Original Paper

Phonetica 2006;63:209–229
DOI: 10.1159/000097306

Received: January 5, 2006
Accepted: July 27, 2006

Is Vowel Normalization Independent of Lexical Processing?

Holger Mitterer

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Abstract

Vowel normalization in speech perception was investigated in three experiments. The range of the second formant in a carrier phrase was manipulated and this affected the perception of a target vowel in a compensatory fashion: A low F2 range in the carrier phrase made it more likely that the target vowel was perceived as a front vowel, that is, with a high F2. Recent experiments indicated that this effect might be moderated by the lexical status of the constituents of the carrier phrase. Manipulation of the lexical status in the present experiments, however, did not affect vowel normalization. In contrast, the range of vowels in the carrier phrase did influence vowel normalization. If the carrier phrase consisted of mid-to-high front vowels only, vowel categories shifted only for mid-to-high front vowels. It is argued that these results are a challenge for episodic models of word recognition.

Introduction

One of the most intriguing problems for spoken-word recognition is the difficulty of finding acoustic properties of a given word that are invariant over pronunciations by different speakers. This problem arises because the speech signal contains at least three types of information. First, there is the linguistic information that conveys the message of the speaker. (Even here, variation arises due to coarticulation of adjacent segments.) Second, the speech signal reflects the anatomical properties of the speaker. Third, the speech signal contains information about the sociophonetic traits as well as the emotional states of a speaker. The variability caused by sociophonetic and anatomical variables gives rise to the invariance problem in recognizing the linguistic message.

Solving the invariance problem is complicated by the fact that all three sources of information influence the same parameters. If one considers vowels, for instance, anatomical, sociophonetic, and linguistic sources all influence their formant frequencies. In their classic study, Peterson and Barney [1952] showed that formant frequencies differ, obviously, as a function of vowel quality, but also as a function of the speaker's gender. Male formant values tend to be lower than female formant values, because men tend to have larger supralaryngeal systems. This male-female difference

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/06/0634–0209
$23.50/0
Accessible online at:
www.karger.com/journals/pho

Holger Mitterer
Max Planck Institute for Psycholinguistics
PO Box 310, NL–6500 AH Nijmegen (The Netherlands)
Tel. +31 (0)24 3521372, E-Mail holger.mitterer@mpi.nl

varies, however, over language communities [Bladon et al., 1984], and is hence not wholly explained by anatomical variability, but also by sociophonetic aspects. Accordingly, the listener must decompose – or demodulate [Traunmüller, 1994] – the signal in order to recover the phonetic, anatomical and sociolinguistic information. Given that the focus of spoken-word recognition research is on the linguistic information, this process has been called 'speaker normalization', and is supposed to filter out speaker-specific influences so that lexical access can be achieved with an abstract, speaker-independent code [e.g., McQueen and Cutler, 2001].

Part of the solution to the 'decomposition problem' in vowel perception lies in adaptation of the listener to a particular speaker, with the classic paper provided by Ladefoged and Broadbent [1957]. They presented listeners with target words with the structure bVt, and listeners had to decide whether the word was *bit* [bɪt], *bet* [bɛt], *bat* [bæt], or *but* [bʌt]. The target bVt was presented after the carrier phrase 'please say what this word is'. The F1 and F2 ranges in the carrier phrase were manipulated. Listeners adapted to this change in range of formant frequencies. The same test word was more likely to be perceived as *bet* [bɛt] rather than [bɪt] if the F1 range in the carrier phrase was lowered. Lowering the F1 in the carrier phrase makes the F1 in the test word relatively higher, and a higher F1 is more appropriate for [ɛ] than for [ɪ]. Subsequent research has replicated the effect with natural speech [Ladefoged, 1989].

Recently, Norris et al. [2003] showed that the listener's adaptation to a speaker utilizes lexical knowledge. In these experiments, listeners first completed a Dutch lexical decision task, which was followed by a two-alternative forced-choice task. In the lexical-decision task, an ambiguous fricative, which could be interpreted as either /s/ or /f/, was embedded at the end of fragments, such as *olij*… or *radij*… For half of the fragments, the fragment plus the ambiguous fricative constituted words only if the fricative was interpreted as /s/, as in *radijs* 'radish'; while the other half constituted words only if the fricative was interpreted as /f/, as in *olijf* 'olive'. Half of the participants heard the /s/ words with a canonical [s], [radɛɪs], and the /f/ words with the ambiguous fricative, which will be transcribed as [$_{s}^{f}$], [olɛɪ$_{s}^{f}$]. The other half of the participants heard the /s/ words with the ambiguous fricative [$_{s}^{f}$], [radɛɪ$_{s}^{f}$], and the /f/ words with the canonical [f], [olɛɪf]. Lexical decision performance indicated a Ganong effect [Ganong, 1980]: The ambiguous fricative [$_{s}^{f}$] was interpreted as /f/ in [olɛɪ$_{s}^{f}$], but as /s/ in [radɛɪ$_{s}^{f}$], evidenced by the fact that most items with the ambiguous fricative triggered 'yes' responses in the lexical-decision task. On the subsequent identification test using an [ɛs]-[ɛf] continuum, in which the endpoints are the letter names of 'f' and 's' in Dutch, participants were more likely to label the ambiguous fricatives as /s/ if they had heard ambiguous fricatives in /s/ words, while participants were more likely to label the same fricatives as /f/, if they had heard ambiguous fricatives in /f/ words. At first glance, this might seem to be another instance of a simple range effect [see Repp and Liberman, 1987, for a review]: Participants adapt to the range of fricatives in the experiment. As a consequence, they label an ambiguous fricative as /f/ after good [s]s and no good [f]s were heard in the lexical-decision task, and vice versa. However, a control experiment by Norris et al. [2003] showed that the effect did not arise if the ambiguous fricative was embedded in nonwords. That is, if the participants heard examples of good [s] in words and the ambiguous fricative [$_{s}^{f}$] in nonwords, no boundary shift was observed in comparison to participants who heard the good [f] in words and the ambiguous fricative in the same nonwords. It seems that the lexicon provides the information about

which phonetic label is appropriate for an ambiguous phoneme, and this information is fed back to prelexical recognition processes, which adjust category boundaries accordingly. If an ambiguous fricative occurs in an /s/-final word, the lexical feedback leads listeners to associate ambiguous fricatives with /s/. Importantly for current purposes, the effect disappears if the test items are fricatives from a different speaker [Eisner and McQueen, 2005]. This indicates that this lexically driven adaptation of phoneme boundaries potentially contributes to speaker normalization. Indeed, Norris et al. [2003] argue that their 'laboratory' finding would be useful 'in the wild' because it allows a listener to adapt to the sociophonetic or idiosyncratic peculiarities of a given speaker.

This lexical effect may also facilitate vowel normalization in the paradigm introduced by Ladefoged and Broadbent [1957]: It is, for instance, rather uninformative that a speaker produces an F2 at 2 kHz, because nearly all speakers sometimes produce F2s at 2 kHz. If, however, the lexicon provides information that this F2 occurred in an /i/, this indicates that this speaker has a below-average F2 range. The literature on vowel normalization shows an equivocal picture with regard to the possibility of higher-level, and especially lexical, influences on vowel normalization. Watkins and Makin [1994] showed for instance that a carrier phrase played backwards gives rise to similar effects as the carrier phrase played forward. This seems to show that it does not matter whether the carrier phrase is meaningful or not. Two provisos have to be noted here, however. First of all, the effect of the backward phrase was smaller, though not significantly, than the effect with the normal phrase. Second, it is possible that listeners spotted words in the utterances played backwards. For instance, the word 'hello', part of the carrier phrase used by Watkins and Makin [1994], seems to contain the word 'well', when played backwards. Nevertheless, additional experiments by Watkins and Makin [1994] showed that vowel normalization can be achieved by a mechanism that evaluates each vowel sound in the light of the long-term average (LTA) spectrum of the preceding carrier phrase. Filtering the target by the inverse of the LTA spectrum of the low and high F1 carriers had a similar effect on perception of the target vowels as actually playing the carrier phrase before the target. Accordingly, vowel normalization may be triggered by adaptation to the LTA spectrum of a speaker.

