Acta Linguistica Hungarica, Vol. 57 (2–3), pp. 239–267 (2010)
DOI: 10.1556/ALing.57.2010.2–3.4

VARIABILITY IN THE ARTICULATION AND PERCEPTION OF A WORD*

MÁRIA GÓSY

Phonetics Department, Research Institute for Linguistics
Hungarian Academy of Sciences
Benczúr u. 33.
H–1068 Budapest
Hungary
gosy@nytud.hu

Abstract: The words making up a speaker's mental lexicon may be stored as abstract phonological representations or else they may be stored as detailed acoustic-phonetic representations. The speaker's articulatory gestures intended to represent a word show relatively high variability in spontaneous speech. The aim of this paper is to explore the acoustic-phonetic patterns of the Hungarian word *akkor* 'then, at that time'. Ten speakers' recorded spontaneous speech with a total duration of 255 minutes and containing 286 occurrences of *akkor* were submitted to analysis. Durational and frequency patterns were measured by means of the Praat software. The results obtained show higher variability both within and across speakers than it had been expected. Both the durations of the words and those of the speech sounds, as well as the vowel formants, turned out to significantly differ across speakers. In addition, the results showed considerable within-speaker variation as well. The correspondence between variability in the objective acoustic-phonetic data and the flexibility and adaptive nature of the mental representation of a word will be discussed.

For the perception experiments, two speakers of the previous experiment were selected whose 48 words were then used as speech material. The listeners had to judge the quality of the words they heard using a five-point scale. The results confirmed that the listeners used diverse strategies and representations depending on the acoustic-phonetic parameters of the series of occurrences of *akkor*.

Keywords: word pronunciation variability, across-speaker differences, durational patterns, formant structures, perception judgments of words

* This research was supported by the Hungarian National Scientific Research Fund (OTKA), project No. 78315.

1. Introduction

Mental representations of linguistic forms contain relevant aspects of the individual's patterns of language knowledge. Certain parts of those mental representations may keep changing or being modified due to diverse factors. The mental representations of individual words can be defined in various ways, including their semantics, grammatical form, as well as phonological and phonetic structures. The difficulty of the exact definition of their mental representations is aggravated by a mismatch between various (phonological, morphological, lexical, and semantic) definitions of the notion of 'word' itself (cf. Zwitserlood 2003; Kenesei 2007). Meaning is supposed to be mentally encoded by humans (Jackendoff 2002). The idea of image schemas (Johnson 1987) seems to be useful for researchers to develop their own hypotheses of how concepts or 'conceptual units' are structured in the mind (Grady 2005). Words are assumed to be stored in the mind either as abstract phonological representations or as detailed acoustic-phonetic representations. (For simplicity, the term 'word' will be used in this paper instead of word form, lexeme, or phonological/morphological word.) Libben and Jarema (2002, 8) claim that "mental representations are metaphors that allow us to capture the nature of lexical knowledge and to test hypotheses of how this knowledge is acquired, organized, employed , and manifested in language breakdown".

The speaker's articulatory gestures (of pronouncing speech sounds) intended to represent a word show relatively high variability within and across the phoneme categories, particularly in spontaneous speech (Rose 1999; Keating et al. 1994; Gósy 2002). Speakers vary the amount of overarticulation vs. underarticulation (cf. Lindblom 1986) they exhibit from time to time but the between-speaker differences have been shown to be greater than within-speaker deviations (Dankovičová–Nolan 1999). The differences among speakers lie in the anatomy of their vocal tracts and vocal cords, their speech characteristics, individual articulatory behaviors, and so on. All these factors are manifested as variations in the speech signal. The fact that repetitions of an utterance, even by the same speaker and on the same occasion, are never exactly the same is called a phonetic truism by Rose (1999). The within-speaker differences lie in speaking style, speech rate, the speaker's physical and mental health, and so on (Krause–Braida 2004). Various phonetic factors in the speech signal may obscure the speaker's production while it still corresponds to the mental representation of the intended phonological word. The English word *the*,

for example, has been found to vary in the exact pronunciation of its vowel (five different vowel qualities were found for the same phoneme), in vowel duration, and in the presence or absence of a glottal stop or laryngealization at the end of the word during reading isolated sentences (Keating et al. 1994).

The variability of articulatory gestures in pronunciation has received a lot of attention in the literature: words in narratives and dialogues, as well as in repetition tasks have been investigated (e.g., Clark–Wasow 1998; Kohler 2000; Krause–Braida 2004; Hazan–Markham 2004; Pluymaekers et al. 2005; Horga 2008). However, variability of pronunciation does not result in frequent misperceptions, a fact that is generally explained primarily by linguistic context, listener's predictions and the informational redundancy of speech. The normalization process both across and within speakers is based on the continuous acoustic waveforms of speech (and of course on shared language knowledge) and results, ideally, in identical representation both for speaker and listener (Nusbaum–Magnuson 1997).

Researchers seem to know the location of some parts of mental images in the human brain (e.g., Dodge–Lakoff 2005; Mildner 2007; Pulvermüller 2007) but they are uncertain about the nature of the mental representations of words in terms of neuronal activities. The earlier term 'neural spectrogram' was used to refer to the correspondence between the acoustic-phonetic properties of speech sounds and their mental representation. As Bishop (1997, 4) writes, "the brain thus maps sound into a neural representation that contains crucial information about the amount of energy in different frequency bands and its rate of change, a so-called neural spectrogram". The basic question is, however, whether the mental representation of a word is a stable phenomenon with only some invariant features or, on the contrary, it is an extremely flexible phenomenon providing control over mapping and selection of the intended word both in articulation and in perception. It seems to be very unlikely that there is a single representation for any given word in the mental lexicon. In addition, there is evidence in support of representation units other than the phoneme (Greenberg 2006). It is widely known that there is no one-to-one correspondence between the acoustic signal and the word's phonological structure. Stevens's (1972) quantal theory claims that the listener is not sensitive to relatively small changes in the speech acoustics. These small changes seem to remain hidden for the listener, and this fact will ensure the correct perception—say of the same word—despite the different

acoustic outcomes of the speakers' pronunciations. What qualifies as a "small" acoustic change differs from language to language. Other theories posit direct mapping from an acoustic representation of the input signal to lexical representations (Andruski et al. 1994; McQueen–Cutler 2002; Pulvermüller 2005).

When the speaker intends to target the idealized articulation of a word s/he accesses its supposed mental representation. Similarly, the listener tries to match the incoming acoustic signal with the supposed idealized mental representation of the word. Since the speaker's articulation is heavily variable, particularly in spontaneous speech, the mental representation should be flexible. There is a growing demand to define the interrelations between linguistics and neuroscience. Poeppel and Embick (2005) relate distinctive features to dendrites, syllables to neurons, and morphemes to cell-assemblies. They hypothesize that two single neurons merging into a cell-assembly are responsible for the word image. One of them contains its phonological structure while the other one contains its meaning. The co-activated neurons develop into cell assemblies. The motor and acoustic representations of a word form are not separate; they are strongly connected so that they form a distributed functional unit (Pulvermüller 1999). The information of these two neurons will result in the mental image of a word like a special hologram.[1] Karl Pribram presented his hologram theory for neuronal storage of memories as early as in the sixties (cf. Pribram 1991). Various pronunciations of the same word may be represented by overlapping cell assemblies, that is, by two word cell assemblies sharing the same phonological structure. The systems responsible for lexical access in comprehension and for lexical retrieval in speech production are claimed to be separate systems (McQueen 2005). Although the topic of this paper is an acoustic-phonetic analysis of a word, the current views on mental representations are important to consider.

These theoretical claims raise further questions. Do speakers (mostly unconsciously) rely on the contextual predictability of word identification in spontaneous speech? Is there any conscious or unconscious control over the articulation gestures in the pronunciation of a word? Does phonologically induced perceptual correction ensure correspondence between

[1] The hologram is an apt metaphor of lexical representations in the brain. A hologram does not record single points of a picture but records the traces of the laser light that scans the object. Each point of the hologram corresponds to many points of the picture, similarly to the assumed word structures of the mental lexicon. The memory traces—and also various memory traces of the language—can be recorded in several neuron-assemblies.

the sound sequence and the mental representation of the word in spontaneous speech? Theoretically, speakers reduce their articulatory efforts spent on words that are predictable for the listener (Lindblom 1990). Listeners are assumed to be able to modify their temporal analysis and their frequency and intensity filtering mechanism in order to normalize the incoming acoustic signal.

