Mária Gósy

ACOUSTIC-PHONETIC VARIABILITY AND THE MENTAL IMAGE OF A WORD

Abstract

The speaker's articulatory gestures intended to represent a word are supposed to show relatively high variability in spontaneous speech. The word forms may be stored as abstract phonological representations or else they may be characterized by detailed acoustic-phonetic patterns. The aim of this paper is to show the acoustic-phonetic patterns of the Hungarian word akkor *(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 be significantly different across speakers. In addition, the results showed considerable individual differences as well. The correspondence between variability in the objective acoustic-phonetic data and the flexibility and adaptive nature of the mental image of a word will be discussed.*

Key words: mental lexicon, acoustic patterns, articulatory variables, Hungarian

1. INTRODUCTION

Linguistic mental representations contain linguistically relevant aspects of the individual's patterns of knowledge. Certain parts of linguistic mental representations may keep changing or being modified depending on diverse factors. Meaning is mentally encoded by humans (Jackendoff, 2002). The mental representation of a word can be defined in various ways: it includes semantics, grammatical form, as well as phonological and phonetic structures. Word forms may be stored as abstract phonological representations or as detailed acoustic-phonetic representations. Experimental results seem to support the existence of prelexical phonological processes in spoken word recognition. Furthermore, there are theories of direct mapping from an acoustic representation of the input signal to the lexical representations (Andruski et al., 1994; McQeen & Cutler, 2002).

Various phonetic factors in a speech signal may obscure the speaker's intended pronunciation while it still corresponds to his/her mental representation of the phonological word. The speaker's articulation gestures intended to represent a word show relatively high variability within and across the phoneme categories particularly in spontaneous speech. This variability of articulatory gestures has received a lot of attention in the literature: words in narratives and dialogues, as well

as repetitions have been investigated (e.g. Kohler, 2000; Pluymaekers et al., 2005); the reduction of words has especially come into prominence. However, variability of pronunciation does not result in frequent misperceptions, a fact that is generally explained primarily by linguistic context and the informational redundancy of speech. However, many novel questions arise. Are there any limits for the variability of articulation? Are speakers aware of the flexibility of the mental representation of a word concerning its acoustic-phonetic features? Do speakers (mostly unconsciously) rely on the contextual predictability of word recognition in spontaneous speech? Is there conscious or unconscious control over the articulation gestures in the pronunciation of a word that corresponds to its mental acoustic-phonetic representation? Does phonologically induced perceptual correction ensure correspondence between the sound sequences and the mental representation of the "phonological word" in spontaneous speech? The explanation might be that speakers reduce their articulatory efforts spent on words that are predictable for the listener. This theory goes back some 20 years (Lindblom, 1990). The listeners are expected to be able to modify their temporal analysis and their frequency and intensity filtering mechanism in order to normalize the incoming acoustic signal. In order to learn more about this assumed human normalization process (cf. Nusbaum & Magnuson, 1997), we have to gather more objective data on the articulatory variability of spoken words.

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. The speakers often want to restore continuity of their speech using filler words (Clark & Wasow, 1998); therefore, such words are relatively frequent in spontaneous speech. We hypothesize that the 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. Therefore, the filler word may show higher variability in articulation than other words do because of its high frequency of occurrence in spontaneous speech. 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. In this paper, we attempt to find out which of these two opposite hypotheses is in fact the correct one.

The aim of this paper is to show the acoustic-phonetic patterns of a Hungarian word *akkor* in both of its functions: as an adverb meaning 'at that time' and as a filler. The disyllabic word *akkor* /ɔkːkor/ (the first syllable carries word stress) has been chosen for the purposes of this paper since this word is frequently used particularly in its filler function in present-day spontaneous speech. Our hypotheses are that (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 and (iii) there are a few invariant features that constitute the interface between the speech sound sequence and the phonological structure of the word.

2. MATERIAL, SUBJECTS, METHOD

The acoustic-phonetic consequences of the pronunciation of the single Hungarian word *akkor* ([ɔkːor]) were analyzed. The original meaning of the word is 'at that time'. This is a double-purpose function word, used either as an adverbial pronoun or as a filler. Ten speakers (5 females and 5 males) from BEA, the Hungarian Spontaneous Speech Corpus were selected. They were native adult speakers of Hungarian (ages ranging from 22 to 28). 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, 2008).

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).

