

Development of self-voice recognition in children

Sofia Strömbergsson

Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm

Abstract

The ability to recognize the recorded voice as one's own was explored in two groups of children, one aged 4-5 and the other aged 7-8. The task for the children was to identify which one of four voice samples represented their own voice. The results indicate that 4 to 5 year-old children perform as well as 7 to 8 year-old children when identifying their own recorded voice. Moreover, a time span of 1-2 weeks between recording and identification does not affect the younger children's performance, while the older children perform significantly worse after this time span. Implications for the use of recordings in speech and language therapy are discussed.

Introduction

To many people, the recorded voice often sounds unfamiliar. We are used to hearing our voice through air and bone conduction simultaneously as we speak, and as the recorded speech lacks the bone conduction filtering, its acoustic properties are different from what we are used to (Maurer & Landis, 1990). But even though people recognize that the recorded voice sounds different from the voice as we normally hear it, people most often still recognize the recording as their own voice. In a recent study on brain hemisphere lateralization of self-voice recognition in adult subjects (Rosa et al, 2008), a mean accuracy of 95% showed that adults rarely mistake their own recorded voice for someone else's voice.

Although there have been a few studies on adult's perception of their own recorded voice, children's self-perception of their recorded voices is relatively unexplored. Some studies have been made of children's ability to recognize other familiar and unfamiliar voices. For example, it has been reported that children's ability to recognize previously unfamiliar voices improves with age, and does not approach adult performance levels until the age of 10 (Mann et al, 1979). Studies of children's ability to identify familiar voices have revealed that children as young as three years old per-

form well above chance, and that this ability also improves with age (Bartholomeus, 1973; Spence et al, 2002). However, the variability among the children is large. These reports suggest that there is a developmental aspect to the ability to recognize or identify recorded voices, and that there might be a difference in how children perform on speaker identification tasks when compared to adults.

Shuster (1998) presented a study where children and adolescents (age 7-14) with deviant speech production of /r/ were recorded when pronouncing words containing /r/. The recordings were then edited so that the /r/ sounded correct. A recording in the listening script prepared for a particular child could thus be either an original recording or a "corrected" recording, spoken either by the child himself/herself or another speaker. The task for the children was to judge both the correctness of the /r/ and the identity of the speaker. One of the findings in this study was that the children had difficulty identifying the speaker as himself/herself when hearing a "corrected" version of one of their own recordings. The author speculates that the editing process could have introduced or removed something, thereby making the recording less familiar to the speaker. Another confounding factor could be the 1-2 week time span between the recording and the listening task; this could also have made the task more difficult than if the children had heard the "corrected" version directly after the recording. Unfortunately, no studies of how the time span between recording and listening might affect children's performance on speaker identification tasks have been found, and any effects caused by this factor remain unclear.

Of the few studies that have been done to explore children's perception of recorded voices – of their own recorded voice in particular – many were done over twenty years ago. Since then, there has been a considerable increase in the number of recording devices that can potentially be present in children's environments. This strongly motivates renewed and deeper exploration into children's self-perception of their recorded voice, and possible developmental changes in this perceptual abil-

ity. If it is found that children indeed recognize their recorded voice as their own, this may have important implications for the use of recordings in speech and language intervention.

Purpose

The purpose of this study is to explore children's ability to recognize recordings of their own voice as their own, and whether this ability varies depending on the age of the child and the time between the recording and the listening. The research questions are:

1. Are children with normal hearing able to recognize their own recorded voice as their own, and identify it when presented together with 3 other child voices?
2. Will this ability be affected by the time span between recording and listening?
3. Will the performance be affected by the age of the child?

It is hypothesized that the older children will perform better than the younger children, and that both age groups will perform better when listening immediately after the recording than when listening 1-2 weeks after the recording.

Method

Participants

45 children with Swedish as their mother tongue, and with no known hearing problems and with no previous history of speech and language problems or therapy were invited to participate. The children were divided into two age groups, with 27 children aged 4-5 years (ranging from 4;3 to 5;11, mean age 5;3) in the younger group and 18 children aged 7-8 years (ranging from 7;3 to 8;9, mean age 8;0) in the older group. Only children whose parents did not know of or suspect any hearing or language problems in the child were invited. All children were recruited from pre-schools in Stockholm.

Material

A recording script of 24 words was constructed (see Appendix). The words in the script all began with /tV/ or /kV/, and all had primary stress on the first syllable.

Three 6-year old children (two girls and one boy, included by the same criteria as the children participating in the study) were recorded

as references. None of the reference children were known to the children in the test groups.

