The effect of speaking rate on the perception of vowel-initial and vowel-final [h]

Jörgen L. Pind

Faculty of Psychology, University of Iceland

Abstract

Numerous studies have shown that the perception of speech segments is often rate-dependent. Thus, with faster speaking rate the boundary between, say, [b] and [p], measured in terms of voice-onset time (VOT), moves to shorter values of VOT. The present paper reports an experiment where the perception of vowel-initial [h] and vowel-final [h], i.e. preaspiration, is compared in words at three different speaking rates. With slower speaking rate the boundaries, as expected, move to longer values of aspiration and preaspiration respectively, though, interestingly, less so in the case of aspiration than in the case of preaspiration.

Introduction

It is common knowledge that speech is a highly variable signal and invariant cues for individual speech segments are not easily established, (though see Pind, 1995b; Stevens and Blumstein, 1981). The sources of variability in the speech signal are many. Some of these sources are extra-linguistic, do not directly concern the linguistic message being conveyed. One such factor is speaking rate which has a significant effect on the duration of individual speech segments. Since some speech cues are primarily temporal in nature, defined by duration or other such temporal measures, speaking rate can obviously affect such cues. To take just one example: Icelandic has long and short vowels and consonants that are for the most part defined by their durations (Einarsson, 1927; Pind, 1986, 1999). Since speech rate affects the segmental durations it can easily happen that a quantitatively long vowel, spoken at a fast utterance rate, will be shorter than a quantitatively short vowel, spoken at a slow rate.

How does the listener deal with such temporal variability? In the main, two different theoretical viewpoints have been advanced, both boasting of long traditions in perceptual psychology. One of them posits a process of *normalization* where the perceptual system is thought to 'take account of' any external factors (for an interesting discussion of this idea in perceptual psychology, see Epstein, 1973). The other viewpoint asserts that invariants can in fact be found in the speech signal, even for speech segments which are highly variable. A case has thus been been made that the ratio of vowel to consonant duration can function as a *higher-order* invariant for quantity in Icelandic (Pind, 1995b), which fits nicely with the perceptual theories of James J. Gibson, e.g. Gibson (1959).

One particular temporal speech cue which has been the focus of great interest is that of VOT or voice-onset-time, the time from the release of a stop consonant to the onset of voicing in the following vowel (Lisker and Abramson, 1964). This speech cue differentiates voiced or unaspirated stop consonants such as [bdg] from voiceless aspirated stops [ptk]. The former series has short VOTs (perhaps 0–20 ms) , the latter longer VOTs (perhaps 50–70 ms). The precise values of VOT are language dependent (Cho and Ladefoged, 1999).

Previous research has shown that VOT is sensitive to speaking rate, with VOTs lengthening at slower speaking rate. This holds especially for the stop category with the longer VOTs, i.e. the voiceless/aspirated series of stops. Perceptual experiments have shown that listener's phoneme boundaries are affected by the duration of surrounding speech segments. Thus, the longer the vowel following a syllable-initial stop, the longer the VOT needed to cue voicelessness/aspiration (Miller et al., 1986; Summerfield, 1981).

Icelandic phonetics has a rather uncommon feature (Ladefoged and Maddieson, 1996) termed *preaspiration*. This is an [h]-like segment at the end of a vowel before stop closure. Consider thus a word-pair like *akur* [aːkʏr] 'field', vs. *akkur* [ahkʏr] 'advantage'. In the first word a long vowel is followed by the closure for the stop [k], in the latter word a short vowel is followed by preaspiration [h] before the closure of the stop. In perceptual research, preaspiration is readily

cued by voice-offset time (VOffT), a speech cue which is in most respects the mirror image of VOT. (Usually VOT is defined to include the initial burst for the stop consonant as well. There is of course no burst following the VOffT, the aspiration is followed by the closure of the stop.)

Previous experiments (Pind, 1995a, 1998) have shown that the perception of preaspiration in Icelandic is sensitive to the duration of the previous vowel. The longer the vowel, the longer the aspiration needs to be for the listener to perceive a pre-aspirated segment.

