INTERNATIONAL JOURNAL OF BEHAVIORAL DEVELOPMENT, 1997, *20* (4), 651–665

Distinguishing One's Own Voice from Those of Others: A Function for Private Speech?

Charles Fernyhough and James Russell

University of Cambridge, Cambridge, UK

The aim of the study reported here was to test a new hypothesis about the function of private speech, namely that children's private speech in social contexts presents them with an opportunity to distinguish their own voices from those of others. Data on the social and private speech of 22 five-year-olds were collected during single hour-long group play sessions. In a later session, children heard extracts from an audio recording of an earlier group discussion, and were asked to state which of each pair of extracts contained their own speech. It was found that children's performance on this speech recognition task was positively correlated with both proportional and frequency measures of private speech. There was no relationship between the speech recognition measure and frequencies of social speech or overall verbosity; and performance on a test of children's ability to recognise their own speech in isolation was no better than chance. We suggest that these findings are consistent with private speech having a specific function in the development of an understanding of oneself as a speaking agent among other such speakers.

Since the first systematic investigations of the phenomenon were undertaken by Piaget and his colleagues in the early part of this century (Piaget, 1926), private speech—speech that is not obviously addressed to another listener—has been the subject of much debate within developmental psychology. In her recent review, Berk (1992) suggests that much of the empirical evidence is supportive of a broadly Vygotskian interpretation of the phenomenon, whereby private speech has a central role in the self-regulation of behaviour (Vygotsky, 1934/1986). Other functions

Requests for reprints should be sent to Dr. Charles Fernyhough, School of Social Sciences, Staffordshire University, College Road, Stoke-on-Trent, ST4 2DE, England.

The research described in this paper was undertaken in partial fulfilment of a PhD degree by the first author at the University of Cambridge, funded by the Medical Research Council. Thanks are due to all the children and parents who agreed to take part in this study; to Kevin Silber for statistical advice; and to David Good, David Wood, David Furrow, and three anonymous reviewers for helpful comments on earlier versions of this paper.

suggested for private speech include emotional expression and role play (Fuson, 1979); language practice (Saville-Troike, 1988; Weir, 1962); practice for communicative encounters (Olszewski, 1987; Rubin, 1982); and establishing the meaning of one's activity for oneself (Kohlberg, Yaeger, & Hjertholm, 1968). In addition to self-regulation, recent research has posited a role for private speech in the acquisition of self-*knowledge* (Morin, 1993; Morin & Everett, 1991). Rubin (1979), following Mead (1934), suggested that comments on one's own actions and feelings might "make young speakers aware of their own actions and of their own separate existences" (Rubin, 1979, p. 291), and Locke (1993) claimed a role for infant babbling in the establishment of vocal self-awareness, in turn related to lexical and phonological learning.

The suggestion that private speech might have a role in the development of the self-concept is, therefore, not a new one. The hypothesis to be tested in this study is that a particular form of private speech, namely, that which takes place in a social context, presents individuals with an opportunity to distinguish between their own speech and that of others. We suggest two main reasons for framing such a hypothesis. First, the experience of hearing one's own speech in a social context may allow individuals to pick out their own utterances from those of others. Second, the social context may present individuals with an opportunity to learn about those aspects of their own speech that are subsumed under the Bakhtinian concept of "voice" (Bakhtin, 1986).

In order to understand how hearing one's own speech in a social context can allow one's own utterances to be picked out from others', we can consider one view of how individuals effect an analogous version of the self-other distinction in the sensorimotor domain. Russell (1996) argues (after Piaget) that infants' endogenous activity allows self-generated changes in the perceptual world to be distinguished from world-generated changes. By virtue of internal monitoring systems, infants are able to determine the extent to which their own actions are effective in altering the visual input, and thus the extent to which the outside world is "refractory" (Baldwin, 1906), or independent of and resistant to their own actions. If a similar system can be described for the monitoring of the sources of the speech the child hears, it might explain how the child first comes to distinguish his/her own words from those of others, and thus effect a linguistic analogue of this primary self-world distinction.

