https://doi.org/10.3176/lu.1998.3.22

NINA ALAROTU, MIETTA LENNES (Helsinki)

VARIABILITY OF FINNISH SPEECH SPOKEN BY HEARING-IMPAIRED INDIVIDUALS

Introduction

When speech recognizers are developed as aids for the hearing-impaired, it is essential to know the typical properties of the speech of the hearing-impaired. A central problem is whether the variability included in the speech produced by hearing-impaired individuals is regular. A great deal of random variability would naturally add to the difficulty of the development process. In case the speech recognizer is based on a neural network, a lot of speech material is needed to cover all the variation, and collecting and preparing the necessary samples costs time, money, and technical resources.

In earlier studies, several tendencies have been observed in the speech production of deaf or hearing-impaired individuals. C. Hudgins and F. Numbers (1942) were the first to present a discussion of the general characteristics of the speech of deaf subjects. They described vowel substitutions, neutralization effects, and errors in the production of diphthongs. C. Hudgins and F. Numbers reported that hearing-impaired individuals have particular difficulties with the production of obstruent clusters (the components of which may be completely dropped) and with the voiced-voiceless distinction. Adventitious segments (usually [ə]) may also be added between sound segments.

J. Subtelny, R. Whitehead and V. Samar (1992) measured the vowel formant frequencies of deaf women. They found that in comparison to normal subjects the F2 was lower for the front vowel [i] and higher for the back vowels [ɑ] and [u]. The ranges of the three lowest formants seemed to be more limited in the productions of the deaf subjects. The study was in accordance with the earlier results of R. Shukla (1989), who also found the phonological space to be reduced for the hearing-impaired, this effect being primarily due to the lower second formant of the vowel [i]. J. Subtelny, R. Whitehead and V. Samar refer to their radiographic data, suggesting that the voice quality typical for deaf subjects, the so-called p h a r y n g e a l r e s o n a n c e, may be caused by the retracted tongue root and the retruded dorsum of the tongue. H. McCaffrey and H. Sussman (1994) reported that the least audible vowel formants F2 and F3 measured from the speech of both normal-hearing and hearing-impaired subjects showed the greatest effects of severe and profound hearing loss. F1 and F0, being more audible, showed interindividual changes and changes connected with the most profound hearing losses.

According to N. McGarr and A. Löfqvist (1982), who recorded speech samples from three deaf speakers, there is variation between speakers with regard to types of errors, variability, and interarticulator coordination. They also noted that hearing-impaired speakers have great difficulty in preserving temporal aspects of speech. In another study, N. McGarr and A. Löfqvist (1988) stated that the articulatory gestures of the stop and fricative productions of hearing-impaired speakers did not have as clear kinematic differences as those of the normal-hearing and, on the other hand, the kinematic measures were more variable for the hearing-impaired subjects. R. Waldstein (1990) claims that postlingual deafness affects all classes of speech sounds, suggesting that auditory feedback is essential for the control of speech production. A review of the literature on the subject of speech deterioration exhibited by postlingually deafened subjects is presented by H. Lane and J. Webster (1991), who also report that there is less differentiation of place of articulation between different fricatives and plosives and greater pitch variation in the speech of the hearing-impaired.

Although most studies in this field have been concerned with the English language, a great deal of the above mentioned tendencies probably occur in other languages, too. For Finnish, very few experiments have been conducted. One of them is a study by O. Aaltonen and J. Suonpää (1982), where the cepstra of the isolated Finnish vowel productions of four hearing-impaired children were investigated. O. Aaltonen and J. Suonpää found great variation in the F0 frequencies of these vowels. The formant structure of the vowels also showed large variability, but this fact was tolerated quite well by the normal-hearing listeners who were asked to identify the vowels. The realizations of /e/ of the hearing-impaired subjects were often confused with /æ/ and /œ/. On the other hand, although there was large variation in the productions of /y/, they were usually identified correctly. Since the normal-hearing speaker that was used as a control subject was an adult male, comparisons were difficult to make between the acoustic properties of the productions of the hearing-impaired children and the control subject.