The pronunciation variability of words in spontaneous speech is explained primarily by the speech production processes that precede articulation. When speakers cannot formulate an utterance properly at once, they may suspend their speech and insert either a pause or a filler before continuing (e.g., Levelt 1983; Shriberg 2001; Fox Tree–Schrock 2002; Horga 2008). Fillers have the advantage that, in a sense, they do not interrupt the speech flow (as do silent or filled pauses) and are not as conspicuous for the listeners as pauses are (Clark–Wasow 1998; Gósy–Horváth 2008). We hypothesize that speakers' articulatory gestures are considerably less controlled when pronouncing a filler word since their speech planning process is simultaneously engaged in another task; for example, in looking for the next intended word in the mental lexicon. On the other hand, the frequent use of a word may result in more automatic articulatory gestures and this might even reduce the variability of its pronunciation. Filler words being frequent in spontaneous speech offer an opportunity to analyze their variability in articulation.

The aim of this paper is to explore the acoustic-phonetic patterns of the Hungarian word *akkor* /ˈɔkːor/ in both of its functions: as an adverbial pronoun meaning 'then, at that time' and as a filler. This disyllabic word is frequently used in younger speakers' spontaneous speech, particularly in its filler function. It is assumed that this word shows high variability in articulation because of its frequency in spontaneous speech (cf. Bybee 2003). In other words: the frequent use of this word does not necessarily lead to a more stable pronunciation.

Our four main hypotheses cover both acoustic-phonetic and perceptual aspects of the analysis: (i) the acoustic-phonetic patterns show evidence for an extremely flexible and adaptive mental image of the word; (ii) the acoustic-phonetic patterns show considerable differences both within and across speakers; (iii) there are a few invariant features that constitute the interface between the speech sound sequence and the phonological structure of the word; and finally, (iv) listeners use a number of different strategies when decoding the same word represented by diverse acoustic structures.

2. Material, subjects, method

2.1. We analyzed the acoustic-phonetic consequences of the pronunciation of the single Hungarian word *akkor* /ˈɔkːor/. The original meaning of the word is 'then, at that time' but it can be used either as an adverbial pronoun or as a filler. Ten speakers (5 females and 5 males) from BEA, the Hungarian Spontaneous Speech Corpus (cf. Gósy 2008b)[2] were randomly selected (only their ages were controlled for). They were young native monolingual adult speakers of Hungarian (ages ranging from 22 to 28). All of them lived in Budapest, spoke the standard dialect and had no speech defects of any kind. The Hungarian Spontaneous Speech Corpus has been designed to record the state of present-day spoken Hungarian in the period starting in 2007 by collecting large amounts of recorded spontaneous speech produced by various speakers in Budapest. Each subject was recorded in a sound-attenuated room using a unidirectional high-quality microphone and a digital recorder connected to a computer. The recording environment and the technical facilities were the same in all cases (Gósy 2008b).

A sample of recorded spontaneous speech (narratives and dialogues), with a total duration of 255 minutes (4.25 hours), was submitted to analysis (136 minutes with female speakers and 119 minutes with males). The topics of the narratives were related to the subjects' work, family and hobbies on the one hand and a selected topic of current interest relevant to the subjects' age and everyday lives (e.g., changes in higher education, protection of animals by law, entertainment of young people, and so on) on the other.

The material selected contained 286 occurrences of the word *akkor*, half of them from males and another half from females. All the words *akkor* were analyzed that occurred in the narratives and dialogues independently of their meaning or function in the given context. The tokens that had no final /r/ were excluded from the analysis (there were only 14 occurrences) in order to have all the four speech sounds in the analyzed words. (Eight tokens of those that had no final /r/ were adverbial pronouns while six of them were fillers. They were followed by a consonant in 57.14% of the cases.) 34 occurrences (11.88%) had the meaning of the adverbial pronoun ('then') while the remaining 252 showed the

[2] For details with respect to the BEA Corpus, see www.nytud.hu/dbases/bea/index.html.

function of a filler. The phonetic context of *akkor* did not show large differences. It occurred after the word *és* /eːʃ/ 'and' in 41.25% and after a pause in 37.06% of all cases. The conjunctions *mert* /mɛrt/ 'because' or *tehát* /tɛhaːt/ 'that is' preceded it in 13.28% of all cases (the remaining 8.41% contained 3 different vowels and 2 different consonants preceding *akkor*). The words occurring after *akkor* had an initial consonant (various types) in 46.85% while they had an initial vowel (various types) in 42.65%. Pauses occurred after the target word in 10.48% of all cases.

The digital recordings were submitted to acoustic-phonetic analysis (Praat 4.2: Boersma–Weenink 2005) using a 44.1 kHz sampling rate with 16-bit resolution. The duration of the words, of the vowels, of the intervocalic velar stops and the VOT of the [kː]'s were measured in order to obtain information about their temporal patterns. The frequency values of the first two formants of the vowels [ɔ] and [o] and the frequency of the burst were also measured. The duration of the words was defined as the interval either from the first glottal pulse or the second formant onset of the first vowel (depending on the preceding sound) to the last glottal pulse of the trill. The duration of the vowels was measured between the first and last glottal pulses of the vowels while the duration of the stops was measured from the last glottal pulse of the preceding vowel to the first glottal pulse of the following vowel. The VOT of the stops was measured as the interval between the beginning of the release and the first pulse of the following vowel. The duration of [r] was measured from the last glottal pulse of the preceding vowel to the last glottal pulse of the trill. The corresponding spectrographic, intensity and waveform displays were consulted when segmenting the words and the speech sounds of the words, and auditory perception was also considered during this process.

The formant values were measured at the midpoint of total vowel duration. The F1 and F2 midpoints were determined by visual inspection using wideband spectrograms. For the burst, the measured value was the frequency of the highest-amplitude peak below 4 kHz. In sum, 11 parameters were analyzed for each token (total word duration, duration of [ɔ, kː, o, r], two formants for the vowels, burst frequency and VOT for the stop), yielding a total of more than 3,000 measurements.

2.2. For the perception test, *akkor* tokens as pronounced by two speakers, one female and one male, were selected from the recording described in the previous section. The first 24 tokens were used from both speakers' material (including also those words in this case that had no final [r]). A carrier sentence was selected where the word *akkor* occurred in the given

speaker's spontaneous speech. The female speaker's carrier sentence was: *ha túlszárnyalod saját magad* ***akkor*** *az csak pozitívum* 'if you outstrip yourself **then** so much the better'. The male speaker's carrier sentence was: *ötéves koromban kezdtem el hegedülni és* ***akkor*** *megszerettem a hegedűt* 'I started to learn playing the violin when I was five years old and **then** I came to like the violin'. All the *akkor* tokens were carefully extracted from their original context and inserted into the carrier sentence. 24 virtual sentences were thus created (for each speaker) and recorded in a random order with 4-second pauses between each pair of sentences.

University students of Budapest (ages between 20 and 22) participated in the experiment in 8 groups. Each group contained 10 or 11 listeners. 4 groups of listeners—altogether 42 subjects—listened to the female speaker's sentences while another 4 groups of listeners—altogether 40 subjects—listened to the male speaker's sentences. Their task was to judge the pronunciation of each instance of *akkor* they heard using a 5-point scale where point 1 meant 'incomprehensible' and point 5 meant 'excellently comprehensible'. The listeners got an answer sheet where there was a clear description of their task with a written example. Their task was also explained in a spoken form. Two sentences served for the "warming up" procedure. These sentences were not part of the experimental material. The 5-point scale corresponded to Hungarian school marks in order to help the subjects to grade the test words. The listeners' attention was drawn to the difficulty of the task in that the same test word would be heard several times. The carrier sentences, however, helped them to focus on the test word. There was a 3-minute break after the 12th virtual sentence in each group. The perceptual test lasted about 10 minutes in each group together with the task explanation.

