THE PROSODIC CORRELATES OF EXPRESSIVE READING

R. Cowie†, E. Douglas-Cowie†, A. Wichmann#, P. Hartley#and C.Smith#
†Queen's University Belfast, UK, #University of Central Lancashire, UK

ABSTRACT

This paper examines the prosodic correlates of expressiveness and fluency in children's reading. Subjects read a passage and were rated on two scales - expressiveness and fluency. Relationships were explored between these ratings and a wide range of prosodic measures. Broadly speaking, fluency is related to measures of temporal organisation , and expressiveness to pitch variation. Closer inspection shows that expressiveness and fluency are interrelated in complex ways. Measures related to expressiveness are text-dependent, and there are gender effects.

1. INTRODUCTION

This study identifies the prosodic correlates of expressive reading in a group of 24 children.

Intuitively, it seems obvious that the distinction between expressive and inexpressive reading is closely related to prosody. One might expect that phoneticians would have confirmed that intuition, and gone on to clarify the particular aspects of prosody that bear on the distinction. In fact, surprisingly little is known at a phonetic level about the prosody of expressive reading. The few experimental studies that there are provide limited information. Expressiveness is often subsumed under a broad distinction between good and bad reading [1] or the terms expressiveness and fluency are banded together with little attempt to separate them empirically. One of the few attempts to distinguish them experimentally was by Perera [2]. Listeners were asked to judge both the expressiveness and the fluency of six children's reading, and the judgements were linked to prosodic measures. However the results were inconclusive.

Following Perera, we consider the relationship between fluency and expressiveness in a systematic way. We take it as read that there is unlikely to be an absolute division between expressiveness and fluency, but empirically we study them together, so that it is possible to observe interactions between them. Our methodology goes beyond Perera's in two important ways. First, we use a substantial sample and a design that facilitates statistical analysis. The other is approach to prosodic analysis. Second, we use a recently developed tool, the ASSESS system [3, 4], which allows us to measure a range of prosodic features in substantial samples.

There are several reasons for trying to clarify the relationship between prosody and expressive reading. On a very practical level, teachers have an interest in understanding what their pupils need to do in order to read expressively [5]. On a more theoretical level, psychological and educational research recognise that expressiveness and fluency are potentially valuable as indicators of underlying processes [6]. The characteristics of expressive reading are also of interest to engineers because there is a growing interest in designing systems that pay attention to the manner of speech as well as the content [7, 8]. Finally, speech scientists and linguists dealing with related topics - such as clinical disorders that affect prosody, or vocal expression of emotion, or prosodic signals in discourse, or social varieties of speech - have an interest in knowing whether variations are specific to the contrasts that they are considering, or whether they are linked to a more general dimension such as expressiveness.

2. METHOD

2.1. Data

Recordings were made of 67 children (aged 8-10) from the Lancashire area of England reading a passage called 'A Trip to the Zoo'. The passage was taken from the Perera study cited above: it had been designed for children of the same age-group and contained textual features (e.g. lists and questions) thought to affect expressiveness. A subsection of the passage (the first paragraph, consisting of 106 words) was subsequently used for prosodic analysis.

Two raters with phonetic training rated each child's reading on two separate scales, one for expressiveness and one for fluency. Two correlations were carried out on the ratings. The first was an inter-rater reliability correlation. This showed good agreement (r=0.8 for both expressiveness and fluency). The second examined the relationship between the measures. By and large high expressiveness and high fluency go together. In particular, it is rare to find high expressiveness without high fluency. Nevertheless there were enough exceptions to fill out a factorial design with fluency on one axis and expressiveness on the other. There were 3 levels of expressiveness and 2 levels of fluency, giving 6 cells in total. A subset of 24 readers was selected, with 4 readers in each cell, 2 male and 2 female. They were chosen to be as closely matched as possible within the basic constraints of the design.

2.2. Speech analysis

Prosodic analysis of the recordings was carried out using the ASSESS system. Entropic WAVES™, operating on a SUN SPARC II workstation was used for basic signal processing. Confidence values and inspection were used to edit out unreliable F0 estimates.

The ASSESS system was then used to derive 2 types of measure from the edited information provided by WAVES. The first type can be called pointwise measures. These are simply summary statistics for the edited data points provided by WAVES™. Measures of this type were mean F0, standard deviation of F0, maximum F0, minimum F0, and the 10th, 50th, and 90th percentile points for F0. The second type can be called piecewise. In the case of F0 these measures described sections of contours obtained by fitting a smooth curve to the edited points and identifying landmarks in the curve - local maxima and minima, and the boundaries of silences. Sections which run between adjacent 'landmarks' on the curve are called 'continuous pitch movements'. Measures extracted for each rising or falling pitch movements were duration, magnitude of the associated pitch change, slope,.and mean amplitude

during the movement. Numbers and durations of pauses were also calculated. Standard statistics were then used to summarise the behaviour of each measure in each sentence.