This study presents a preliminary auditory analysis of the production errors and the qualitative variability observed in the 400 isolated Finnish words recorded from hearing-impaired subjects. It was hypothesized that the quality and the amount of errors are affected by such factors as time elapsed from the loss of hearing, the degree of hearing-impairment, and the age of the subject. Since the aim was to find basic information for a diphone-based speech recognizer, only the qualitative variability occurring in the segmental level is more closely investigated.

Methods

Spoken utterances of 400 isolated Finnish words were recorded from eight hearing-impaired adult subjects (age 35—64 years; 3 males). One male subject was congenitally deaf, other subjects had postlingually lost their hearing either gradually or abruptly. Five subjects were able to at least partially hear their own voice; six subjects were able to hear some other sounds. Two subjects were profoundly deaf, and another two wore multichannel cochlear implants.

The word list was built in order to cover the most frequently used words of Finnish. Previously recorded, transcribed samples of spontaneous speech were available from 11 normal-hearing Finnish speakers. 200 most frequently used words were picked out from these samples, and another 200 words were added to the list from the frequency dictionary of Finnish. Almost all words were in their basic forms. A small amount of words were replaced in order to increase the diphone diversity of the word list.

The recordings were to be used for training and testing a neural network speech recognizer. Since it was therefore desirable to pursue as natural speech samples as possible, the recordings were performed in an ordinary conference room. The room was arranged so that only minor background noise was present. The utterances were recorded using a Sennheiser high-quality miniature microphone attached on a head set and placed a few centimeters to the right from the right corner of the subject's mouth, ensuring that the distance between the mouth and the microphone was kept nearly constant.

A special computer program was used to present each subject with the words and to record the speech signals directly to the hard disk of a Macintosh computer at a sampling rate of 22 050 Hz. The subject was asked to read the word appearing on the screen and to produce it three times consecutively with a small pause between each utterance. The subject was also instructed to speak in the rate, pitch and level that he/she would normally use. The next word was presented when the subject was ready. The presentation of words was controlled with the mouse by the experimenter and the subject was allowed to have short breaks and at least one longer break during the recording session. From each subject, all 400 words were recorded in a single session. Each session lasted approximately 2 hours in total.

Results

A qualitative auditory analysis was performed by two phoneticians on all the recorded sound files. Since variable amounts of audible noise were present in the sound signals, different spectral measurements would probably be unreliable. All three instances of a few randomly selected words spoken by each subject were roughly transcribed to find general trends in the changes occurring from utterance 1 to utterance 3. All subjects appeared to exhibit normal-like reduction effects between the three utterances. However, the first utterance was sometimes less carefully pronounced than the second one — despite the fact that it usually carried a primary "sentence" stress. The prosodic details will not be further considered here.

For all subjects, the second utterances of each word were then picked out as prosodically similar instances for a comparison between the different words and the different subjects. These instances were generally found to be the clearest and the most successful productions of the words. Each of the selected instances was transcribed for an analysis of sound errors. In the following, the level and duration of the hearing-impairment of each subject is briefly described along with the individual properties observed in the subject's speech.

PP, female, 62: The subject started to lose her hearing when she was 21 years old and became deaf ten years later. Four years prior to the experiment she had received a multichannel cochlear implant. PP hears many kinds of sounds, including her own speech, but does not comprehend other people's speech without seeing the speaker's lips. PP's speech was completely intelligible and she did not use significantly deviant sound articulations. However, her voice quality was very creaky and glottalized.

HN, male, 51: The subject completely lost his hearing when he was 19 years old. Despite this fact, he has maintained very good speaking skills and even his original dialect. HN's productions of the words were very clear indeed, but some minor peculiarities could be noted. However, it is not clear whether these phenomena really are abnormal or just part of the "normal" variation in speech. HN was in the habit of starting and ending each word with a voiced or voiceless bilabial nasal [m]. The vowel of the second syllable tended to be slightly lengthened, which

was probably due to HN's dialect. His productions of /æ/ were often lowered to [a]. HN usually produced voiceless fricatives instead of [k]: [ç] in front vowel environments and [x] in others — this may not be unusual in normal-hearing speakers.