The evidence presented by Watkins and Makin [1994] shows that auditory processes contribute to speaker normalization. However, some effects on speaker normalization cannot be captured easily by an auditory account. It is plausible that auditory processes can compensate for speaker differences based on anatomical differences, because perceptual processes have evolved to cope with such kinds of physically driven variability [cf. Kluender et al., 2001]. It is, however, less conceivable that auditory processes would be shaped just so as to compensate for sociophonetic differences, which entail more degrees of freedom than anatomically based variation. This is buttressed by an analysis by Adank et al. [2004]. They investigated how the different algorithms that have been proposed for vowel normalization fare when confronted with sociophonetic and physiological differences in vowel spaces. They found that vowel-normalization methods were able to map male and female vowel spaces, which differ due, partly at least, to anatomical reasons, onto each other, while, at the same time, regional differences between speakers were still present after normalization. Nevertheless, listeners normalize for accent as well. Evans and Iverson [2004] presented vowel targets after a carrier sentence spoken in different accents of British English. The results demonstrated that some listeners adjusted their vowel categorizations

based on the accent of the carrier sentence. Differences in language background, for instance, whether listeners grew up in the north or south of England, affected the patterns of normalization. This pattern of individual differences makes it unlikely that auditory processes can explain the result. Moreover, another nonauditory influence on speaker normalization has been reported by Johnson et al. [1999]. They presented listeners with audiovisual vowels. The visual speaker could either be male or female. Vowel categorization was influenced by this manipulation. The same vowel sound was more likely to be perceived as /ʊ/ as in *hood*, with a lower F1, than as /ʌ/ as in *hud*, with a higher F1, if the video depicted a female speaker. This mirrors the fact that women produce on average higher F1 values, so that a physically higher F1 is still perceived as comparably low, as F1 should be in /ʊ/. A similar finding was also obtained if the listeners had to imagine that a vowel, presented only acoustically, was produced by a female or a male speaker. The results obtained by Johnson et al. [1999] show that auditory processes alone cannot account for vowel normalization, because vowel normalization was induced by visual and conceptual inputs. Accordingly, both the acoustic aspects, such as the range of formant frequencies, as well as wider contextual aspects, such as the regional background or the implied gender of the speaker, are used in speaker normalization. These latter results indicate that a lexical modulation of vowel normalization is conceivable. In the present paper, therefore, I tested in two experiments whether the lexical context effect observed by Norris et al. [2003] does contribute to speaker normalization as observed in a paradigm introduced by Ladefoged and Broadbent [1957]. If this is the case, the influence of a carrier phrase on a target vowel should be larger if the constituents of the carrier phrase are words, and listeners can thus attach phonetic labels to the vowels in the carrier phrase.

Experiment 1

In this experiment, vowel normalization was tested in target words that appeared in a carrier phrase that contained a wide variety of vowels (/u, i, ɑ, ɛ/). These vowels occurred in a carrier phrase that was either a meaningful sentence or a sequence of phonologically similar nonwords. Table 1 shows all carrier phrases and targets used in this and the following experiment, with the Dutch orthographic transcription, a gloss and an IPA transcription. The task of the subject was to indicate whether the target was *keer* /ker/ ‘time’, *keur* /kør/ ‘choice’, or *koor* /kor/ ‘choir’ (three-alternative forced choice). These three vowels have a similar typical F1 and mainly differ in F2 (and F3) with /ker/ having the highest F2 and /kor/ having the lowest F2. The carrier sentence was presented with F2 as in the original utterance, with F2 increased or decreased by 20%. Given the results of Ladefoged and Broadbent [1957] and Watkins and Makin [1994], a target should be perceived as containing a vowel with a lower F2, that is, leading to more /kør/ and /kor/ responses, if the carrier phrase has an increased F2. The more interesting question is whether normalization is more effective if the vowels are embedded in words. Therefore, the lexical status of the constituents of the carrier phrase was varied. As table 1 indicates, the identical vowels were presented in different carrier phrases with different consonants. The consonants determined whether the constituents of the carrier phrase were words or not. One problematic aspect of this design is that the change of the consonantal context between word and nonword carrier phrases may also affect the perceived vowel quality. In a classic paper, Lindblom and Studdert-Kennedy [1967] showed that more members of

Table 1. Stimuli used in experiments 1 and 2

	Carrier before		Target	Carrier after
Experiment 1				
Words	toen	was hier	keer/keur/koor	gezegd
IPA	t u n ʊ	ɑ s h i ʁ	keʁ/køʁ/koʁ	g ə z ɛ xt
Gloss	then	was here	time/choice/choir	said
Nonwords	noet	fas tier		ketegd
IPA	n u t f	ɑ s t i ʁ		k ə t ɛ xt
Experiment 2				
Words	weer	is hier	keer/keur/koor	gezegd
IPA	ʋ e ʁ	ɪ s h i ʁ	keʁ/køʁ/koʁ	ɡ ə̯ z ɛ xt
Gloss	again	is here	time/choice/choir	said
Nonwords	beeg	it tier		ketegd
IPA	b e g	ɪ t t i ʁ		k ə̯ t ɛ xt

Stimulus materials for all experiments are available at: http://www.mpi.nl/world/persons/private/holmit/phonetica.html.

an /i/-/u/ continuum were identified as /u/ in the context of a /d/ than in the context of a /b/, mirroring the coarticulatory influences of the consonants on the vowels. Therefore, the consonant in the word carrier phrase and the consonant nonword carrier phrases adjacent to a given vowel had the same or a similar place of articulation (table 1), so that the influence of consonantal context on perceived vowel quality should be similar in the word and nonword carriers.

Method

Participants

Eight members of the participant pool of the Max Planck Institute for Psycholinguistics took part in the experiment. All reported normal hearing and were native speakers of Dutch.

Materials

A male native speaker of Dutch was recorded saying multiple instances of the phrases 'tun ʊɑs hir TARGET xəzext' and 'nut fɑs tir TARGET kətext', in which TARGET could be either /ker/, /kør/, or /kor/. Table 1 presents the Dutch and English orthographic transcriptions of these phrases with the different targets. In order to achieve a /ker/-/kør/-/kor/ continuum, the vocalic parts of the three different target words were excised from a word carrier and formant frequencies were analyzed. This analysis (fig. 1) indicated that all vowels had a similar F1 trajectory starting from 380 Hz after the stop rising to 550 Hz into the final consonant. F2 midpoint ranged from 2,200 Hz in the vowel [e] to 800 Hz in [o]. The transition into the last consonant converged to, but did not reach, a locus of 1,500 Hz for all three vowels. F3 was stable for [ø] and [o] at about 2,350 Hz. For [e], F3 started at 2,850 Hz and fell to 2,620 Hz. Based on this analysis, 11 targets with a duration of 170 ms were generated using a Klatt synthesizer [Klatt, 1980]. In all targets, F1 rose from 380 Hz at 0 ms to 440 Hz at 50 ms and to 550 Hz at 130 ms in all targets. F4, F5, and F6 were fixed at 3,385 Hz, 4,720 Hz, and 6,000 Hz, respectively. F2 and F3 were manipulated as indicated in table 2. The first target (target 0) had formant frequencies that corresponded to the natural vowel [o], the last target (target 10) had formant frequencies that corresponded to the natural vowel [e]. Interpolation of F2 in bark distances showed that an F2 corresponding to the natural vowel [ø] was reached for target 6. Accordingly, F3 was kept constant from target 0 to target 6 – emulating F3 in [o] and [ø], and then interpolated to the onset and offset values observed in the natural vowel [e] in target 10. Following the natural utterances, formant frequencies were stable at