2.3. Data reduction and inferential statistics

The fact that ASSESS generates a large number of measures, makes it necessary to extract a core subset of measures which adequately reflects the information in the full set. That was done using factor analysis along with a priori preferences for measures which were simple and formed a cohesive set.

Analysis of variance was then applied to the core measures to test the significance of differences between levels of expressiveness and fluency. Each ANOVA had three between variables (expressiveness, fluency and gender). Sentence was a within variable. Where measures were logically related (e.g. rise magnitude and fall magnitude), they were paired and treated as levels of a second within variable.

3. RESULTS

3.1. Data reduction and the core measures

Factor analysis identified seven factors. A core set of measures was then selected from the variables which loaded highly on each factor, in a way that preserved distinctions but avoided redundancy in the data. Pitch related measures were-mean F0, standard deviation of F0, 90th, 50th and 10th percentile of F0, mean magnitude of rising and falling pitch movements, standard deviation of magnitude for rising pitch movements and mean length of rising and falling pitch movements. Amplitude measures were mean and standard deviation of amplitude. Timing-related measures were intra and inter sentence pause numbers and length, numbers of rising and falling pitch movements, and syllable time (i.e. overall time to read a sentence upon the number of syllables).

3.2. ANOVAS: core effects and the main points

It is useful to present results in a progression from coarse to fine. At the coarsest level, the ANOVAS showed that a few, simple effects, which might be anticipated intuitively, did actually occur. Finer levels of description showed that these effects had subtler properties, some of which might not be foreseen.

Two main groups of variables emerged, one related mainly to temporal organisation, the other related to pitch variation. Speakers who differed in fluency tend to differ in temporal organisation, and speakers who differed in expressiveness tended to differ in pitch variation. A third group of variables related to amplitude. These showed no main effects, suggesting that control of amplitude is not simply related to expressiveness or fluency. As might be expected from the fact that there are more than two factors, these generalisations are qualified in various ways.

MEASURE	MAIN EFFECTS OF FLUENCY and FLUENCY x SENTENCE INTERACTIONS	MAIN EFFECTS OF EXPRESSIVENESS and EXPRESSIVENESS x SENTENCE INTERACTIONS
Normed duration	Main p=0.0009 Flu*Sent p= 0.025	Exp*Sent p=0.010
Numbers of rises & falls	Main p= 0.001	Main p=0.016 Exp*Sent p=0.020
Numbers of pauses	Main p=0.007	Main p=0.038 Exp*Sent p=0.032
Mean length of between sentence pauses	Main p=0.0017	
Mean length of intra sentence pauses	Main p=0.05 Flu*Sent p=0.037	
Mean length rises & falls	Main p=0.0001	

Table 1. Basic effects associated with temporal organisation

MEASURE	MAIN EFFECTS OF EXPRESSIVENESS and EXPRESSIVENESS x SENTENCE INTERACTIONS	MAIN EFFECTS OF FLUENCY and FLUENCY x SENTENCE INTERACTIONS
Stand dev of F0	Main p= 0 Exp*Sent p=0	Flu*Sent p=0.003
Mean magnitude of rises & falls	Main p=0 Exp*Sent p=0	Main p=0.014 Flu*Sent p=0.007
Stand dev of rise & pause magnitude	Main p=0 Exp*Sent p=0	
Mean length rises & falls	Main p=0.0001	

Table 2. Basic effects associated with pitch variation

Table 1 summarises the basic effects associated with temporal organisation. High fluency was associated (as might be expected) with relatively short syllable duration, relatively low numbers of inflections in the pitch contour, and a low

level of pausing - with respect to both numbers of pauses and their lengths, between and within sentences. The interactions shown in the table indicate that qualifications are needed. They will be considered later.

Table 2 summarises the basic effects associated with pitch variation. Most simply, expressive speakers showed a relatively high standard deviation of F0. The same pattern applied to both magnitude and standard deviation of segment magnitudes, indicating that - to a first approximation, at least - inexpressive speakers produced pitch contours that were simply compressed in pitch space relative to expressive speakers, rather than using pitch contours with a fundamentally different shape. The table also shows a main effect of length of pitch movements. Basically expressive speakers had pitch movements that were stretched not just in height but also in time. However this generalisation should not be taken too far: the main effect of length is mostly due to the behaviour of a particular subgroup of expressive speakers. We return to that point below.