To test statistical significance, analysis of variance (ANOVA), *t*-tests, correlation analysis and a linear regression analysis were used (SPSS, version 14.0). In all cases, the confidence level was set at the conventional 95%.

3. Results

3.1. Acoustic-phonetic properties of *akkor*

Subjects produced 1.12 *akkor* words per minute. The mean occurrence of *akkor* was 1.05 tokens per minute in female subjects' speech (min.: 19, max.: 35) and 1.2 tokens per minute in that of males (min.: 20, max.: 38). 88.12% of all occurrences were identified as fillers (no differentiation was

made in the analysis in terms of the diverse functions of the filler). The following examples show utterances containing *akkor* as an adverbial pronoun and as a filler.

In the function of an adverbial pronoun (□ = silent pause):

(1) (a) amikor leérettségiztem *akkor* □ műszaki rajzolónak mentem
'when I had finished high school *then* □ I became a draftsman'

(b) a könyvet én a nagynénémtől kaptam és *akkor* még a nyolcadikos írásommal beírtam neki valami ajánlást
'I got the book from my aunt and *then* I wrote some dedication for him in my childish writing'

In the function of a filler:

(2) (a) és minden évben így hívnak hova is Balaton nem Ábrahámháza igen oda és onnan kiindulunk és *akkor* így körbe így majdnem körbe a Balatonon
'I am invited every year where to Lake Balaton no to Ábrahámháza yes there and we start from there and *then* so round so almost round Lake Balaton'

(b) ott állt mellettem a srác és így *akkor* ilyen ragasztóval így firkálta össze az üveget
'the guy was standing next to me and so *then* with this adhesive so he was smudging the glass'

Figure 1 (overleaf) shows the acoustic structures of two pronunciations of *akkor* by the same female subject. The relatively large differences in the acoustic properties of the tokens can be clearly seen. There are considerable differences in the durational patterns of the vowels and the velar stops as well as in the intensity structures of the release bursts of the stops. The formants are radically different in both vowels, especially with the stressed vowels. Although both tokens contain the word-final [r], its acoustic manifestations are different: it is followed by a schwa in the first case while it is a vocalized realization of the phoneme in the other case.

The total duration of the word *akkor* shows enormous differences across speakers that were confirmed by statistical analysis (one-way ANOVA for total word duration: $F(9, 285) = 3.615$, $p = 0.001$). The word durations are more variable with females than with males (the shortest word is 136 ms while the longest one is 580 ms in the case of females and the shortest one is 144 ms and the longest one is 430 ms for males). There was no significant difference across genders, though. There was no great difference in the subjects's speech tempi, either (mean speech tempo for females was 160.4 words/min while the mean value for males was 157.6 words/min).

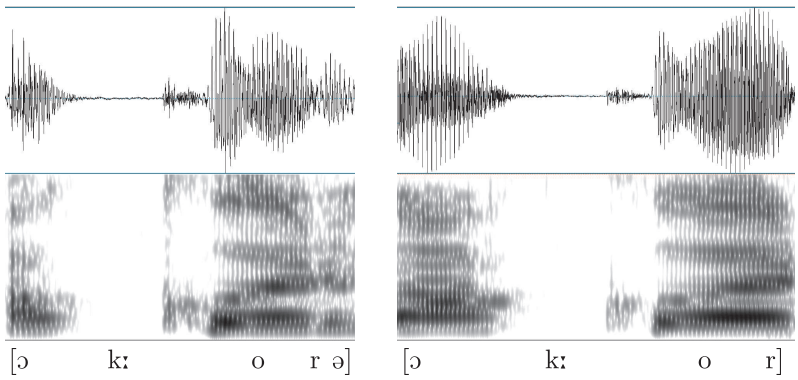

Fig. 1

Acoustic-phonetic properties of the word *akkor* pronounced by the same female speaker in two different contexts

Less variability of word durations had been expected within speakers than across speakers (cf. Dankovičová–Nolan 1999). The word duration values that we found contradicted this assumption: within-speaker variability is very similar to across-speaker variability (cf. Figure 2).

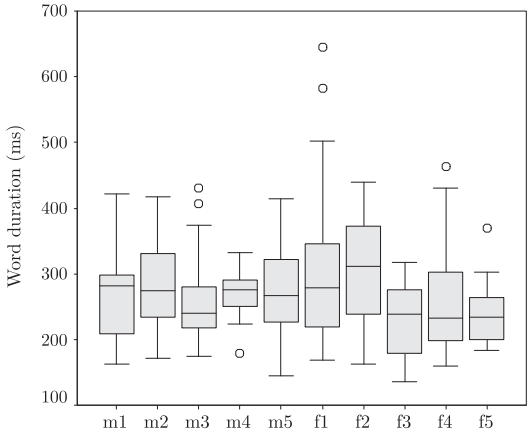

Fig. 2

Word duration values of *akkor* with all subjects, medians and ranges (m = male, f = female)

The durations of the two vowels of the words *akkor* are significantly different from each other with all subjects ($F(1, 571) = 4.368$, $p = 0.037$). The stressed vowel is longer than the unstressed vowel but the difference

is not large. Both vowels are significantly different across speakers (for [ɔ]: $F(9, 285) = 4.029$, $p = 0.001$ while for [o]: $F(9, 285) = 2.751$, $p = 0.004$), cf. Figures 3 and 4. The values of the velar stops significantly differ across speakers ($F(9, 285) = 8.478$, $p = 0.001$); however, the range of the durations is wider in females than in males. No significant difference was found in the case of the durations of word final trills (cf. Table 1).

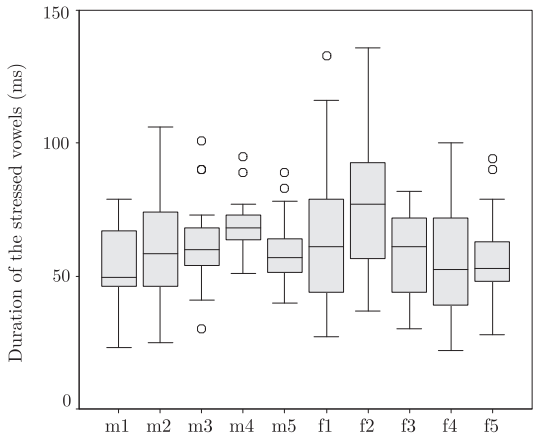

Fig. 3

Durations of the stressed vowel of *akkor* across speakers, medians and ranges (m = male, f = female)

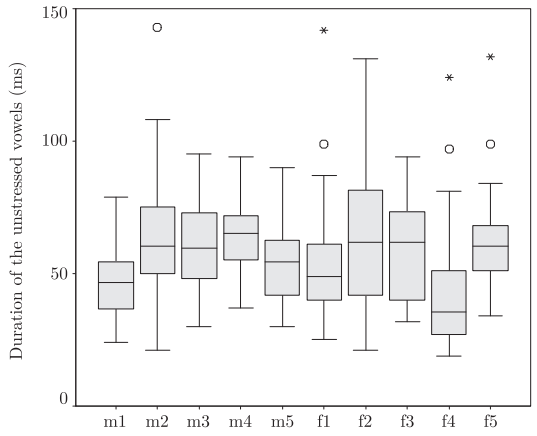

Fig. 4

Durations of the unstressed vowel of *akkor* across speakers, medians and ranges (m = male, f = female)

Table 1
Durational patterns of the word *akkor* (SD = standard deviation)

akkor	Durations (ms)					
	females		males		all subjects	
	mean	SD	mean	SD	mean	SD
word	278	109	271	60	274	88
[ɔ]	64	25	61	15	63	21
[o]	59	28	59	19	59	24
[kː]	120	47	110	27	115	38
VOT	48	16	34	11	41	14
[r]	34	29	38	27	36	28

The duration of the phonologically long voiceless stop ranged from 30 ms to 300 ms in our material. The difference based on gender turned out to be significant (one-way ANOVA: $F(1, 285) = 4.499$, $p = 0.035$). The durational values show large individual differences both within and across speakers (Figure 5). The velar stop /kː/ appears in a wider range of duration with the females than with the males. The data confirmed large differences both within and across speakers (cf. also Crystal–House 1990).