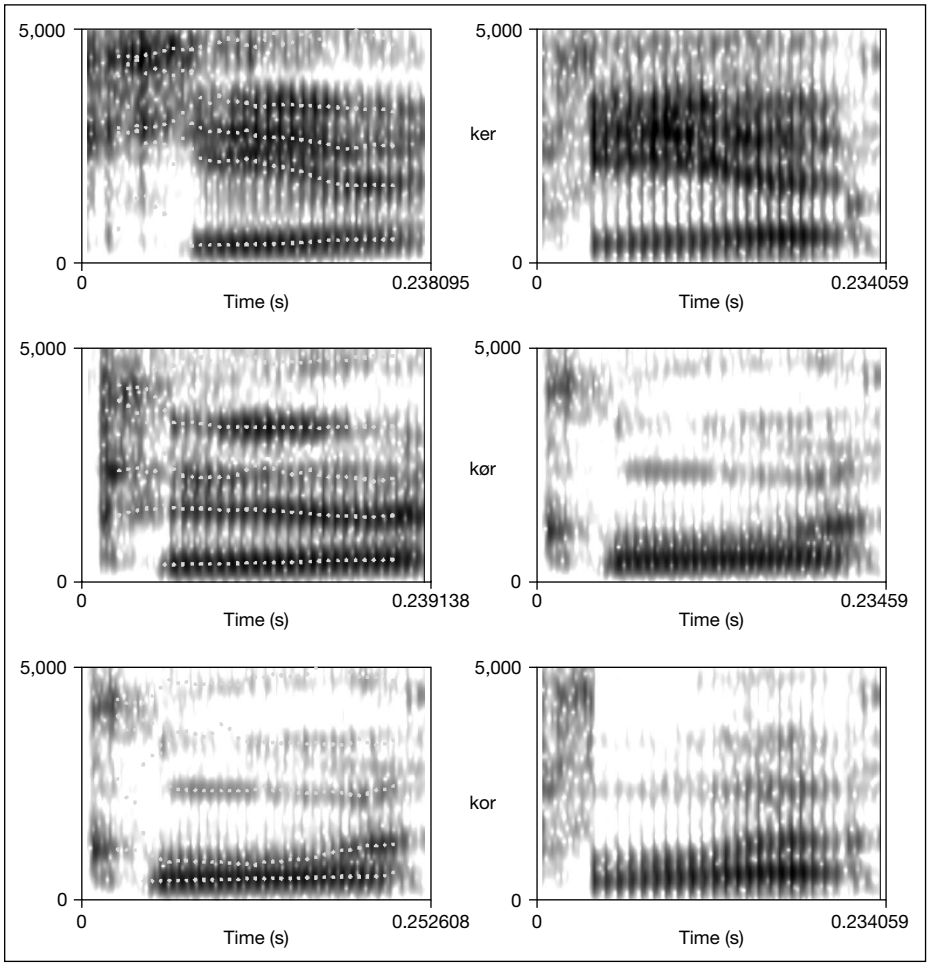

Fig. 1. Natural utterances of [ker], [kør], and [kor] produced by a male speaker of Dutch with overlaid formant traces (left) and targets with altered natural consonants and re-synthesized vowels (right).

their onset positions for the first 50 ms, then changed to their offset positions at 130 ms, where they remained for the last 40 ms.

In order to have nonbiasing consonants, we did not use the consonantal parts of one of the target utterances. For the onset, the different waveforms for the [k] releases from all three target words were excised, attenuated by one third, and mixed together aligned at release, giving rise to an average /k/-release burst. The final consonant /ɾ/ was realized mostly as an uvular fricative [ʁ] or as an approximant. In order to get a stable percept of a word-final /ɾ/ in all targets, an [ʁ] frication was appended from a different utterance of [hiʁ]. The eleven synthesized vowels were concatenated with the edited natural consonantal onset and coda to yield the eleven targets.

The carrier phrases were constructed using the natural consonantal portions from one of the speaker's tokens of the word or the nonword carrier sentence. Formant trajectories were analyzed in the natural vowels and used as templates for vowel synthesis. Parameters for the neutral carrier imitating the natural utterance are displayed in table 3. For the low- and high-F2 version of the carrier, F2

Table 2. Parameters for target synthesis (in Hertz)

Target No.	F2 onset	F2 offset	F3 onset	F3 offset
0	800	1,233	2,350	2,350
1	896	1,265	2,350	2,350
2	999	1,300	2,350	2,350
3	1,110	1,336	2,350	2,350
4	1,229	1,376	2,350	2,350
5	1,358	1,419	2,350	2,350
6	1,498	1,466	2,350	2,350
7	1,651	1,517	2,464	2,414
8	1,817	1,572	2,586	2,480
9	1,999	1,633	2,714	2,549
10	2,200	1,700	2,850	2,620

Table 3. Parameters for vowel synthesis in the neutral carrier phrase in experiment 1

Vowel		Time	F1	F2	F3	F4	F5
[u]	onset	0	280	1,400	2,500	3,600	4,900
	mid	25–40	380	1,100	2,400	3,300	…
	offset	70	…	1,300	2,200	…	…
[ɑ]	onset	0	500	1,220	2,450	3,700	4,700
	mid	20–35	585	1,300	…	…	…
	offset	70	550	1,400	2,500	3,800	…
[i]	onset	0	300	2,200	2,800	3,500	4,700
	mid	45–90	320	2,150	…	…	…
	offset	155	410	1,700	2,350	3,200	…
[ə]	onset	0	450	1,550	2,400	3,400	4,300
	mid	20	400	…	2,500	…	…
	offset	55	380	…	…	…	…
[ɛ]	onset	0	350	1,680	2,500	3,650	4,650
	mid	30	450	1,500	2,400	3,450	…
	offset	70	500	1,560	2,300	…	…

Continuation points '…' indicate no change.

was increased or decreased by 20%. In order to prevent unnatural formant constellations, F3 was set at 1.2 times the F2 value, if the original value was within a 20% range of the manipulated F2. Otherwise, F3 remained unchanged. These synthesized vowels were concatenated with the natural consonant onsets and codas to produce the preceding and following parts of the carrier phrases. Informal listening tests indicated that this F2 change did not lead to the perception of a phonemically different vowel. With three versions of the vowels (low, medium, and high F2) and the two sets of consonantal portions (forming words and nonwords), this gives rise to six carrier phrases.

Procedure

Experiments were run on a standard PC with the NESU (Nijmegen Experimental Set-Up) package. Participants were wearing headphones and faced a computer screen with a four-button response box in front of them. They were told that they would hear a sentence in which the penultimate word was either *keer*, *keur*, or *koor*. It was stressed that the sentences could contain nonsense words. They were told to press the leftmost button if they perceived the word as *keer*, the left middle button if they perceived *keur*, and the right middle button if they perceived *koor*. These response allocations were fixed and appeared on the screen at the beginning of each experimental trial.

The trials had the following structure. After 150 ms of blank screen, the three response alternatives were presented on the screen. After another 450 ms, the sentence was played. From the onset of the target word, participants had 2.5 s time to respond. After a response, the visual display of the response alternatives disappeared. If a participant failed to respond within 2.5 s, a stopwatch appeared as a sign to speed up responses.

Each of the 66 sentences – six carrier phrases crossed with eleven targets – was presented six times to each participant. Participants had the opportunity to take short breaks after 50 trials. An experimental session lasted about 25 min.

Design

The experiment entailed three independent variables: F2 range (low, medium, high), Lexical Status of the constituents of the carrier phrase, and Target (11 levels). The dependent variable was the degree of backness of the vowel choice with three ordinal levels (front: /ker/, medium: /kør/, back: /kor/).

Results and Discussion

Figure 2 displays, in line with the ordinal regression reported below, the aggregated proportion of perceived vowel frontness for each combination of Lexical Status (words, nonwords), F2 range (different symbols), and Target vowel (ordinate). The continuous lines represent the likelihood that the vowel was perceived as either /ø/ or /e/, that is more front than /o/, or 100% minus percentage /o/ responses. The dotted lines represent the likelihood that the vowel was perceived as /e/, that is more front than /ø/ and /o/. Accordingly, the areas under the dotted lines represent the proportion of /e/ responses, the areas between the dotted and the continuous lines the proportion of /ø/ responses, and the areas above the continuous lines the proportion of /o/ responses. The results show that the continuum endpoints were identified as intended /o/ and /e/, respectively. Moreover, targets in the middle of the continuum are recognized almost exclusively as /ø/.

Results were analyzed with ordinal logistic regression [Agresti, 1989], in which perceived vowel frontness was coded as 0 (=/o/), 1 (=/ø/), or 2 (=/e/). This analysis generates an aggregate of two logistic regressions in which the three ordinal response categories are differently assigned to a binary coding: once for responses /o/ vs. not /o/, that is, either /ø/ or /e/, and once for responses /e/ vs. not /e/, that is, either /ø/ or /o/. Predictors in these analyses were target F2 (F2 midpoint in bark), F2 range (mean F2 in bark) as ordinal covariates, and Lexical Status as a categorical factor. Besides main effects, also the interaction between Lexical Status and F2 range was entered in the models. A significant positive β-weight for this interaction would indicate that the effect of the F2 range in the carrier sentence is larger in the carrier consisting of words. Ordinal regression models were applied to the data of each participant, and the significance of each parameter for the whole group was assessed with a t test of the individual parameters against 0.

