4 Motivating the durational alternations

Why are there so many non-contrastive vowel duration degrees in Finnish, alternations that partly interfere with the optimal realisation of the quantity opposition? The answer seems to be provided by the particular combination of prosodic properties in the language. Given the uniformity of the accentual tune across different word structures, and given the moraic alignment of the accentual tune, the durational alternations discussed above are necessary. If the durational alternations did not exist but accent nevertheless had the moraic alignment that it has, the uniformity of the accentual tune would not be possible. Why the tonal uniformity exists is not clear, but there it is. It is somewhat paradoxical that, in a full-fledged quantity language in which segment durations signal phonemic distinctions, segment durations nevertheless also vary extensively to serve tonal purposes, while in non-quantity languages like Greek the segmental composition of the accented syllable determines the tonal realisation.

The durational alternations are also observable in unaccented words. But this does not undermine the motivation just suggested, because unaccented and moderately accented words do not differ from each other durationally, and the alternations are directly motivated in moderately accented words. Thus unaccented words are as if prepared for being accented. A conceivable alternative would be that unaccented words would lack the alternations present in accented words, but this state of affairs would further complicate the durational system.

To summarise, beyond the loci in which stress and accent are realised, i.e. when vowels do not constitute M_1 or M_2, single vowels are "extra short" and double vowels "very long", which results in their clear separation. In (C)V.CV(X) words, the tonal rise is realised during the initial syllable and it is sufficient that the vowel is "short". The long fall is realised during the second syllable, and therefore the vowel must be "long". In (C)VV.CV(X) words, both the rise and most of the fall is realised during the initial syllable, and therefore both segments in the VV sequence must be "longish". This paper is not about consonant durations but in (C)VC.CV(X) words, in which M_2 is a consonant, it too has to be "longish"; if the consonant has relatively short intrinsic duration elsewhere, it is lengthened in this position. As a consequence of these alternations, the accentual rise-fall can be uniform across different word structures, and at the same time, the quantity oppositions are not jeopardised.

References

Arvaniti, A., D.R. Ladd & I. Mennen, 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics 26,* 3-25.

Karlsson, F., 1969. Suomen yleiskielen segmentaalifoneemien paradigma. *Virittäjä 73,* 351-362.

Ladd, D.R., 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Lehtonen, J. 1970. *Aspects of quantity in standard Finnish*. Jyväskylä: Jyväskylä University Press.

Suomi, K., 2005. Temporal conspiracies for a tonal end: segmental durations and accentual f_0 movement in a quantity language. *Journal of Phonetics 33,* 291-309.

Suomi, K., in press. On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*.

Suomi, K., in preparation. Durational elasticity for accentual purposes (working title).

Suomi, K., J. Toivanen & R. Ylitalo, 2003. Durational and tonal correlates of accent in Finnish. *Journal of Phonetics 31,* 113-138.

Suomi, K. & R. Ylitalo, 2004. On durational correlates of word stress in Finnish. *Journal of Phonetics 32,* 35-63.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics 133
Working Papers 52 (2006), 133–136

Phonological Demands vs. System Constraints in an L2 Setting

Bosse Thorén
Dept. of Linguistics, Stockholm University
bosse.thoren@skola.sundsvall.se

Abstract

How can system constraints and phonological output demands influence articulation in a L2-speaker? When measuring durations and articulator movements for some Swedish /V:C/ and /VC:/ words, pronounced by a Swedish and a Polish speaker, it appeared that phonological vowel length was realized very similarly by both speakers, while complementary consonant length was applied only by the native Swedish speaker. Furthermore, the tendency for increased openness in short (lax) vowel allophones was manifested in analogous jaw and lip movements in the Swedish speaker, but followed a different pattern in the Polish speaker.

1 Introduction

How is articulation influenced by system-based constraints and output-based constraints, when a person uses a second language? According to the Hyper & Hypo speech theory (Linblom 1990) the degree of articulatory effort in human speech is determined by mainly two factors: 1) The limitations that inertia in the articulators poses upon speech, including the tendency for economy in effort. 2) The demands of the listener, e.g. sufficient phonological contrast. The former is assumed to result in unclear speech, or "under shoot", and the latter to "over shoot" or "perfect shoot" (clear speech). According to the H&H-theory, the output demands vary depending on e.g. contextual predictability and the acoustic channel being used, the presence of noise etc.

