GABI – HUNGARIAN CHILD LANGUAGE AND SPEECH DATABASE AND INFORMATION REPOSITORY

Judit Bóna, Vakula Tímea, Váradi Viola
ELTE Eötvös Loránd University, Budapest, Hungary
bona.judit@btk.elte.hu, vakula.timi@gmail.com, varadiviola@gmail.com

Abstract

GABI (Gyermeknyelvi beszédAdatBázis és Információtár; Child Language and Speech Database and Information Repository) is a multi-purpose speech database which is currently under development. This database contains recordings of speech production by 420 children / adolescents ages 3-18 years of age. Two age-specific variations of linguistic material were developed. One protocol was developed for children between the ages of 3 and 9 years, and the other one for adolescents between the ages of 9 and 18 years. In the case of 9 year old children, recordings were using both protocols.
Speech tasks included the following: sentence repetition, spontaneous speech, definition of words, narrative recall of two listening materials, oral reading, story-telling based on a picture sequence, and a debate between two speakers about various topics (for children 9 years of age and older).

Keywords: child speech database, language acquisition, phonetic transcription, phonetic and psycholinguistic analyses

1 Introduction

Speech databases are essential for research and many practical applications. The structure (recorded linguistic material, number and characteristic features of speakers, circumstances of the recordings etc.) of a particular speech database depends on its aim. Speech databases exist for phonetic and psycholinguistic analysis, for speaker or language identification, for the analysis of pathological speech, or for applications of speech technology (speech recognition, speech synthesis) (e.g., Greenbaum & Svartvik, 1990; Godfrey & Holliman, 1993; Raab et al., 2007; Gósy, 2012). The majority of speech databases only contain speech samples of adults, but the number of speech databases which also contain child speech samples is increasing (see examples below).

Precise preparations and great attentiveness are needed to design and develop a child speech database. For instance, children will vary by native language, number of languages known, cognitive ability, and various psychological characteristics. When designing a child language database, the following considerations should be considered:

1) how cognition and speech develop in a child of a particular age: this affects the choice of speech tasks, the selection of pictures, texts etc.,
2) how children communicate and with whom. It is important to know what characterizes a child's speech and language ability at a given age. Although there are big individual differences, generally, younger children prefer to communicate with family members rather than with unfamiliar adults, and
3) how children are expected to respond to the recording situations and tasks. Interviewers have to be prepared to respond appropriately when a child doesn't want to speak, or if he/she moves away from the microphone, etc.

One of the greatest challenges is to handle the intra- and interspeaker variability which significantly influences children's speech. The speech of the infants and toddlers (before the age of 3) is characterized by their specific articulation, vocabulary and use of language which is better comprehended by their closest caretakers (parents, etc.) than unfamiliar adults. These phenomena make it almost impossible for database developers to annotate or recognize the speech automatically (Beckman et al., 2017).

Another difficulty could be that speech aptitude might be different depending on the age and gender of the child. Younger children are often embarrassed in experimental situations (like a speech recording), and the interviewer is often unfamiliar to them. However, adolescents can also be quite non-communicative or unenthusiastic about the task (Vakula & Váradi, 2017).

When planning the recording protocol of the recordings for the speech of older children (after the age of 3), the developmental level of their speech should be considered. It is known that vocabulary and grammatical knowledge are different in case of preschoolers, lower elementary students or adolescents, and this fact should to be considered in the compilation of linguistic stimuli. The recording of oral reading is only possible over when a child reaches a certain age and level of schooling. To analyze the phonetic characteristics of oral reading, we need recordings from children who can read almost fluently. Teaching reading will vary depending upon the reading technique used by speakers of a particular language and country of origin. In Hungary, the reading instruction begins in primary school. Usually around the second grade, the phonetic parameters of oral reading can be analyzed.