The second reason for claiming such a role for private speech in a social setting stems from the Bakhtinian concept of "voice". According to Bakhtin (1986), a voice is a way of speaking which makes manifest a particular

perspective on reality through specificities of tone, prosody, content, and so forth. In so far as it encompasses the phenomenon of the "speaking consciousness" (Holquist, 1981), the individual's voice is specific to his/her point of view. As Wertsch (1991) has pointed out, the notion of voice has much in common with Mead's (1934) and Habermas' (1984) notions of "social role". The voice of the individual represents the role he/she adopts in a particular verbal exchange. To recognise one's own voice, in this case, is to recognise the role one is capable of playing in interactive speech. In the light of observed commonalities between private and social speech (Berk, 1992; Vygotsky, 1934/1986), we suggest that the individual's voice will continue to be heard in noncommunicative speech, and that engaging in such speech in social settings allows children a specific opportunity to become familiar with their own ways of speaking in social contexts.

This hypothesis presents us with some clear empirical predictions. If private speech in a social context has a role in enabling children to distinguish their own voices from those of others, one would expect a positive correlation between levels of private speech and subjects' ability to distinguish their own voices from those of others. Further, if this function is specific to private speech, one would expect no such relationship between levels of *social* speech and contextual speech recognition abilities. The study reported here was designed to test the prediction that both coefficients of private speech and frequency measures of private speech would correlate positively with children's ability to recognise their own spontaneous social speech, when heard in the context of other children's spontaneous social speech on an audio recording. It was not expected that there would be any relationship between contextual speech recognition and general verbosity, nor between contextual speech recognition and frequency of social speech. In order to determine the importance of the social context for these speech recognition abilities, measures were also taken of children's ability to recognise their own speech when heard in isolation.

To date, there has been no systematic investigation of young children's ability to recognise recordings of their own speech. Martin and Clark (1982) found that 16-hour-old infants could distinguish their own cry from that of another. Bartholomeus (1973) reported that some preschoolers identified the speech of their classmates as accurately as their teachers, whereas Mann, Diamond, and Carey (1979) showed that six-year-olds' recognition of *unfamiliar* speech to which they had previously been exposed was not significantly better than chance. The present study employs a new technique for assessing speech recognition abilities, by asking children to choose which of a pair of extracts from a recording is the one that contains their own speech.

METHOD

Subjects

Subjects were 22 five-year-old children (mean age: 64.9 months; range: 60–70 months) from predominantly middle class backgrounds. There were 12 boys and 10 girls. All of the children were familiar with the experimenter and had visited the experimental room at least once before.

Procedure

The three stages of the study were as follows:

1. A small group of children was left to play together with minimal adult intervention, while video recordings were made. These recordings were later coded for private and social speech.
2. The children were then brought together by the experimenter for the purpose of a "group discussion session", during which a high-quality audio recording was made.
3. Extracts from this audio recording were played back to individual children after a time lapse of one week. Measures of contextual speech recognition were derived by asking children to indicate which of a pair of extracts contained their own speech.

Phase 1: Collecting Measures of Private Speech

The first two stages of the study were carried out with four groups of four children and two groups of three,[1] balanced for age and gender. Each group of children was brought to the observation room by their parents and given about five minutes to settle in. The room was equipped with various toys and wall friezes. When the last parent had left the room, the video recording was started. Adult intervention in the play session was kept to a minimum. Three times throughout the hour-long session, at roughly 15-minute intervals, the experimenter would enter the room for a brief period and ask the children how they were getting on. If any child looked bored the experimenter would suggest another activity, as well as helping out with any problems. In order to preserve the naturalistic setting, the experimenter would answer when spoken to by a child; otherwise his speech was kept to a minimum.