The analysis revealed a significant *positive* β-weight for Target F2 (mean $\beta = 4.76$, SD = 1.62, $t(7) = 7.76$, $p < 0.001$), and a significant *negative* β-weight for F2 range ($\beta = -0.47$, SD = 0.30, $t(7) = -4.15$, $p < 0.01$). Neither the main effect of Lexical Status ($\beta = 0.44$, SD = 2.99, $t^2 < 1$) nor its interaction with F2 range ($\beta = -0.03$, SD = 0.27, $t^2 < 1$) reached significance. The positive β-weight for Target F2 indicates that a higher F2 triggers more front-vowel identification. The negative β-weight for F2 range in the carriers means that the current experiment replicates Ladefoged and Broadbent [1957]: The F2 range in the sentence has a compensatory influence on the vowel decision. The lower the mean F2 in the sentence, the more likely

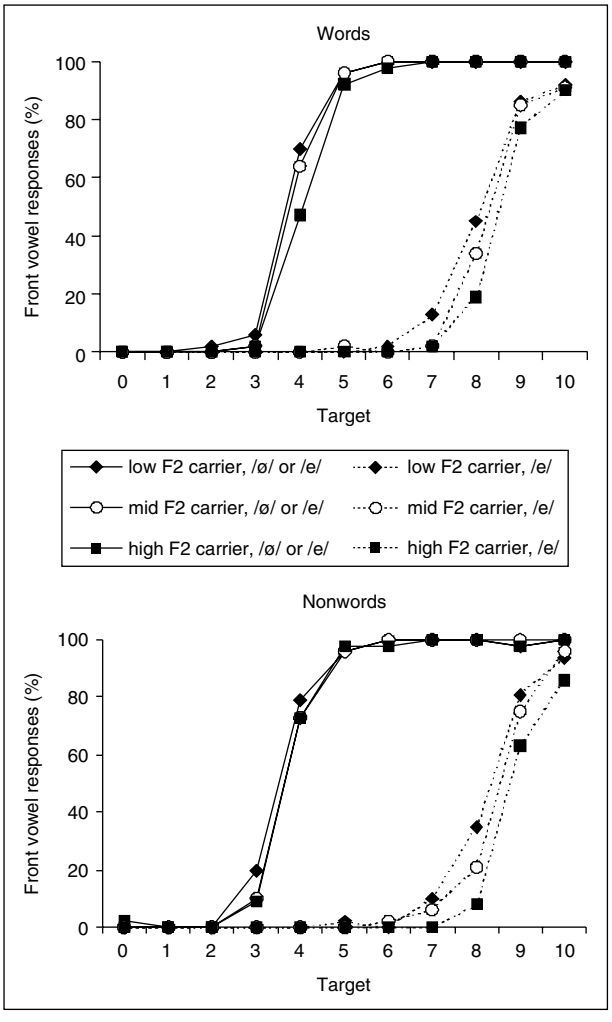

Fig. 2. Vowel identifications in experiment 1 depending on the lexical status of the constituents of the carrier phrase (upper and lower panels), the F2 range in the carrier phrase (symbols) and the target (abscissa). Solid lines show the proportion of /ø/ or /e/ responses, dotted lines the proportion of /e/ responses.

the target vowel is to be perceived as more front, that is, with a higher apparent F2. This effect was not modulated by the Lexical Status of the words containing the vowels with varying F2 range. The perception of the target word depends on the F2 in the carrier phrase equally after words as after nonwords.

In order to evaluate whether the effect of the carrier F2 range was restricted to a subset of the target vowels, simple logistic-regression models were used to determine the perceptual boundary between /o/ versus /ø/ (excluding the /e/ responses) and the perceptual boundary between /ø/ versus /e/ (excluding the /o/ responses). As predictors, Target F2 and F2 range were used. Models were applied to the individual data sets and the significance of the parameters assessed with t tests. This procedure revealed a significant effect of F2 range for the /o/-/ø/ boundary (mean $\beta = -0.35$, SD = 0.35, $t(7) = -2.69$, $p < 0.05$) and the /ø/-/e/ boundary (mean $\beta = -0.63$, SD = 0.61, $t(7) = -2.75$, $p < 0.05$).

The current experiment did not provide evidence that the lexical status of the carrier sentence moderates the effect of vowel normalization. At the same time, the current experiment showed a vowel normalization effect. It was noted in the 'Introduction' that vowel normalization probably rests on both auditory as well as higher-level mechanisms [see Watkins and Makin, 1994, and Johnson et al., 1999, respectively]. The current experiment may, however, have been designed so as to stack the deck in favor of signal-based normalization, and, as a consequence, leave little room for higher-level mechanisms to moderate the auditory effects. This argument is based on the standard finding in phonetic categorization experiments that extraauditory information – such as lexical or visual information – tends to have the strongest influence if the auditory signal is ambiguous [Massaro, 1998]. In the current case, the carrier phrase contained the two corner vowels [i] and [u], as well as [a]. This allows listeners to get a good estimate of both F1 and F2 range of the speaker, so that phonetic labels provided by the lexicon could not further help vowel normalization. A lexical effect may nevertheless be obtainable if the carrier phrase contains a narrower sample of vowels. I therefore ran another experiment in which the carrier phrase contained mid-to-high front vowels only. In this case, a signal-based strategy may not be completely successful if there is no information about the phonetic labels of the received mid-to-high front vowel signals. Embedding the vowel in words may provide the listener with those phonetic labels, which in turn could help the listener to interpret the small range of F1 and F2 encountered in the carrier phrase. Accordingly, a lexical effect on vowel normalization may occur if the carrier phrase contains only a small subset of the vowel space.

Experiment 2

In this experiment, the same targets as in experiment 1 were used, while the carrier phrases were different. The carrier phrases now only contained the mid-to-high front vowels [i], [ɪ], [e], and [ɛ], as the vocalic, i.e. voiced part, of the schwa in the first syllable after the target was cut out. This is indicated by the diacritic mark in table 1 that indicates reduction. Such reductions are commonplace in speech production [Kohler, 2000; Shockey, 2003], but do not inhibit word recognition, as the coarticulatory cues in the surrounding consonants are sufficient for listeners to perceive the presence of a schwa [Manuel, 1992]. As regards the effect of carrier phrases with a restricted vowel repertoire, it is noteworthy what Ladefoged and Broadbent [1957, p. 102] pointed out: 'Nor do we know to what extent it is necessary to use an introductory sentence containing a wide variety of vowels which may serve as reference points'. But this is still an open issue. Verbrugge et al. [1976] tested whether the corner vowels have a special status in vowel normalization, by using either three peripheral vowels /I, a, u/ or the more central /ɪ, œ, ʌ/ as precursors. They failed to find evidence for a special status of the former in vowel normalization. It should be noted, however, that they still used a wide variety of vowels dispersed over the vowel space in all conditions. It is still unclear how vowel normalization is affected if only vowels from the same region of the vowel space, such as mid-to-high front vowels, are presented in a carrier phrase. This experiment investigates this question, and, in addition, tests whether the lexical status of the constituents of the carrier phrase influences the amount of vowel normalization when the range of vowels in the carrier is smaller than in experiment 1.

Table 4. Parameters for vowel synthesis in the neutral carrier phrase in experiment 2

Vowel		Time	F1	F2	F3	F4	F5
[e]	onset	0	450	1,800	2,450	3,450	5,500
	mid	50–140	480	2,000	2,666	3,500	…
	offset	205	430	2,160	2,600	…	…
[ɪ]	onset and mid	0–45	350	2,100	2,800	3,800	4,700
	offset	65	…	1,800	2,600	…	…

Parameters for [i] and [ɛ] as in experiment 1.

Method

Participants

Ten members of the participant pool of the Max Planck Institute for Psycholinguistics took part in the experiment. All reported normal hearing and were native speakers of Dutch. None of them took part in experiment 1.

Materials

The same target stimuli as in experiment 1 were used. In order to construct new carrier phrases, the same male native speaker of Dutch as in experiment 1 was recorded saying multiple instances of the phrases 'ʋeʁ ɪs hir TARGET xəzext' and 'bex ɪt tir TARGET xəzext'and 'bex ɪt tir TARGET kətext', in which TARGET could be either /ker/, /kør/, or /kor/. Table 1 presents the Dutch and English orthographic transcriptions of these phrases. From two of these utterances, the vocalic parts of the carrier portions were excised and formant frequencies were analyzed. The carrier phrase was again constructed using the natural consonantal portions for both the word and nonword carrier sentence. Formant trajectories were analyzed in the natural vowels and used as templates for the vowel synthesis. Parameters for the neutral carrier imitating the natural utterance are displayed in table 4. For the low- and high-F2 versions of the carrier, F2 was increased or decreased by 20%. In order to prevent unnatural formant constellations, F3 was set at 1.2 times the F2 value, if the original value was within a 20% range of the manipulated F2. Natural consonantal parts and synthesized vowels were concatenated as in experiment 1.