Brian MacWhinney and Catherine Snow began to develop the most well-known international child language speech database, CHILDES (MacWhinney & Snow, 1990) in 1984 with the aim of documenting the acquisition of the first language with speech samples. The database includes speech recordings of children with different mother tongues, thus providing an opportunity to describe universal phenomena, and to analyze the characteristics of vocabulary and grammar. The database contains 26 different languages at present, including materials from Hungarian-speaking children. Speech samples stored in CHILDES were recorded by multiple interviewers under various circumstances, in diverse speech situations, with children of various ages. Most of the recordings are spontaneous interactions and conversations. The database is freely accessible and usable (http://childes.psy.cmu.edu), and it is possible to upload new recordings. Currently, it is available as part of TalkBank with several other databases.

Several other (mainly American and British English) child speech databases exist. For example, Kids' Audio Speech Corpus is a collection of speech samples of American children (Eskenazi, 1996); and speech databases for speech recognition were also created (Shobaki et al., 2000; D'Arcy et al., 2004; Kazemzadeh et al., 2005). Using several speech situations (e.g., narratives, conversations and fairy tales), monolingual and bilingual children were also recorded in the Systematic Analysis of Language Transcripts (SALT) databases. Conversational dialogues were collected from 3rd, 4th and 5th grades (i.e., children who are 7 to 11 years old) for the Boulder Learning – MyST (My Science Tutor) Corpus in which students talk to a virtual tutor on scientific topics (Ward et al., 2013; Ward & Cole, 2015). The Providence Corpus contains mother-child spontaneous speech interactions of six children from southern New England. This is a longitudinal corpus which contains recorded speech samples from children between approximately one and three years (Demuth et al., 2006). There are also databases of languages other than English, as well as multi-lingual child speech databases which contain speech samples from a child's second language (L2) (Children News, Raab et al., 2007). INFANTRU and CHILDRU are Russian databases (Lyakso, 2010). The CHIEDE database contains conversations and interviews from Spanish children (Garrote, 2008). French mother-child spontaneous speech interactions are recorded in the Lyon Corpus (Demuth & Tremblay, 2008). The CASS_CHILD (Gao et al., 2012) and SingaKids-Mandarin (Chen et al., 2016) databases collect speech samples from Mandarin-speaking children. There are also speech databases in less commonly studied languages, like the children's Filipino speech corpus (CFSC, Pascual & Guevara, 2012) and the Demuth Sesotho Corpus (Demuth, 1992).

The earliest Hungarian child language recordings can be found in the Hegedűs Speech Database (Menyhárt, 2012). In this repository, there are 125 minutes of speech samples from children living in a village. Their age was between 10 and 16 years. The recordings were made between 1950 and 1956. The SPECO children's speech database contains speech samples from 76 children (5 to 10 years of age), which were recorded for speech technology applications (Csatári et al., 1999). The Hungarian preschool speaker's speech database (MONYEK) contains speech samples of 62 children aged between 4;6 and 5;6 performing several different speech tasks (Mátyus & Orosz, 2014). TiniBEA contains recordings from 16 to17 year old adolescents and is a database that is well suited for phonetic examinations (Gyarmathy & Neuberger, 2015). In addition, many other speech corpora have been developed for different research purposes with kindergarten, schoolchildren and teenagers (e.g., Gósy, 1984; Horváth, 2006; Laczkó, 2009; Neuberger, 2013). However, there is no children's speech database available to researchers that contains a large number of speech samples that could be used for a wide range of research purposes.

The Hungarian child speech database which contains the most extensive amount of speech material across a wide variety of ages is the GABI – a Hungarian Child Language and Speech Database and Information Repository. The GABI is presented in the following section.

2 The GABI speech database

GABI database is being developed in the Phonetics Department of Eötvös Loránd University by the members of the Child Language Research Group – comprised of faculty members, doctoral students and graduate students of the department. The development of this speech database started in February of 2013. The developers received a research grant to study the speech of 5 to10 year old children, recorded from 2016 to 2020. For this reason, the recordings and the annotations from children in this age range are the most represented within the database.