Coding Children's Utterances. An utterance was defined as a unit of speech containing no temporal or semantic discontinuities. A temporal

[1]Although the intention was to use only groups of four children, last-minute cancellations meant that two groups were a child short.

discontinuity was defined as a pause of more than 2 seconds, and a semantic discontinuity included any change of subject, whether or not preceded by a pause of 2 seconds. All utterances made in the hour-long play session were then classified as social or private, according to the following objective criteria (C = child, O = other; adapted from Diaz, 1992; Furrow, 1992; Goudena, 1992):

1. *Eye Contact:* If C showed sustained eye contact with O during or within two seconds of an utterance, the utterance was coded as social. It was not necessary for the eye contact to be reciprocated by O.
2. *Behavioural:* The utterance was coded as social if, within two seconds of the utterance: (a) C's behaviour involved O (through physical contact, or approach, or extension of arms toward O); (b) O's behaviour involved C (through physical contact or an action attracting C's gaze).
3. *Content Markers:* The utterance was coded as social if: (a) the utterance had the same topic as O's preceding utterance (one that ended no more than two seconds before C's began); (b) the utterance was a question directed to O, where an answer appeared to be expected (indexed by rising intonation); (c) if the utterance contained a vocative or name.
4. *Temporal Contiguity:* The utterance was coded as social if it occurred less than 2 seconds after any Social Utterance.

Any utterance that did not meet the criteria for social speech was classified as private. A random quarter of the videotapes (five subjects) was coded by a second rater. For the distinction between social and private speech, the percentage agreement was 93.8%.

Phase 2: Group Discussion Session

After one hour of free play had elapsed (Phase 1), the experimenter re-entered the observation room and asked the subjects to join him at the table. The children were told that they were going to talk about a soft toy called Charlie the Crocodile. They were then told that today was a special day for Charlie, and invited to guess why. Once it was agreed that today was Charlie's birthday, the experimenter announced that Charlie was to have a party, and that the children's task was to decide what should happen at Charlie's party. The experimenter kept his own comments to a minimum, so that the voices recorded were primarily those of the children. Children were also asked what songs they thought should be sung at the party, and then

encouraged to sing the songs in question individually. The entire session typically lasted 20–25 minutes.

As well as being video recorded, the discussion was audio recorded using a DAT (Digital Audio Tape) recorder (Sony DTC 750) and two wall-mounted PZM (Pressure Zone) microphones. The result was a high-quality stereo recording on which all speech could be clearly distinguished.

Phase 3: Measures of Contextual Speech Recognition (CSR)

The aim of this phase of the study was to produce, for each child, a DAT cassette containing edited extracts from the recordings made in Phase 2, some of which contained the child's own speech. The video recording was used in tandem with the audio system to make sure that the voices on the tape were accurately identified.

The playback session took place in the observation room, one week after the discussion session. The DAT player and hi-fi amplifier were placed on a table in the centre of the room, with the two speakers placed on small chairs about one metre apart. The speakers were covered with red and white crepe paper (red on the child's left, white on the right) in order to aid left-right discrimination. The experimenter sat on the other side of the table, with access to the controls of the DAT recorder and amplifier. The subject sat about a metre back from the two speakers, and the subject's mother sat nearby but outside the child's field of vision.

Subjects were first asked whether they could remember what had happened the previous week in the discussion session. All children remembered the basic details of the earlier session. Next the playback equipment was introduced, with the child being shown that there were two speakers (referred to as "boxes") out of which sound might come.

The playback session was made up of three parts. In each part, a number of pairs of extracts were played to the child. The child heard one of the pair of extracts through one speaker and the other through the second. An electronic switching device allowed the output of the amplifier to be routed through either one or the other of the speakers.

A. Orientation Task. The first part of the playback session used pairs of familiar animal noises to make sure that the subject understood the set-up and could accurately locate a sound as coming from one or the other speaker. Subjects were told: "*First I'm going to play you some animal noises. See if you know which animals they are*." After each recording, the subject was asked: "*What animal do you think that was?*" For example, the subject might first hear a dog barking in the white (right-hand) speaker. After identifying the animal, the subject was asked: "*Which box did the dog come*

from?" The subject was then asked to indicate (by pointing or saying "white" or "right") which speaker the sound had come from. The second member of the pair (a cow) was then played, and the same questions asked. The child was then asked which speaker the dog had come from and which speaker the cow had come from. Each of the children answered all these questions correctly. The order of presentation of the animal noises was counterbalanced with the order of speakers used.