The material selected contained 286 occurrences of the word, half of them from males and the other 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 phonetic context of *akkor* does not show large differences. It occurs after the word *és* ('and') in 41.25% and after a pause in 37.06% of all occurrences. The conjunctions *mert* and *tehát* ('because' and 'that is') preceded it in 13.28% of all occurrences (the remaining 8.41% contains 3 vowels and 2 consonants preceding *akkor*).

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 stop and the VOT of the [kː] 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] 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 corresponding spectrographic, intensity and waveform displays were consulted and auditory perception was considered.

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. Narrowband fast Fourier transforms (FFT) were also considered. In sum, nine parameters were analyzed for each token, yielding a total of 2,574 measurements. To test statistical significance, analysis of variance (ANOVA) and correlation analysis were used (SPSS, Version 8.0). In all cases, the confidence level was set at the conventional 95%.

3. RESULTS

Subjects produced 1.12 *akkor* words per minute. The mean occurrence of *akkor* was 1.05 words per minute in female subjects' speech (min.: 19, max.: 35) and 1.2 words per minute in that of males (min.: 20, max.: 38). 68% of all occurrences were identified as fillers (though no differentiation was made in the analysis depending on functions). Figure 1 shows the acoustic structures of two pronunciations of *akkor* by the same female subject. The relatively large differences in the acoustic properties of the two tokens can be visually inspected.

There are considerable differences in the durational patterns of the vowels and the velar stop as well as in the intensity structure of the release burst of the stop. The formants are basically different with both vowels, particularly with the stressed one. Although both tokens contain the word-final [r], its acoustic manifestation is different: it is a real trill followed by a schwa in the first case while it is a vocalized realization of the phoneme in the second case (the acoustic-phonetic properties of the phoneme /r/ will not be further analyzed here). The total duration of the word *akkor* and the durations of the two vowels, of the intervocalic long velar voiceless stop and the voice onset time of the stop will be analyzed on the one hand, while the measurement results of the formant structures of the vowels will be discussed, on the other.

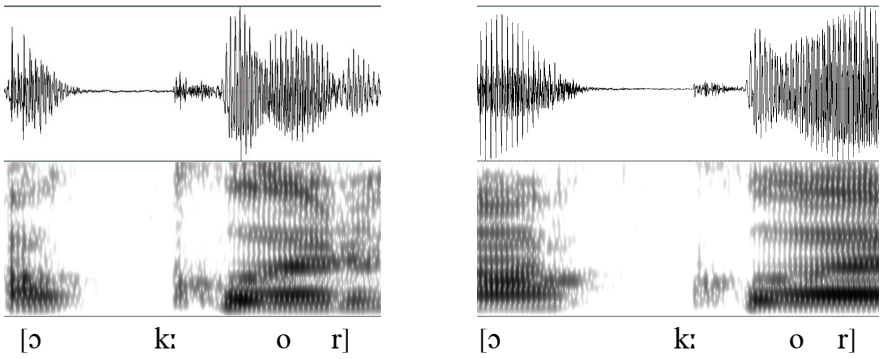

Figure 1. Acoustic-phonetic properties of the word *akkor* pronounced by the same female speaker in two different contexts
Slika 1. Akustičko-fonetska obilježja riječi *akkor* koju izgovara ista ispitanica u dva različita okruženja

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 500 ms in the case of females whereas the shortest one is 144 ms and the longest one is 430 ms for males). There was no significant difference depending on gender, though (cf. Table 1). Less variability of word durations was expected within speakers than across speakers (Dankovičová & Nolan, 1999). The word duration values contradicted this assumption: within-speaker variability is very similar to across--speaker variability (cf. Fig. 2).

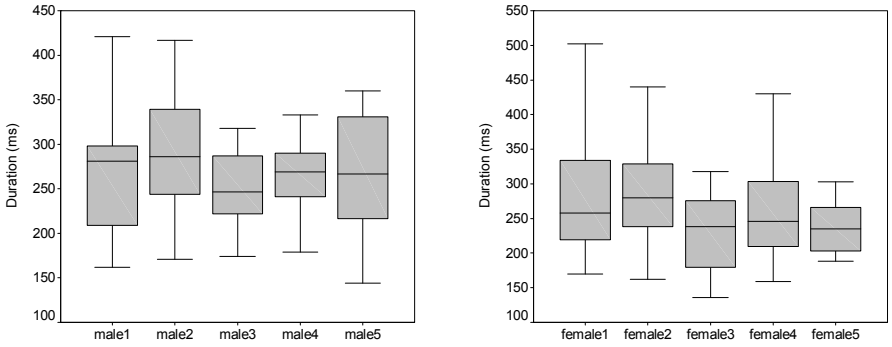

Figure 2. Word duration values of *akkor* with all subjects
Slika 2. Vrijednosti trajanja riječi *akkor* za sve ispitanike