The development of the recording protocol was preceded by extensive professional consultation and detailed processing of relevant international and Hungarian literature. We have consulted with many colleagues who are familiar with the recordings of children's speech as well as with the developers of BEA Hungarian Spoken Language Database (Gósy, 2012; Neuberger et al., 2014). We also have made trial recordings with children of different ages. During this preliminary phase, a number of texts and picture sequences were tested to determine the age-specific linguistic materials and series of pictures suitable for use with children between the ages of 3-18 years.

Since speakers of the database represented a large age range, two age-specific variations of linguistic material were developed for the protocol. One of the protocol versions was developed for children between the ages of 3 and 9 years, and another one for adolescents between the ages of 9 and 18 years. In the group of 9-year-old children, recordings were made using both protocols.

2.1 Subjects

The database contains recordings of speech production by children/ adolescents aged between 3 and 18 years. Currently, this includes 442 children (Table 1). In this database, the majority of the recorded speech is from monolingual Hungarian-speaking children, but the speech of 10 Hungarian-German and 15 Hungarian-English bilingual children have been included. Speech samples of two twin pairs were also recorded and two children participated in the recording process twice. The second recordings were made two years after the initial ones.

The age group of 5 to 10 year old children is a special one, so we focused on collecting and analyzing speech recordings for this age range. As a consequence, most of our recordings are in this age range. This period is especially important for us, because institutional learning in Hungary starts at the age of 6 years. Younger children attend kindergarten, which is an educational institution for children between the ages 3 and 6 in Hungary. There is an obligatory oral linguistic test at the age of 5. If a disorder is revealed, there is time to provide intervention before the child starts primary schooling. For this reason, the speech of lower primary students is highlighted in our current focus, because producing coherent connected speech might predict academic success in schooling. In addition, the level of acquisition of written language can be assessed using oral reading tasks.

Table 1: Number of recordings at each age

Age	Number of recordings
3	5
4	20
5	31
6	46
7	57
8	41
9	66
10	23
11	23
12	20
13	28
14	23
15	15
16	14
17	25
18	5
All	442

2.2 Data collection

Recordings were made in an everyday environment, in a quiet room in the kindergarten or school or in their own home. Excluding a completely unfamiliar speech situation, we can minimize the Observer's Paradox (i.e., the child’s speech is influenced by the presence of an observer; Labov, 1972). Before starting the recording, the interviewer usually showed the children how the recorder worked, thereby reducing the frustration caused by this special speech situation. The interviewer made the child feel comfortable with the recording situation, thereby increasing his/her willingness to speak. The speech samples were recorded in a digital format, 44.1 kHz, 16 bit, 86 kbit/s, mono.

2.3 Linguistic material

The speech tasks for children aged between 3 and 9 years included the following:

1. Recall of 15 sentences (10 declarative sentences and 5 interrogative sentences) with various lengths and phonetic properties.
2. Spontaneous speech about different topics like nursery/school, free time, family celebrations, events, travels.
3. Definition of 20 words/phrases.

4. Recall of a text presented auditorially to the child.
5. Reading aloud sentences and a text (starting in grade 2).
6. Story-telling based on a series of pictures.

The speech tasks for speakers aged between 9 and 18 years included the following:

1. Recall of 15 sentences (10 declarative sentences and 5 interrogative sentences) with various lengths and phonetic properties.
2. Spontaneous speech sample using different topics, like school, free time, events, travels.
3. Definition of 20 words/phrases.
4. Recall of two listening texts.
5. Reading aloud of sentences and a text.
6. Story-telling based on a series of pictures.
7. A debate between two speakers about various topics.

Some examples from the sentence repetition task (recall of sentences) include:

A gyermekek bukfencezni is megtanulnak testnevelésórán.
'Children learn to perform a somersault in P. E. lessons, too.'

Nemsokára odaérünk, ugye?
'We're almost there, aren't we?'

A gyöngyhalászok hosszú ideig képesek a víz alatt maradni.
'Pearl-divers are able to stay underwater for a long time.'