After the successful completion of the orientation task, the rest of the session was introduced with the following general preamble: "*When you were here last week with the other children, I recorded everything that we said. Today I'm going to play back some of the pieces of the recording*." All the children were sufficiently familiar with the concept of recording, and so no further explanation was necessary. The subject's task was then described as follows: "*We're going to hear some children's voices, and I want you to help me by telling me when you can hear YOUR voice. Sometimes you'll hear your voice coming from the red box, and sometimes you'll hear your voice coming from the white box. I want you to listen hard and tell me which box you can hear your voice coming from*."

Throughout the following three tasks (B, C, and D), a forced-choice procedure was used. No feedback was given on whether the subject's choice was correct. If a response was not made after the first playing of the pair of extracts, the extracts were repeated in the same order. If children were still unsure, they were encouraged to say which speaker they *thought* their voice had come from, and, failing that, to guess. In the few cases where the child was still unable or unwilling to answer, no response was recorded. The entire session (Tasks A–D) lasted about 30 minutes, with a 5-minute break halfway through.

B. Isolation Recognition Task. This task was designed to assess children's ability to recognise their own voices when heard in isolation. The aim was to present each child with a pair of recordings which were identical in content, prosody, intonation, and so on, such that the only way children could discriminate their own speech from that of another would be through the acoustic characteristics of the speech. An opportunity for collecting such recordings was available during Phase 2, where children were encouraged to give examples of songs, nursery rhymes, etc. that might be sung at Charlie's birthday party.

Using the instructions given earlier, subjects were told that they were about to hear two recordings of children singing or reciting, and that they were to indicate which of them was their voice. Because of compliance problems, only one pair of extracts could be presented for each child. The subject's utterance was paired with a rendition of the same material by

another child in an earlier session. Care was taken to match the utterances for age, gender, pitch, and loudness. Order of presentation was counterbalanced with the order in which the speakers were used.

C. Content Recognition Task. Subjects then heard a representative (i.e. one in which the themes of the discussion were clear) 30-second extract from the discussion at which they had been present, matched with a similar extract from another discussion at which they were not present. Neither extract contained the child's speech. The aim was to ascertain whether children recognised the content of the discussion and/or the voices of their co-discussants. Due to time limitations, only one such pair was presented, and performance on this measure was assessed across subjects.

Instructions to the subject were as follows: "*This time you're going to hear some children talking. From one of the boxes you'll hear the children who were here with you last week, and from the other you'll hear some other children. You may not hear your own voice this time, but I want you to tell me which of the conversations you remember.*" Presentation was counterbalanced across subjects, and the same forced-choice procedure used.

D. Contextual Speech Recognition. The subject was then played 10 pairs of 20-second-long extracts from the discussion at which he or she was present. In each pair, one of the extracts contained the child's own social speech (a "target" extract), whereas the other (a "foil" extract) did not. Each target extract was edited so that the subject's speech began exactly 10 seconds into the extract, in order to give subjects time to orient themselves to the playback and establish attention to the appropriate speaker. Presentation of the extracts was fully randomised and counterbalanced across and within subjects. Precise instructions to subjects were as follows: "*This time I want you to listen out for YOUR voice. Which box can you hear YOUR voice coming from?*"

Extracts from the recorded discussion to be used in this part of the study were selected according to the following criteria: (1) All target utterances were classified as "social" according to the criteria outlined earlier. (2) When choosing target utterances, content cues (names, personal information, etc.) were avoided as far as possible. (3) Both target and foil extracts included at least two children's voices, including, where appropriate, that of the subject. (4) Both types of extract were chosen evenly from throughout the whole discussion. (5) In a target extract, the subject's own utterance always contained more than one word. (6) If possible, extracts were chosen to begin with the experimenter's voice asking a question (e.g. "What will we have to eat at Charlie's party?"), in order to make the extract easier to follow. Otherwise the experimenter's presence on the tape was kept to a minimum. (7) In matching target and foil extracts, care was taken to make sure that the

extracts came from the same part of the discussion, while avoiding actual overlap. (8) In the few cases where the child did not stop talking for long enough to allow the production of 10 foil extracts, some of the existing foil extracts were used twice, but never consecutively. The number of foil extracts that were repeated in this way never exceeded three. (9) Occasional raucous interludes were avoided.