From a cross-linguistic point of view, the demands of a listener are to a high degree determined by the phonologic system of the language in question. These demands are supposed to be intuitively inherent in the native speaker of the language, i.e. the speaker has a clear but probably unconscious picture of the articulatory goal. What happens to a L2-speaker in this perspective? We can assume that the L2-speaker is influenced both by L1 and L2 demands on the output, as well as by system-based constraints.

Swedish has a quantity distinction in stressed syllables, manifested in most varieties as either /V:(C)/, or /VC:/. Elert (1964) has shown that the Swedish long-short phonological distinction is accompanied by analogous differences in duration for the segments involved. His study also shows that the differences in duration between long and short Swedish vowel allophones are significantly greater (mean 35%) than durational differences between closed and open vowels (5-15%). This predicts that output constraints for Swedish segment durations would override the system constraints, i.e. the inherent differences in duration between open and closed vowels.

Polish on the other hand, is a language without phonological quantity, and is not expected to involve any output constraints on the duration of segments. Duration differences in Polish are assumed to result mainly from vowel openness, in accordance with the "Extent of Movement Hypothesis" (Fischer-Jörgensen, 1964). A native polish speaker, who speaks

Swedish as a second language, is therefore expected to show more influence from the system constraints in his/her Swedish production than a native Swedish speaker. In addition to the longer inherent duration in open vowels, there is a clear connection between long/tense and short/lax vowels, resulting in Swedish short vowel allophones being pronounced more open than their long counterparts (cf. Fant 1959).

The present study examines what happens when a native speaker of Swedish and a native speaker of Polish pronounce test words containing the following combinations in trisyllabic nonsense words: long open vowel /ɪpɑ:pɪ/, short open vowel /ɪpap:ɪ/, long closed vowel /ɪpi:pɪ/ and short closed vowel /ɪpɪp:ɪ/.

In this study, the movements of mandible and lips are measured in addition to segment durations, in order to compare the two speakers with regard to patterns of articulatory gestures as results from output demands and system constraints respectively.

The question is: Will the duration of segments, produced in Swedish /V:C/ and /VC:/-contexts differ significantly when pronounced by a native Swedish speaker, and a native Polish speaker? And will the timing and magnitude of lip- and jaw-movements differ in a significant way between the two speakers, indicating more influence from output demands or system constraints?

2 Method

Two adult male subjects, one native Swede and one native Pole, who had lived in Sweden for 22 years, were recorded when pronouncing the nonsense words [ɪˈpɑ:pɪ], [ɪˈpap:ɪ], [ɪˈpi:pɪ]and [ɪˈpɪp:ɪ], all of which are possible Swedish words according to Swedish phonotactics and prosody. The Swedish speaker read the sequence of test words five times, and the Polish speaker read it three times.

Measurements of lip and mandible movement as well as speech signal were carried out by means of Move track: a magnetometer system with sender coils attached to the speaker's lips and lower incisors, and receiver coils placed in a light helmet on the speaker's head. The device measures variation in magnetic field that can be directly related to distance between coils. The system produces data files with articulator movements synchronized with the speech signal.

3 Results

3.1 Segment durations

The two speakers realized phonological vowel length in a similar way, making clear temporal differences between long and short vowel allophones, as shown in Figure 1. The complementary long consonant after short vowel in stressed syllables in Swedish is very clear in the native Swedish speaker, but non-existent and even shorter in the native Polish speaker.

Differences in vowel duration ratios, are illustrated in Figure 1b, where the differences in vowel length are seen as functions of phonological demands and system constraints respectively. Both speakers realize phonologically long vowels with more than the double duration of short vowels, the Polish speaker having even more difference than the Swedish speaker. The Polish speaker made a greater duration difference between /a/ and /i/ than the Swedish speaker did. This latter difference in duration ratios between speakers is significant ($p < 0.05$ ANOVA) whereas the inter-speaker difference in V:/V ratios is not.