Some examples from the definition task: *lift* 'elevator', *mérleg* 'scales', *szirom* 'petal', *lóvá tesz* 'take for a ride'

Two children were involved in a debate discussing a topic provided by the interviewer. Speakers received the topic, and then their first task was to decide whether they agreed with the statement or not. Then the debate started and the participants discussed their opinions. The goal was to reach a common understanding. If the speakers had the same opinion, one of them was appointed to represent the pro and the other to represent the contra position. The goal of this task was to record a conversation between speakers.

Topics for the debate included the following:
Jó-e az ötfokozatú értékelés?
'Do you agree with the grading system at school?'

Jó-e az iskolai egyenruha?
'Do you like/would you like to wear a school uniform?'

Az autó helyett vissza kellene térni a kerékpárhoz.
'Instead of driving a car, we should ride a bike more often.'

2.4 Annotation of the recordings

The speech samples included in the database are being annotated with Praat software (Boersma & Weenink, 2008) using specific rules. The first step in the annotation was the manual segmentation of the recordings. This meant that we annotated the boundaries of speech units. Speech units were defined as the sections of speech between two silent or filled pauses or between a pause and an instance of turn-taking. Non-lexical backchannel responses including hemmings were also taken as speech units (see Figure 1). Speech units with background noise were also annotated if the speech was comprehensible.

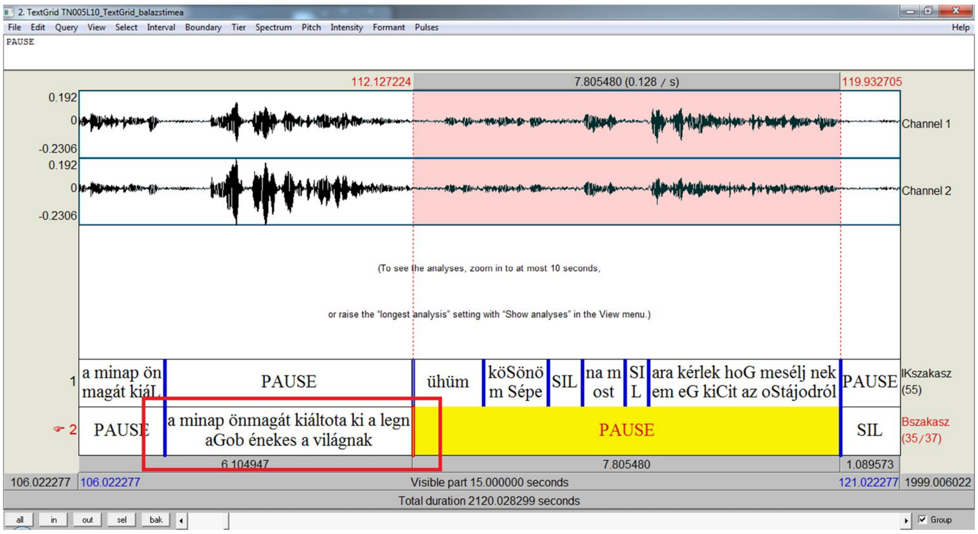

Figure 1.
A segment of the annotation of a speech sample (Meaning of the highlighted section: 'he called himself the greatest singer in the world recently')

Only one tier belongs to each speaker. Speech samples were annotated into two tiers (one tier belonged to the interviewer and the other tier belonged to the child). In case of the debate task, the interlocutor (the second child who took part in the debate) was annotated in a third tier.

The order of the tiers is the following:

1. interviewer
2. child1
3. child2

It was of essential importance that we keep this order independent from the order of the speakers.

We use a simple annotation system based on the "one sound is one character" principle to label the segmented speech units. Vowel length is indicated by the equivalent Hungarian letters (long *i* is *í*, long *o* is *ó*). Two-character consonants are indicated by an upper case letter (e.g., *gy* will be G, *sz* will be S). Silent and filled pauses (e.g., SIL, Ö, ÖHM, etc.) and the non-verbal vocalizations are indicated by appropriate abbreviations (e.g., KÖH stands for cough which *köhögés* in Hungarian). This kind of annotation makes it easy to read, does not use punctuation and focuses only on pronunciation.