The experiment reported in this paper compares the influence of speaking rate on the perception of preaspiration and initial [h]. This is done using the word-pair [aːka]–[ahka] on the one hand with the word-pair [aːka]–[haːka] on the other hand. It is hypothesized that in both cases, the slower the speaking rate, the longer the aspiration needs to be to cue the perception, either of vowel-initial aspiration [h-] or of vowel-final aspiration [-h], i.e. preaspiration.

Experiment 1

Method

Participants

Ten undergraduate students of psychology at the University of Iceland participated in the experiment. They were all native speakers of Icelandic and reported normal hearing.

Stimuli

A total of six synthetic speech continua were used in the experiment, made with the Sensyn speech synthesizer, a version of the Klatt-synthesizer (Klatt, 1980; Klatt and Klatt, 1990). Three of the continua were of the type [aːka]–[haːka], three of the type [aːka]–[ahka]. The three continua were distinguished by the length of first vowel and following closure in the first syllable. In the first continuum the vowel was 150 ms long and the closure 75 ms, in the second the vowel was 200 ms long and the closure 100 ms, and in the third and final continuum the vowel was 250 ms long and the closure 125 ms long. In all cases the ratio of vowel duration to the duration of the following closure is thus constant and typical for a phonemically long vowel, followed by a phonemically short stop. The VC durations of the first syllable thus range from 225 ms, through 300 ms to 375 ms. This lengthening of the initial syllable is perceived as a slowing of speaking rate.

In all cases the second syllable started after the closure with a 25 ms long burst and aspiration, followed by a 75 ms long voiced vowel, the second [a] vowel in the word.

The steady state vowel formants for both [a] tokens were set at 770 Hz for F1, 1280 Hz for F2 and 2425 Hz for F3. For the transitions into the [k]-closure the formants moved linearly to respectively 200 Hz, 1600 Hz and 2000 Hz over 45 ms. The transitions into the second vowel were a mirror image of the transitions into the closure.

Within each stimulus continuum aspiration (or preaspiration) was varied in 8 ms steps from 0 to 80 ms. Each continuum thus contained 11 stimuli, bringing the total number of stimuli to 66 for the whole experiment. Aspiration was cued by disabling voicing (synthesis parameter **AV** set to 0 dB), turning on aspiration (**AH** = 40 dB) and frication (**AF** = 55 dB, **A2F** = 50 dB, and **A3F** = 40 dB), and increasing the bandwidth of the first formant (**B1**) from 60 to 200 Hz. The fundamental frequency during the voiced portion of the stimuli was fixed at 125 Hz and the synthesizer was set to use an update interval of 4 ms. The sampling frequency was 11.025 Hz. Figure 1 shows spectrograms of three stimuli from the experiment.

The stimuli were recorded onto two tapes, by playing them through a Turtle Beach Multisound card. One tape contained the stimuli for the [aːka]–[haːka] continua, the other the tapes for the [aːka]–[ahka] continua. Each tape started with a practice block consisting of all 33 stimuli played in randomized order. This was followed by five blocks, each of which contained two tokens of each stimulus. These were presented randomly, a total of 66 stimuli in each block. The interstimulus interval was 2.5 seconds.

Procedure

Five participants listened to the [aːka]–[haːka] tape followed by the [aːka]–[ahka] tape, for the other five participants the order was reversed. The testing took place in a quiet room. The participants listened to the stimuli, which were played at a comfortable listening level, over Sennheiser HD-530-II circumaural headphones.

