Children's perception of direct and indirect reported speech

Nigel Hewlett[†], Cherry Kelsey[‡] and Robin Lickley[†]
† Queen Margaret University College, Edinburgh
[‡]University of Edinburgh
E-mail: nhewlett@qmuc.ac.uk, ckelsey@staffmail.ed.ac.uk, rlickley@qmuc.ac.uk

ABSTRACT

This study investigated the abilities of adults and children to distinguish direct reported speech from indirect reported speech in sentences read aloud by a native English speaker. The adults were highly successful, the older children less so and the younger children were relatively unsuccessful. Indirect reported speech appeared to be the default category for the children. Potential prosodic cues were identified and measured from waveforms and pitch contours of the stimulus sentences. Statistical analysis was applied with a view to ascertaining which (combination of) cues best predicted the listener responses. The results suggest that pitch movement and duration both provided important cues to distinguishing the sentence types. The analysis also revealed a learning effect by all groups.

1. INTRODUCTION

The English sentences in 1, below, are syntactically ambiguous. In 1a, it was Mary who was driving the car. In 1b, it was John who was driving the car. Sentence 1a is an example of indirect reported speech, while sentence 1b is an example of direct reported speech. In writing, the two sentences are disambiguated by their punctuation. In speaking, such sentences may be disambiguated by their prosodic characteristics. Various prosodic devices, local or global, may be used to signal that the speaker is reproducing another person's utterance [1,2,3,4].

1a John said Mary was driving the car.

1b "John", said Mary, "Was driving the car."

Sentence 1b would probably be considered unusual in everyday spoken English. The speaker would usually be specified before or after their actual words, as, for example, in *John said "Mary was driving the car"*, or, in the frequent usage of younger generation English speakers, by using the construction *be like*, as in *John was like "Mary was driving the car"*. However, constructions like that in 1b routinely occur in written English, including children's books, and therefore in loud reading. Children are exposed to such constructions and their prosodic characteristics, through having stories read to them and through reading aloud themselves. Children listen to stories throughout the primary school years and when reading aloud themselves, they are encouraged to employ the same prosodic devices that adults use. However, the prosodic characteristics of this rather literary form might take some time to be acquired and are unlikely to be fully developed in many younger school age children.

The experiment reported here was designed to test the abilities of children of various ages to distinguish between pairs of sentences of the sort exemplified in 1, above, on the basis of their prosodic characteristics alone. Adults were tested as well, in order to verify that the prosodic distinction is indeed generally acquired. Adults, however, tend not to perform at ceiling on tests of prosodic distinctions [5] and so their patterns of responses are also of interest in themselves. Of course, in order for listeners to make reliable judgements, sentence pairs such as those in 1, above, must remain acoustically distinct. Comparison of acoustic cues with listener judgements might reveal the combinations of cues which contributed to the listeners' perceptions. A further aim of this study was to discover the acoustic basis for listeners' responses.

2. METHOD

Participants

The participants were one hundred and six children and forty-two adults. The children were between the ages of four and thirteen years and they were tested in groups according to school year, with the result that there were occasional slight overlaps in age between one year group and the next (see Table 1). The groups are labelled by their mean age in years, in this report. All participants were native speakers of English.

Group	***Mean Age***	***Age Range***	***No. of Participants***
5 yrs	5;3	4;11–5;8	13
7 yrs	7;4	6;11–8;8	19
9 yrs	9;3	8;7–10;1	23
11 yrs	11;8	11;0–12;4	23
12 yrs	12;9	12;2–13;4	28
Adults	-	18– 60	42

Table 1. Details of participant groups.

Preparation of Stimuli

The stimuli were twenty-four syntactically ambiguous sentences, read aloud and recorded on audio tape. Twelve of the sentences were intended to convey indirect reported speech (IDRS) and the other twelve were intended to convey direct reported speech (DRS). Examples are given in 2, below.

2a Pikachu said Squirtle was chasing the cat. (IDRS)

2b "Pikachu", said Squirtle, "Was sleeping in the classroom." (DRS)

The same two names (in varying order) were used throughout: either Pikachu was reporting an action of Squirtle's or *vice versa*. However, sentences forming exact minimal pairs were avoided. For the purposes of the recording, all 24 sentences were printed in the form of a list, in random order. Following the method used by Cruttenden [5], a linguist (the first author) read the test sentences aloud, in a manner indicated by the punctuation but without employing exaggerated intonation. Cool Edit software was used to produce three replicas of each sentence, with a three second gap between each. For the benefit of the younger children, two A5 picture cards were prepared for each sentence, one depicting the 'right' character engaged in the relevant action, the other depicting the 'wrong' character engaged in the same action. The name of the character was written on the card and the Pikachu character was always depicted on yellow card and Squirtle on blue.