RESULTS

Table 1 shows the mean utterance frequencies from the hour-long free play period in Phase 1. Coefficients of spontaneous private speech (CPS) were obtained by dividing the frequency of private speech by the total frequency of utterances ($\Sigma PS/\Sigma U$). Contextual speech recognition (CSR) is given as the number (out of 10) of correct forced-choice responses made in Phase 3 (Task D). In three cases where subjects refused to make a choice, or where the full 10 extract-pairs could not be obtained (see earlier), the CSR score was derived by dividing the number of correct choices by the total number of choices for which an answer was given, and then multiplying by 10. Scores of 8 and above were shown by binomial analysis to be significantly higher than chance. Twelve of the 22 subjects scored 8 or higher on the CSR measure.

A possible confounding factor may have been the differing sizes of the groups in Phase 1. A 2(group size) × 2(speech type) ANOVA showed a main effect of group size on number of utterances produced [$F(1,20) = 4.53$, $P < .05$], with the children in the smaller groups producing more utterances overall ($M_3 = 291.2$, $M_4 = 161.4$), but no interaction with speech type ($F < 1$). Unpaired t-tests showed no difference between the two group sizes in the coefficient of private speech [$M_3 = 0.36$, $M_4 = 0.29$, $t(20) = 0.93$, n.s.] or CSR performance [$M_3 = 7.97$, $M_4 = 7.43$, $t(20) = 1.06$, n.s.]. Thus, although subjects in the smaller groups produced more utterances overall, there was no effect of group size on the proportion of private utterances or later CSR performance.

A series of Pearson's product moment correlation coefficients was then calculated, partialling out both age and group size ($df = 18$). For the relationship between CSR and coefficient of private speech, the correlation

TABLE 1
Measures of Private and Social Speech

Measure	*Mean*	*(SD)*
Total utterances (ΣU)	196.70	(137.7)
Social utterances (ΣSS)	130.10	(80.7)
Private utterances (ΣPS)	66.60	(71.3)
Coefficient of private speech (ΣPS/ΣU)	0.30	(0.16)
Contextual speech recognition (CSR)	7.58	(1.06)

was positive and significant ($r_{xy.ab} = .39$, $P < .05$, one-tailed). There was also a significant positive correlation between CSR and frequency of private utterances ($r_{xy.ab} = .40$, $P < .05$, one-tailed). There was no correlation between CSR and the total frequency of utterances ($r_{xy.ab} = .22$, n.s., two-tailed), nor between CSR and the frequency of social utterances ($r_{xy.ab} = .01$, n.s., two-tailed).

The study was also designed to test whether children who recognise their own speech in the context of other speech can do so when the speech is presented in isolation. The *isolation recognition task* (Task B) was included as a test of this hypothesis. Of the 22 children who took part in the experiment, three failed to produce any utterances that could be used in this task. Thirteen out of the remaining 19 children answered the isolation question correctly, a success rate of 68%. A binomial analysis showed that children's performance on this task was not significantly better than chance.

Using success/failure on the isolation recognition task as an independent variable, we asked whether there was any relationship between this measure and the other measures described earlier. A series of unpaired *t*-tests showed no differences between the two groups in: the total number of utterances produced [$M_{pass} = 196.2$, $M_{fail} = 220.3$, $t(17) = 0.33$, n.s.]; the frequency of social utterances [$M_{pass} = 134.5$, $M_{fail} = 128$, $t(17) = 0.15$, n.s.]; the frequency of private utterances [$M_{pass} = 61.7$, $M_{fail} = 92.3$, $t(17) = 0.82$, n.s.]; the coefficient of private speech [$M_{pass} = 0.29$, $M_{fail} = 0.38$, $t(17) = 1.04$, n.s.]; or CSR performance [$M_{pass} = 7.91$, $M_{fail} = 7.68$, $t(17) = 0.54$, n.s.]. In order to control for age and group size, the relationships between performance on the isolation recognition task and each of the above measures were further investigated using point-biserial correlations with age and group size partialled out. None of the resulting correlation coefficients approached significance.

The *content recognition task* (Task C) had been designed to assess children's memory for the discussion session. Of the 22 subjects in the sample, 3 failed to discriminate correctly between a recording of a session at which they were present and an extract from another session. These findings are important in showing that the majority of children remembered enough about the content of the discussion session to make the presentation of edited extracts from this discussion meaningful.

