8)	692.4 (164.5)
Brain-damaged control	793.7 (177.3)	840.3 (212.2)
AOS	996.8 (331.7)	1089.2 (433.4)

Table 2: Mean and standard deviation values for all three groups by frequency: utterance duration.

Group	Word Frequency	
	High	Low
Non-brain damaged control	572.0 (114.8)	598.6 (140.9)
Brain-damaged control	574.3 (102.5)	634.5 (220.0)
AOS	657.5 (180.9)	615.0 (160.2)

Table 3: Mean and standard deviation values for all three groups by frequency: word duration.

Repeated measures ANOVA tests for the response latency data of all three groups combined indicated that there were no significant differences by frequency. There were, however, significant frequency-by-group interactions ($F(2,76)=4.29$, $p<.05$) and significant group differences in the response latency duration values ($F(2,76)=18.82$, $p<.001$). Post-hoc Scheffé tests showed significant group differences ($p<.05$) between the AOS and NBDC groups for the low frequency words, with the AOS group having longer response latencies. In addition, significant differences were found between the BDC and NBDC groups for the low frequency words, with the former group showing longer response latencies. In the case of the high frequency words, significant differences ($p<.05$) were found between the AOS group and both the BDC and NBDC groups, with the AOS group having the longest response latency values.

In addition, paired *t*-tests indicated that there were significant differences between the response latency values of the high and low frequency data of the non-brain damaged controls ($t(29)=-4.56$, $p<.001$), with the low frequency words displaying longer duration values. This pattern was replicated in the data of the brain damaged controls, who also showed a significant difference between the response latency values of the high and low frequency data using a paired t-test ($t(19)=-2.50$, $p<.05$), in the predicted direction. In contrast to this, the data of the AOS group showed the reverse trend of the high frequency words having significantly longer response latency values. However, when the response latency data of the high and low frequency data were compared using a paired t-test, this difference did not reach significance.

Repeated measures ANOVA tests for the utterance duration values showed significant differences by frequency ($F(1, 57)=5.72$, $p<.05$), with low frequency words displaying longer durations. In addition, significant group differences ($F(2, 57)=12.99$, $p<.001$) were found and post-hoc Scheffé tests showed significant differences ($p<.05$) between the AOS group and both the BDC and NDBC groups for both the low and high frequency words. No frequency-by-group interactions were found.

A repeated measures ANOVA showed no overall frequency, frequency-by-group or group differences for the word duration values. However, appreciably longer word duration values were observed for the high frequency words of the AOS group. Further to this, a Least Significant Difference post-hoc test, indicated a significant difference ($p<.05$) between the AOS and NBDC groups.

3.2. Coarticulation Measures

The mean and standard deviation values for the F2 change and CV timelag parameters are given by word frequency in Tables 4, 5 and 6, for all three groups respectively. A repeated measures ANOVA for the F2 change parameter data showed no overall significant differences for any of the comparisons (namely frequency, frequency-by-group, and group comparisons).

	Word Frequency	
	High	Low
F2 Change (Hz)	109.2 (282.6)	141.7 (324.8)
CV Timelag (ms)	124.3 (29.4)	117.1 (14.7)

Table 4: Mean and standard deviation values for the degree of F2 change from vowel onset to mid vowel (in Hz) and the CV timelag values (in ms) for the first repetition bilabial data: non brain damaged controls.

	Word Frequency	
	High	Low
F2 Change (Hz)	280.7 (362.3)	240.8 (210.3)
CV Timelag (ms)	122.7 (23.4)	119.8 (19.3)

Table 5: Mean and standard deviation values for the degree of F2 change from vowel onset to mid vowel (in Hz) and the CV timelag values (in ms) for the first repetition bilabial data: brain damaged controls.

	Word Frequency	
	High	Low
F2 Change (Hz)	330.3 (371.8)	198. 3 (331.4)
CV Timelag (ms)	193.8 (41.7)	188.8 (84.9)

Table 6: Mean and standard deviation values for the degree of F2 change from vowel onset to mid vowel (in Hz) and the CV timelag values (in ms) for the first repetition bilabial data: apraxic speakers.

In addition, a repeated measures ANOVA for the CV timelag values showed no significant frequency or frequency-by-group interactions. However, significant group differences ($F(2, 15)=8.82$, $p<.005$) were indicated. Subsequent post-hoc Scheffé tests showed significant differences ($p<.05$) between the AOS group and both the BDC and NDBC groups, for the high frequency words. The post-hoc tests also indicated significant differences between the AOS and NBDC groups for the low frequency words. In all group comparisons the CV timelag values were longer for the AOS group.