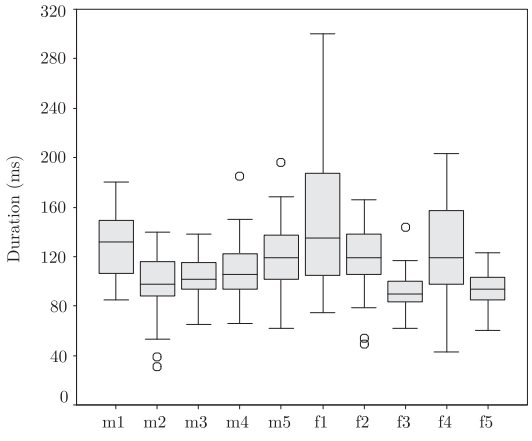

Fig. 5
The durations of the voiceless stop with females (f) and males (m) (medians and ranges)

Voice onset time (VOT) is generally assumed to be an invariant feature for the voiceless stops, including the velars as well. In Hungarian, the

mean VOT of phonologically short voiceless velars in females' spontaneous speech is 35.31 ms (Gósy 2001) but there is no data for either the long stops or for male pronunciation. The [kː]'s in the present material had a mean 41.03 ms of their VOTs which may be interrelated with the longer duration of the stop itself. There is no significant difference between females and males; however, a significant difference could be seen across speakers (one-way ANOVA: $F(9, 285) = 10.880$, $p = 0.001$). Figure 6 shows the medians and the ranges of VOT in each speaker, demonstrating the wide range of the values within speakers as well.

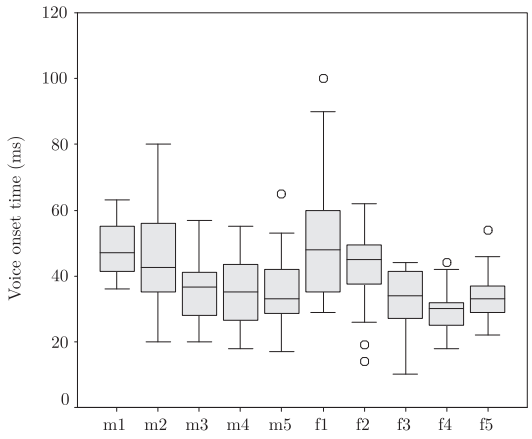

Fig. 6

The durational differences in the speakers' voice onset time of the [kː] (m = male, f = female), medians and ranges

The occurrence of the schwa after the trill (Vago–Gósy 2007) was not frequent, 9.44% of all trills (altogether 27 schwas with various occurrences of each speaker). There was no statistical difference in the durations of the trill based on gender. (The acoustic-phonetic properties of the phoneme /r/ will not be further analyzed here).

The analysis of the formants of the /ɔ/ realizations in females revealed that both F1 and F2 are significantly different across speakers (for F1: $F(4, 142) = 5.390$, $p = 0.001$ and for F2: $F(4, 142) = 7.352$, $p = 0.001$). The same results were found for their /o/ realizations (for F1: $F(4, 142) = 8.250$, $p = 0.001$ and for F2: $F(4, 142) = 6.068$, $p = 0.001$). The pronunciation of the same vowels in the word *akkor* by male subjects seems to be somewhat different, an auditory impression which is supported by the acoustic-phonetic correlates. Both first and second

formants of /ɔ/ showed significant differences across subjects (for F1: $F(4, 142) = 2.857$, $p = 0.026$ and for F2: $F(4, 142) = 13.691$, $p = 0.001$). However, there were no significant differences in the first formants of /o/. The second formants of /o/ realizations, however, showed significant differences across the male subjects ($F(4, 142) = 3.034$, $p = 0.020$). Table 2 summarizes the formant values characteristic of the two vowels in the realizations of the word *akkor*.

Table 2

Formant frequencies of the vowels in the word *akkor* (SD = standard deviation)

	Formant frequency values (Hz)			
	females		males	
Formants	mean	SD	mean	SD
F1 [ɔ]	623	94	567	80
F2 [ɔ]	1445	219	1261	150
F1 [o]	546	58	477	45
F2 [o]	1299	199	1123	140

On the basis of formant frequency values it can be seen that females pronounced both vowels with higher across-speaker variability than our male subjects. The pronunciation of the latter did not show large differences concerning the articulation of the unstressed vowel. The formant structures of the males' [o] vowels are similar to, or even coincide with, those of the neutral vowel. This means that males tend to pronounce a schwa in the unstressed position of the word. Figures 7 and 8 demonstrate the F1/F2 patterns of the two vowels both for females and males. The frequency values show considerable scatter along the axes representing the first and second formants. The tokens representing the phonemes /ɔ/ and /o/ overlap the frequency space of other Hungarian vowels, including [aː, ɛ, ø, ə].

The stressed vowels' formants show a larger range than those of the unstressed vowels, particularly with females. This can be explained by the more frequent realizations of the unstressed vowels as the neutral schwa. This is confirmed by some other measurements for Hungarian as well (Gósy 2004; Beke–Szaszák 2010). Figure 9 shows the within-speaker formant values of the stressed vowel in *akkor* while Figure 10 displays the within-speaker formant values of the unstressed vowel in the same word.

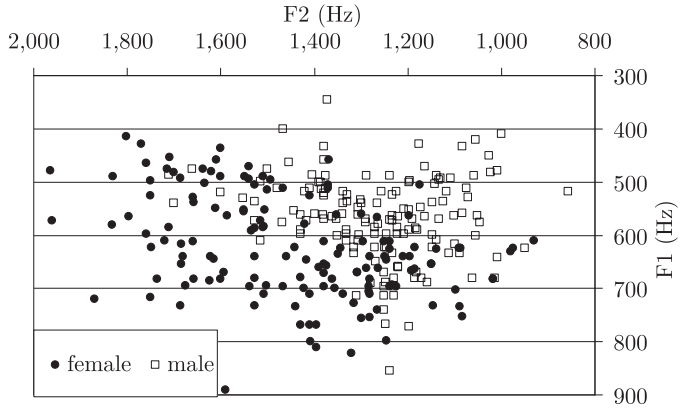

Fig. 7

The F1/F2 space of the realizations of the phoneme /ɔ/ in the word *akkor* (black circles represent the females' data while the squares represent the males' data)

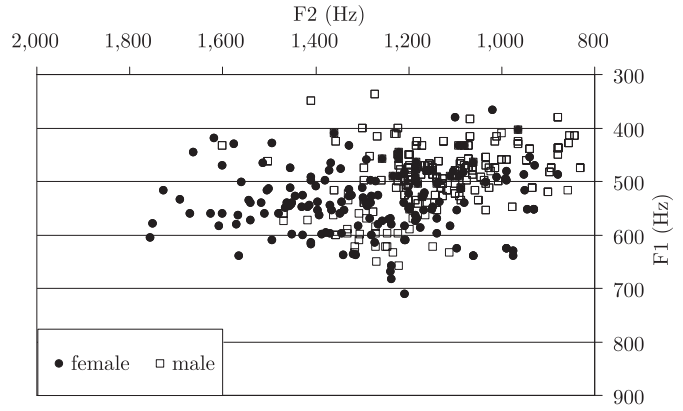

Fig. 8

The F1/F2 space of the realizations of the phoneme /o/ in the word *akkor* (black circles represent the females' data while the squares represent the males' data)

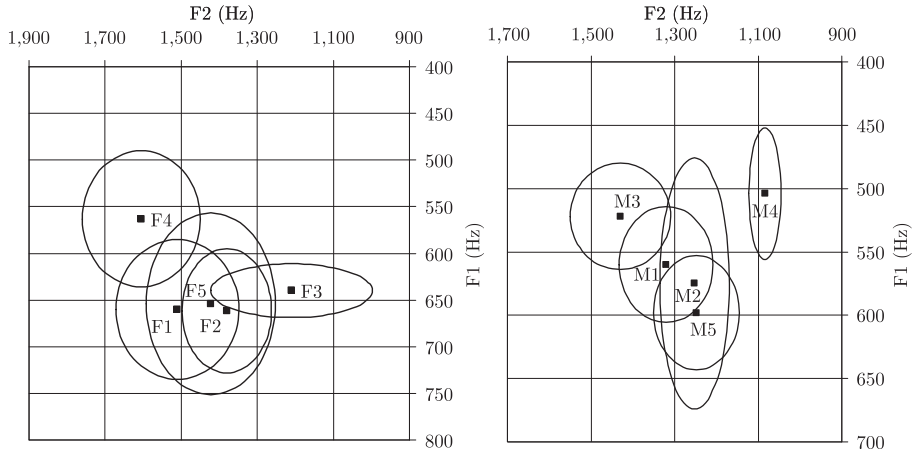

Fig. 9

The F1/F2 space of the individual realizations of the phoneme /ɔ/ in females (left) and males (right)