2.4.1 Difficulties in annotation

Some factors make annotation difficult. These are described below:

1) One of the greatest individual characteristics of the speech of children is the accuracy of articulation. Inaccurate articulation can be observed particularly in the younger ages, which is explained in many cases by physiological reasons, requiring special attention from the annotator. In addition, besides the typical articulation inaccuracies, the speech of the youngest age groups was characterized by inaccuracies involving sound addition, or sometimes even speech disorders. The annotation of the speech of such children took twice as long on average as the annotation of the speech of a child with typical language development. It is also difficult that the perceptual mechanism of the annotator might correct the articulation errors after a while (i.e., they might get used to the child's speech). This was corrected by another annotator as each annotation is checked by a second person.
2) The annotation was made more difficult by the fact that non-verbal vocalizations were frequent, which interfered with the annotator's attention to the task. Because of the unfamiliar speech situation, children often squeaked with the chair, fidgeted, etc. These sounds were often recorded along with the speech, increasing the difficulty of the annotation process. In addition, several children began to sing during the recording process or spoke in a melodic form, as if almost singing.
3) The annotation of spontaneous speech was time-consuming because of unclear utterances, as well as due to the frequent turn-taking between examiner and child. Since children are not used to producing longer narratives by themselves, it was very common for children to finish their turn after 5 to10 seconds of speech. Therefore, the annotators needed to segment speech attempts more often, requiring frequently refocusing of their attention. The experience of the annotators, who also participated in developing the BEA adult speech database, showed that the longer units were easier for the child to remember.
4) Finally, in cases of longer pauses, the audio recordings did not support the decision-making of the examiner, i.e., whether the silence was actually a long pause in the child's speech or he/she lost his/her concentration. Occasionally,

the child would stand up from the table, walk to the far end of the room while continuing to speak. In order to overcome this problem, the interviewers were asked to verbalize during these events so that the annotator would be informed of what was happening.

3 Research on the speech samples of the GABI Speech Database

Literature on first language acquisition is very rich, although in some respects under-represented both in the Hungarian and international research. For example, there is much more research on the first stage of language acquisition than on the later stages. More studies report grammatical features, such as phonetic or pragmatic aspects and strong emphasis has been placed on the research of children between the ages of 1 and 2 years, neglecting older age groups. The majority of these studies are based on the observation of only a few children, and many times this information is based on parental observations. Earlier (very valuable) phonetic examinations were made with a relatively small number of speakers and were usually limited to examining only one speech type. The scientific value of the GABI speech database is that it contains recordings of several different speech types from many speakers, using the same recording protocol.

After the first recordings of GABI, researchers immediately started using them. The early studies focused on the specificity of the definition of words (Nagy-Varga, 2014; Bóna & Imre, 2015). These studies analyzed the strategies used by children of various ages as they defined word meanings. Since then, several publications and theses have been prepared considering the temporal analysis of children's speech productions and the examination of disfluencies (e.g., Vakula, 2016; Gósy & Krepsz, 2019; Vakula & Krepsz, 2018). These researchers found that the change of speech and articulation rate is not linear, and that there are large individual differences in temporal processing of speech among children. In addition, the inner time structure (the ratio of stem and suffix durations) of suffixed words were noted to change as children aged. In addition, using the framework of the above mentioned four-year NKFIH project (between 2016 and 2020), the temporal characteristics of the speech of children aged 5 to 10 years are being analyzed using different speech tasks taken from speech samples in the GABI speech database. This is a novel research task because it compares speech tasks which require different cognitive loads.