Table 1. Durational patterns of the word *akkor* (SD = standard deviation)
Tablica 1. Obrasci trajanja riječi *akkor* (SD = standardna devijacija)

akkor	Durations (ms)					
	females		males		all subjects	
	mean	SD	mean	SD	mean	SD
word	277.52	109.38	271.37	60.44	274.45	88.26
[ɔ]	64.34	25.29	61.16	15.41	62.75	20.96
[o]	58.90	27.53	58.77	19.26	58.84	23.72
[k:]	119.75	46.57	110.23	26.71	114.99	38.19

The durations of the two vowels 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]: $F(9, 285) = 4.029$, $p < 0.000$ while for [o]: $F(9, 285) = 2.751$, $p < 0.004$).

The values of the velar stops are significantly different across speakers ($F(9, 285) = 8.478$, $p = 0.001$); however, the range of the durations is greater in females than in males. Voice onset time (VOT) is generally assumed to be an invariant feature for the voiceless velar [k]. The results confirmed that assumption. The mean value is 39.45 ms (SD=13.69) for all subjects while it is 39.84 ms (SD=12.75) for females and 39.06 ms (SD=14.59) for males (with no significant difference across speakers).

The analysis of the **formants** of [ɔ] vowels in females revealed that both F1 and F2 are significantly different across speakers (for F1: $F(4, 142) = 5.390$, $p < 0.000$ and for F2: $F(4, 142) = 7.352$, $p = 0.001$). The same results were found for [o] vowels (for F1: $F(4, 142) = 8.250$, $p < 0.000$ 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, which is supported by the acoustic-phonetic correlates. Both first and second formants of the vowel [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 formant of the vowel [o]. On the other hand, the second formant of [o] 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)

Tablica 2. Frekvencije formanata za vokale u riječi *akkor* (SD = standardna devijacija)

Formants	Formant frequency values (Hz)			
	females		males	
	mean	SD	mean	SD
F1 [ɔ]	623.13	94.44	567.37	79.76
F2 [ɔ]	1444.68	219.40	1261.00	149.90
F1 [o]	546.03	57.99	476.59	44.54
F2 [o]	1298.90	199.48	1123.21	139.70

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 3 and 4 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].

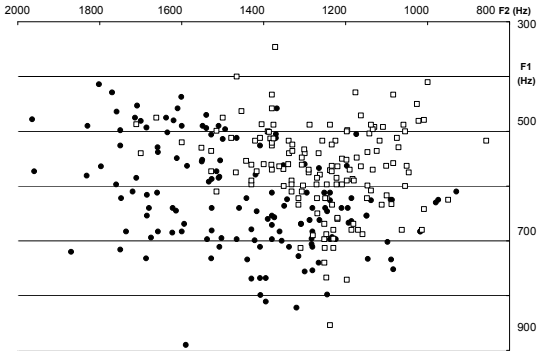

Figure 3. The F1/F2 space of the realizations of the phoneme /ɔ/ in the word *akkor* (black circles represent the females' data while the squares represent the males' data)

Slika 3. F1/F2 vokalski prostor u realizaciji fonema /ɔ/ u riječi *akkor* (crni kružići označavaju podatke za govornice a kvadratići za govornike)

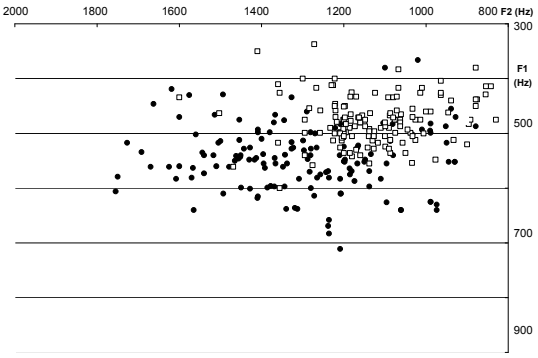

Figure 4. The F1/F2 space of the realizations of the phoneme /o/ in the word *akkor* (black circles represent the females' data while the squares represent the males' data)

Slika 4. Formanti F1 i F2 fonema /o/ u riječi *akkor* (crni kružići označavaju podatke za govornice a kvadratići za govornike)

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 occurrences of *akkor* follow a voiceless fricative 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 originates in the characteristics of the speech planning process.