Procedure and Design

The procedure and design were the same as in experiment 1, with three independent variables: F2 range (low, medium, high), Lexical Status of the constituents of the carrier phrase, and Target (11 levels). The dependent variable was the degree of frontness of the vowel choice with three ordinal levels (front: /ker/, medium: /kør/, back: /kor/).

Results and Discussion

Figure 3 displays the aggregated proportion of perceived vowel frontness for each combination of Lexical Status (upper panel: words, lower panel: nonwords), F2 range (different symbols), and Target vowel (ordinate). As in figure 2, the continuous lines represent the likelihood that the vowel was perceived as 'more front' than /o/, that is, as either /ø/ or /e/. The dotted lines represent the likelihood that the vowel was perceived as 'more front' than /ø/, that is, as /e/. Accordingly, the areas under the dotted lines represent the proportion of /e/ responses, the areas between the dotted and the continuous lines the proportion of /ø/ responses, and the area above the continuous lines the proportion of /o/ responses.

As in experiment 1, results were analyzed with ordinal logistic regression [Agresti, 1989], in which perceived vowel frontness was coded as 0 (=/o/), 1 (=/ø/), or

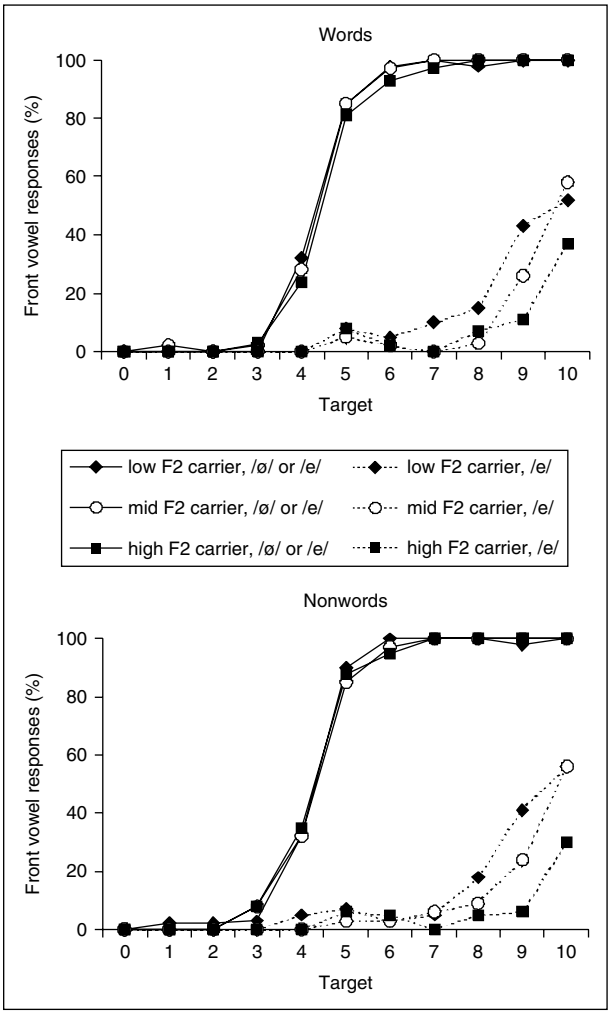

Fig. 3. Vowel identifications in experiment 2 depending on the lexical status of the constituents of the carrier phrase (upper and lower panels), the F2 range in the carrier phrase (symbols) and the target (abscissa). Solid lines show the proportion of /ø/ or /e/ responses, dotted lines the proportion of /e/ responses.

2 (=/e/). Target F2 (F2 midpoint in bark), F2 range (mean F2 in bark) as covariates, and Lexical Status as a categorical factor were used as predictors. Besides main effects, the interaction between Lexical Status and F2 range was also entered in the model. The analysis revealed a significant *positive* β-weight for Target F2 (mean $\beta = 3.45$, $SD = 0.99$, $t(9) = 10.45$, $p < 0.001$), and a significant *negative* β-weight for F2 range ($\beta = -0.36$, $SD = 0.41$, $t(9) = -2.64$, $p < 0.05$). Neither the main effect of Lexical Status ($\beta = 0.49$, $SD = 2.33$, $t^2 < 1$) nor its interaction with F2 range ($\beta = -0.03$, $SD = 0.29$, $t^2 < 1$) reached significance.

In order to evaluate whether the effect of the carrier F2 range was restricted to a subset of the target vowels, simple logistic-regression models were used to determine the perceptual boundary between /o/ versus /ø/ (excluding the /e/ responses) and the perceptual boundary between /ø/ versus /e/ (excluding the /o/ responses). As predictors, Target F2 and F2 range were used. Models were applied to the individual data sets and

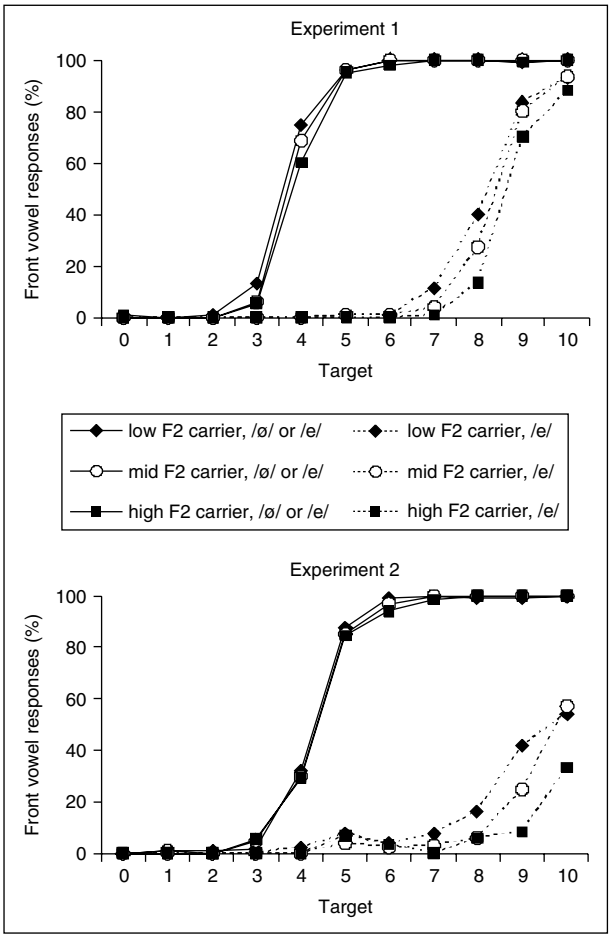

Fig. 4. Vowel identifications in experiment 1 (upper panel) and experiment 2 (lower panel) aggregated over the lexical status of the carrier phrase depending on the F2 range in the carrier phrase (symbols) and the target (abscissa). Solid lines show the proportion of /ø/ or /e/ responses, dotted lines the proportion of /e/ responses.

the significance of the parameters assessed with t tests. This procedure revealed no significant effect of F2 range for the /o/-/ø/ boundary (mean $\beta = 0.01$, SD $= 0.50$, $t^2 < 1$), but a significant effect on the /ø/-/e/ boundary (mean $\beta = -0.65$, SD $= 0.69$, $t(6) = -2.51$, $p < 0.05$, 3 participants were excluded because they gave no or less than 10 /e/ responses, preventing a valid model fitting for these participants).

In comparison with experiment 1 (see figure 4 for a compact comparison of the results of both experiments), the current experiment replicates the vowel-normalization effect. In contrast to experiment 1, however, the F2 range does not have an influence on the /o/-/ø/ boundary, but only on the /ø/-/e/ boundary. This shows that vowel normalization is vowel-dependent. If the carrier phrase only consists of mid-to-high front vowels, phoneme boundaries are only adjusted for mid-to-high front vowels. Even if phonetic labeling of the altered vowels in the carrier phrase is supported by using words in the carrier phrase, listeners do not adjust all vowel boundaries, but only those in the vicinity of the vowels encountered in the carrier phrase.