The GABI speech database makes it possible to describe the characteristics of spontaneous speech and oral reading (starting at the age of 7 years) of Hungarian-speaking children with typical development across several different speech tasks. We can also learn more about the features of speech planning in the late stages of first language acquisition, the age-related changes in the acoustic characteristics of speech, and the specific integration of suprasegmental factors across different speech types. In addition, the findings based on the analysis of the GABI speech materials provide an opportunity to characterize the speech production of those bilingual children whose one of the languages is Hungarian. For example, Jordanidisz et al. (2015) found differences between native Hungarian monolingual children and Hungarian–English bilingual children in voiceless plosives when producing Hungarian speech. Finally, these recordings would be useful for the comparison of the speech of atypical children to the speech of typically developing monolingual children.

4 Conclusions

Based on our experiences during the recording and annotation of children's speech samples, we have concluded that recording and annotating children's speech presents many challenges. Therefore, even experienced interviewers and annotators should be prepared for the challenging task of working with children's speech.

Research using speech samples from the GABI speech database can provide valuable results describing the speech of typically developing children. In addition to providing insights into speech and language development, large amounts of manually annotated speech samples might be suitable for developing a child speech automatic recognition system.

5 Acknowledgements

The authors wish to thank Mária Gósy, Angéla Imre and Alexandra Markó for their help in the initial development of the GABI speech database. The authors also wish to thank Zsófia Koren-Dienes and Ágnes Jordanidisz for their help in preparing this paper. This paper was supported by the Hungarian National Research, Development and Innovation Office of Hungary, project No. K-120234.

References

Beckman, M. E., Plummer, A. R., Munson, B., & Reidy, P. F. 2017. Methods for eliciting, annotating, and analyzing databases for child speech development. *Computer speech & language*, *45*, 278-299.

Boersma, P. & Weenink, D. 2008. Praat: doing phonetics by computer (Version 5.0.1). Available from: http://www.fon.hum.uva.nl/praat/download_win.html.

Bóna, J. & Imre, A. 2015. Szójelentések meghatározása 5 és 9 éves kor között [Definition of word meanings between 5 and 9 years]. *Beszédkutatás, 25*, 185-203.

Chen, N. F., Tong, R., Wee, D., Lee, P. X., Ma, B., & Li, H. 2016. SingaKids-Mandarin: Speech Corpus of Singaporean Children Speaking Mandarin Chinese. In: *Proceedings of Interspeech 2016.* 1545-1549.

Csatári, F., Bakcsi, Zs., & Vicsi, K. 1999. A Hungarian child database for speech processing applications. In: Olaszy, G., Németh, G., & Erdőhegyi, K. (eds.): *Proceedings of Sixth European Conference on Speech Communication and Technology*. Budapest. 2231-2234.

D'Arcy, S. M., Wong, L. P., & Russel, M. J. 2004. Recognition of read and spontaneous children's speech using two new corpora. In: *Proceedings of Interspeech 2004 - ICSLP.* 1473-1476.

Demuth, K. 1992. Acquisition of Sesotho. In: Slobin, D. I. (ed.): *The Cross-Linguistic Study of Language Acquisition, 3*. Hillsdale, N.J.: Lawrence Erlbaum Associates. 557-638.

Demuth, K., Culbertson, J., & Alter, J. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137-173.

Demuth, K., & Tremblay, A. 2008. Prosodically-conditioned variability in children's production of French determiners. *Journal of child language*, *35*(1), 99-127.

Eskenazi, M. S. 1996. KIDS: a database of children's speech. *The Journal of the Acoustical Society of America, 100,* 2759.

Gao, J., Li, A., & Xiong, Z. 2012. Mandarin multimedia child speech corpus: Cass_Child. In: *2012 International Conference on Speech Database and Assessments.* IEEE. 7-12.

Garrote, M. 2008. *CHIEDE: A spontaneous child language corpus of Spanish*. Unpublished Ph.D. thesis. Madrid: Universidad Autónoma de Madrid.

Godfrey, J. & Holliman, E. 1993. *Switchboard-1 release 2 LDC97S62.* DVD. Philadelphia: Linguistic Data Consortium.