4. CONCLUSIONS

Although some pronunciation differences had been expected, the results obtained 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 image of a word. The stimulus itself carries a sufficient amount of invariant features to map the acoustic signal onto the phonological word 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". The intensive part of the release burst occurs in all cases below 1,500 Hz, which ensures the feature "velar". The closure part together with the low-frequency burst is characteristic of stops. The majority of the formants of the

stressed vowels point to a low back vowel. The word-final [r], independently of its actual realization, narrows down the competing word forms in the mental lexicon. The statistical analysis confirmed significant correlations (Pearson's analysis) between the durations of the speech sounds and that of the total word (between [ɔ] and the word: $r = 679$, between [o] and the word: $r = 689$ and between [kː] and the word: $r = 688$, $p < 0.000$ in all cases). These interrelations can be seen as some invariant cue rescuing the important durational patterns of the word that help the human normalization process.

The spoken sound sequence of /ɔkːor/ – in terms of the acoustic-phonetic patterns – looks like a puzzle: [ɔ] or [o] + [k] or [kː] + [ə] or [ø] + [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.

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.

REFERENCES

Andruski, J. E., Blumstein, S. E., Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition* **52/3**, 163-187.

Boersma, P., Weenink, D. (2005). *Praat: Doing phonetics by computer*, http://www.praat.org/ [visited March 12, 2005].

Clark, H. H., Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology* **37**, 201-242.

Dankovičová, J., Nolan, F. (1999). Some acoustic effects of speaking style on utterances for automatic speaker verification. *Journal of the International Phonetic Association* **29**, 115-229.

Fox Tree, J. E., Schrock, J. C. (2002). Basic meanings of *you know* and *I mean*. *Journal of Pragmatics* **34**, 727-747.

Gósy, M. (2008). Magyar spontánbeszéd-adatbázis – BEA [Hungarian spontaneous speech data base]. *Beszédkutatás* 2008, 194-207.

Horga, D. (2008). Repetitions in interrupted speech production. *Beszédkutatás* 2008, 157-171.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Kohler, K. (2000). Investigating unscripted speech: Implications for phonetics and phonology. *Phonetica* **57**, 85-94.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition* **14**, 41-104.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling*, 403-440. Dordrecht: Kluwer.

McQeen, J. M., Cutler, A. (eds.) (2002). *Spoken word access processes*. Special Issue of Language and Cognitive Processes. London: Taylor and Francis.

Nusbaum, H., Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullenix (eds.), *Talker variabitlity in speech processing*, 109-132. San Diego: Academic Press.

Pluymaekers, M., Ernestus, M., Baayen, H. R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* **62**, 146-159.

Shriberg, E. (2001). To "errrr" is human: Ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* **31**, 153-169.

AKUSTIČKO-FONETSKA VARIJABILNOST I MENTALNA SLIKA RIJEČI

Sažetak

Oblici riječi koji čine mentalni leksikon (govornika) mogu se pohraniti kao apstraktne fonološke reprezentacije ili kao detaljne akustičko-fonetske reprezentacije. Govornikovi artikulacijski pokreti koji bi trebali predstavljati riječi, u spontanom govoru pokazuju relativno velik stupanj varijabilnosti. Cilj je ovog rada pokazati akustičko-fonetske obrasce mađarske riječi akkor '*tada*'. *Snimljen je i analiziran spontani govor deset govornika ukupnog trajanja 255 minuta, tijekom kojeg se riječ* akkor *pojavila 286 puta. Programom PRAAT izmjereni su obrasci trajanja te frekvencijski obrasci. Rezultati pokazuju viši stupanj varijabilnosti od očekivanog i u govoru pojedinca i među govornicima. Trajanje riječi i trajanje glasova jednako kao i formanti značajno se razlikuju među govornicima. Nadalje, rezultati također pokazuju zamjetne razlike u govoru pojedinačnih govornika. U radu će se razmatrati veza između varijabilnosti objektivnih akustičko-fonetskih podataka i fleksibilnosti odnosno adaptivnosti mentalnih predodžbi riječi.*

Ključne riječi: mentalni leksikon, akustički obrasci, varijable artikulacije, mađarski jezik