Data collection

Adults were tested individually or in small groups. The 9, 11 and 12 year-olds were tested in their classrooms. Each participant was provided with an answer sheet and instructed to tick the name of the character that they believed was *doing the activity*, such as chasing the cat. The 5 and 7-year-olds were tested individually, and gave their responses by pointing to a choice of two pictures. The test took about 15 minutes to complete.

Acoustic analysis of stimulus sentences

Audio waveforms and fundamental frequency (f0) waveforms of each of the stimulus sentences were prepared, using Multispeech™ software. This software allows editing of the pitch contour with reference to the audio waveform, in order to achieve maximum accuracy in the pitch analysis. The goal of the acoustic analysis was to identify and measure potential acoustic cues to the distinction between the two types of reported speech. Sentences which had achieved high rates of identification by the adult listeners were used in this analysis. The focus was on features of duration and f0, particularly surrounding the first name (N1) and the second name (N2). (In 2, above, for example, Pikachu is N1 and Squirtle is N2, in both sentences.)

Analysis of participants' responses

Sentence scores (the percentage of correct responses to each sentence) and participant scores (the number of correct responses by each participant) were calculated.

Statistical analysis was firstly directed towards a comparison of performance by age group and sentence type. Additional analysis addressed the role of acoustic factors in distinguishing the two types of utterance.

3. RESULTS

Participant responses

Figure 1 shows the percentage of correct responses by each participant group on each of the sentences. In this figure, the order of the sentences is different from their original order of presentation: the IDRS sentences are grouped together on the left of the figure (sentences 1-12) and the DRS sentences on the right (sentences 13-24). Within each type, the original order of precedence has been preserved (sentence 13 came earlier in the presentation than sentence 14, for example). Performance by the adults was high, although not at ceiling. The five, seven and nine year old groups had much lower scores, with patterns broadly similar to each other. In between were the 11 and 12 year olds, who were also broadly similar to each other. Table 1, which shows mean scores, standard deviations and ranges, by group, demonstrates the wide range of performance found among the participants within all groups, including the adults.

	5-yrs-old	***7-yrs-old***	***9-yrs-old***	***11-yrs-old***	***12-yrs-old***	***Adults***
Mean (SD)	58 (21)	60 (17)	63 (24)	76 (18)	77 (19)	88 (11)
Range	46–92	42–92	46–83	54–100	50–100	46–100
Range SS	23–92	32–89	17–100	21–100	30–100	40–98

Table 2. Mean score (SD in brackets), range of participant scores and range of sentence scores, by group. 'SS' = Sentence Scores. All values are percentages.

The IDRS scores of the children were closer to those of the adults than were their DRS scores and the 12-year-olds achieved higher scores than adults on some IDRS utterances. Sentence 6 (target IDRS) stands out as having a low score, notably by the adults. The adults scored higher than each of the child groups on every DRS sentence. There was a tendency for performance on the DRS sentences, by all the child groups, to improve with ascending order of presentation, suggesting a learning effect over the time scale of the test.