The points made in this section provide a basic frame of reference. They show that the two dimensions - fluency and expressiveness - were associated in a broad, intuitive way with the two main kinds of measurement that we have described. The interactions, though, indicate that there is more detail to be assimilated. Three main kinds of detail will be considered - (i) the relationship between expressiveness and fluency, (ii) the interaction of expressiveness with sentence, (iii) gender effects.

3.3. Expressiveness and fluency

It is tempting from the picture given above to think that there is a neat line dividing expressiveness and fluency. Certainly the temporal measures seem to be broadly related to fluency and the pitch measures broadly related to expressiveness. But as we noted in the introduction, it would be surprising if expressiveness and fluency were not related. In fact a closer the data shows that they do relate, and suggests a number of points about the nature of the relationship.

The first point comes from looking at the three temporal measures, syllable duration, numbers of pauses and numbers of rises and falls (see Table 1). These measures all showed strong main effects of fluency, but they also showed effects of expressiveness. Of particular interest are the expressiveness by sentence interactions which characterised all three. The interactions reflect the fact that all the expressive groups were relatively similar on these measures for about the first half of the passage , but pulled apart thereafter, with the least expressive group showing markedly higher scores (i.e. longer syllable duration, more pauses, more rise/falls).

It is not trivial to integrate this pattern with the main effects of fluency on the same variables, which are (essentially) time-independent. The most obvious possibility is that two types of skill are at work, variations in the first being associated primarily with variations in fluency, variations in the second being associated primarily with variations in expressiveness. The first, mechanical type of skill controls performance on these variables directly: the second, meaning-oriented type provides support which enables control to be achieved with less effort. It is possible to operate for a while in a mechanical way, without support from the meaning-oriented systems, but the effort of operating that way cannot be sustained for long, and so there is a gradual degradation of performance among subjects who initially try to operate in that way.

The second point comes from looking at the statistical interactions between fluency and expressiveness. Table 3 summarises the most straightforward of these. Sometimes there is a straight expressiveness by fluency interaction; sometimes the pattern of variation that occurs with sentence is dependent on fluency and expressiveness.

These interactions arise from the fact that expressiveness seems to be tempered by fluency. There is a multiplicative pattern here. Fluent readers in expressive group1 (the highest expressive group) achieve markedly higher scores for the relevant measures than non fluent readers in the same expressive group. The same trend is just discernible in expressive group 2, but fluency seems to have virtually no effect by the time we reach expressive group 3.

The broad conclusion from these interactions is that fluency and expressiveness are interlinked. In particular, both aspects of reading performance are impaired if readers lack either of the skill types involved.

MEASURE	INTERACTIONS	P values
pitch change in rises&falls - mean	ex*fl	p= 0.002
	ex*fl*sent	p=0.000
pitch change in rises & pauses - standard deviation	ex*fl*sent	p=0.014
F0 mean	ex*fl*sent	p=0.025
F0 centiles	ex*fl*sent*centile	p=0.002

Table 3. Expressiveness by fluency interactions

3.4. Sentence effects

Tables 1 and 2 show significant interactions with sentence, particularly for expressiveness. The main pattern that explains these interactions is one of high variation across sentences for highly expressive readers and very little variation for inexpressive readers. This phenomenon is reflected in strong expressiveness/sentence interactions for the pitch variability measures (mean magnitude of rises and falls, standard deviation of f0, standard deviation of rise and pause magnitude). The most expressive readers show a pattern of dramatic fluctuation by sentence. The measures for inexpressive readers stay relatively constant across the text.

To some extent this fluctuation can be linked to particular textual explanations. Expressive readers mark the initial sentence of the passage with higher values on all measures, show raised 10% and 50% F0 points for direct questions, and have steeply rising pitch slopes for lists.. However not all intersentence variation can be easily related to textual features. Subjectively, there is a regularity in the way expressive readers vary a range of pitch measures that suggests they may be trying to achieve global patterning - particularly an up/down pattern in pitch-related measures which suggests that a preset rhythmic pattern may be interacting with particular textual features.