Recording/Identification procedure

A computer program was used to present the words in the scripts in random order, and for each word

1. Play a reference voice (adult) that reads a target word, while displaying a picture that illustrates the word.
2. Record the subject's production of the same word (with the possibility of listening to the recording and re-recording until both child and experimenter are satisfied).
3. Play the subject's production and 3 reference children's productions of the same word, in random order, letting the subject select one of these as his/her own. (See Figure 1.)

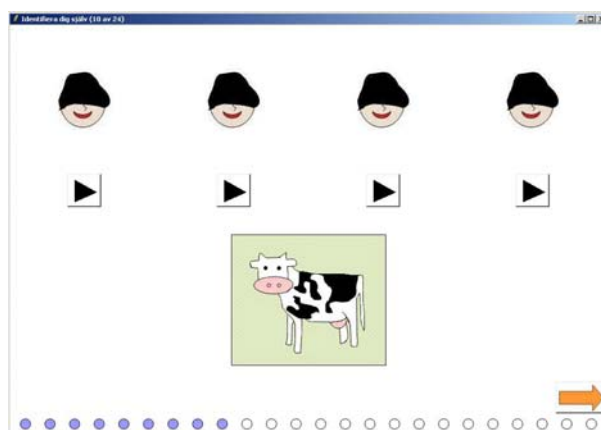


Figure 1: The listening/identification setup.

In both test sessions, the children were fitted with a headset and the experimenter with headphones to supervise the recordings. The children were instructed to select the character they believed represented their own voice by pointing at the screen; the actual selection was managed by the experimenter by mouse clicking. The children were given two introductory training items, to assure understanding of the task.

In the first test session, the children performed both the recording and the voice identification task, i.e. step 1-3. For the recordings, all children were instructed to speak with their normal voice, and utterances were re-recorded until both child and experimenter were satisfied. In the second test session, after a period of 1-2 weeks, the children performed only the identification task, i.e. step 3. Apart from general encouragement, the experimenter provided

no feedback regarding the children's performance during the voice identification task. All actions – recording, listening and selecting – were logged by the computer program.

Results

Table 1 displays the mean correct own-voice identification for all 45 children on both test occasions. The standard deviation reveals a large variation within the groups; the performance varies between 4 and 24 in both the first and the second test. However, the average results on both test occasions are well above chance level. A closer look at the individual results reveals that two children perform at chance level (or worse), while 12 children (27% of the children) perform with more than 90% accuracy.

Table 1. Mean correct responses on the first and second test, for both age groups (max score/test = 24).

	<i>First test</i>	<i>Second test</i>
Younger	18.8 (SD: 5.5)	17.9 (SD: 6.2)
Older	21.0 (SD: 2.2)	16.6 (SD: 5.5)
<i>Mean</i>	19.7 (SD: 4.6)	17.3 (SD: 5.9)

No difference was found between the younger and the older children's performance on the first test ($t(37.2) = 1.829$, $p = 0.076$) or on the second test ($t(43) = 0.716$, $p = 0.478$).

For the older children, a significant difference was found between the children's performance on the first test and their performance on the second test ($t(17) = 4.370$, $p < 0.001$), while for the younger children, no significant difference was found between the results from the two tests ($t(26) = 1.517$, $p = 0.141$).

Discussion

The high average performance rates confirm that children in the ages 4-5 years old and 7-8 years old are indeed able to recognize their recorded voice as their own. However, large variation was found among the children, with a few children performing at chance level (or worse) and more children performing with 90% accuracy or more.

Contrary to the hypothesis, no significant difference was found between the age groups. Thus, no support was found for a development in children's self-voice recognition ability between the ages of 4-8 years.

A significant difference was found between the older children's performance on the first and the second test. The hypothesis that children would perform better when listening immediately after the recording than after a period of 1-2 weeks could thus be confirmed for the older children, but not for the younger children. This might suggest a developmental aspect to what cues children use when identifying their own voice; younger children might attend to more stable characteristics, while older children recognize their own voice by other, more time-sensitive, features. As the age of the older children in this study match the age of the children in Shuster (1998), the results presented here supports the suggested interpretation of Shuster's results; children's difficulties to identify themselves as the speaker in "corrected" versions of their own recordings could be explained by the time span between recording and identification.

In this study, the children's speech production was only controlled to the extent that the experimenter instructed the children to speak with their normal voice, both when introducing the children to the task and whenever the experimenter judged that the child was somehow "playing" with his/her voice. However, some children tended to be more playful than others, and it is unlikely that all recordings reflect the children's normal speech behavior (whatever that is). Although this might certainly have an impact on the results – the children might recognize their speaking behavior rather than their own voice – this would have been difficult to avoid. Moreover, considering that speech play is often encouraged in clinical settings, one could argue that this is also ecologically valid.

