



THE PUNCH AND JUDY MAN: A STUDY OF PHONOLOGICAL / PHONETIC VARIATION

Gavin J Dempster, Sheila M Williams, Sandra P Whiteside

Centre for Language Engineering, Department of Computer Science and Speech Science Unit,
University of Sheffield, UK S10 2TN

ABSTRACT

This paper describes an approach to the analysis of speaker variation through a contrastive study of the voices of characters in a puppet show, produced by a single speaker. A recording of an impromptu performance of an unscripted Punch and Judy show is analysed to compare variability within and between the three principle characters. Here we investigate some of the phonological and acoustic-phonetic factors that lead to perceptual discrimination between the character voices. The effects of context, including phonological boundaries, are considered in investigating variations in pronunciation of common words and phrases found within in each character. Acoustic correlates of some of these variants are also discussed. The distribution of phonemic realizations of word-final underlying /t/ segments in a set of common words is found to vary substantially between two of the characters. Differences in the acoustic realizations of the same vowel from each of the two characters supports earlier evidence that voice quality is also a distinguishing factor. In addition, some evidence is presented that the need to emphasise such distinctions may influence lexical choice.

1. INTRODUCTION

We investigate speaker variation by contrasting the voices of characters in an unscripted Punch and Judy show created by a single puppeteer (Bob Arkley), without the use of any mechanical or other voice transformation devices.

The three characters studied are Barnacle Bill the sailor, Mister Punch and Judy. The two investigators who were not present at the show experienced no difficulty in distinguishing auditorily between the characters in contrastive contexts, apparently on the basis of accent and voice quality.

The character voices display variations in accent and voice quality normally indicative of the social or regional origins of the speaker. We informally attributed Mister Punch to the British Midlands (Birmingham) and Barnacle Bill to Bristol. Judy was assigned to the non-regional RP variant of British English. Informal assessments of the characters' voice quality supported these attributions [1], with Judy apparently having a relatively deep voice and slow enunciation, Mister Punch a much more rapid and higher pitched voice and Barnacle Bill perceived as somewhere between these two extremes.

We present a study of some aspects of segment distribution to identify the characteristics that contribute to the individual perceived identities of the characters.

2. BACKGROUND

This study forms part of a much more comprehensive investigation into speech and speaker characterisation (ESPRIT Working Group No. 6298, VOX). It is intended partly as an exercise in the integration of articulatory, perceptual and acoustic approaches. The work reported here is an exploratory study of a data set that is supplementary to a core speaker style database [2].

Following a preliminary general analysis of the data in which we applied a variety of perceptual and acoustic techniques [1] and a more detailed examination into the acoustic correlates of some of the pitch contrasts that distinguish the character voices [3], we now examine segmental distribution in a little more detail.

Phonemic investigations of inter-speaker variation are complicated by the systemic nature of phonology. At a phonemic level, clearly audible differences between the phonetic quality of segments from different speakers portraying the same utterance need not be distinguished. For different languages, the same symbol may be employed to represent sounds of the different languages for which there is little overlap in phonetic realization. However, for the purposes of contrastive study between languages or between speakers or speech styles, the relationship of the realization of a word to the idiolect and style of the speaker requires that segments from different speakers that would be phonemically distinct if found in the speech of a single speaker also need to be identified.

We are therefore interested in deriving an appropriate level of detail for transcription (of vowel sounds, for example) that will reflect contrastive differences between speakers as well as significant differences of meaning while maintaining the abstract symbolic representation of language as a system that allows us to capture significant generalisations.

3. DATA ANALYSIS

In this paper we analyse phonemic transcriptions of parts of the recording to compare some aspects of segment distribution. A set of common words with underlying word-final /t/ are examined and the surface variants identified are compared and discussed with respect to the different character voices and to perceptually derived prosodic boundaries. In addition, phonemic representations of realizations of the phrase 'boys and girls' and the word 'just' are compared with acoustic measurements.

3.1 The Data

The speech data examined here was recorded on DAT tape at an impromptu performance of a Punch and Judy show. The show took place in a large rehearsal room during a break in rehearsals of an amateur dramatic production. The audience were members of the drama group, their families and friends, and ranged in age from about 8 to 70 years. The show is adaptive and performed unscripted following "story-lines". Only the voices of characters created by the puppet-master are analysed here.

3.1.1 The Speaker

The puppet-master is aged 44 and uses no artificial voice transformation devices. He has been performing Punch and Judy shows and developing the various characters since about the age of 7.