The clearest difference with the results of experiment 1 is, nevertheless, the small percentage of /e/ responses in this experiment. While the percentages of /e/ responses

approached ceiling level in experiment 1, no more than 60% of /e/ responses were observed in any of the cells in this experiment. There are at least two forces that may bias listeners against perceiving the target vowel as /e/. First of all, the /k/ release for the synthesized target words was generated by averaging the three /k/ releases from the words [ker], [kør], and [kor]. Two of these words contain rounded vowels, and accordingly the average /k/ release contains some cues for anticipatory lip rounding. Listeners use such coarticulatory cues to identify a following vowel as rounded or not [see, e.g., Gow, 2001; Mitterer, in press], and this leads to a bias against identifying the vowel in the target words as the unrounded /e/. This alone, however, cannot be sufficient to explain the small percentage of /e/ responses in this experiment, because a higher percentage of /e/ responses was observed with the same targets in experiment 1. In this second experiment, an additional factor might be that the listeners heard three mid-to-high front vowels in the carrier phrase, which also biases against the perception of the target vowel as high-front, akin to the selective adaptation effect [Eimas and Corbit, 1973]. It needs to be stressed, however, that this adaptation may not occur on a phonological level of processing – as the wording above implies – but may just as well be explained by adaptation on an auditory level to high F2 frequencies [see Remez, 1987, for a discussion on the levels of selective adaptation].

Similar to the question of the level of adaptation, it also needs to be considered whether the effects of vowel normalization arise at an auditory or a more abstract level of processing. There is at least one predecessor to the current finding that context effects in vowel perception are moderated by the distance in vowel space, which would point to a possible auditory effect. Thompson and Hollien [1970] tested whether identification of a second vowel in a vowel-vowel sequence was influenced by the first vowel. They found a contrastive effect, such that a vowel was identified as higher (i.e., with a lower F1) if the preceding vowel had a higher F1 than the target vowel. This contrastive effect got smaller as the vowels in a sequence were more dissimilar. Such a simple contrastive effect cannot explain the current pattern of results for the simple reason that the immediately preceding context in experiment 1 and experiment 2 was identical (table 1). The conclusion would have to be that the first vowel /u/ in the carrier phrase in experiment 1 influenced the perception of the target vowel despite the three intervening vowels. Such a long-distance effect points towards some form of more abstract normalization, in which the speaker's vowel space is warped onto a 'standard' vowel space [for an overview and comparison of such methods, see Adank et al., 2004]. However, Holt [2005] recently presented evidence that auditory aftereffects may come in shapes more complex than local contrast effects. She introduced an 'acoustic-history' design, in which a series of pure tones is presented prior to a stop-vowel syllable to be identified. All acoustic histories used in these experiments ended on the same tone and only differed with respect to the mean frequency of the earlier tones. Still, the identification of the speech syllable was influenced contrastively by the acoustic history. In order to test the possibility that the pattern of results in experiment 1 could be due to such an 'acoustic-history' effect, an additional experiment was run.

Experiment 3

The purpose of this experiment was to test whether adaptation to 'acoustic history' [Holt, 2005] may explain the vowel normalization observed in experiment 1, especially

the nonlocal effect the first back vowel had on the /o/-/ø/ boundary. Therefore, nonspeech analogues of the carrier phrases in experiment 1 were generated. These analogues had the same long-term spectral properties and the same amplitude contour and overall amplitude as the speech sounds. If target identification is influenced by these nonspeech analogues, just as by speech, one may conclude that auditory processing is instrumental in achieving the vowel normalization observed in experiment 1. It may be argued that by the introduction of nonspeech material, the task may be so different from the one in the previous experiments to make them incommensurable. Indeed, some have questioned the usefulness of such speech-nonspeech comparisons [see especially Fowler, 1990, 2006]. Nevertheless, a theoretical and an empirical argument may be brought forward to foster the logic of the speech-nonspeech comparison. First of all, the task in the current experiment is basically the same as in the previous experiments: phonetic identification of speech materials. While the nature of the context sounds changes from speech to nonspeech, the participants perform a rather similar task on the same target stimuli. Secondly, it is the strength of the auditory approach that despite the dramatic change in the quality of the context sounds, the context effects caused by nonspeech sounds are sometimes comparable to the context effects caused by speech sounds [see, e.g., Lotto and Kluender, 1998; Mitterer et al., 2006]. Hence, it is worthwhile to investigate the effect of nonspeech context sounds on the perception of the speech-target sounds in this third experiment.

Method

Participants

Fifteen members of the participant pool of the Max Planck Institute for Psycholinguistics took part in the experiment. All reported normal hearing and were native speakers of Dutch. None of them took part in experiment 1 or 2.

Materials

The same target stimuli as in experiment 1 were used. Nonspeech carrier phrases were constructed on the basis of the LTA spectrum (bandwidth = 5 Hz) of the word carriers used in experiment 1. This gives rise to six LTA spectra: one for each F2 range (low, medium, high) for both parts of the carrier phrase, preceding and following the target (table 1). Stretches of white noise with the same durations as the two parts of the carrier were then generated and converted to dft spectra. The noise spectra were filtered by one of the LTA spectra in the following way. A filter coefficient for each bin (e.g., at 102 Hz) of the noise spectrum was calculated from the two values of the LTAs spectra of the speech sound using the bins above or below the frequency of the bin (100 and 105 Hz) with linear interpolation. After this filtering, the spectra were again converted to sound, multiplied by the intensity contour of the original speech sounds and amplitude-scaled so the overall amplitude (i.e., root-mean-square value of the sound samples) was the same for speech and nonspeech carriers. These filtered sounds still have the quality of noise stimuli. Visual inspection confirmed that the nonspeech carriers had very similar LTAs spectra as the speech carriers. Thirty-three experimental 'sentences' were then constructed embedding one of the 11 targets in one of the three noise carrier phrases derived from low, medium, and high F2 speech carrier phrases.

Procedure and Design

The procedure and design were similar to experiment 1. Given the nonspeech nature of the carrier, there was obviously no manipulation of the Lexical Status of the constituents of the carrier phrase, leaving two independent variables: F2 range (low, medium, high), and Target (11 levels). The dependent variable was the degree of frontness of the vowel choice with three ordinal levels (front: /ker/, medium: /kør/, back: /kor/).

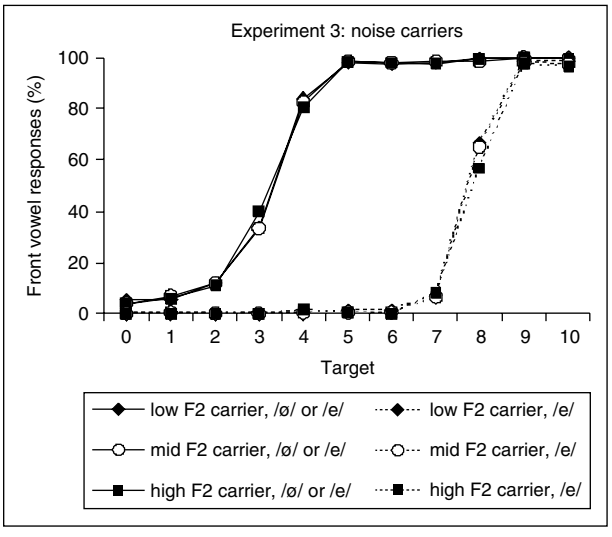

Fig. 5. Vowel identifications in experiment 3. Solid lines show the proportion of /ø/ or /e/ responses, dotted lines the proportion of /e/ responses.

Results and Discussion

Figure 5 shows the results of experiment 3. As in the previous figures, the continuous lines represent the likelihood that the vowel was perceived as 'more front' than /o/, that is, as either /ø/ or /e/. The dotted lines represent the likelihood that the vowel was perceived as 'more front' than /ø/, that is, as /e/. The areas under the dotted lines hence represent the proportion of /e/ responses, the areas between the dotted and the continuous lines the proportion of /ø/ responses, and the area above the continuous lines the proportion of /o/ responses. The descriptive data do not reveal any clear effect of the nonspeech carriers on vowel identification.

The ordinal-logistic-regression analysis was performed as in the previous experiments, vowel frontness as dependent variable was coded as 0 (=/o/), 1 (=/ø/), or 2 (=/e/), and Target F2 (F2 midpoint in bark), and F2 range (mean F2 in bark) were used as predictors. The analysis revealed a significant *positive* β-weight for Target F2 (mean $\beta = 4.35$, $SD = 2.21$, $t(14) = 7.34$, $p < 0.001$), but no significant β-weight for F2 range ($\beta = -0.05$, $SD = 0.14$, $t(14) = -1.56$, $p = 0.14$).