DISCUSSION

The present study demonstrates a relationship between private speech and contextual speech recognition abilities. Proportional and frequency measures of private speech were positively correlated with subjects' ability to recognise their own spontaneous social speech in the context of the social speech of others, when heard on an audio recording. There was no such

relationship between contextual speech recognition (CSR) performance and measures of social speech. In what follows, we consider some of the possible interpretations of this finding, considering how future research might shed light on: (1) what features of their own speech children are recognising in the CSR task; and (2) the direction of any casual relationship between CSR performance and private speech.

Let us approach the first of these questions by considering three main types of cue that might be involved in the recognition process. First, subjects might be basing their judgements primarily on structural or acoustic information: in other words, recognising the *sound* of their own speech. Second, they might be utilising features of the semantic *content* of the utterance, i.e. information about what is being said. Third, children might through private speech, come to recognise those aspects of their own speech that can be subsumed under the category of *voice* (Bakhtin, 1986).

To begin with the first of these three hypotheses, the present study attempted, despite serious methodological difficulties, to assess children's ability to recognise the *sound* of their own speech when presented in isolation. Subjects' performance on this task was not significantly better than chance. There was no relationship between scores on this measure and any of the speech measures, nor was there any relationship with CSR performance. These findings might be taken to imply that the acoustic properties of speech do not play a major part in contextual speech recognition. Indeed, one possible reason for children's lack of success on the isolation recognition task is that the acoustic properties of one's own speech differ when heard from "within the head" to when heard on a recording (Maurer & Landis, 1990) because of the effects of bone conduction. For this reason, it might seem unlikely that children showing higher levels of private speech would be better at recognising their own speech in the CSR test simply because they are more talkative, and therefore have "heard themselves speak" more often.

However, there are two main reasons why it would be unwise to exclude this possibility at this stage. First, serious methodological problems with this task (including low power) mean that replication and further research is needed before any conclusions can be drawn about isolation recognition abilities. Second, a possible confound is the fact that recitation and singing were used in the isolation recognition task (in an attempt to control for differences in intonation, prosody, and so on), rather than spontaneous social speech. Further research is required to determine how well children can recognise ordinary speech when presented in isolation.

The second possible explanation of subjects' success on the CSR task is that children are recognising the *content* of their own utterances. Although every effort was made to ensure that highly salient utterances and those with obvious content cues were excluded from the playback, subjects' success on

this task might be due in part to the fact that they remembered what they had said the week before. One way of controlling for this possibility would be to manipulate the delay between phases 2 and 3. A longer time delay would be expected to reduce children's memory for verbatim information (Sachs, 1967), thus minimising the possibility that they are recognising their own speech by simply remembering what they said. Another way of settling the issue might be to test children's recognition of the content of their own utterances in the absence of any other markers, for example, by presenting them with their own utterances spoken by an actor. If children were equally capable of recognising their own utterances in this situation, it would be a strong indication that they were recognising the content of their own utterances, rather than using acoustic or other cues.

The third source of information available in the CSR task is that which relates to Bakhtin's (1986) concept of *voice*. As mentioned in the Introduction, a voice is a way of speaking which makes manifest a particular perspective on reality through features of tone, prosody, content, and so forth. This third type of information therefore incorporates both types of information mentioned earlier. However, in discussing this source of information separately, we note that Bakhtin's conception of voice goes some way beyond acoustic features and utterance content to consider how perspectives on the world are manifested semiotically in concrete social exchanges. Although it may be difficult to see how such features of one's "speaking identity" might be operationalised (and thus controlled for experimentally), we can certainly hope that future research will allow us to be more specific about which aspects of contextual speech recognition depend on acoustic structure, intonation, memory for content, prosody, and so forth. For example, a correct answer on the CSR task might have less to do with the subject recognising *what* was said as the fact that he/she had *chosen* to say it. Comparison of recall for scripted and spontaneous utterances might demonstrate an element of utterance recognition which goes beyond the mere content of the utterance to the issue of how it expresses the child's point of view: in other words, the extent to which it is an example of the child's voice. Studies such as this might allow us to be more specific about how different informational sources are integrated in the CSR task, and thus increase our understanding of what it means to possess a "way of speaking" in social contexts.