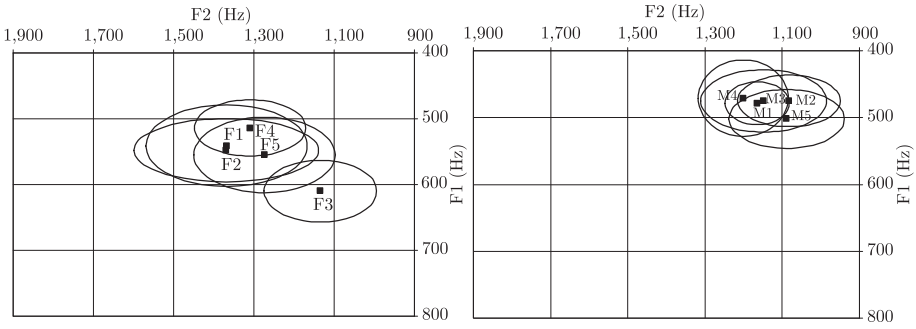

Fig. 10

The F1/F2 space of the individual realizations of the phoneme /o/ in females (left) and males (right)

The formant values seem to confirm two main facts. (i) The stressed vowels are realized in a wider range than the unstressed vowels. The values of the unstressed vowels show a tendency toward the schwa pronunciation. (ii) The second formants of the unstressed vowels are more scattered than those of their first formants. This means that tongue height is more variable than the horizontal position of the tongue. Context effects might be expected to explain the variability of the formants of the stressed vowels. However, the target word, *akkor* is preceded by a voiceless fricative ([ʃ]) or a pause in 78.31% of all instances in our material. Therefore, the

phonetic context effect explanation seems to be inadequate. Instead, we assume that the acoustic-phonetic variability of the word analyzed here can be explained by the active planning processes that are engaged in fulfilling other tasks while the speaker articulates the (mostly filler) word *akkor*. The speakers do not spend as much effort (and perhaps attention) on the proper articulation gestures of this word as it would be needed in order to result in similar pronunciations of all of these words. Accepting this fact we might also assume that various contextual perseveration and anticipation effects (across several syllables) influence the articulation.

Analysis was carried out concerning the burst frequencies of the velar stops. The mean value for the females turned out to be 1023.41 Hz (std. dev.: 136.37 Hz) while 850.93 Hz was the mean value (std. dev.: 117.60 Hz) for the male speakers. These data again confirmed both intra- and inter-speaker variability in the velar stop articulation (Figure 11). Statistical analysis showed significant difference depending on gender (one-way ANOVA: $F(1, 285) = 79.812$, $p = 0.001$). Although the release of a velar stop is influenced by the adjacent vowel in a CV context (Keating et al. 1994), this was not the case in our study since the [kː] always occurred in the same phonetic context. This means that the wide range of the burst frequency values can be attributed, in our case, to nothing but the speakers' diverse articulation habits.

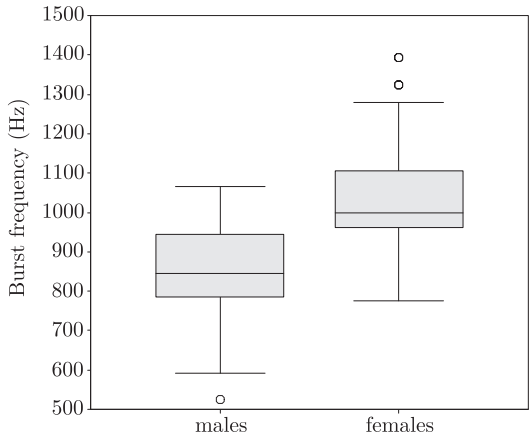

Fig. 11

Burst frequencies of the velar stops (medians and ranges)

The burst appears mainly once but 18.9% of all occurrences had two bursts in the analyzed intervocalic positions (Figure 12). They merged

in one velar burst in perception. There was no interrelation between the number of bursts and the duration of the stop consonant. The bursts seem to be characteristic of the speaker: there was a speaker with only one occurrence of two bursts while another speaker produced two bursts in 45.94% of all his stops.

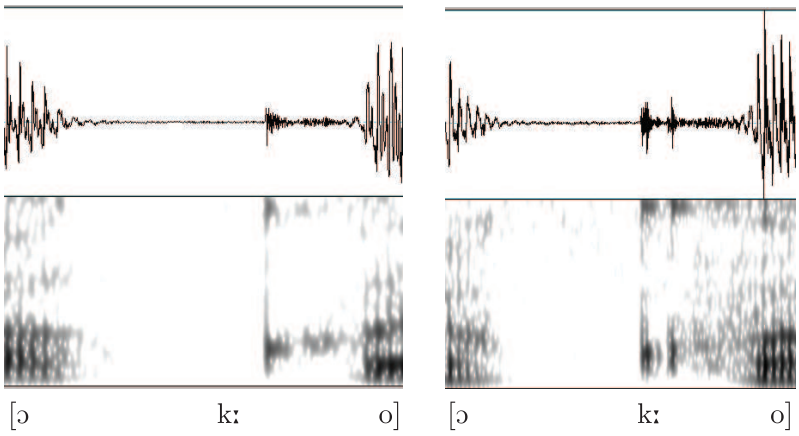

Fig. 12
The articulation of the voiceless velar stops with one or two bursts (female speaker). (The pictures contain parts of the preceding and the following vowels.)

3.2. Perceptual judgments of *akkor*

Both speakers show a wide variety of the articulation of the word *akkor* with respect to both the durational patterns and the pronunciation of the vowels and consonants the word consists of. Although there is no statistical difference between the female and the male subject's total word durations, the female speaker's realizations show a wider range of word durations than those of the male speaker (Figure 13). The shortest *akkor* in the female speaker's rendering is 170 ms while it is 148 ms in the male speaker's material. The longest word is 580 ms among the female and 360 ms among the male tokens. Table 3 summarizes the temporal data of the 48 test words.

There were no significant differences between the durations of the two vowels pronounced by either speaker. The duration differences of the [kː] and the [r] consonants pronounced by the selected female and male did not turn out to be significant, either. Our two subjects' voice onset time values, however, were significantly different (one-way ANOVA:

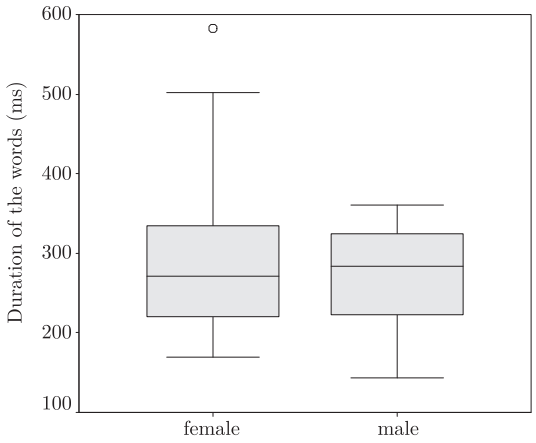

Fig. 13
The durations of the test words
with the two speakers in the perception experiments

Table 3
The descriptive data of the temporal patterns of the test words

akkor	Durational patterns (ms)			
	female speaker		male speaker	
	mean	SD	mean	SD
word	295	101	272	68
[ɔ]	64	25	58	14
[o]	54	29	55	17
[kː]	140	49	121	30
VOT	48	16	34	11
[r]	37	28	38	25

$F(1, 47) = 11.94$, $p = 0.002$). The voice onset time of the voiceless velar stop shows that the female's articulation strengthened the voiceless character of the consonant more than the male's articulation since the female's VOTs were longer than those of the male (Table 3). The longer the VOT-value, the more expressed the lack of the vocal fold vibration during the consonant articulation. The burst frequencies—as expected—significantly differed between the two speakers being higher in the female's case (mean: 1022.12 Hz, std. dev.: 182.15 Hz) while lower in the male's case

(mean: 813.95 Hz, std. dev.: 135.24 Hz). (The one-way ANOVA's result: $F(1, 47) = 24.435$, $p = 0.001$.)