The large variation among the children could be due to differences in attention, concentration or understanding of the task, but may also be explained by a difference in aptitude for the task at hand. A closer inspection of the recordings and results of the three children with the worst results (with a total score of 15 or below) revealed that two of these children actually produced slightly deviant speech (despite their parents' assurance that their children had normal speech and language development). This was also noted by the experimenter at the time of the recordings, judging both from the recordings and the children's spontaneous speech. One of the children (a girl aged 5;8) produced [j] for /r/. Another child (a boy aged 5;4) exhibited the same /r/-deviation, together

with a few cluster simplification patterns, such as [ta:vla] for “tavla” (*picture*). For the third of these children (a boy aged 4;4), and for all of the other children included in the study, no speech production deviations was noted or could be detected in the recordings. This might suggest a correlation between deviant speech production and difficulties of recognizing the recorded voice as one’s own. However, a contradictory example was also found that had to be excluded from the study. Dentalisation (i.e. systematic substitution of [t], [d] and [n] for /k/, /g/ and /ng/, respectively) was noted for one girl who could not participate for a second test, and who was therefore excluded from this study. Interestingly, this girl scored 23 of 24 on the first test. These single cases do certainly not present a uniform picture of the relation between deviant speech production and the ability to recognize the recorded voice as one’s own, but rather illustrate the need for further investigation of this relation.

The results in this study give support to the use of recordings in a clinical setting, e.g. when promoting awareness in the child of deviations in his/her speech production. An example of an effort in this direction is presented in Shuster (1998), where children were presented with original and “corrected” versions of their own speech production. The great variation between children in their ability to recognize their recorded voice as their own requires further exploration.

Conclusions

The findings in this study indicate that children in the ages of 4-5 and 7-8 years can indeed recognize their own recorded voice as their own; average performance results are well above chance. However, there is a large variability among the children, with a few children performing at chance level or worse, and many children performing with more than 90% accuracy. No significant difference was found between the younger and the older children’s performance, suggesting that self-voice perception does not improve between these ages. Furthermore, a time span of 1-2 weeks between recording and identification seems to make the identification task more difficult for the older children, whereas the same time span does not affect the younger children’s results. The findings here support the use of recordings in clinical settings.

Acknowledgements

This work was funded by The Swedish Graduate School of Language Technology (GSLT).

References

- Bartholomeus, B. (1973) Voice identification by nursery school children, *Canadian Journal of Psychology/Revue canadienne de psychologie* 27, 464-472.
- Mann, V. A., Diamond, R. and Carey, S. (1979) Development of voice recognition: Parallels with face recognition, *Journal of Experimental Child Psychology* 27, 153-165.
- Maurer, D. and Landis, T. (2005) Role of bone conduction in the self-perception of speech, *Folia Phoniatrica* 42, 226-229.
- Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P. and Belin, P. (2008) Investigations of hemispheric specialization of self-voice recognition, *Brain and Cognition* 68, 204-214.
- Shuster, L. I. (1998) The perception of correctly and incorrectly produced /r/, *Journal of Speech, Language, and Hearing Research* 41, 941-950.
- Spence, M. J., Rollins, P. R. and Jerger, S. (2002) Children’s Recognition of Cartoon Voices, *Journal of Speech, Language, and Hearing Research* 45, 214-222.

Appendix

	Orthography	Transcription	In English (the letter k)
1)	k	/ko:/	
2)	kaka	/ka:ka/	<i>cake</i>
3)	kam	/kam/	<i>comb</i>
4)	karta	/ka:ʔa/	<i>map</i>
5)	katt	/kat/	<i>cat</i>
6)	kavel	/ka:vəl/	<i>rolling pin</i>
7)	ko	/ku:/	<i>cow</i>
8)	kopp	/kɔp/	<i>cup</i>
9)	korg	/korj/	<i>basket</i>
10)	kula	/ku:lɑ/	<i>marble</i>
11)	kulle	/kələ/	<i>hill</i>
12)	kung	/kɛŋ/	<i>king</i>
13)	tåg	/to:g/	<i>train</i>
14)	tak	/ta:k/	<i>roof</i>
15)	tant	/tant/	<i>lady</i>
16)	tavla	/ta:vla/	<i>picture</i>
17)	tidning	/ti:nɪŋ/	<i>newspaper</i>
18)	tiger	/ti:gr/	<i>tiger</i>
19)	tomte	/tɔmtə/	<i>Santa Claus</i>
20)	topp	/tɔp/	<i>top</i>
21)	tub	/tʉ:b/	<i>tube</i>
22)	tumme	/tɔmə/	<i>thumb</i>
23)	tunga	/tɔŋɑ/	<i>tongue</i>
24)	tupp	/tɔp/	<i>rooster</i>