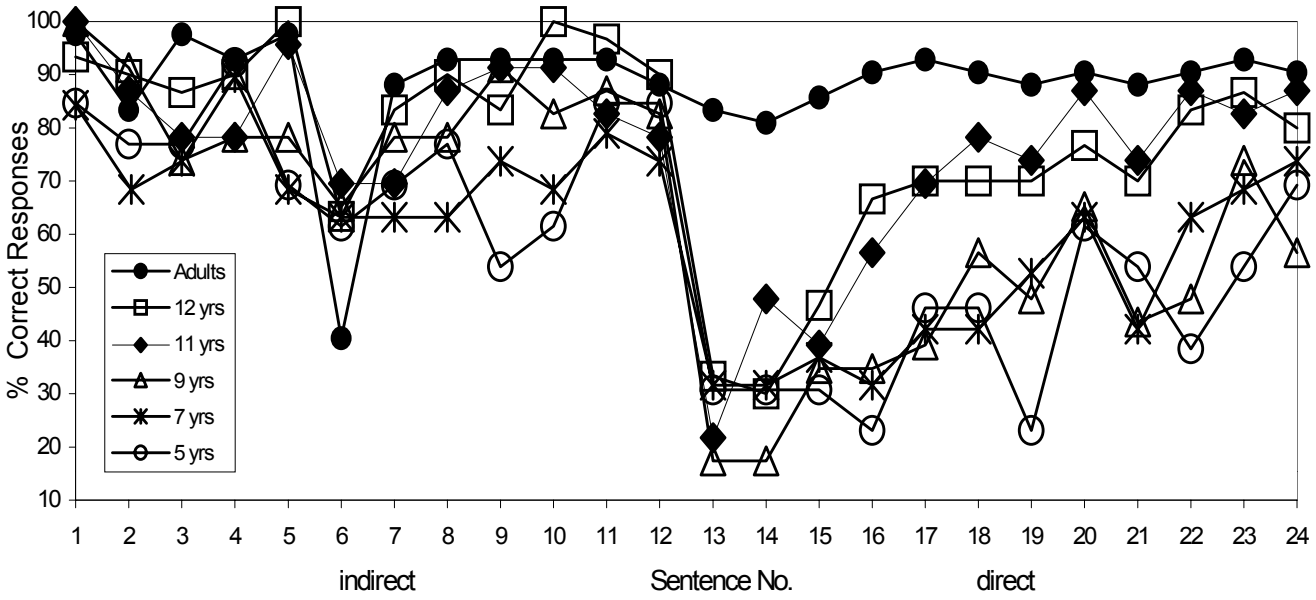

Figure 1. Mean score for each sentence, by group. Sentences 1-12 = IDRS, sentences 13-24 = DRS.

Acoustic analysis

Following inspection and comparison of the duration and f0 characteristics of those IDRS and DRS sentences which received high identification scores by the adult listeners, the following measures were decided: i) the difference in f0 between the first syllable and final syllable of the first name in each sentence; ii) the presence or absence of a rise in fundamental frequency on the final syllable of the first name and of the second name; iii) the length of any pause after the second name; iv) the duration from the onset of the sentence to the end of the second name. Table 3 shows the values for each of these measures for each stimulus sentence, along with its percentage correct identification by the adult group. Most IDRS sentences had a rise in mean f0 between the first and last syllable of the first name. DRS sentences, in contrast, had a drop in mean f0 between the first and last syllable of the first name, but with a rising contour over the final syllable. DRS sentences were also distinguished by a pause of around 400–550 ms after the second name and by a longer duration from sentence onset to the beginning of this pause. Sentence 6 (IDRS) is exceptional in having a fall between the first and last syllable of the first name which is comparable in size to that of a typical DRS sentence.

Analysis of acoustic-perception relationships

Presence of a final f0 rise on N1 and N2 and presence of pause after N2 are clear candidates for predictors of listener judgements: the binary division of these factors between the utterance types makes statistical examination redundant. For analysis of other variables, we took mean judgements of "indirect" for each age group as the dependent variable. With three exceptions (S6, S9 and S13), the difference in f0 between the first and second syllable of N1 seems to be directly related to presence of f0 rise on the final syllable of the word. As one would expect, this factor correlates negatively with judgements of "indirect" for each age group for the whole set of stimuli at $p<0.03$ or lower. However for separate groups of stimuli, no systematic pattern of correlation by age groups can be discerned.

Duration from onset of the utterance to the end of N2 (before the silent pause in DRS) is negatively correlated with "indirect" judgements for each age group (summary: $r = < -0.642$, $N = 24$, $p<0.01$). The longer duration from onset to N2 in the DRS utterances (mean = 1705 ms, sd = 139) versus the IDRS utterances (mean = 1439 ms, sd = 82) may be a further factor. Pause duration within DRS utterances, where pause was always greater than zero (mean = 453 ms) did not correlate with mean "indirect" judgements for any age group.

Figure 1 suggests that there was a learning effect within the DRS utterances. A positive correlation was found for all age groups between mean "direct" judgements and item number. An alternative explanation would be that an acoustic factor varied with item number. Although a positive correlation was found between pause length and item number ($r = 0.649$, $N = 12$, $p<0.03$), no correlation was found between pause length and mean judgements and we therefore discount this possibility.

4. DISCUSSION

The overall adult score was comparable to that of Cruttenden's subjects [5], on a different test of perception of prosody. Clearly, the prosodic characteristics used by the speaker were familiar to most of the adults. The markedly greater success of the adults as compared with children of even five years (a comparatively late age in terms of language acquisition) and more, is also in line with previous findings.