The formant frequencies show a wide variety of the articulation of the two vowels, particularly in the case of the female speaker. Table 4 summarizes the mean and the standard deviation values of the vowel formants.

Table 4
The descriptive data of the formants of the test words

akkor	Formants (Hz)			
	female speaker		male speaker	
	mean	SD	mean	SD
F1 [ɔ]	656	89	577	55
F2 [ɔ]	1537	199	1245	120
F1 [o]	532	53	482	49
F2 [o]	1365	224	1126	164

Statistical analysis showed that there are significant differences between the first formants of the two vowels in the case of both the female and the male speaker (paired samples t-test for the former: $t(23) = 6.064$, $p = 0.001$; for the latter: $t(23) = 3.408$, $p = 0.002$). No such difference was found with the male speaker's second formant frequencies between the two vowels while the second formant frequencies turned out to be significantly different with the female speaker ($t(23) = 2.758$, $p = 0.011$). This means that the male's vowel distinction is based primarily on the vertical movement of the tongue resulting in the realizations of the /ɔ/ and the /o/ phonemes. The distinction of these two vowels is based on both the vertical and horizontal movements of the tongue in the case of the female speaker. The formant frequencies of the two vowels confirm a clearer articulation in the case of the female speaker. The data of the test words show a remarkable variety of the realizations of /ɔ/ and /o/ phonemes they contain. The unstressed vowels' formant frequencies overlap across genders to a certain extent. This, again, can be explained by the schwa realizations of the /o/ phonemes in both speakers.

The results of the perceptual test show that the male's words were judged to be better articulated than the female's words. The mean of the points awarded is 3.49 (std. dev.: 1.16 points) for the test words in

the male pronunciation while it is 2.95 points (std. dev.: 0.91 points) in the female pronunciation. There is no significant difference in the ratings between the two speakers. The question is what parameters led to the listeners' judgments.

In the first series of analyses, all test words were divided into two categories on the basis of the listeners' judgments. The first category was made up by test words that had been awarded 1 or 2 points while the second category was made up by ones that had been given 4 or 5 points. The first category means "incomprehensible or poor pronunciation" while the second category means "good pronunciation or excellently comprehensible". We analyzed the acoustic-phonetic parameters of all words falling either into the first or into the second category (no statistical analysis was made between the values of the two categories because of the relatively low number of instances).

(i) The male speaker's words were judged to be poorly pronounced when the mean value of their total word durations was 172.5 ms, the mean duration of the intervocalic stop was 79.5 ms and the mean duration of the [r] consonant was 12.3 ms. The mean burst frequency was 791.66 Hz. The mean frequency value of the first formants of the stressed vowel was 536.5 Hz while that of the F2 was 1260.16 Hz. The mean frequency value of the first formants of the unstressed vowels was 471.2 Hz while its F2 was 1451.5 Hz. The male speaker's words were judged to be pronounced excellently when the durations were longer than those in the first category. The mean duration of the words in this category was 315.5 ms, the mean duration of the intervocalic stop was 142.3 ms, and the mean duration of [r] was 57.9 ms. There is no difference in the burst frequencies, the mean value in the first category was 795.8 Hz. The frequency values of the stressed vowels show larger differences in the second formant values, their mean frequency was 1096 Hz (the mean value of the F1 was 593.9 Hz). Similarly, the second formants of the unstressed vowels differed between the words in the two categories; the mean of the F2 here was 1108.6 Hz (the first formants' mean value was 507.3 Hz).

(ii) The female speaker's words were judged to be poorly pronounced when the mean value of their total word durations was 252.5 ms, the mean duration of the intervocalic stop was 125.2 ms, and the mean duration of the [r] was 34.0 ms. The mean burst frequency was 1032.12 Hz. The mean value of the first formants of the stressed vowels was 611.87 Hz while its F2 was 1700.12 Hz. The mean values of the unstressed vowels in this category were 500.87 Hz for F1 and 1242.62 for F2. The female speaker's words were judged to be pronounced excellently when the durations of the words in this category were longer than those in the first category. The mean value of their total word durations was 489.6 ms, the mean duration of the intervocalic stop was 225.3 ms, and the mean duration of the [r] was 72.6 ms. The mean burst frequency is similar to that of the first category, here this was 974 Hz. The values of the second formants of the stressed vowels differed largely between the two categories. The mean value of F2 was 1348.6 Hz while the mean value of F1 was 705.3 Hz. The values of the first formants of the unstressed vowels were almost the same as in the first category (511.3 Hz) while the F2's mean value was lower, 1113.6 Hz in this category.

The listeners judge the quality of the words in terms of a total impression of all acoustic-phonetic parameters. Further statistical analysis was carried out to find out the interrelations of the phonetic properties of the words and listeners' judgments. High correlation was found between the total durations of the words and the listeners' quality judgments in the case of the male speaker (Pearson's test: $r = 0.742$ at 99% confidence level). Beside total word duration, the temporal patterns of the [o] ($r = 0.564$) and the [k] ($r = 0.710$), as well as the first formant value of the [o] ($r = 0.588$), are of crucial importance (in all cases the confidence level is 99%). To explain these data from the aspect of perception, it can be claimed that there are four parameters that affect the quality of the words in question. They are total word duration, the duration of the unstressed vowel, the duration of the intervocalic stop, and the F1 of the unstressed vowel. Although the duration of the [r] is important for total word duration, it is not significantly correlated with the listeners' quality judgments. The male speaker's stressed vowels did not show large differences in their F1/F2 parameters (mean value for F1 when the words were judged positively is 591.42 Hz and for F2 1241.71 Hz, while the F1 mean value is 536.5 Hz and F2 mean value is 1260.16 Hz when the words were judged negatively).

Judgments made on the basis of the female speaker's words show a different correlation pattern (all the values are at the 99% confidence level). Total word duration seems to be important again but the value of Pearson's rho is lower ($r = 0.571$) than in the male speaker's case. The correlation between the listeners' judgments and the acoustic-phonetic parameters of the female speaker's production shows that mainly three parameters affect the listeners' perception: total word duration ($r = 0.571$), the second formant frequency of the stressed vowel ($r = 0.755$) and the duration of the [k] ($r = 0.562$).

The larger the differences of the temporal structures of the words, the fewer parameters the listeners need to make their judgments. The duration of the female speaker's test words shows high variability as opposed to the male speaker's word durations. This might explain why three further parameters seem to play an important role in listeners' judgments in the case of the male speaker and only two further parameters were needed in the case of the female speaker.

There can be no doubt that the listeners' perceptual judgments are affected, more or less, by all parameters that the word they hear contains. The data show, however, that these parameters are changeable according

to the actual input (as the perceptual differences based on the female and the male pronunciations show). The question is whether the parameters that contribute more than others can be defined using the listeners' judgments (Picheny et al. 1985).

A linear regression model (cf. Tacq 1997) was used for this purpose (using SPSS software). The behavior of the dependent variable can be explained by some independent variables in this model. We have used the Stepwise method that provides an opportunity for all independent variables to be equally considered in the model. The results show that total word duration on its own accounts for 51% of the perceptual judgments concerning the female speaker's words while 68% of judgments are explained if both total word duration and the second formant value of [ɔ] are considered. If the model considers a third parameter—in this case, the first formant of the [ɔ]—, the explanation for the perception judgments hardly increases (by a mere 6%). So, the importance of the first formant value of the stressed vowel is slight.

The results in the case of the male speaker's words show that total word duration on its own accounts for 54% of the perceptual judgments while 66% are explained if both total word duration and the first formant value of [o] are considered. This means that these two acoustic-phonetic parameters of the test words are mainly responsible for their perceptual ratings though these two parameters are different from those in the female speaker's case. An interpolation was made by means of the equation of the linear regression in order to demonstrate the results. Figures 14 (in the case of the female speaker) and 15 (in the case of the male speaker) show the interrelations of the decisive parameters affecting the listeners' perception (overleaf).

