Van Hoof, S. & Verhoeven, J. (2011). Intrinsic vowel F0, the size of vowel inventories and second language acquisition. Journal of Phonetics, 39(2), pp. 168-177. doi: 10.1016/j.wocn.2011.02.007

Intrinsic vowel F0, the size of vowel inventories and second language acquisition

Sarah Van Hoof[a]*, Jo Verhoeven[b]

*Corresponding author: sarah.vanhoof@ua.ac.be, Tel: +32 3 265 42 70, Fax: +32 3 265 42 59
[a] University of Antwerp, Department of Linguistics, Prinsstraat 13, 2000 Antwerpen, Belgium
[b] City University London, Department of Language and Communication Science, Northampton Square, London EC1V 0HB, United Kingdom, jo.verhoeven@city.ac.uk

Abstract
The phenomenon of intrinsic vowel F0 (IF0), in which high vowels exhibit higher F0 than low vowels, has been widely attested in languages of the world. Most often, IF0 is regarded as an automatic, physiologically determined phenomenon, whereas others claim that IF0 is a controlled feature, introduced to enhance vowel contrasts. This paper presents new evidence on this issue by means of a cross-linguistic investigation of the influence of vowel inventory size on IF0 and a study of IF0 in second language (L2) acquisition. IF0 was measured in three language varieties: Arabic (a language with 3 vowels), Dutch (a 12-vowel system), and Dutch spoken by native Arabic-speaking learners. IF0 was significantly larger in Dutch than Arabic, but did not differ significantly between Arabic and Dutch produced by L2 learners. No spectral differences between the corresponding vowels of the three language varieties were found. While confirming the universality of IF0, these results also suggest that the size of IF0 may be language-specific, depending on the need to enhance vowel contrasts. Thus, these results agree well with a mixed physiological-enhancement account, which assumes that IF0 is physiologically determined, but also at least in part the effect of an interacting, controlled mechanism.

Keywords
Intrinsic F0, vowels, Arabic, Dutch, second language acquisition

Intrinsic vowel F0, the size of vowel inventories and second language acquisition

1.0 Introduction

Although individual vowels can be produced with a large range of fundamental frequencies (F0), different vowels are typically characterized by different F0 values when pronounced in identical phonetic environments. More specifically, F0 has been found to correlate with vowel height: high vowels such as [i] and [u] tend to have higher F0 values than low vowels such as [a] and [Æ]. This phenomenon of 'intrinsic vowel F0'[1] (henceforth IF0) has been attested in a wide range of languages. A cross-linguistic survey was provided by Whalen and Levitt (1995), who analysed the results of all languages known to have been investigated for IF0 up to that date. All 31 languages in their study, representing 11 language families and exhibiting a typologically representative range of pitch functions (stress, contour and register tone, and pitch accent), showed a significant F0 difference between the high vowels [i] and [u] and the low vowel [a], averaging 15.3 Hz. Later studies also uncovered an effect of IF0 for all languages investigated in them (see e.g. Connell (2002), for four African tone languages, and Verhoeven, Connell & Swerts (forthcoming), for the Belgian Limburg dialect of Hamont) – although Mambila, a tone language spoken in Nigeria and Cameroon, is a notable exception (Connell, 2002). Since an IF0 effect has been found in all languages investigated so far except one, IF0 appears to be a language universal.

The mechanism that causes IF0 has been the subject of debate. Because of its apparent universality, IF0 has mostly been assumed to be the consequence of vowel articulation, caused by an indirect biomechanical link (the nature of which remains as yet unclear) between the articulatory and the phonatory systems. Detailed discussions of the various hypotheses that have been put forward in this perspective can be found elsewhere (see e.g. Sapir (1989), Fischer-Jørgensen (1990) and Dyhr (1990)). In the context of this paper, it is sufficient to point out that most of these explanations are variants of the so-called tongue pull hypothesis, which assumes that the movements of the tongue during vowel articulation exert force on the larynx, thus affecting the tension of the vocal folds. Consequently, different tongue pull in high and low vowels causes F0 differences between these vowels. Explanations like these can be considered variants of what in the remainder of this paper will be referred to as the physiological hypothesis. Essential to this hypothesis is the automaticity of IF0 as a by-product of differing biomechanical forces involved in vowel articulation.