TK, female, 35: The subject has gradually lost her hearing since the age of 16. TK is still able to hear some sounds and her own voice, but it is difficult for her to control the loudness of her speech. She usually speaks with a very quiet, breathy voice. This resulted in a poor signal-to-noise ratio and sometimes impeded the transcription work. TK's vowels were usually centralized: her [o] was advanced, [u] became [ʉ], and in fact, /ɑ/ was mostly realized as [a]. The front vowels [e] and [i] were usually palatalized. She also produced very open variants of /e/ and raised variants of /æ/ to the point where the two articulations were practically reversed. TK's [s] was consistently "dark", i.e., retracted, [k] was slightly aspirated, and before close front vowels [k] was palatalized.

HM, female, 65: The subject began to lose her hearing at the age of 6 months and became severely hearing-impaired at the age of 50 years. With the hearing aid, she is able to hear her own voice and some other sounds, too. However, before the experiment, her hearing aid was turned off. HM's vowels were often centralized, even in stressed positions. The articulations of /e/ and /æ/ tended to be confused with each other. She also produced advanced, dental [s̪] sounds and had difficulties with the Finnish tremulant /r/, sometimes realizing it as [ɹ], sometimes as [rˡ].

HK, female, 61: The subject has gradually lost her hearing since the age of four years. She still hears her own voice and some other sounds with the hearing aid. HK tended to produce an [ʉ] instead of [u]. The articulations of /e/ and /æ/ were often similar or reversed. HK tended to velarize unstressed vowels, especially /ɑ/ and /i/. Her /l/-s were realized as velar [ʟ]-s in most environments. HK produced /d/ as a tap or flap [ɾ]. The sibilant /s/ had many variants: [ʃ], [θ], [s̪], and even [f], and this variation was not found to be rule-governed but rather free.

UL, male, 65: The subject was hard-of-hearing until the age of 16 years, when he lost the rest of his hearing. He received a cochlear implant eight months before the experiment and hears some noises but cannot quite comprehend speech. UL used the variants [θ], [sʲ], and a frontal [s̪] in free variation in spite of the sound [s]. In the beginning of a word, [b] became voiceless, but surprisingly, a word-starting /p/ was realized as [b]. In this data, UL's /d/ was always realized as [ɾ] and his /æ/-s were lowered to [ɑ]. The productions of /e/ and /æ/ tended to be confused.

KK, female, 65: The subject lost hearing almost completely at the age of 12. KK's word-final [ɑ] and [æ] had a tendency to be retracted, velarized or even pharyngealized. KK realized /s/ as [ʃ] after [o] or adjacent to [u] and as [θ] in other environments. Her productions of /m/ were quite tense with [b]-like releases and tended to spread over the following vowel. For KK's diphthongs, there was regularity in the sense that either component of closing diphthongs tended to be retracted, whereas in opening diphthongs there was a tendency of centralization.

PK, male, 51: PK was the only congenitally hearing-impaired subject. PK is able to hear some speech, but mainly reads from lips. PK tended to produce all voiced sounds with a creaky voice quality. In general, PK's speech was not very intelligible. There was a lot of variation in the prosodic features of his speech: especially the components of diphthongs seemed to be too separate. Particularly in vowels of long quantity, nasality could often be heard. In general, the subject's vowels had a tendency of strong centralization and often also velarization. There was a tendency of confusion between the realizations of /e/ and /æ/. In combinations of two plosives, eg., [t] and [k], the sounds had separate releases. In the beginning of the word, /m/ was realized as [b] or [mb], and sometimes this phenomenon also occurred in other

environments. [æ] was often replaced by [ɑ] and [ʋ] by [β]. PK usually added a schwa vowel after [l], [r], and all nasal consonants (cf. Hudgins, Numbers 1942). He also used nasal consonants as variants of /l/.

In general, the centralization of vowels was quite common in the data. One of the more stable vowels was /y/, which was usually produced quite accurately even by the less intelligible subjects (cf. Aaltonen, Suonpää 1982). It was sometimes observed that the subjects had difficulties in controlling the relative durations of sounds. However, this seemed to affect the intelligibility of the speech only in a few cases.

Discussion

Since the variability between subjects was considerable, a general summary of all subjects turned out to be difficult to build. This supports previous results (e.g., McGarr, Löfqvist 1982). It must be stated that, although some of the sound errors produced by the eight hearing-impaired subjects were regular within an individual, general trends for all subjects are rather difficult to identify. Another problem was that in a few cases the background noise of the recording prevented an accurate transcription.