3.1.3 Selection and Transcription of target items

During a preliminary study, we had produced a full orthographic transcription of the full 35 minute show [1]. Minor and major prosodic boundaries had been assigned perceptually using Reyelt's method [4], displaying a high level of agreement between all three authors.

For this analysis, the target words 'that(s)', 'what(s)' and 'it(s)' were located in the orthographic transcription. The words all contain word-final /t/ and are among the 50 most common lemmas in English usage [5].

An orthographic sequence containing each target word, together with a left and right context wherever possible, was listed. Each sequence was then located in the recording and the portion of the utterance containing the sequence was transcribed into a broad phonetic representation.

In these transcriptions, the following phonetic realizations, corresponding to the word-final underlying /t/ segment, were found:

[d]	voiced dental/alveolar closure
[t]	voiceless dental/alveolar closure
[r]	unreleased voiceless dental/alveolar closure
[ʔ]	glottal stop
0	no corresponding segment to the /t/
[p]	voiceless bilabial closure
[p̚]	unreleased voiceless bilabial closure
[ɹ]	intervocalic approximant

After, identifying all the target segments, the transcribed portions containing the target words were each attributed to one of the puppet characters by cross-reference to the orthographic transcription.

For each of the three characters, Barnacle Bill (BB), Mister Punch (MP) and Judy (J), several examples of the phrase 'boys and girls' and all examples of the word 'just' were also located and transcribed in a similar way.

3.2 The Analysis Techniques

Target items belonging to the /t/-final word set and total words per utterance were counted, acoustic measurements of overall duration and formant frequency values were estimated for some target items and the perceived phonetic variants were assigned to phonemic categories. Then the overall frequency of target items and of each variant realization were calculated for each of the character voices.

3.2.1 Word counts

First, the number of 'that(s)', 'what(s)' and 'it(s)' target words was counted for each of the character voices. Rapid repetitions of the same word several times, which occur in a few instances throughout the performance and sung words, were excluded from the analysis as being special cases, not common to fluent speech.

Then approximate totals of the number of words in each utterance were estimated and summed for each character. Total word counts are approximate only, as it is not possible to define precisely what constitutes a separate word in spontaneous speech, due to the presence of various contracted forms and portmanteau words. In general, if there exists an orthographic representation of a contracted expression which is commonly found in written texts (such as 'couldn't', for example) then this has been used in the orthographic transcription in preference to separate words, wherever it provided a closer match to the perceived utterance.

3.2.2 Acoustic Measures

The duration of each of the selected 'boys and girls' phrases was estimated using a Kay CSL Model 4300 speech analysis workstation. Similarly, the length of each example of the word 'just', together with the frequency values of the first and second formants of the vowel realization were assessed.

3.2.3 Phonemic attributions

It was decided that significant distinctions in underlying /t/ realizations, for phonological purposes were presence/absence of audible closure and evidence of place of closure. On this basis unreleased closures were grouped with full closures to obtain the phonemic categorization.

3.2.4 Calculation of distributions

The frequency of usage of target items were calculated as percentages of the number of words uttered by each puppet character. The instances of each variant found were calculated as percentages of the total number of examples from the target set for each character voice.

3.2.5 Phonemic Contexts

The phonemic contexts, including prosodic boundaries, of the variants were examined for evidence of systematic effects. Phonemic transcriptions were compared with acoustic measurements of the selected phrases to identify any correlations.

3.2.6 Vowel Formant level comparisons

An analysis of vowel formant frequencies found in each unobscured instance of the word 'just' was performed on FFT spectrograms derived by the Kay system, with vowel formant midpoint values estimated visually and measured using the cursor.

3.3 The Results

The results of all the analyses are collated here according to the different target data which was selected for closer examination.

3.3.1 Word set distributions

Distributions of the target words for each character from the set of words, 'that(s)', 'what(s)' and 'it(s)', are shown in Table I below. The percentage

distribution represents the ratio of the target words found to the total word count for all utterances for that character.

Table I: Distribution of target words

	BB	MP	J
targets	80	190	19
total	1301	3264	687
%age	6.15	5.82	2.77

3.3.2 Distribution of word-final variants

The distribution of each phonetic variant found corresponding to the word-final underlying /t/ segment is summarised below, for each character.