The alternative account focuses on the perceptual relevance of IF0 (Diehl, 1991; Diehl & Kluender, 1989a, b; Kingston, 1992; Kingston & Diehl, 1994). This account is based on research in Traunmüller (1981) and Syrdal & Gopal (1986), who found that listeners do not judge vowel height solely on the basis of F1 frequency, but rather on the basis of the auditory distance between F1 and F0: the smaller this distance, the greater the perceived vowel height. Diehl et al. argued that IF0 might be caused by speakers actively raising F0 in the production of high vowels (thereby diminishing the distance between F1 and F0), and lowering F0 in low vowels (thereby enlarging the F1-F0 distance), with the purpose of rendering the perceptual distinctiveness of high and low vowels more salient (Diehl, 1991, p.126). This enhancement hypothesis thus holds that IF0 is a feature which languages implement to perceptually enhance vowel contrasts and which speakers may actively and independently control, whereas the physiological hypothesis considers IF0 to be a purely phonetic effect that speakers are unable to exert independent control over.

As to the evidence for both approaches, it is clear that the strongest argument in favour of the physiological hypothesis is distributional in nature and relates to the apparent universality of IF0. Whalen and Levitt (1995) argued that if IF0 were an enhancement feature, one would

[1] It should be noted that the term intrinsic vowel F0 is used in the literature in different ways. Some authors (e.g. Reinholt Petersen, 1978; Dyhr, 1990) use the term IF0 to refer to the typical F0 value of individual vowels. Others (e.g. Whalen & Levitt, 1995) use the term to indicate the characteristic difference between F0 values of vowels of different articulatory heights, typically [i] versus [a] and [u] versus [a]. In the present article, the term IF0 will be used in the latter way, to refer to the F0 difference between high and low vowels.

expect it to be optional, i.e. a variable which languages can, but need not implement. There should then be at least some languages that do not exhibit IF0. Nevertheless, IF0 occurs in all languages but one that have thus far been investigated for the feature. The one language which failed to exhibit IF0 is Mambila, an African tone language (Connell, 2002). All other tone languages investigated do exhibit IF0, whereas, as Connell (2002) and Whalen and Levitt (1995) pointed out, IF0 would not be expected to occur in these languages, since it may put the perception of tonal contrasts at risk.

Secondly, Kingston (2007) argued that physiological determination of IF0 would imply that it should invariably occur in any prosodic context, whereas if IF0 were an enhancement feature, it would be expected to occur specifically in prosodically prominent contexts, enhancing vowel contrasts at locations where information content is high (2007, p. 172). In a small experiment with four speakers, Kingston found F0 differences between high and low vowels in both unaccented (non-focus position) as well as accented syllables (focus position), which suggests that IF0 is automatic rather than controlled.

A final argument in favour of physiological determination derives from speech development data in infants. Since infants as young as 6 months do not yet have vowel categories to enhance, the finding of Whalen, Levitt, Hsiao and Smorodinsky (1995) that these infants do have IF0 is consistent with the hypothesis that IF0 is an automatic consequence of vowel articulation.

Most evidence for the enhancement hypothesis focuses on the perceptual relevance of IF0. Gandour & Weinberg (1980) and Pettorino (1987) found that oesophageal speech produced by laryngectomized speakers exhibits IF0, comparable in size to IF0 in the speech of laryngeal speakers. Since tongue pull on the phonatory system must be ruled out as the cause of IF0 in oesophageal speech, this finding suggests that IF0 in oesophageal speech is actively and deliberately produced. Whalen & Levitt (1995) and Whalen et al. (1998) counter-argued that this does not necessarily mean that laryngeal speakers introduce IF0 deliberately, because oesophageal speakers may very well reproduce in a different way what comes automatically to laryngeal speakers, thus making their speech sound more natural. Still, this would imply that speakers are aware that vowels of different heights are accompanied by typical F0 differences, as Kingston notes (1992, p. 101). Moreover, the fact that oesophageal speakers make the effort to imitate IF0, suggest that IF0 is not just a phonetically redundant feature of vowels but is important to their perceptual identity (Silverman, 1984). Suggestive evidence that IF0 might indeed function as a perceptual cue to vowel height was provided by Reinholt-Petersen (1986). In his small pilot study on the role of F0 in the perception of vowels with ambiguous formant values, IF0 was used by participants to disambiguate the vowels. In a larger study by Katz & Assmann (2001), however, IF0 was found not to contribute to identification accuracy in open-set vowel identification tasks.