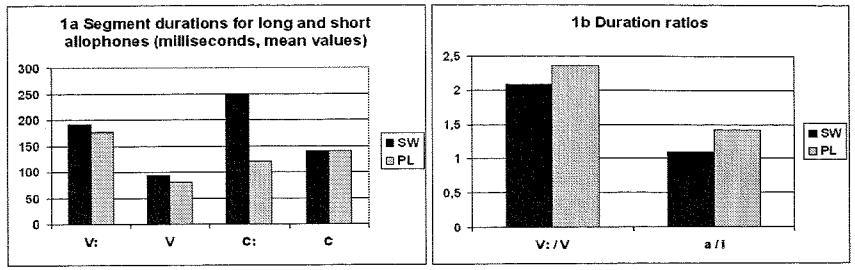

Figure 1a and 1b. a) Durations of long and short allophones produced by the Swedish (black columns) and the Polish speaker (gray columns). Mean values from 10 realizations by the Swedish speaker and 6 realizations by the Polish speaker. b) Inter-speaker differences for long/short vowel ratios and open/closed vowel ratios (mean values).

3.2 Vowel durations and articulator movements

Two principal measures of articulator movements were taken to show possible differences between the speakers; 1) vertical mandible displacement in relation to vowel openness and phonological length, 2) vertical lower lip depression in relation to vowel openness and phonological length. The pattern of jaw opening in the two speakers is shown in Figure 2a. The Swedish speaker follows an expected pattern, where the jaw movement seems to reflect vowel openness, with greater openness for /a/ than for /i/, but also more open articulation for short allophones than for long allophones. The Polish speaker also shows greater jaw lowering for /a/ than for /i/, but the smaller opening for short allophones compared to long allophones, does not reflect the spectral vowel quality, i.e. the fact that at least for /a/, the Polish speaker produces higher F1 for [a] than for [ɑ:]. Inspection via listening and spectral analysis, shows that both speakers produce very similar F1 and F2 values.

The pattern of lip aperture, as shown in Figure 2b, follows roughly the pattern of jaw lowering gestures, except for the Swedish speaker's smaller lip aperture for short /i/ compared to long /i:/.

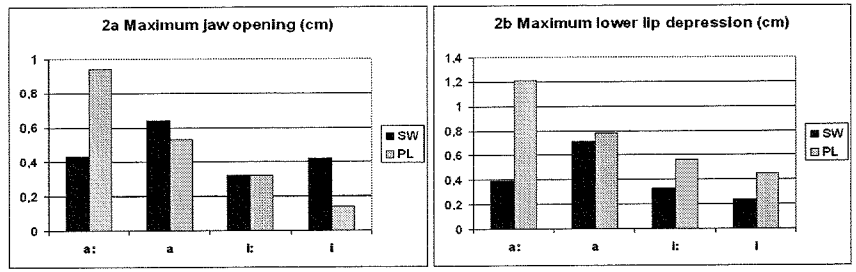

Figure 2a and 2b. Mandible (2a) and lower lip (2b) depression for long-short and open-closed vowels, produced by the Swedish and the Polish speaker.

The timing pattern in terms of lip aperture duration related to vowel duration, and time laps from vowel end to maximal lip closure, did not show any systematic differences between speakers or vowel types.

4 Discussion

The segment duration patterns produced by the two speakers are not surprising. Starting with vowel duration, the phonological vowel length is a well established and well known property of Swedish, both as an important feature of Swedish pronunciation, and a way of accounting for the double consonant spelling. As seen in Figure 1, both speakers realize long and short vowel allophones quite similarly. The Swedish speaker, as shown in Figure 1, demonstrates in addition a substantial prolonging of the /p/ segment after short vowel, which the Polish speaker does not. The Polish speaker reports having encountered rules for vowel length as well as consonant length while studying Swedish, implying that mere ignorance does not account for his lack of complementary long consonant. Literature in phonetics, e.g. Ladefoged & Maddieson (1996), gives the impression that phonological vowel length is utilized by a greater number of the world's languages than is consonant length. This suggests that phonological consonant length is a universally more marked feature than is vowel length, and hence more difficult to acquire.

The somewhat greater difference between long and short vowel allophone, demonstrated by the Polish speaker, can be interpreted as a compensation for the lack of complementary consonant length, which is demonstrated to serve as a complementary cue for the listener, when segment durations are in the borderland between /V:C/ and /VC:/ (Thorén 2005).