4. DISCUSSION

The results suggest that there is some evidence for dual-route phonetic encoding. Although there are relatively small numbers of observations, there are trends evident in the data that suggest high frequency words are encoded differently from low frequency phonetically-matched cognates. Trends for high frequency words to have shorter response latencies, utterance and word durations than their low frequency cognate are evident in the data from non-phonetically disordered subjects (NBDC and BDC). These trends reached significance for both groups on the response latency and utterance duration measures. Additional word duration data would be valuable to assess

whether the trend in this fine durational measure could be raised to significance. The response latency difference might be attributable to a linguistic lexical access factor, for example, low frequency words require a greater level of activation and thus take longer to reach threshold than higher frequency counterparts. Hence frequency effects can be observed on many linguistic tasks, whether or not they involve phonetic encoding. It is more difficult, however, to explain utterance and word duration differences in this way as these measures appear to address post-access and phonetic parameters more narrowly. The pattern of durational data from the non-phonetically disordered subjects therefore suggests some support for the hypothesis that high frequency words are retrieved as wholes, whilst lower frequency words require more phonetic planning.

The durational data from the AOS subjects is suggestive of abolition of any frequency effect and provides preliminary support for the reconceptualisation of AOS as an impairment of direct route speech encoding [10]. The utterance duration measures revealed a pattern of performance similar to that of the control groups, with high frequency words having shorter durations, but on the remaining measures, the data from the AOS showed the reverse trend with shorter durations on low frequency words. We advance the hypothesis that the fundamental impairment in AOS is the inability to access, or the loss, of stored word schema. As a result, all speech output, independent of its frequency of occurrence, would have to be encoded via less computationally efficient indirect route encoding resources. This entails the assembly of the syllable/word from sub-component gestures. Indirect route encoding is likely to be less efficient for speech assembly for all speakers in that that it is slower and more error prone, but for speakers with AOS it maybe particularly problematic due to the presence of a lesion in the motor control regions of the language dominant hemisphere. Further validation of this hypothesis depends partly upon the inclusion of a phonetically disordered, but non-AOS group (subjects with dysarthric disorders) within the experiment. This would establish whether the abolition of frequency effects observed in this study are specific to phonetic encoding disorders (AOS), or are a general consequence of the perturbation introduced into speech production by both encoding and implementation disorders.

The coarticulation measures produced a less coherent pattern of results than the corresponding durational data. It was hypothesized that the storage of information into a syllable/word gestalt would result in a greater degree of intrasyllabic cohesiveness and would be reflected in increased degrees of coarticulation. There is some temporal evidence of this from the CV timelag data of the AOS group, who displayed longer durations than the control groups. There is, however, no evidence of this from either the temporal data of the two control groups, or the F2 frequency change data of all three groups. There are two possible reasons for this. First, only a small subset of data was used in the coarticulation analyses. The small number of observations might therefore be a factor in the failure to identify potentially subtle patterns within the data. Second, the sensitivity of the F2 and CV timelag measures may be a relevant factor. The measures that were used are likely to be too gross to capture any subtle coarticulation patterns that may be differentiating high/low frequency words. Metrics that could be used for measuring subtle coarticulation patterns need further development.

5. ACKNOWLEDGMENTS

The authors wish to thank the participants, and our appreciation is also extended to H. Luff for contributions to earlier phases of the investigation.

6. REFERENCES

1. Kent, R. D. and Rosenbek, J. C. "Acoustic patterns of apraxia of speech", *Journal of Speech and Hearing Research*, 26, 1983, 231-249.

2. Kent, R. D. and Rosenbek, J. C. "Prosodic and neurologic lesion", *Brain and Language*, 15, 1982, 259-291.

3. Hardcastle, W. J. "Electropalatographic study of articulation disorders in verbal dyspraxia", In: Ryalls, J. H. (Ed.), *Phonetic Approaches to Speech Production in Aphasia and Related Disorders*, College-Hill Press, Boston, 1987.

4. Zeigler, W. and Von Cramon, D. "Disturbed coarticulation in apraxia of speech: acoustic evidence", *Brain and Language*, 29, 1986, 34-47.

5. Varley, R. A. and Whiteside, S. P. "Voicing in severe apraxia of speech: perceptual and acoustic analysis of a single case", *Journal of Neurolinguistics*, 11(3), 1998, 259-273.

6. Shattuck-Hufnagel, S. "The role of word-onset consonants in speech production planning: New evidence from speech error patterns", In: Keller, E. and Gopnik, M. (Eds.), *Motor and Sensory Processes of Language*, Lawrence Erlbaum Associates, Hillsdale New Jersey, 1987.

7. Dell, G. "The retrieval of phonological forms in production: tests of prediction from a connectionist model", *Journal of Memory and Language*, 27, 1988, 124-142.

8. Levelt, W. J. M. and Wheeldon, L. "Do speakers have access to a mental syllabary?", *Cognition*, 50, 1994, 239-269.

9. Keller, E. "The cortical representations of motor processes of speech", In: Keller, E. and Gopnik, M. (Eds.), *Motor and Sensory Processes of Language*, Lawrence Erlbaum Associates, Hillsdale New Jersey, 1987.

10. Whiteside, S. P. and Varley, R. A. "A new conceptualisation of apraxia of speech: a synthesis of evidence", *Cortex*, 34, 1998, 221-231.

11. Thorndike, E. L. and Lorge, I. *The Teacher's Wordbook of 30,000 Words*, Teachers College Columbia University, New York, 1944.