Table II: Distribution of phonetic realizations

	BB	BB%	MP	MP%	J	J%
[d]	3	3.75	13	6.84	1	5.26
[t]	19	23.75	148	77.89	17	89.47
[ɾ]	2	2.50	7	3.68	0	0.00
[ʔ]	3	3.75	5	2.63	0	0.00
0	48	60.00	16	8.42	1	5.26
[p]	1	1.25	0	0.00	0	0.00
[p̚]	4	5.00	0	0.00	0	0.00
[ɹ]	0	0.00	1	0.53	0	0.00
total	80		190		19	

3.3.3 Phonemic distribution

Similarly, the phonemic categories applied to the variants are summarised below.

Table III: Phonemic distributions

	BB	BB%	MP	MP%	J	J%
/d/	3	3.75	13	6.84	1	5.26
/t/	21	26.25	155	81.58	17	89.47
/ɹ/	3	3.75	5	2.63	0	0.00
0	48	60.00	16	8.42	1	5.26
/p/	5	6.25	0	0.00	0	0.00
/ɹ̥/	0	0.00	1	0.53	0	0.00
total	80		190		19	

3.3.4 Analysis of the phrase 'boys and girls'

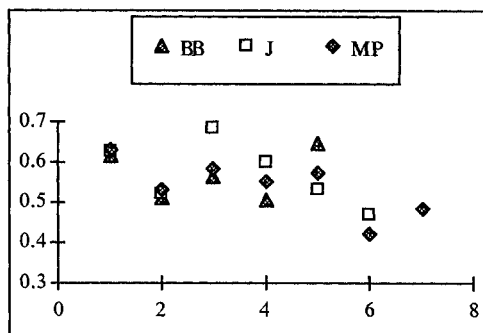
The phonemic transcriptions of the phrase 'boys and girls' varied little between the three characters. In one example from Judy's speech, the first vowel was not diphthongised and for all three characters the pronunciation of 'and' varied audibly. For Barnacle Bill, /d/ deletion was constant but the vowel was variously realised as /æ/, schwa or deleted. Mister Punch also deleted the /d/ regularly but tended to retain the /æ/ in almost every case.

Judy was the only character to retain /d/ and this occurred in 2 out of the 6 examples processed. Oddly, the Judy voice either retained the /æ/ vowel or deleted it. No schwa forms were found. In cases where the vowel was absent, the following nasal was sometimes syllabified.

Phrase lengths varied from 0.419 to 0.682 secs. However the average length of the measureable examples for each character varied from 0.539 secs for Mister Punch (over 7 examples) to 0.566 secs for

Barnacle Bill (5 examples) and 0.571 secs for Judy (6 examples).

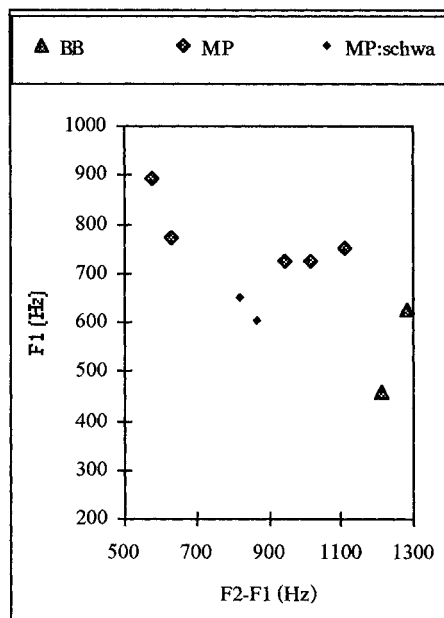
Graph I: Phrase length 'boys and girls'



3.3.5 Analysis of the word 'just'

The distribution of the lexeme 'just' was uneven over the three characters with no examples in Judy's speech and, of the 10 examples for which duration was estimated, only three came from Barnacle Bill. One of these was partially obscured by audience noise and so its formant values could not be ascertained. For the rest, we display the F1/F2 distribution relative to the phonemic attribution of the vowel perceived. Average word-length for Barnacle Bill was 0.217 and for Mister Punch, 0.242 secs. When collated by vowel quality, the words containing schwa averaged 0.223 secs, whereas those containing the mid back vowel (all from Mister Punch) averaged 0.246 secs.

Graph II: F1/F2 Distribution



A total of 4 examples from Barnacle Bill and 13 from Mister Punch were transcribed phonemically. The transcriptions varied mostly in the category of vowel perceived but again Barnacle Bill retained the word-final /t/ in only 1 of the 4 examples whereas Mister Punch

retained it in 12 out of the 13 examples. The word-onset was realised as a voiced affricate in all cases except one where Barnacle Bill produces /t/ immediately following a series of snoring sounds from Mister Punch. This may be due to starting from a whispered voice. Barnacle Bill's vowels were both reduced to schwa whereas Mister Punch also had a mid, back, unrounded form which predominated although it's phonetic realization varied, being somewhat more close in some cases.