In the current experiment, the different carrier phrases failed to influence vowel identification significantly. Can one then accept the null hypothesis that auditory processes do not contribute to vowel normalization? Certainly not, because the nonsignificant effect of the nonspeech carriers has the same direction as the effect of the speech carriers in experiment 1. Nevertheless, the effect failed to reach significance in this experiment despite the fact that the sample size was larger than in experiment 1. Moreover, the effect of the nonspeech carriers is much and significantly smaller than the effect of the speech carriers (nonspeech: $m = -0.056$; speech: $m = -0.467$; $t(24) = -4.11$, $p < 0.001$). So, it seems safe to conclude that auditory processes alone cannot account for the vowel-normalization effect in experiment 1. (It needs to be noted that researchers arguing for the usefulness of auditory processing in overcoming the invariance problem never argued that auditory processing was sufficient [see, e.g., Holt et al., 2001, Mitterer et al., 2006].)

General Discussion

In two experiments, possible lexical effects on vowel normalization [Ladefoged and Broadbent, 1957] were tested. In experiment 1, a carrier phrase with a diversity of vowels was used, and manipulating F2 in this carrier led to a compensatory vowel-normalization effect on a F2 test continuum. Listeners were more likely to perceive a vowel which contains – within a given speaker's utterances – a lower F2, that is either /o/ or /ø/ rather than /e/ if the carrier phrase contained an elevated F2 range. This effect was independent of the lexical status of the constituents of the carrier phrase. In experiment 2, the carrier phrase again consisted of either words or nonwords, but only contained high-front vowels rather than a diversity of vowels. This design change altered two aspects of the results. First of all, a vowel-normalization effect was only observed for high-front vowels. This indicates that vowel normalization is vowel-specific: If only high-front vowels are encountered in a carrier phrase, only phoneme boundaries of high-front vowels are adjusted. Secondly, presenting a series of high-front vowels in the carrier phrase triggered selective adaptation, so that overall fewer high-front vowels were reported. This selective adaptation did, however, not obliterate the effect of F2 range in the carrier phrase on the /ø/-/e/ boundary. A third control experiment showed that the results cannot be attributed to auditory processing alone.

The main impetus of these experiments was to explore a possible relation between the 'adaptation-to-speaker' effects reported by Norris et al. [2003] on the one hand and Ladefoged and Broadbent [1957] on the other. The results indicate that these short-term adaptations [Ladefoged and Broadbent, 1957] and medium-term adaptations [Norris et al., 2003] are independent. Lexical feedback to prelexical processing seems not to help in vowel normalization. So with regard to the current results, we can answer the question whether vowel normalization is independent of lexical processing with 'no'. It is important to note, however, that the manipulation of F2 range in the current experiment more or less resembles anatomically grounded speaker variation. It is possible, if the manipulation of the carrier phrases differed in more idiosyncratic ways, resembling sociophonetic variation, that a lexical effect may nevertheless still be observed. But the current experiments show that the effects of formant range observed by Ladefoged and Broadbent [1957] are independent of the lexical status of the constituents of the carrier phrase.

The current results also speak to the cue-weighting of intrinsic and external cues in vowel normalization. Nearey [1989] conducted an experiment investigating the cue-weighting for intrinsic cues, such as f0 and higher formants, and extrinsic cues, such as F2 range, in vowel normalization. He concluded that both cues contribute to vowel normalization, but 'it is clear the extrinsic ensemble effect dominates the changes' [Nearey, 1989, p. 1201]. The current results indicate that the cue-weighting for the extrinsic factor is, however, dependent on the similarity of the vowels in the carrier phrase and the target vowel. Extrinsic factors may only play a role in vowel perception if the listener has been exposed to similar vowels previously, so that the cue-weighting strategies for extrinsic and intrinsic cues are in fact dynamic and not static.

One noteworthy aspect of the current results is that a difference in results occurred between experiment 1 and experiment 2, although the context immediately preceding the target word is identical in both experiments (tables 1, 4). This indicates that not only the immediate context is used in vowel normalization but also the wider sentence context. This is a necessary correlate of the finding that the range of vowels in the carrier

phrase influences vowel normalization. This result has a wider implication for current theories of speech perception which embrace an exemplar approach [e.g., Goldinger, 1998]. The basic tenet of this approach may best be explained in the light of the trade-off between storage and computation in any cognitive model. This trade-off may be best exemplified with an example from morphological processing. In order to generate a derived form, for instance a past tense of a verb, a cognitive model may argue that the complete form is either stored, requiring little computation, or computed from the base form, requiring little storage. The focus within linguistics has classically been on 'computation', assuming a rather limited capacity for storage. More or less in parallel with the development of computers, however, experimentation showed that human storage capacity is much larger than previously envisioned [Landauer, 1986] and that long-term memories often entail surface details of presented stimuli [Snodgrass et al., 1996]. There is also clear evidence that listeners retain detail episodes of spoken words in long-term memory [Goldinger, 1996]. Based on these findings, it has been proposed that the mental lexicon is nothing more than the collection of these episodic traces. In such models, speaker normalization may seem unnecessary, because the listener is not trying to access an abstract, invariant code. Similarly as in morphological processing, the computation of speaker normalization is replaced by the explicit storage of episodic traces of words or vowels stored with all fine phonetic detail encountered. The current results pose three problems for models that assume that the mental lexicon is composed of nothing but episodes consisting of the raw perceptual input without normalization; two problems of a more general nature, and one with a specific, but rather prominent, implementation of an episodic model. First of all, it is difficult to account for the fact that vowel normalization occurs at all in the paradigm of Ladefoged and Broadbent [1957]. How can the recognition of a given word depend on formant frequencies of vowels in surrounding words if lexical representations in existing episodic models see the word as the basic unit of storage? No normalization is supposed to occur, and the raw perceptual input is compared with previous episodes. In order to account for normalization effects as in the present experiment, it becomes necessary to store the context along with the target vowel, which would quadruple the 'head-filling problem', that is, the problem of storage of a vast amount of data. This is, because not only would at least the two previous vowels need to be stored with every target vowel, but also the following vowel [Watkins and Makin, 1996], so that for the long-term representation of each vowel encountered, four would have to be encoded. A second problem arises from the fact that episodic models are driven by similarity between representations. This makes it difficult to account for generalized normalization processes. The episodic memory of the word *keer* in one carrier phrase would be rather dissimilar to an episode with the same word in another carrier phrase. In an episodic model, the relationship between the formants in different surrounding vowels could not be recognized. Accordingly any vowel normalization that was acquired by experience with a certain carrier phrase cannot easily be generalized to another carrier phrase. Note that this argumentation is not targeted at the idea of episodic storage, but rather at the idea that normalization is superfluous. Models of episodic storage that assume unnormalized spectrograms as the input to the lexicon still have not faced the challenge that Klatt [1989, pp. 169–170] recognized for – nota bene – his own [Klatt, 1979] model of speech perception: '…was discouraged by the behavior of the distance metrics available to compare spectra. These metrics were as sensitive to irrelevant spectral variability as to cues to fine phonetic distinctions.'

A more specific problem arises with the precise implementation of the most widely cited episodic model of Goldinger [1998]. At a closer look, it becomes apparent that this model turns out to imply some process of speaker normalization after all. Words are assumed to be represented by a vector with 100 scalars representing the 'name' of the word, 50 scalars representing the voice, and 50 scalars representing the situational context. Interestingly, 'the name elements were identical for all 20 tokens of each word; voice and context elements were generated randomly' [Goldinger, 1998, p. 254]. The model was thus provided with an invariant code for word meaning that was independent of the speaker. The raison d'être for normalization procedures is, however, that such an invariant code does not exist because, as laid out in the 'Introduction', formant frequencies are influenced simultaneously by speaker differences and vowel differences. Stated otherwise, the name and voice elements in episodic instantiations of a given word are not independent as assumed in Goldinger's [1998] model. The nature of the input to the model thus in fact necessitates that some 'decomposition' of the raw perceptual data takes place before lexical access is attempted. Accordingly, the statement that 'although many theories consider normalization a logical necessity, episodic models provide an alternative' [Goldinger, 1998, p. 264] should be supplemented by the proviso 'if an invariant code can be provided for lexical access'. This invariant code, however, cannot be provided without speaker normalization.