These findings support the view that listeners are capable of accurate identification of the speech signal while ignoring speaker-induced variations. However, they also show that the human decoding mechanism is capable of using different mental representations of the same words depending on speaker-induced variations.

4. Conclusions

This study looked at the range of pronunciation variants of a frequent Hungarian word in a spontaneous speech corpus, and asked what factors enter into determining the variability of the analyzed words. Although some pronunciation differences had been expected, the results obtained

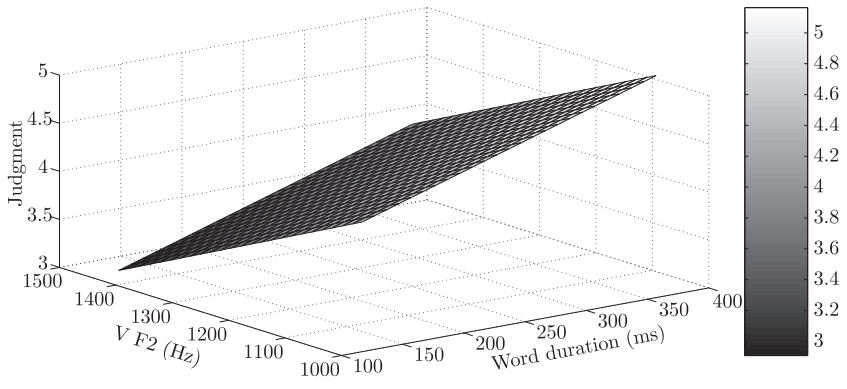

Fig. 14

Three-dimensional representation of the two decisive parameters (word duration and F2 of [ɔ]) affecting the listeners' judgments in the case of the female speaker. The lighter the space the more intelligible the word is

show unexpectedly high variability both within and across speakers in the case of this extremely frequent word. The data support our hypothesis concerning the flexibility and adaptive nature of the mental representation of a word. In addition, the production and perception results seem to support the hypothesis of word hologram as a theory of the lexical representations in the mind based on an assumed neuronal mechanism. Admittedly, this is speculative but the experimental findings seem to support the view.

The data obtained have confirmed that the frequent use of the analyzed word did not result in more automatic articulatory gestures and did not reduce the variability of its pronunciation.

As is well known, there is a great deal of redundancy in the speech signal, and therefore a multiplicity of acoustic parameters defines the phonetic identity of speech sounds (Scott 2005). The word *akkor* itself carries a sufficient amount of invariant features to map the acoustic signal onto the phonological word (as lexical representation) in the mental lexicon. What are these invariant features for *akkor*? The relatively stable voice onset time of the stop ensures that this consonant has the feature 'voiceless' (though the values are different between the females and the males). The intensive part of the release burst occurs in all cases far below 1500 Hz, which ensures the feature 'velar'. The closure part together with the low-frequency burst with one or two (and rarely more) releases is characteristic of these stops in Hungarian. The majority of the formant values of the stressed vowels point to a low back vowel. The word-final

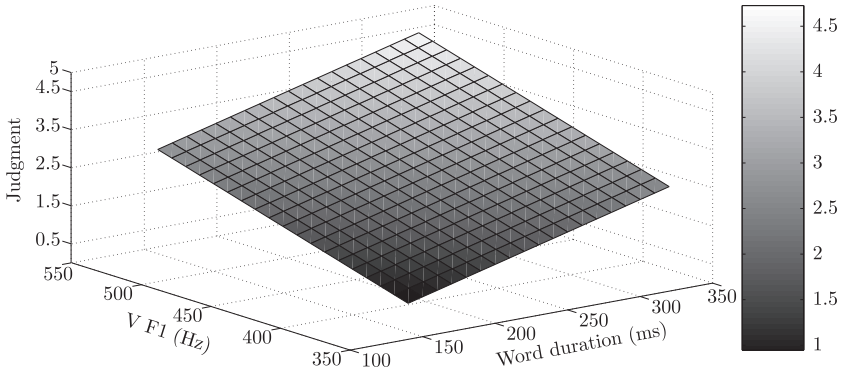

Fig. 15

Three-dimensional representation of the two decisive parameters (word duration and F1 of [o]) affecting the listeners' judgments in the case of the male speaker. The lighter the space the more intelligible the word is

/r/, independently of its actual realization (whether it is a trill escorted by a schwa or a vowel-like realization), narrows down the range of possible competing word forms in the mental lexicon.

The variability of speech as a consequence of diversity in pronunciation can be seen as a perceptual problem to be solved by listeners. In this experiment, listeners used different parameters of words they heard in order to make their judgments on their quality. This means that there are no general factors affecting the perceptual judgments. The actual judgments are based on the listeners' specific perceptual strategies that are flexible enough to be able to consider the demands of the properties of the incoming speech signal. If all speakers spoke in the same way, the perception mechanism would not need to be flexible (McQueen et al. 2006). The flexibility of the perception mechanism means that all traces are activated in proportion to the familiarity to the incoming speech signal (Goldinger 1997). There must be a process of comparison going on in the perception mechanism that maps the stored lexical representation and the present acoustic-phonetic parameters. This mapping procedure considers the differences of the individual pronunciations (and employs the theory of more than one word forms that are temporarily activated by the speech input). For intelligibility, however, what is important is not how specific acoustic-phonetic characteristics are produced by a speaker but the degree of internal consistency in the production of speech sounds (cf. Hazan–Markham 2004).

The spoken sound sequence representing /ˈɔkːor/—in terms of acoustic-phonetic patterns—looks like a puzzle: [ɔ] or [o] or [ə] + [k] or [kː] +

[ə] or [ø] or [o] + [r]. Obviously, the success of mapping between the acoustic signal and the phonological form of the word in question (from the listener's perspective) depends on the number of possible competitors the language offers, on the context in which the word occurs and on the guessing strategy of the listener. For illustration, here are some Hungarian words sharing a similar phonetic structure: *akar* [ˈɔkɔr] 's/he wants', *ókor* [ˈoːkor] 'antiquity', *ökör* [ˈøkør] 'ox', *a kör* [ɔˈkør] 'the circle', *a kór* [ɔˈkoːr] 'the illness', *a kor* [ɔˈkor] 'the age', etc. The listeners in this experiment used different strategies based on different cues in their perception to make judgments on the quality of the test words (cf. Boothroyd–Nittrouer 1988). The word hologram theory seems to provide an acceptable explanation for the quick operations and decisions among lexical representations in one's mental lexicon (which, of course, needs further experimental neurolinguistic support).

The present results support the view that speakers exercise control over their articulatory gestures. Even in cases where they can be assumed to pay no conscious attention to the pronunciation of a word and therefore there is considerable variability in articulation, important invariant cues remain available for the word as a gestural unit.

The results of such phonetic research could prove useful for psycholinguistic aspects of the study of the mental lexicon, as well as for text-to-speech and speech recognition systems. The implication of the results for speech recognition, for instance, is that all frequent variants of frequent words like *akkor* need to be included in a recognition lexicon in order to achieve a better recognition result.

References

Andruski, Jean E. – Sheila E. Blumstein – Martha W. Burton 1994. The effect of subphonetic differences on lexical access. In: Cognition 52 : 163–87.

Beke, András – György Szaszák 2010. Automatic recognition of schwa variants in spontaneous Hungarian speech. In: Acta Linguistica Hungarica 57 : 329–53.

Bishop, Dorothy V. M. 1997. Uncommon understanding. Development and disorders of language comprehension in children. Psychology Press, London.

Boersma, Paul – David Weenink 2005. Praat: Doing phonetics by computer. (Version 4.2) [Computer program]. Retrieved March 12, 2005, from http://www.praat.org/.

Boothroyd, Arthur – Susan Nittrouer 1988. Mathematical treatment of context effects in phoneme and word recognition. In: Journal of the Acoustical Society of America 84 : 101–14.

Bybee, Joan 2003. Mechanisms of change in grammaticalization: The role of frequency. In: Brian D. Joseph – Richard D. Janda (eds): The handbook of historical linguistics, 602–23. Blackwell, Malden MA & Oxford.