4. DISCUSSION

The number of examples of the target words varied considerably as did the total number of words produced. The analysis of the /t/ set of target words indicated that pronunciation of word-final /t/ distinguishes the speech of Barnacle Bill from that of Mister Punch and Judy, with Barnacle Bill manifesting 60% deletions compared with only 8.42% and 5.26% for the other two characters. Interestingly, this factor seems to be particularly discriminatory between Barnacle Bill and Mister Punch, since there are far fewer examples of the target words in Judy's utterances, suggesting the possibility that distinguishing factors may be emphasised by lexical choice.

The variant /p/ was found only the speech of Barnacle Bill and in contexts where it immediately preceded /m/, /b/, /w/ or /f/ in the onset of the following word, with no major prosodic boundaries intervening. Similarly, /d/ for Barnacle Bill, only appeared preceding an unstressed word-initial vowel. It should be noted that both /t/ and 0 realizations were also found preceding word-initial /w/ and although the /t/ variant was found in much lower proportions than segment deletion, no context effect on distribution was apparent.

Again, for Mister Punch the /d/ variant occurs mostly before an unstressed vocalic onset although there are two cases where it precedes a nasal (in both cases in the context 'what noise').

Judy displays too few variants to analyse, presumably at least partially due to the low incidence of the target words.

Although the sample sizes are small, the phrase length data does tend to support our impressionistic findings that one of the factors which distinguishes the characters is speech speed, with Mister Punch's speech being generally most rapid and Judy slowest. There was no clear correlation between phrase length and number of segments present in the phonemic transcription.

Our previous informal accent attributions also seem to be more strongly supported by the contrastive word-final /t/ retention and deletion [6] between Mister Punch and Barnacle Bill than was apparent from our initial analysis [1].

The contrast between /t/ retention and deletion was also maintained in the evidence from the word 'just', although here the durational relationship seems to be reversed between the two characters represented. However, with only 3 examples for which duration could be successfully measured for Barnacle Bill this may be due to the sparseness of the data. The perceptual category attributions between the schwa vowels and the mid back vowels are supported by the distribution of their formant values. Further, the F1/F2

graph also displays a distinction between the schwa realization for Mister Punch and Barnacle Bill. This may be due to larynx raising to obtain the voice quality distinctions perceived.

Given the high levels of background noise, it was often difficult to locate formant frequencies and, in cases of considerable overlap by audience response, even word and phrase boundaries were difficult to locate with any confidence, thus reducing the amount of acoustic measurements we could extract from the recording.

5. SUMMARY

This is a preliminary phonological analysis of a relatively large data set, examining a few potential contexts for variation. Here we have demonstrated how phonological aspects of the utterances help to support voice identity in distinguishing the characters produced by the same speaker for a puppet show. We plan more detailed analysis of this data as part of our ongoing work on speech and speaker characterization.

ACKNOWLEDGEMENTS

The authors thank Bob Arkley ("Professor Bauble") and "the Dream Team" for granting permission to record and analyse the performance of the Punch and Judy show. The work was carried out in connection with EC Working Group 6298 (VOX: speech/speaker characterization).

REFERENCES

- [1] Sheila M Williams, Sandra P Whiteside, Gavin J Dempster, *The Punch and Judy Man: Speaker or Speakers?*, ESCA, ISSN: 1018-4554, pp131-4, 1994.
- [2] Vincent Péan, Sheila Williams and Maxine Eskénazi, *The Design and Recording of ICY, a Corpus for the Study of Intraspeaker Variability and the Characterization of Speaking Styles*, Proc. EUROSPREECH '93, Berlin, Germany, 1993.
- [3] Sandra P Whiteside, Gavin J Dempster, Sheila M Williams, *The Punch and Judy man: One or more speakers*, (to appear in: Proc. IOA Autumn meeting on Speech and Hearing, Windermere, 1994).
- [4] Matthias Reyelt, *Experimental investigation on the perceptual consistency and the automatic recognition of prosodic units in spoken German*, Proc. ESCA Workshop on Prosody, Lund, Sweden, Lund University Working Papers 41. 1993.
- [5] W. Nelson Francis, Henry Kucera, *Frequency Analysis of English Usage*, Boston: Houghton Mifflin Company, 1982.
- [6] J. C. Wells, *Accents of English* (Volumes. 1 and 2), Cambridge University Press, 1982