Gósy, M. 1984. *Hangtani és szótani vizsgálatok hároméves gyermekek nyelvében* [Phonetic and vocabulary examinations in the language of three-year-old children]. Budapest: Akadémiai Kiadó.

Gósy, M. 2012. BEA – A multifunctional Hungarian spoken language database. *Phonetician*, *105*, 50-61.

Gósy, M. & Krepsz, V. 2019. *Stem and suffix durations in words of increasing length across the lifespan. Alkalmazott Nyelvtudomány, 19*(1), 1-23.

Greenbaum, S. & Svartvik, J. 1990. *The London-Lund corpus of spoken English, 7*. Lund: Lund University Press.

Gyarmathy, D. & Neuberger, T. 2015. Egy hiánypótló adatbázis: a Tini BEA [A niche speech database: TiniBEA]. *Beszédkutatás, 23*, 209-222.

Horváth, V. 2006. A spontán beszéd és a beszédfeldolgozás összefüggései gyerekeknél [Relationships between spontaneous speech and speech processing in children]. *Beszédkutatás, 14,* 134-146.

Jordanidisz, Á., Auszmann, A., & Bóna, J. 2015. Voice onset time of the voiceless alveolar and velar stops in bilingual Hungarian-English children and their monolingual Hungarian peers. In: Babatsouli, E. & Ingram, D. (eds.): *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015*. Chania: Institute of Monolingual and Bilingual Speech (ISMBS). 105-111.

Kazemzadeh, A., You, H., Iseli, M., Jones, B., Cui, X., Heritage, M., Price, P., Anderson, E., Narayanan, S., & Alwan, A. 2005. Tball data collection: the making of a young children's speech corpus. In: *Proceedings of Interspeech 2005*. 1581-1584.

Labov, W. 1972. *Sociolinguistic patterns. No. 4.* Philadelphia: University of Pennsylvania Press.

Laczkó, M. 2009. Tizenévesek beszédének fonetikai és stilisztikai elemzése [Phonetic and stylistic analysis of teens' speech]. *Anyanyelv-pedagógia, 2*(1), Available from: http://www.anyanyelv-pedagogia.hu/cikkek.php?id=151

Lyakso, E. E. 2010. Russian infants and children's sounds and speech corpuses for language acquisition studies. In: *Eleventh Annual Conference of the International Speech Communication Association.* 1878-1881.

MacWhinney, B. & Snow, C. 1990. The child language data exchange system: An update. *Journal of Child Language, 17*(2)*,* 457-472.

Mátyus, K. & Orosz, Gy. 2014. MONYEK – Morfológiailag egyértelműsített óvodai nyelvi korpusz [Morphologically annotated kindergarten language corpus]. *Beszédkutatás, 22*, 237-245.

Menyhárt, K. 2012. A beszéd temporális jellemzői 60 évvel ezelőtti gyermek beszélőknél [Temporal characteristics of speech in child speakers 60 years ago]. *Beszédkutatás, 20*, 246-259.

Nagy-Varga, Zs. 2014. Definícióalkotási stratégiák tizenéves diákok körében [Definition strategies of teenage students]. *Anyanyelv-pedagógia 7*(3), http://anyanyelv-pedagogia.hu/cikkek.php?id=525 A let

Neuberger, T. 2013. A spontán beszéd temporális sajátosságai 6–14 év közötti gyermekeknél [Temporal characteristics of spontaneous speech in children aged 6–14 years]. *Anyanyelv-pedagógia 6*(2), Available from: http://www.anyanyelv-pedagogia.hu/cikkek.php?id=451

Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., & Beke, A. 2014. Development of a large spontaneous speech database of agglutinative Hungarian language. In: *International Conference on Text, Speech, and Dialogue.* Cham: Springer. 424-431.

Pascual, R. M. & Guevara, R. C. L. 2012. Developing a children's Filipino speech corpus for application in automatic detection of reading miscues and disfluencies. In: *TENCON 2012 IEEE Region 10 Conference*. IEEE. 1-6.