Considering that the evidence to either perspective seems compelling, it is useful to point out that both hypotheses are not necessarily mutually exclusive. In view of the overwhelming evidence on the universality of IF0, it is quite likely that IF0 is fundamentally physiologically determined and thus has an automatic basis. However, as Fischer-Jørgensen (1990) pointed out, accepting this does not rule out the possibility that IF0 may in some languages be subject to phonological influence and be "embodied in the knowledge of the speaker and included in the norm of the language, because [it is] physiologically economical and perceptually useful" (Fischer-Jørgensen, 1990, p. 135; cf. also Silverman, 1984 and Connell, 2002, p. 103). Consequently, IF0 may come automatically to all languages, but may be enhanced in some languages or restrained in others, which implies the involvement of more than one mechanism in the production of IF0 (Silverman, 1984).

Minimizing IF0 differences may be particularly relevant to tone languages, since IF0 may interfere with tonal F0 distinctions in these languages. An indication that IF0 may be constrained in tone languages can be found in Whalen and Levitt's (1995) survey, in which the mean IF0 value for tone languages was substantially smaller (1.10 ST) than for non-tone languages (1.60 ST). Connell (2002) has further provided evidence indicating that IF0 may be constrained in the four African tone languages he investigated, particularly in Mambila, the one language which did not exhibit a significant IF0 effect.

Maximizing IF0 differences, on the other hand, may be particularly relevant to languages with large vowel systems, in which there may be a greater need to perceptually enhance vowel contrasts than in languages with small vowel systems. IF0 would be at least one of the options available to speakers to maximize the perceptual difference between high and low vowels, thus freeing up more perceptual space for the vowels with intermediate degrees of opening.

The hypothesis that IF0 is basically an automatic, biomechanically caused by-product of vowel articulation, but one that is available to manipulation through an interacting, controlled mechanism, implies that IF0 itself is language-universal, but that the size of IF0 is language-specific, depending on the extent to which languages restrain the feature, to preserve tonal contrasts, or exploit the feature, to enhance vowel contrasts. The latter assumption was addressed by Whalen and Levitt (1995), who examined the relation between vowel inventory size and IF0 by classifying the 31 languages in their survey into small inventories (4-5 vowels), medium-sized inventories (6-11 vowels), 12-vowel-systems and large inventories (13 vowels or more). Expressed in semitones, the mean IF0 values in the four language groups amounted to 1.17, 1.33, 1.70 and 1.64, respectively. In the absence of any statistical effect Whalen and Levitt concluded that vowel inventory size does not affect the size of IF0 (1995, p. 375). Nevertheless, the figures above show an interesting upward trend in the size of IF0 between the small and the large inventories, and Whalen and Levitt in fact argued the need for a more balanced dataset to ensure that the statistics do not hide a small effect. The first objective of this study therefore is to provide such a balanced dataset and to examine the possible influence of vowel inventory size on IF0, by comparing IF0 in a language with a small vowel inventory, viz. Arabic (3 vowels), with IF0 in a language with a large inventory, viz. Dutch (12 vowels). It is expected that both languages will show IF0, but that they will differ significantly in the size of the effect. Because the need to perceptually enhance differences between vowels will be greater in the large vowel system of Dutch than in the small vowel system of Arabic, larger IF0 values in Dutch compared to Arabic can be expected.