Sent-ence	*f0 drop (Hz) within N1*	*Rise on final syll-able of N1?*	*Rise on final syll-able of N2?*	*Length of pause after N2 (ms)*	*Length from onset to end of N2 (ms)*
IDRS					
S1	-10	no	no	0	1346
S2	-9	no	no	0	1623
S3	-12	no	no	0	1366
S4	-19	no	no	0	1454
S5	-12	no	no	0	1416
S6	47	no	no	0	1492
S7	-22	no	no	0	1360
S8	-1	no	yes	0	1450
S9	27	no	no	0	1526
S10	-8	no	no	0	1471
S11	-14	no	no	0	1356
S12	-16	no	no	0	1407
DRS					
S13	44	no	yes	415	1657
S14	61	yes	yes	565	1553
S15	59	yes	yes	392	1583
S16	71	yes	yes	541	1658
S17	55	yes	yes	411	2040
S18	42	yes	yes	344	1658
S19	56	yes	yes	460	1571
S20	31	yes	yes	435	1818
S21	42	yes	yes	422	1698
S22	42	yes	yes	524	1706
S23	37	yes	yes	423	1667
S24	45	yes	yes	510	1856

Table 3. Acoustic measures of stimuli. 'N1' = first name in sentence, N2 = second name. Column 2 shows the difference in f0 between the first and last syllable of N1.

A similar level of performance on DRS and IDRS sentences was another factor that marked the adults out from all the child groups. The greater sureness of the adult responses is also testified by the large dip in their performance on sentence 6, which turned out to be acoustically ambiguous. All the child groups however demonstrated a greatly increased understanding of the distinction over the time scale of the test. What the learning effect was, precisely, must be a matter of speculation. In some cases it may have been that a subject was already aware of the intonation-meaning links involved and merely required a few examples in order to 'tune in'. On the other hand, it may have been possible for a subject to improve performance over time simply by detecting a 'marked' intonation pattern and identifying it with the 'marked' interpretation. Whatever it was, this learning effect demonstrates the sensitivity of children to the meaning-bearing possibilities of prosody.

Twelve years seems rather late for non-adult-like ability even on a prosodic variable. However, as has been pointed out, this particular variable is of a literary, rather than everyday, nature and this is likely to influence its age of acquisition. It may indeed be a feature that is variably acquired, with 'exposure to print' being a significant factor. The adult subjects of this study were from a range of backgrounds but we do not have the information to make any association between social background and test score. The children were from a private school and can therefore be presumed, for the most part, to come from homes with higher than average print exposure. However, the direct speech construction investigated here is common in story books for children of all ages, as well as in adult fiction, and these findings should be of interest in an educational, as well as a linguistic, context.

We would not wish to make any strong claims about the precise nature of the normal prosodic cues to this distinction, on the basis of data from a single (and non-naïve) speaker. Indeed, they doubtless vary to some extent according to speaker and circumstances and one might expect listeners to need a few tokens to become fully familiar with the relevant acoustic cues (thus even the adults in this experiment showed a mild learning effect). Rather, we would emphasize that most adults successfully recognized the acoustic cues and that the children recognized – or acquired – them within the time scale of the test, with a success that varied according to age. This outcome itself supports the validity of the method used. Furthermore, the number of subjects tested, and the acoustic analysis adopted, allowed us to gain strong evidence on which cues were most salient. We suggest that this experiment offers an efficient paradigm for investigating acoustic-perceptual relations in cases in which syntactic ambiguities are potentially disambiguated by means of prosody.

REFERENCES

[1] G. Klewitz and E. Couper-Kuehlen. "Quote – Unquote? The role of prosody in the contextualization of utterances". *Pragmatics*, vol. 9, pp. 459-485, 1999.

[2] F. Coulmas (Ed.). *Direct and Indirect Speech.* Berlin: Mouton de Gruyter, 1986.

[3] E. Holt. "Reporting on talk: The use of direct reported speech in conversation". *Research on Language and Social Interaction*, vol. 29, pp. 219-245, 1996.

[4] M. Lind. "The use of prosody in interaction: observations from a case study of a Norwegian speaker with a non-fluent type of aphasia," in *Investigations in Clinical Linguistics and Phonetics*, F. Windsor, M.L. Kelly and N. Hewlett , Ed., pp. 373-390.

[5] A. Cruttenden. "Intonation comprehension in ten-year-olds". *Journal of Child Language*, vol. 12, pp. 643-661, 1985.