MEASURE	INTERACTION	PROBABILITY VALUE
mean length intra sentence pause	ex*fl*gender*sent	p=0.001
mean magnitude rises and falls	Ex*gender*sent	p=.003
stand dev f0	Ex*fl*sent*gender	p=.011
stand dev rise and pause magnitude	Ex*gender*sent	p=0.0268
mean length rises and falls	Ex*fl*gender	p=0.021

Table 4. Interactions involving gender

3.5. Gender effects

Table 4 summarises interactions involving gender.

The main trend that underlies these interactions is for female readers in the most expressive group (group 1) or the top expressive/fluency group (expressive group 1, fluency group 1) to have the highest scores i.e. to give the most extreme performances. This pattern tends to be sentence-related, as seen by the fact that most of the interactions involve sentence. Sentence differentiation takes several forms, but particularly characteristic is one in which a marker of highly expressive readers is even more extreme when highly expressive (and fluent) female readers read a sentence that lends itself to expressive performance.

The last measure in the table (mean length of rises and falls) shows a different gender-related pattern. On that mesasure, males who were highly expressive and highly fluent had higher scores than females in the same group. The effect was very marked, and made a large contribution to the main effect of expressiveness for this measure. The effect reflects the way that males in that group achieved the increase in magnitude of rises and falls that characterised all expressive readers (as indicated by main effect of expressiveness for rise/fall magnitude). Males matched increased pitch change with increased length, whereas females appear simply to have changed pitch more quickly.

These gender effects are interesting, but they should be treated with caution. The numbers of each sex involved are few. In the case of the highly expressive, highly fluent group, for example, the effects are based on comparison of means for cells containing two readers. Hence the patterns suggest a direction for further study rather than firm conclusions.

4. CONCLUSION

The study confirms and clarifies what we might suspect intuitively. We can say categorically from this study that expressive, fluent reading is characterised by short duration and high pitch movement, whereas inexpressive, non fluent reading is characterised by long duration and little pitch movement. At a finer level, inexpressive non fluent reading is characterised by many pauses, long pauses, fragmentation and very little pitch movement, whereas expressive fluent reading is characterised by fewer pauses, shorter pauses, smoothing of the signal rather than fragmentation, and wide pitch movement, particularly in the top end of the pitch range. Expressiveness is broadly pitch related and fluency is broadly time-related. However it is clear that the two are interrelated in complex ways.

The study raises a number of more general and challenging points. The first of these is what it suggests about the relationship between expressiveness and fluency. On an empirical level the study shows that these two are interlinked in rather complex ways. These links may help to clarify the roles and interactions of the different mechanisms involved in skilled reading.

The second point concerns how much of the prosody of expressive reading is textually driven and how much it is more globally determined or preset. The data suggest that while textual features may play a part in determining what happens, there are other more global mechanisms at work.

The third point concerns the relationship of prosodic patterns in reading to gender. The sociolinguistics of prosody have received very little attention. The findings of this study are in line with general sociolinguistics where females tend to 'outperform' the males.

REFERENCES

[1] Clay, M. and Imlach, R. 1971. Juncture, pitch and stress as reading behavior variables. *Journal of Verbal Learning and Verbal Behavior*, 10, 133-139.

[2] Perera, K. 1989. The development of prosodic features in children's oral reading. Unpublished Ph.D dissertation, University of Manchester, UK.

[3] Cowie, R., Sawey. M. and Douglas-Cowie, E. 1995. A new speech analysis system: ASSESS (Automatic Statistical Summary of Elementary Speech Structures). In Elenius, K. and Branderud, P. (eds.), *Proceedings of X111th International Congress of Phonetic Sciences*, 3, 278-281. Stockholm:KTH and Stockholm University.

[4] Douglas-Cowie, E. and Cowie, R. 1998. Intonational settings as markers of discourse units in telephone conversations. *Language and speech*, 41 (3-4), 347-370.

[5] Allington, R. 1983. Fluency: The neglected reading goal. *The Reading Teacher*, 36, 556-561.

[6] Young, A. and Bowers, P. 1995. Individual difference and text difficulty determinants of reading fluency and expressiveness. *Journal of Experimental Child Psychology*, 60, 428-454.

[7] Murray, I. and Arnott, J. 1996. Synthesizing emotions in speech: is it time to get excited? In Bunnell, T. and Idsardi, W. (eds.) *Proceedings of Fourth International Conference on Spoken Language Processing*, 3, 1816-1818. Delaware, USA.

[8] Kollias, S. et al. 1998. Review of exisiting techniques for human emotion understanding and their applications in human-computer interaction. Report to the EC for the TMR PHYSTA project 'Principled Hybrid systems: Theory and Applications', Research contract FMRX-CT97-0098 (DG 12 - BDCN).