We have suggested that the findings of the present study are consistent with the view that private speech provides individuals with a variety of cues which allow them to distinguish their own voices from those of others. Bearing in mind the correlational nature of these findings, an obvious objection is that the direction of causation might have been wrongly described. That is, some children might start off with a clearer idea of how their voices sound, perhaps because of a more developed self-awareness,

and these qualities might be related to elevated levels of private speech.

One way of settling this issue would be to take independent measures of self-awareness and see how they correlate with CSR performance. For example, in response to Duval and Wicklund's (1972) theory of self-awareness, Fenigstein, Scheier, and Buss (1975) developed the Self-consciousness Scale, designed to measure individuals' disposition to focus inwards or be self-attentive. However, Duval and Wicklund's theory proposes the existence of "standards" of values for certain behaviours and self-dimensions, which form the basis for self-evaluation and self-criticism. It is perhaps unlikely that such "standards" would be held by children of the ages studied here, particularly as attempts to induce a state of self-awareness in children aged nine or below have proved unsuccessful (Beaman, Klentz, Diener, & Svanum, 1979; Morin & Everett, 1991).

A more promising route to independent measures of self-awareness might be to analyse the content of children's private and social speech, with a view to determining the frequency of self-referring utterances. Future studies might investigate the relationship between the frequency of such utterances and the ability to recognise one's own speech. In particular, a longitudinal study would allow us to determine how private speech and CSR abilities develop over time, and how their emergence is related to the child's developing awareness of self.

Finally, we should consider whether the proposed function of private speech in distinguishing one's own voice might apply to other forms of intentional behaviour, both nonverbal (e.g. symbolic play) and verbal (e.g. social speech). For instance, it may be that children are able to compare information about the social speech they have produced with what they hear around them, and thus come to distinguish their own voices from those of others. That said, social speech has an obvious function (communication) which is not shared by private speech. The puzzle presented by private speech is this: Why do children engage in noncommunicative speech when they have an obvious opportunity to communicate? The answer, we suggest, is that private speech in social contexts has a specific function in self-discrimination. In support of this view, the present study has shown no evidence of a link between social speech and CSR. Children's performance on the task was related neither to the total frequency of utterances, nor to the frequency of social utterances. We therefore conclude that, if social speech does perform the same function as private speech in self-discrimination, this fact is not reflected in the correlation data.

Another possible objection might run as follows: Given that it is engaging in *private* speech, which is supposed to aid the development of voice self-recognition, how does this help children to recognise their own *social* speech on an audio recording? The answer to such an objection lies in what Kohlberg et al. (1968) termed the "parasociality" of private speech. Rather

than representing an entirely different form of linguistic behaviour, private speech is a form of communicative behaviour which does not fully distinguish between self-as-auditor and other-as-auditor, and which therefore represents the "insufficient differentiation of speech for oneself from speech for others" (Vygotsky, 1934/1986, p. 231). One might therefore think of such speech as being private only in a behavioural sense, in its lack of any obvious direction towards another listener, and in its frequent failure to take the perspective of the other into account. This "parasocial" nature of private speech suggests that the characteristic features of an individual's voice will be manifested in both social and private speech.

We have presented evidence that private speech may have a function in development which is additional to those described by previous theories. Our study has demonstrated a positive relationship between levels of private speech and a measure of contextual speech recognition, but no such relationship with social speech or overall verbosity. We have suggested that private speech might have a role in allowing individuals to distinguish their own voices from those of others in a multi-voiced situation, and thus come to understand themselves as speaking agents in a world of other such agents. It is possible, therefore, that private speech has a role not only in using words in the self-regulation of behaviour (Vygotsky, 1934/1986), but also in drawing the very distinction between self and other on which such verbal self-regulation can be assumed to depend.

Manuscript received June 1995
Revised manuscript received December 1995

REFERENCES

Bakhtin, M.M. (1986). *Speech genres and other late essays* (C. Emerson and M. Holquist, ed., V.W. McGee, trans.). Austin, TX: University of Texas Press.