The between-speaker difference is not surprising, since the phonological quantity in Swedish is a predominant phonetic feature, and can be expected to influence the temporal organization of the native Swede's speech from early age. The Polish speaker came to Sweden as an adult and has acquired one important temporal feature, but his overall temporal organization may still bear strong traces of the system constraints, concerning the duration of segments.

The differences in lip and mandible movements between the speakers could be interpreted as follows: Both speakers produce a higher F1 for short [a] than for long [ɑ:] (e.g. Fant 1959), which typically correlates with lower tongue and mandible. The Polish speaker however, shows a clearly greater jaw and lip opening for long [ɑ:] than for short [a], which suggests that the Polish speaker has a compensatory tongue height in [ɑ:], to maintain correct spectral quality. The greater mandible excursion in [ɑ:] can not be the result of an articulatory goal for this vowel, but could possibly be interpreted as an inverse "Extent of Movement Hypothesis" (Fischer-Jörgensen 1964), letting the mandible make a greater excursion owing to the opportunity offered by the long duration of the [ɑ:].

References

Elert, C-C., 1964. *Phonologic studies of Swedish Quantity*. Uppsala: Almqvist & Wiksell.

Fant, G., 1959. Acoustic analysis and synthesis of speech with application to Swedish. *Ericsson Technics. No. 15*, 3-108.

Fischer-Jörgensen, E., 1964. Sound Duration and Place of articulation. *Zeitschrift für Sprachwissenschaft und Kommunikationsforschung 17*, 175-207.

Ladefoged, P. & I. Maddieson, 1996. *The sounds of the World's Languages*. Oxford: Blackwell publishers.

Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle & Marchal (eds.), *Speech production and speech modeling*. Dordrecht: Kluwer, 403-439.

Thorén, B., 2005. The postvocalic consonant as a complementary cue to the quantity distinction in Swedish – a revisit. *Proceedings from FONETIK 2005,* Göteborg University, 115-118.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 137–140

Cross-modal Interactions in Visual as Opposed to Auditory Perception of Vowels

Hartmut Traunmüller
Department of Linguistics, Stockholm University
hartmut@ling.su.se

Abstract

This paper describes two perception experiments with vowels in monosyllabic utterances presented auditorily, visually and bimodally with incongruent cues to openness and/or roundedness. In the first, the subjects had to tell what they heard; in the second what they saw. The results show that the same stimuli evoke a visual percept that may be influenced by audition and may be different from the auditory percept that may be influenced by vision. In both cases, the strength of the influence of the unattended modality showed between-feature variation reflecting the reliability of the information.

1 Introduction

Nearly all research on cross-modal interactions in speech perception has been focused on the influence an optic signal may have on auditory perception. In modeling audiovisual integration, it is common to assume three functional components: (1) auditory analysis, (2) visual analysis and (3) audiovisual integration that is assumed to produce an 'amodal' phonetic output. Although details differ (Massaro, 1996; Robert-Ribes et al., 1996; Massaro & Stork, 1998), the output was commonly identified with what the subjects *heard,* not having been asked about what they *saw*. This experimenter behavior suggests the amodal representations of phonetic units (concepts), which can be assumed to exist in the minds of people, to be closely associated with auditory perception. The *seen* remains outside the scope of these models unless it agrees with the *heard*.

The present experiments were done in order to answer the question of whether a visual percept that may be influenced by audition can be distinguished from the auditory percept that may be influenced by vision and whether the strength of such an influence is feature-specific. Previous investigations (Robert-Ribes et al., 1998; Traunmüller & Öhrström, in press) demonstrated such feature-specificity in the influence of optic information on the auditory perception of vowels: the influence was strongest for roundedness, for which the non-attended visual modality offered more reliable cues than the attended auditory modality. In analogy, we could expect a much stronger influence of non-attended acoustic information on the visual perception of vowel height or "openness" as compared with roundedness.

2 Method

2.1 Speakers and speech material

For the two experiments performed, a subset of the video recordings made for a previous experiment (Traunmüller & Öhrström, in press) was used. It consisted of the 6 incongruent auditory-visual combinations of the nonsense syllables /gi:g/, /gy:g/ and /ge:g/ produced by each one of 2 male and 2 female speakers of Swedish. Synchronization had been based on the release burst of the first consonant. In Exp. 1, each auditory stimulus was also presented alone and in Exp. 2 each visual stimulus instead.