In addition, the mixed physiological-enhancement account also has important implications in contexts of second language (L2) learning. It can be argued that if languages differ in IF0 and if IF0 is not entirely an automatic consequence of vowel articulation, a native-like pronunciation of L2 vowels by L2 learners does not entail a native-like IF0 production in L2: in order to achieve the latter, L2 learners will have to learn to what extent the L2 exploits IF0 and modify their L1 phonatory mechanism to match the L2 IF0 value. This issue of the 'learnability' of IF0 is the focus of the second research objective of this paper, which is addressed by examining IF0 in the Dutch speech of native speakers of Arabic who are learning Dutch as an L2.

2.0 Materials and methods

This study aimed to investigate the relationship between vowel inventory size and IF0, by examining IF0 in Arabic and Dutch, two languages of which the vowel systems differ considerably in size. Modern Standard Arabic has a vowel system consisting of 3 distinctive monophthongal vowels qualities /i/, /u/ and /a/. Length in the Arabic vowel systems is phonologically distinctive, but since the long vowels /i:/, /u:/ and /a:/ do not differ qualitatively from their short counterparts (Mitchell, 1990), Standard Arabic can be considered to have a small vowel inventory consisting of three vowels. The vowel system of Belgian Standard Dutch distinguishes between 12 qualitatively different monophthongs /I ɪ E ɛ A ɑ O ɔ U Y Y Ø/ (Verhoeven, 2005). In the particular variety of Standard Dutch investigated in this paper, i.e. the variety spoken in the region of Antwerp, the vowels /i/, /u/ and /a/ are typically pronounced long (Verhoeven & Van Bael, 2002). This made it possible to compare IF0 in the long vowels [i:], [u:] and [a:] in the 3-vowel system of Arabic with IF0 in the corresponding vowels in the 12-vowel system of the Antwerp variety of Dutch.

The second aim of this study was to investigate the acquisition of IF0 in L2 language learning. This was done by including a group of native speakers of Arabic who were learning Dutch as an L2. IF0 in their Dutch vowels [i:], [u:] and [a:] was compared to the reference values of Arabic and Dutch provided by the native speaker groups.

2.1 Materials

The data were collected in a reading task, the materials of which were obtained by inserting the vowels [i:], [u:] and [a:] in monosyllabic non-words with a CVC structure. The consonants of these words were selected on the basis of phonological and morphological information about Modern Standard Arabic and Moroccan Arabic in Mitchell (1990) and Caubet (1993). 10 consonantal contexts were used, viz. [b_n], [m_s], [f_t], [d_k], [n_s], [s_f], [z_t], [l_m], [k_f] and [h_k]. The three vowels were inserted between the consonants, yielding a total of 30 non-words which were phonologically and morphologically well-formed in both languages.
For Dutch, these words were inserted in the first slot of the carrier phrase [ɪN _ STA:T əN _] ("In _ staat een _", literally: "In _ stands a _"). This places the target word in [+FOCUS, -FINAL] position so that it is likely to be realised with a clear sentence accent. The second slot contained an orthographic transcription of the vowel contained in the stimulus word. The stimuli were placed in the reading list in random order, with each stimulus occurring twice, resulting in a total of 60 stimuli (i.e. 3 vowels x 10 consonantal contexts x 2 repetitions).
For Arabic, the target words were inserted in the first slot of an Arabic translation of the Dutch carrier phrase, [FI: _ _]. Also in Arabic, the [+FOCUS, -FINAL] position of the target word in the carrier phrase is conducive for sentence stress. The target word's vowel was repeated in the second slot of the carrier phrase. Although vowels in Arabic texts are not usually transcribed orthographically, the vowel length markers are. Whereas in principle these were sufficient to ensure a correct pronunciation of the words, it was nevertheless decided to also orthographically transcribe the vowels in some nonsense words, to avoid problems with their correct recognition. The 60 stimuli in the Arabic reading list occurred in the same sequential order as the ones in the Dutch list in order to avoid potential order effects.