Participants were provided with response sheets which contained two fields for each stimulus presented. In the [aːka]–[haːka] test the fields were marked with the words *aka* on the one hand and *haka* on the other. In the [aːka]–[ahka] test the words were *aka* on the one hand and *akka* on the other using the normal Icelandic orthography. Participants were instructed to mark the appropriate box for each stimulus presented and to guess if they were not sure which response was appropriate. Participants did not report any difficulty in carrying out this task. With the first 33 trials

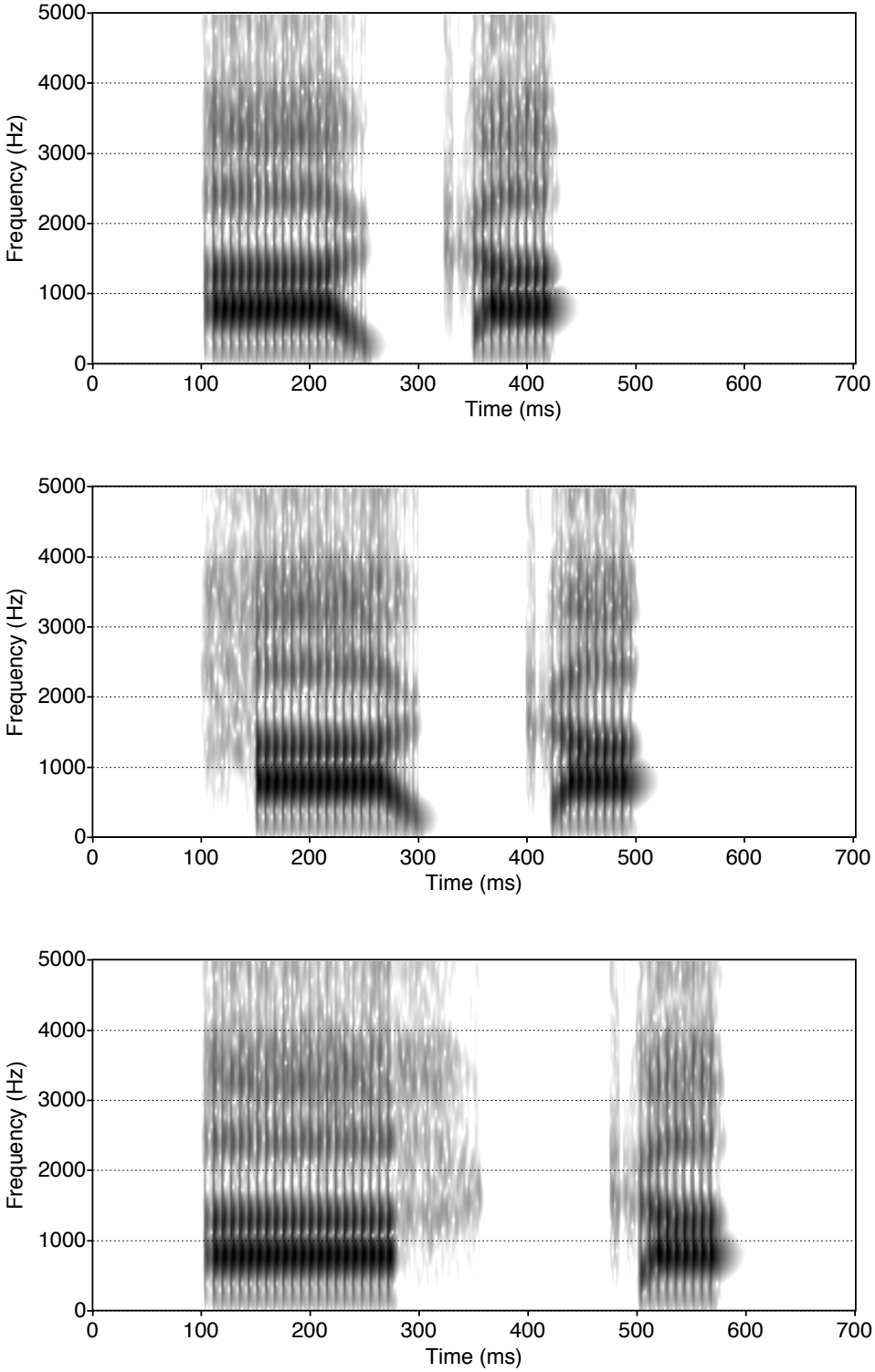

Figure 1: Spectrograms of three stimuli (out of a total of 66) used in the experiment. The topmost spectrogram shows a stimulus with an initial 150 ms long vowel followed by a 75 ms long closure. The stimulus contains neither aspiration nor preaspiration. In the middle spectrogram the first vowel is 200 ms long (including 48 ms of VOT (aspiration) at the beginning of the vowel) followed by a 100 ms long closure. The final spectrogram shows a stimulus with a 250 ms long vowel (including 80 ms of VOffT (preaspiration) at the end of the vowel) followed by a 125 ms long closure.