Raab, M., Gruhn, R., & Noeth, E. 2007. Non-native speech databases. In: *2007 IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU)*. IEEE. 413-418.

Shobaki, K., Hosom, J-P., & Cole, R. A. 2000. The OGI kids' speech corpus and recognizers. In: *Proceedings of the Sixth International Conference on Spoken Language Processing.* Beijing. 259-261.

Vakula, T. 2016. Óvodás és kisiskolás gyermekek interpretált beszédének vizsgálata [Examining the interpreted speech of pre-school and schoolchildren]. Paper presented at the 10th Conference for PhD Students of Applied Linguistics. RIL HAS, Budapest.

Vakula, T. & Krepsz, V. 2018. Egyéni sajátosságok vizsgálata a spontán beszéd temporális szerkezetében [Investigation of individual characteristics in the temporal structure of spontaneous speech]. In: Magyari, S. & Bartha, K. (eds.): *Nyelv – nyelvközösség – közösségi perspektíva.* Nagyvárad: Partium Kiadó. 45-58.

Vakula, T. & Váradi, V. 2017. Gyermeknyelvi hangfelvételek rögzítésének és lejegyzésének tapasztalatai [Experiences in recording and annotation of children's speech]. In: Bóna, J. (ed.): *Új utak a gyermeknyelvi kutatásokban.* Budapest: ELTE Eötvös Kiadó. 51-64.

Ward, W. & Cole, R. 2016. Developing Conversational Multimedia Tutorial Dialogs. In Sottilare, R., Graesser, A., Hu, X., & Brawner, K. (eds.): *Design recommendations for intelligent tutoring systems, Vol 3. Authoring tools and expert modelling techniques. Chapter 20.* U.S. Army Research Laboratory. 243-254.

Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., & Weston, T. 2013. My science tutor: A conversational multimedia virtual tutor. *Journal of Educational Psychology*, *105*(4), 1115.

CASE STUDY

VOICE PATTERNS AS CHARACTERISTIC FEATURES IN PRIMARY PROGRESSIVE APHASIA

Mária Gósy
Research Institute for Linguistics, ELTE Eötvös Loránd University, Budapest
gosy.maria@nytud.hu

Abstract

This paper intends to evaluate the voice quality of a Hungarian-speaking female patient diagnosed with a non-fluent type of primary progressive aphasia. The patient's subjective complaints on the worsening of her voice quality were confirmed by objective acoustic data. Progressive deterioration of jitter, shimmer, signal-to-noise ratio, glottalization and fundamental frequency as well as the fundamental frequency range of the patient's voice was noted during a sentence repetition task in a longitudinal analysis. Worsening of voice quality features may signal the presence of primary progressive aphasia prior to appearance of the usual symptoms in certain cases. We suggest that a subjective complaint on voice quality is worth considering as a characteristic symptom of neurological decline that can be diagnostically useful.

Keywords: voice quality, fundamental frequency, aphasia

1 Introduction

The voice contains hidden information about neurological disorders. For instance, aphasia can affect the brain, spinal cord, nerves and muscles, and limits the typical movements of the speech organs, resulting in declines in intelligibility and changes in voice quality. There are cases when voice function is the leading deficit of the neurological disease, affected and impaired to a greater extent in some cases than is usual (Simpson & Woodson, 2003). In this study we report on an acoustic analysis of voice quality of a female patient diagnosed with the non-fluent type of primary progressive aphasia (nfvPPA); however, she showed symptoms characteristic of logopenic type of PPA, as well (Hoffman et al., 2017). Her primary subjectively identified disorder was the change of her voice quality from the very beginning of the disease. Our hypothesis was that the measured voice parameters would support the patient's own impression about the worsening of her voice quality and would be a preliminary symptom of the disease in her case.

The term 'primary progressive aphasia' refers to various types of neurodegenerative disorders where language impairment is the primary symptom