2.2. Speakers
33 subjects participated voluntarily in the experiment. They constituted three groups on the basis of their experience and competence in Dutch and Arabic. Although the subjects were not instrumentally assessed for hearing defects, no hearing problems were observed and no one had any difficulties in completing the task.
A first group consisted of 5 male and 5 female native speakers of the Belgian variety of Dutch. They were all students or lecturers at the University of Antwerp and had all been born of Dutch-speaking parents and raised in the Dutch speaking part of Belgium (Flanders), more specifically in the region of Antwerp, where they had been living most of their lives and were still living at the time of the study. Mean speaker age was 22 years. All speakers in this group reported to have no knowledge of Arabic, apart from one, who had attended an introductory Arabic course, but reported to be unable to engage in even a simple conversation in Arabic.
The second group consisted of 6 male and 5 female native speakers of Moroccan Arabic, who had all been born of Arabic-speaking parents and raised in various regions in Morocco, and had immigrated to Flanders no earlier than 2 years before participating in the experiment. At the time of the study they were all living in the Antwerp region. Their average age was 27 years. They were recruited at Antwerp centres for adult education, where they were registered for elementary courses of Dutch. From the researchers' unsuccessful attempts to converse with them in Dutch before the experiment, it was clear that their knowledge of Dutch was virtually non-existent.
The third group consisted of speakers from Moroccan origin, who had grown up with Moroccan Arabic as L1, but had acquired Belgian Dutch as an L2. This group was recruited through personal contacts in the Antwerp Moroccan community and advertisements circulated via the Antwerp city service for newcomers and Antwerp student unions for students with an immigrant background. Initially, a pool of 12 candidates was recruited, based on two criteria: (1) acquisition of Moroccan Arabic as an L1 from birth, and (2) speakers' self-assessment of both their Arabic and Dutch abilities as 'advanced'.
For the purpose of selecting participants from the pool and matching them according to L2 competence, each participant was asked to fill out a language background questionnaire, which contained questions about the extralinguistic factors mentioned most commonly in the L2 acquisition literature as having an influence on the degree of foreign accent in L2 production: age of learning (AOL) of Arabic and Dutch, length of residence (LOR) in the Arabic and the

Dutch-speaking area, and patterns of language use (Piske, MacKay & Flege, 2001). Age of arrival was used as an index of AOL; estimates of overall percentages of use of Arabic and Dutch served as indexes of L1 and L2 use. Self-estimates of language dominance and performance were elicited by asking candidates to indicate which language they considered their "best language" and to rate their speaking abilities in Arabic and Dutch on a scale ranging from 1 ("very poor") to 5 ("excellent").

Based on the results of the questionnaire, 8 speakers emerged as late L2 learners (AOL ≥ 25), who had been born in Morocco and had learned Dutch in early adulthood after immigrating to Flanders. They all reported fairly high percentages of Arabic use (mean: 59%), considered Arabic to be their dominant language and rated their speaking ability in Arabic higher than in Dutch. The remaining 4 candidates were early L2 learners, who had all been born in Flanders and had learned Dutch in early childhood (AOL ≤ 3). They all reported fairly low percentages of Arabic use (mean: 36%). 3 of them considered Dutch to be their best language, and all 4 of them rated both their Dutch and Arabic speaking abilities as 4 or higher. It was concluded that these 4 speakers were presumably balanced or Dutch dominant bilinguals (Flege, MacKay & Piske, 2002). It was decided to exclude them from the current study, and to only included the 8 late learners, whose characteristics are summarized in table 1. All speakers lived and worked in the Antwerp region at the time of the study, apart from one speaker, who worked in Antwerp but lived in the Brussels region.

Table 1 Characteristics of the 8 late L2 learners. AR = Arabic; DU = Dutch; Age = chronological age, in years; LOR = length of residence in the Dutch-speaking part of Belgium, in years; AOL = age of learning Dutch, as indexed by the subjec*ts' age of arrival in the Dutch-*speaking part of Belgium, in years; Ability = self-*rating of speaking ability on a scale ranging from 1 ("very poor") to 5 ("very good"); Use = reported overall percentage of use of the language.* Use of Arabic and use of Dutch do not add up to 100% for participants who reported also using other languages in their daily lives, such as French.