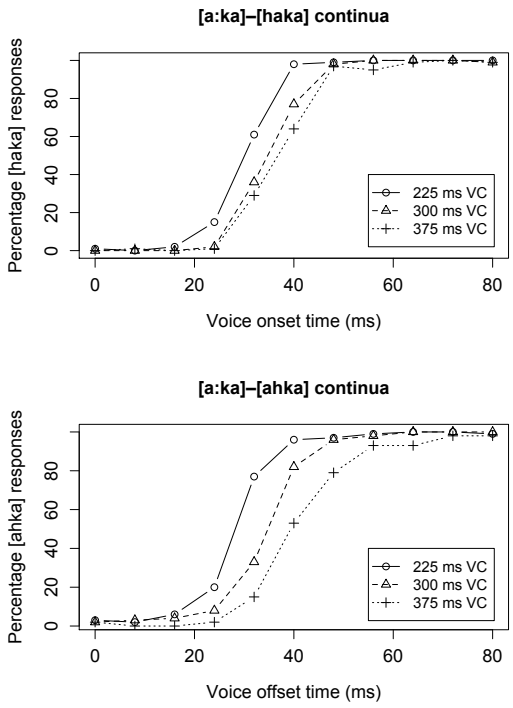

Figure 2: Pooled identification curves for all 10 participants in the present experiment. The upper graph shows the results for the stimulus continua with vowel-initial aspiration. The lower graph shows the results for the three preaspiration continua.

of each run used for familiarization, 10 responses were tabulated for each participant per stimulus.

Results and discussion

Pooled identification curves for all 10 participants are shown in figure 2. The figure clearly show that the percentage of aspiration responses ([haka] or [ahka]) increased as the duration of vowel-initial aspiration, voice-onset time, or vowel-final aspiration, voice-offset time, increases. Additionally the figures clearly show the effect of VC duration on the percentage of aspiration responses. As VC duration increases the phoneme boundaries for the perception of aspiration or preaspiration move to longer values of VOT and VOffT respectively.

Phoneme boundaries were calculated for individual participants using the method of probits (Finney, 1947). The calculation were done in R with the help of the MASS library (Venables and Ripley, 1999). The average phoneme boundaries are shown in Table 1.

A two-factor repeated measures ANOVA (stimulus series × VC duration) shows that the effect of stimulus series is not significant, $F(1,9) = 0.028$, whereas the effect of VC duration is significant, $F(2,18) = 92.25$, $p < 0.001$. The interaction of series and VC duration is significant, $F(2,18) = 7.83$, $p < 0.01$.

Table 1: Average phoneme boundaries (10 participants) for the six stimulus continua of the present experiment. The values denote milliseconds of aspiration.

VC duration (ms)	[a:ka]–[ahka] continua	[a:ka]–[ha:ka] continua
225	28.30	29.69
300	33.85	35.10
375	41.58	37.48

Repeated paired *t*-tests show that average phoneme boundary in the words with the 300 ms long VC is significantly longer (34.46 ms) than in the words with the 225 ms long VC (29 ms), $t(38) = 3.66$, $p < 0.001$. Again the average phoneme boundaries in the words with the 375 ms long VC is significantly longer (39.53 ms) than in the words with the 300 ms long VC, $t(38) = 2.674$, $p < 0.05$.