Clark, Herbert H. – Thomas Wasow 1998. Repeating words in spontaneous speech. In: Cognitive Psychology 37 : 201–42.

Crystal, Thomas H. – Arthur S. House 1990. Articulation rate and the duration of syllables and stress groups in connected speech. In: Journal of the Acoustical Society of America 88 : 101–12.

Cutler, Anne (ed.) 2005. Twenty-first century psycholinguistics: Four cornerstones. Lawrence Erlbaum, Mahwah NJ.

Dankovičová, Jana – Francis Nolan 1999. Some acoustic effects of speaking style on utterances for automatic speaker verification. In: Journal of the International Phonetic Association 29 : 115–229.

Dodge, Ellen – George Lakoff 2005. Image schemas: From linguistic analysis to neural grounding. In: Hampe (2005, 57–92).

Fox Tree, Jean E. – Josef C. Schrock 2002. Basic meanings of *you know* and *I mean*. In: Journal of Pragmatics 34 : 727–47.

Goldinger, Stephen D. 1997. Words and voices: Perception and production in an episodic lexicon. In: Johnson – Mullenix (1997, 33–66).

Gósy, Mária 2001. The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. In: International Journal of Speech Technology 4 : 75–85.

Gósy, Mária 2002. Long-term within-speaker and between-speaker differences in phonetic output: Evidence from Hungarian. In: Angelika Braun – Herbert R. Masthoff (eds): Phonetics and its applications. Festschrift for Jens-Peter Köster on the occasion of his 60th birthday, 75–85. Steiner, Stuttgart.

Gósy, Mária 2004. The manifold function of schwa. In: Grazer Linguistische Studien 62 : 15–26.

Gósy, Mária (ed.) 2008a. Beszédkutatás 2008 [Speech research 2008]. MTA Nyelvtudományi Intézet, Kempelen Farkas Beszédkutató Laboratórium, Budapest.

Gósy, Mária 2008b. Magyar spontánbeszéd-adatbázis – BEA [Hungarian spontaneous speech corpus – BEA]. In: Gósy (2008a, 194–207).

Gósy, Mária – Viktória Horváth 2008. Acoustic-phonetic analysis of two words on the way to becoming fillers. In: Rudolph Sock – Susanne Fuchs – Yves Laprie (eds): Proceedings of the 8th International Seminar on Speech Production 2008, 153–7. University of Strasbourg, Strasbourg.

Grady, Joseph E. 2005. Image schemas and perception: Refining a definition. In: Hampe (2005, 35–56).

Greenberg, Steven 2006. A multi-tier framework for understanding spoken language. In: Steven Greenberg – William A. Ainsworth (eds): Listening to speech. An auditory perspective, 411–34. Erlbaum, Mahwah NJ & London.

Hampe, Beate (ed.) 2005. From perception to meaning. image schemas in cognitive linguistics. Mouton de Gruyter, Berlin & New York.

Hazan, Valerie – Duncan Markham 2004. Acoustic-phonetic correlates of talker intelligibility for adults and children. In: Journal of the Acoustical Society of America 116 : 3108–18.

Horga, Damir 2008. Repetitions in interrupted speech production. In: Gósy (2008a, 157–71).

Jackendoff, Ray 2002. Foundations of language: Brain, meaning, grammar, evolution. Oxford University Press, Oxford.

Johnson, Keith – John W. Mullenix (eds) 1997. Talker variability in speech processing. Academic Press, San Diego.

Johnson, Mark 1987. The body in the mind: The bodily basis of meaning, imagination and reason. The University of Chicago Press, Chicago.

Keating, Patricia A. – Dani Byrd – Edward Flemming – Yuichi Todaka 1994. Phonetic analyses of word and segment variation using the TIMIT corpus of American English. In: Speech Communication 14 : 3–142.

Kenesei, István 2007. Semiwords and affixoids: The territory between word and affix. In: Acta Linguistica Hungarica 54 : 263–93.

Kohler, Klaus 2000. Investigating unscripted speech: Implications for phonetics and phonology. In: Phonetica 57 : 85–94.

Krause, Jean C. – Louis D. Braida 2004. Acoustic properties of naturally produced clear speech at normal speaking rates. In: Journal of the Acoustical Society of America 115 : 362–78.

Levelt, Willem. J. M. 1983. Monitoring and self-repair in speech. In: Cognition 33 : 41–103.

Libben, Gary – Gonia Jarema 2002. Mental lexicon research in the new millennium. In: Brain and Language 81 : 2–11.

Lindblom, Björn 1986. On the origin and purpose of discreteness and invariance in sound patterns. In: Joseph S. Perkell – Dennis H. Klatt (eds): Invariance and variability of speech, 493–510. Lawrence Erlbaum, Hillsdale NJ.

Lindblom, Björn 1990. Explaining phonetic variation: A sketch of the h&h theory. In: William J. Hardcastle – Alain Marchal (eds): Speech production and speech modeling, 403–40. Kluwer, Dordrecht.

McQueen, James M. 2005. Spoken word recognition and production: Regular but not inseparable bedfellows. In: Cutler (2005, 229–44).

McQueen, James M. – Anne Cutler (eds) 2002. Spoken word access processes. Special issue of Language and Cognitive Processes. Taylor and Francis, London.

McQueen, James M. – Dennis Norris – Anne Cutler 2006. The dynamic nature of speech perception. In: Language and Speech 49 : 101–12.

Mildner, Vesna 2007. The cognitive neuroscience of human communication. Lawrence Erlbaum, New York & Abingdon.

Nusbaum, Howard – James Magnuson 1997. Talker normalization: Phonetic constancy as a cognitive process. In: Johnson – Mullenix (1997, 109–32).

Picheny, Michael A. – Nathaniel I. Durlach – Louis D. Braida 1985. Speaking clearly for the hard of hearing 1. Intelligibility differences between clear and conversational speech. In: Journal of Speech and Hearing Research 28 : 96–103.

Pluymaekers, Mark – Mirjam Ernestus – Harald R. Baayen 2005. Articulatory planning is continuous and sensitive to informational redundancy. In: Phonetica 62 : 146–59.

Poeppel, David – David Embick 2005. Defining the relation between linguistics and neuroscience. In: Cutler (2005, 173–89).

Pribram, Karl H. 1991. Brain and perception: Holonomy and structure in figural processing. Lawrence Erlbaum, Hillsdale NJ.

Pulvermüller, Friedemann 1999. Words in the brain's language. In: Behavioral and Brain Sciences 22 : 253–336.

Pulvermüller, Friedemann 2005. Brain mechanisms linking language and action. In: Nature 6 : 576–82.

Pulvermüller, Friedemann 2007. Word processing in the brain as revealed by neurophysiological imaging using EEG and MEG. In: Gareth M. Gaskell (ed.): The Oxford handbook of psycholinguistic, 119–39. Oxford University Press, Oxford.

Rose, Phil 1999. Long- and short-term within-speaker differences in the formants of Australian *hello*. In: Journal of the International Phonetic Association 29 : 1–33.

Scott, Sophie K. 2005. The neurobiology of speech perception. In: Cutler (2005, 141–56).

Shriberg, Elisabeth 2001. To 'errrr' is human: Ecology and acoustics of speech disfluencies. In: Journal of the International Phonetic Association 31 : 153–169.

Stevens, Kenneth N. 1972. The quantal nature of speech: Evidence from articulatory–acoustic data. In: Peter B. Denes – Edward E. David Jr. (eds): Human communication: A unified view, 51–66. McGraw Hill, New York.

Tacq, Jacques 1997. Multivariate analysis techniques in social science research. From problem to analysis. Sage, London.

Vago, Robert M. – Mária Gósy 2007. Schwa vocalization in the realization of /r/. In: Jürgen Trouvain – William J. Barry (eds): Proceedings of the 16th International Congress of the Phonetic Sciences, 505–9. Saarbrücken University, Saarbrücken.

Zwitserlood, Pienie 2003. The internal structure of words: Consequences for listening and speaking. In: Niels O. Schiller – Antje S. Meyer (eds): Phonetics and phonology in language comprehension and production, 79–114. Mouton de Gruyter, Berlin & New York.