A tendency of vowel centralization was quite evident in the data and cannot be accounted for by referring to normal reduction, because the centralization effect seemed to occur even in stressed syllables. This could have been expected on the basis of previous studies (Shukla 1989; Subtelny, Whitehead, Samar 1992). The vowels [e] and [æ] tended to get confused with each other in the productions of five subjects, which is well in accordance with O. Aaltonen and J. Suonpää's results.

No clear correspondence between the amount of deviant articulations and the subject's age or the time elapsed from the loss of hearing could be found on the basis of this data. Subject HN was quite fluent both prosodically and articulatorily, although he had been completely deaf for 32 years. Several other subjects were almost totally intelligible even after more than 30 years of severe hearing impairment. However, the age at which the subject had become severely hearing-impaired (or started to lose hearing) seems to be of some significance: the speakers with sound qualities closest to normal, i.e., HN and PP, had lost their hearing as late as at the age of 19 and 21 years. As expected, the subject PK, who had been almost completely deaf from birth, had the most difficulties in his speech. It should be noted that no information about the speech training (e.g., speech therapy) histories of the subjects is available at this point. Thus, the subjects may have had different amounts of practice and may therefore not be comparable with each other. Moreover, no audiograms or other clinical evaluations of the subjects' degree of hearing impairment were obtained, so that nothing more than preliminary and suggestive conclusions can be drawn here.

The velarization and/or palatalization of sounds were typical for four of the subjects. This result is in accordance with previous studies (see, e.g., Subtelny, Whitehead, Samar 1992). These effects may have something to do with the audibility of the sounds. One sometimes hears the claim that the prominent feature in the speech of hearing-impaired individuals is an overall nasality. In the present data, however, there was no evidence of any strong tendency to this direction, except possibly in the case of subject PK. It is possible that nasality has been confused with the velarization and/or palatalization effects, which were much more common. The lack of auditorily perceivable nasality may also be partly due to noisy signals and the existence of this feature needs to be studied acoustically.

The general question raised by studies of the speech of the hearing-impaired is, which properties of speech can be considered as "normal". It must not be regarded as self-evident that all the errors made by hearing-impaired subjects occur in deviant speech alone. There are a great deal of reduced forms and natural variability in the speech of normal-hearing individuals as well. Answering this question requires a comparison with normal-hearing speakers, more spoken material, some acoustic analyses, and statistical analyses that can only be made from segmented and labelled speech samples.

REFERENCES

Aaltonen, O., Suonpää, J. 1982, Cepstrum Analysis of Vowels Produced by Profoundly Hearing-Impaired Speakers of Finnish. — University of Jyväskylä, Institute of Finnish Language and Communication, Papers in Speech Research 30, Jyväskylä, 168—180.

Hudgins, C., Numbers, F. 1942, An Investigation of the Intelligibility of the Speech of the Deaf. — Genetic Psychology Monographs 25, 289—392.

Lane, H., Webster, J. 1991, Speech Deterioration in Postlingually Deafened Adults. — Journal of the Acoustical Society of America 89, 859—866.

McCaffrey, H., Sussman, H. 1994, An Investigation of Vowel Organization in Speakers With Severe and Profound Hearing Loss. — Journal of Speech and Hearing Research 37, 938—951.

McGarr, N., Löfqvist, A. 1982, Obstruent Production by Hearing-Impaired Speakers. Interarticulator Timing and Acoustics. — Journal of the Acoustical Society of America 72, 34—42.

—— 1988, Laryngeal Kinematics in Voiceless Obstruents Produced by Hearing-Impaired Speakers. — Journal of Speech and Hearing Research 31, 234—239.

Shukla, R. 1989, Phonological Space in the Speech of the Hearing-Impaired. — Journal of Communication Disorders 22, 317—325.

Subtelny, J., Whitehead, R., Samar, V. 1992, Spectral Study of Deviant Resonance in the Speech of Women Who Are Deaf. — Journal of Speech and Hearing Research 35, 574—579.

Waldstein, R. 1990, Effects of Postlingual Deafness on Speech Production. Implications for the Role of Auditory Feedback. — Journal of the Acoustical Society of America 88, 2099—2114.