The interaction of stimulus series and VC duration can be inferred from Figure 2. Here it can be seen that the phoneme boundaries move to steadily longer values of voice-offset time in the [a:ka]–[ahka]] series, in approximately equal steps, confirm also Table 1. In the [a:ka]–[ha:ka] series there is a clear movement of the phoneme boundaries as VC duration increases from 225 to 300 ms (change of 5.41 ms), with a much smaller increase from 300 to 375 ms (change of 2.38 ms). In the [a:ka]–[ahka] series the corresponding changes in the phoneme boundaries are 5.55 ms and 7.73 ms. So in the case of preaspiration the movement of the phoneme boundaries increases as the vowel is lengthened, in the case of vowel-initial aspiration it is the other way around, the movement of the phoneme boundaries decreases with increasing vowel length.

How can this be explained? From earlier experiments it is clear that vowel quantity has a decisive influence on the perception of preaspiration and this is so regardless of whether quantity is cued by vowel duration (Pind, 1996a) or the spectrum of the vowel (Pind, 1998). (The vowel spectrum plays an important role in the perception of quantity of the three central vowels of Icelandic, Pind (1996b).) Since care was taken in this experiment to keep the quantity constant over the different stimulus continua by keeping the vowel to closure duration ratio fixed the effects of quantity would seem to be ruled out, though admittedly

a direct test of perceived vowel duration was not undertaken. This needs to be addressed in further experiments. If there is a tendency for the stimuli with the longest VC duration to be perceived as perhaps having a more robust – or more prototypical – long vowel than in the stimuli with the shortest VC then this would undoubtedly influence the perception of preaspiration more than the perception of initial [h]. If this is not the case, then an explanation needs to be sought elsewhere. Perhaps the weak [h] sound following the loud vowel is more susceptible to masking than the word-initial [h] (cf. Bladon, 1986, for a similar hypothesis). Further experiments are needed to explore this issue in greater detail.

References

Bladon A (1986). Phonetics for hearers. In G McGregor, ed., *Language for hearers*, 1–24. Oxford: Pergamon Press.

Cho T and Ladefoged P (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27:207–229.

Einarsson S (1927). *Beiträge zur Phonetik der isländischen Sprache*. Oslo: A. W. Brøgger.

Epstein W (1973). The process of 'taking-into-account' in visual perception. *Perception*, 2:267–285.

Finney D J (1947). *Probit analysis*. Cambridge: Cambridge University Press.

Gibson J J (1959). Perception as a function of stimulation. In S Koch, ed., *Psychology: A study of a science. Volume I. Sensory, perceptual and physiological formulations*, 456–501. New York: McGraw-Hill Book Company.

Klatt D H (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67:971–995.

Klatt D H and Klatt L C (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87:820–857.

Ladefoged P and Maddieson I (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lisker L and Abramson A S (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20:384–422.

Miller J L, Green K P and Reeves A (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, 43:106–115.

Pind J (1986). The perception of quantity in Icelandic. *Phonetica*, 43:116–139.

Pind J (1995a). Constancy and normalization in the perception of voice offset time as a cue for preaspiration. *Acta Psychologica*, 89:53–81.

Pind J (1995b). Speaking rate, VOT and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics*, 57:291–304.

Pind J (1996a). Rate-dependent perception of aspiration and pre-aspiration in Icelandic. *Quarterly Journal of Experimental Psychology*, 49A:745–764.

Pind J (1996b). Spectral factors in the perception of vowel quantity in Icelandic. *Scandinavian Journal of Psychology*, 37:121–131.

Pind J (1998). Auditory and linguistic factors in the perception of voice offset time as a cue for preaspiration. *Journal of the Acoustical Society of America*, 103:2117–2127.

Pind J (1999). Speech segment durations and quantity in Icelandic. *Journal of the Acoustical Society of America*, 106:1045–1053.

Stevens K N and Blumstein S E (1981). The search for invariant acoustic correlates of acoustic features. In P D Eimas and J L Miller, eds., *Perspectives on the study of speech*, 1–38. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Summerfield Q (1981). On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7:1074–1095.

Venables W N and Ripley B D (1999). *Modern applied statistics with S-PLUS*. Springer, third edn.