Baldwin, J.M. (1906). *Thought and things: Vol. 1. Functional logic.* London: Swan & Sonnenschein.

Bartholomeus, B. (1973). Voice recognition by nursery school children. *Canadian Journal of Psychology, 27*, 464–472.

Beaman, A.L., Klentz, B., Diener, E., & Svanum, S. (1979). Self-awareness and transgression in children: Two field studies. *Journal of Personality and Social Psychology, 37*, 835–846.

Berk, L.E. (1992). Children's private speech: An overview of theory and the status of research. In R.M. Diaz and L.E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 17–53). Hove, UK: Lawrence Erlbaum Associates Ltd.

Diaz, R.M. (1992). Methodological concerns in the study of private speech. In R.M. Diaz and L.E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 55–81). Hove, UK: Lawrence Erlbaum Associates Ltd.

Duval, S., & Wicklund, R.A. (1972). *A theory of objective self-awareness.* New York: Academic Press.

Fenigstein, A., Scheier, M.F., & Buss, A.H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 36*, 1241–1250.

Furrow, D. (1992). Developmental trends in the differentiation of social and private speech. In R.M. Diaz and L.E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 143–158). Hove, UK: Lawrence Erlbaum Associates Ltd.

Fuson, K.C. (1979). The development of self-regulating aspects of speech: A review. In G. Zivin (Ed.), *The development of self-regulation through private speech* (pp. 135–217). New York: Wiley.

Goudena, P.P. (1992). The problem of abbreviation and internalization of private speech. In R.M. Diaz and L.E. Berk (Eds.), *Private speech: From social interaction to self-regulation* (pp. 215–224). Hove, UK: Lawrence Erlbaum Associates Ltd.

Habermas, J. (1984). *The theory of communicative action. Vol. 1: Reason and the rationalization of society*, (T. McCarthy, trans.). Boston, MA: Beacon Press.

Holquist, M. (1981). Glossary to Bakhtin (1981), *The dialogic imagination: Four essays by M.M. Bakhtin* (M. Holquist, ed.; C. Emerson and M. Holquist, trans.; pp. 423–434). Austin, TX: University of Texas Press.

Kohlberg, L., Yaeger, J., & Hjertholm, E. (1968). Private speech: Four studies and a review of theories. *Child Development, 39*, 691–736.

Locke, J. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27*, 153–165.

Martin, G.B., & Clark, R.D. (1982). Distress crying in neonates: Species and peer specificity. *Developmental Psychology, 18*, 3–9.

Maurer, D., & Landis, T. (1990). Role of bone conduction in the self-perception of speech. *Folia Phoniatrica, 42,* 226–229.

Mead, G.H. (1934). *Mind, self and society from the standpoint of a social behaviourist*. Chicago, IL: University of Chicago Press.

Morin, A. (1993). Self-talk and self-awareness: On the nature of the relation. *Journal of Mind and Behaviour, 14*, 223–234.

Morin, A., & Everett, J. (1991). Self awareness and 'introspective' private speech in six-year-old children. *Psychological Reports, 68*, 1299–1306.

Olszewski, P. (1987). Individual differences in preschool children's production of verbal fantasy play. *Merrill-Palmer Quarterly, 33*, 69–86.

Piaget, J. (1926). *The language and thought of the child*. London: Kegan Paul, Trench, Trubner & Co.

Rubin, K.H. (1979). The impact of the natural setting on private speech. In G. Zivin (Ed.), *The development of self-regulation through private speech* (pp. 265–294). New York: Wiley.

Rubin, K.H. (1982). *The private speech of preschoolers who vary with regard to sociability*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Russell, J. (1996). *Agency: Its role in mental development*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics, 2*, 437–442.

Saville-Troike, M. (1988). Private speech: Evidence for second language learning strategies during the 'silent' period. *Journal of Child Language, 15*, 567–590.

Vygotsky, L.S. (1986). *Thought and language* (A. Kozulin, ed. and trans.). Cambridge, MA: MIT Press. (Original work published 1934.)

Weir, R. (1962). *Language in the crib*. The Hague: Mouton.

Wertsch, J.V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Hemel Hempstead, UK: Harvester Wheatsheaf.