The fact that linguistic, anatomical, and sociolinguistic variables all simultaneously influence vowel formant frequencies does indeed logically necessitate some form of normalization in order to receive any of the linguistic, anatomical, or sociolinguistic information in the signal. This, however, does not lead to a general plea against episodic models, but rather as a plea against models that do not assume normalization. Johnson [1997] proposed an episodic model in which the storage of the raw perceptual input is connected to speaker identity and category identity, and vowel normalization is achieved by raising the base activation level for all vowels associated with a given speaker, once the speaker is recognized. In the implementation of this model, it first had to be trained on exemplars for which vowel and speaker identity were provided by the training regime. Hence, this model, despite being episodic, incorporates some form of speaker normalization. Goldinger [1998] also sees the possibility of a hybrid model, in which normalization occurs, but a procedural record is stored alongside the normalized value. This would mean that a vowel is mapped onto a kind of canonical vowel space and the correction applied to the raw formant frequencies is retained. Such a model would obviously be in a much better position to actually store episodic traces of fine phonetic detail: Adank et al. [2004] showed that regional variation can in fact be appreciated better after normalization for vocal-tract differences has occurred. This again follows quite logically from the fact that formant frequencies carry multiple sources of information. In order to appreciate that a given speaker fronts a given vowel, it is necessary to know the F2 range of a speaker. Accordingly, a 'raw' episodic model of speech perception is unlikely to succeed in storing fine phonetic detail of episodes in a useful manner.

In summary, the current results replicated the effect that speaker normalization occurs in utterances which are more natural than those of Ladefoged and Broadbent [1957; see also Ladefoged, 1989]. The current results add that this speaker normalization effect is independent of the lexical status of the constituents of the carrier phrase and is, moreover, vowel-specific. If only high-front vowels are encountered, only the phoneme boundaries of high-front vowels are adjusted depending on the range of formant frequencies in the carrier phrase. The fact that speaker normalization occurs

obviously contradicts the assumption of pure episodic models of speech recognition, which assume that normalization is not necessary. I argued that the fact that formant frequencies are determined simultaneously by linguistic, anatomical, and sociophonetic variables makes such a solution unlikely. An episodic model is more likely to succeed if it makes use of the extra information provided by normalization processes.

Acknowledgments

The author wishes to thank James McQueen, Klaus Kohler, Randy Diehl, and an anonymous reviewer for comments on an earlier version of this article and Marloes van der Goot, Laurance Bruggeman, and Jet Sueters for running the experiments.

References

Adank, P.; Smits, R.; Hout, R. van: A comparison of vowel normalization procedures for language variation research. J. acoust. Soc. Am. ***116:*** 3099–3107 (2004).

Agresti, A.: Tutorial on modeling ordered categorical response data. Psychol. Bull. ***105:*** 290–301 (1989).

Bladon, A.; Henton, C.; Pickering, J.B.: Towards an auditory theory of speaker normalization. Lang. Commun. ***4:*** 59–69 (1984).

Eimas, P.D.; Corbit, J.D.: Selective adaptation of linguistic feature detectors. Perception Psychophysics ***4:*** 99–109 (1973).

Eisner, F.; McQueen, J.M.: The specificity of perceptual learning in speech processing. Perception Psychophysics ***67:*** 224–238 (2005).

Evans, B.G.; Iverson, P.: Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. J. acoust. Soc. Am. ***115:*** 352–361 (2004).

Fowler, C.A.: Sound-producing sources as objects of perception: Rate normalization and nonspeech perception. J. acoust. Soc. Am. ***89:*** 2905–2909 (1990).

Fowler, C.A.: Compensation for coarticulation reflects gesture perception, not spectral contrast. Perception Psychophysics ***68:*** 161–177 (2006).

Ganong, W.F.: Phonetic categorization in auditory word perception. J. exp. Psychol. hum. Perception Performance ***6:***110–125 (1980).

Goldinger, S.: Words and voices: Episodic traces in spoken word identification and recognition memory. J. exp. Psychol. Learning Memory Cognition ***22:*** 1166–1183 (1996).

Goldinger, S.: Echoes of echoes? An episodic theory of lexical access. Psychol. Rev. ***105:*** 251–279 (1998).

Gow, D.W.: Assimilation and anticipation in continuous spoken word recognition. J. Memory Lang. ***45:*** 133–159 (2001).

Holt, L.L.: Temporally non-adjacent non-linguistic sounds affect speech categorization. Psychol. Sci. ***16:*** 305–312 (2005).

Holt, L.L.; Lotto, A.J.; Kluender, K.R.: Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? J. acoust. Soc. Am. ***109:*** 764–774 (2001).

Johnson, K.: Speech perception without speaker normalization: An exemplar model; in Johnson, Mullennix, Talker variability in speech processing, pp. 145–165 (Academic Press, San Diego 1997).

Johnson, K.; Strand, E.A.; D'Imperio, M.: Auditory-visual integration of talker gender in vowel perception. J. Phonet. ***27:*** 359–384 (1999).

Klatt, D.: Speech perception: A model of acoustic-phonetic analysis and lexical access. J. Phonet. ***7:*** 279–312 (1979).

Klatt, D.: Software for a cascade/parallel formant synthesizer. J. acoust. Soc. Am. ***67:*** (1980).

Klatt, D.: Review of selected models of speech perception; in Marslen-Wilson, Lexical representation and process, pp. 169–226 (MIT Press, Cambridge 1989).

Kluender, K.R.; Coady, J.A.; Kiefte, M.: Sensitivity to change in perception of speech. Speech Commun. ***41:*** 59–69 (2001).

Kohler, K.J.: Investigating unscripted speech: Implications for phonetics and phonology. Phonetica ***57:*** 85–95 (2000).

Ladefoged, P.: A note on 'Information conveyed by vowels'. J. acoust. Soc. Am. ***85:*** 2223–2234 (1989).

Ladefoged, P.; Broadbent, D.E.: Information conveyed by vowels. J. acoust. Soc. Am. ***27:*** 98–104 (1957).

Landauer, T.K.: How much do people remember? Some estimates of the quantity of learned information in long-term memory. Cognitive Sci. ***10:*** 477–493 (1986).

Lindblom, B.; Studdert-Kennedy, M.: On the role of formant transitions in vowel recognition. J. acoust. Soc. Am. ***43:*** 830–843 (1967).

Lotto, A.J.; Kluender, K.R.: General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. Perception Psychophysics ***60:*** 602–619 (1998).
Manuel, S.Y.: Recovery of 'deleted' schwa. Perilius: Papers from the Symposium on current phonetic research paradigms for speech motor control, p. 115–118 (University of Stockholm, Stockholm 1992).
Massaro, D.: Perceiving talking faces: From speech perception to a behavioral principle (MIT Press, Cambridge 1998).
McQueen, J.M.; Cutler, A.: Spoken word access processes: An introduction. Lang. Cognitive Processes ***16:*** 469–490 (2001).
Mitterer, H.: On the causes of compensation for coarticulation: Evidence for phonological mediation. Perception Psychophysics (in press).
Mitterer, H.; Csépe, V.; Blomert, L.: The role of perceptual integration in the perception of assimilated word forms. Q. Jl. exp. Psychol. ***59:*** 1305–1334 (2006).
Nearey, T.D.: Static, dynamic, and relational properties in vowel perception. J. acoust. Soc. Am. ***85:*** 2088–2113 (1989).
Norris, D.; Cutler, A.; McQueen, J.M.: Peceptual learning in speech. Cognitive Psychol. ***47:*** 204–238 (2003).
Peterson, G.E.; Barney, H.: Control methods used in a study of vowels. J. acoust. Soc. Am. ***24:*** 369–381 (1952).
Remez, R.E.: Neural models of speech perception: A case history; in Harnad, Categorical perception: The groundwork of cognition, pp. 199–225 (Cambridge University Press, Cambridge 1987).
Repp, B.H.; Liberman, A.M.: Phonetic categories are flexible; in Harnad, Categorical perception: The groundwork of cognition, pp. 89–112 (Cambridge University Press, Cambridge 1987).
Shockey, L.: Sound patterns of spoken English (Blackwell, Cambridge 2003).
Snodgrass, J.G.; Hirshman, E.; Fan, J.: The sensory match effect in recognition memory: Perceptual fluency or episodic trace? Memory Cognition ***24:*** 367–383 (1996).
Thompson, C.L.; Hollien, H.: Some contextual effects on the perception of synthetic vowels. Lang. Speech ***13:*** 1–13 (1970).
Traunmüller, H.: Conventional, biological and environmental factors in speech communication: A modulation theory. Phonetica ***51:*** 170–183 (1994).
Verbrugge, R.R.; Strange, W.; Shankweiler, D.P.; Edman, T.R.: What information enables a listener to map a talker's vowel space? J. acoust. Soc. Am. ***60:*** 198–212 (1976).
Watkins, A.J.; Makin, S.J.: Perceptual compensation for speaker differences and for spectral envelope distortion. J. acoust. Soc. Am. ***96:*** 1263–1282 (1994).
Watkins, A.J.; Makin, S.J.: Effects of spectral contrast on perceptual compensation for spectral-envelope distortion. J. acoust. Soc. Am. ***99:*** 3749–3757 (1996).