	Sex	Age	LOR	AOL	Ability AR	Ability DU	Use AR	Use DU
1	M	52	20	32	5	3	40	10
2	M	38	12	26	5	4	60	40
3	M	27	3	24	5	2	80	5
4	M	35	4	31	4	3	70	5
5	F	45	20	25	5	4	70	30
6	F	47	8	39	5	3	50	20
7	F	30	6	24	5	3	60	10
8	F	32	4	28	5	4	40	30
mean		38	10	29	5	3	59	19

2.3. Recording procedure

The native speakers and L2 learners of Dutch were given the Dutch reading list; the native speakers of Arabic the Arabic reading list. The reading list contained an introductory paragraph, in which the speakers were asked to read the sentences as naturally as possible and were told that they were allowed to repeat a sentence if they were not satisfied with their pronunciation. The recordings were made in quiet surroundings with no disturbing background noise. The recording equipment consisted of a TASCAM DAT recorder (DA-P1) and an AKG head-mounted microphone (CLL 444).

2.4. Analysis procedures

The recordings were digitized. In order to measure the F0 of the vowels, each vowel was manually selected in PRAAT (Boersma & Weenink, 2006) on the basis of a broadband spectrogram which was time-aligned with the sound wave. Subsequently, F0, F1 and F2 of each vowel were measured as the average value in the vowel's middle third portion. The F0 analysis used PRAAT's standard autocorrelation algorithm optimized for intonation analysis. The formants were extracted by means of PRAAT's standard LPC-based method. Depending on the speaker, the formant tracking conditions were set to be appropriate to a male or female voice. The selection of the middle third portion of the vowel and the acoustic analyses were carried out automatically by means of a PRAAT script written for this purpose.

3.0 Results

In the experiment, a total number of 1740 observations were obtained, i.e. 60 stimuli x 29 speakers. 600 observations pertained to Dutch, 660 observations related to Arabic and 480 related to the learners of Dutch as an L2.

In the first instance it was investigated to what extent the actual F0 (in Hz) in the high vowels was different from that in the low vowels. For this purpose, a repeated measures ANOVA was carried out with Language (3 levels: native Arabic, native Dutch, Dutch as an L2) and Sex (2 levels: male vs. female) as between-subject variables. The within-subjects variables in this analysis were Vowel (3 levels: /i/, /u/, /a/), Context (10 levels: k_f, m_s, d_k, z_t, n_s, s_f, b_n, l_m, h_k, f_t) and Repetition (2 levels). The analysis was carried out in SPSS (version 17.0).

One of the issues associated with a repeated measures ANOVA with within-subject effects with three or more levels, relates to the sphericity assumption: this is met when the “variances of the difference scores for all pairs of treatment levels are homogeneous” (Max & Onghena, 1999, p. 262). Sphericity could thus be an issue in this analysis for the within-subjects factors Vowel and Context. There are various ways in which to deal with this issue, but in this analysis sphericity was investigated by means of Mauchly’s test (Field, 2009). This test indicated that the assumption of sphericity had been violated for the factors Vowel ($\chi^2(2) = 11.465$; $p = 0.003$) and Context ($\chi^2(44) = 102.859$; $p = 0.0001$). In addition, sphericity had been violated for the following interactions: Vowel*Context ($\chi^2(170) = 311.072$; $p = 0.0001$), Context*Repetition ($\chi^2(44) = 75.746$; $p = 0.003$) and Vowel*Context*Repetition ($\chi^2(170) = 402.548$; $p = 0.0001$). In order to provide protection against type 1 errors in these instances the degrees of freedom were adjusted by means of the Greenhouse-Geisser correction factor Epsilon: Vowel ($\varepsilon = .72$), Context ($\varepsilon = .50$), Vowel*Context ($\varepsilon = .32$), Vowel*Repetition ($\varepsilon = .85$), Context*Repetition ($\varepsilon = .57$) and Vowel*Context*Repetition ($\varepsilon = .24$). The Greenhouse-Geisser correction factor is more conservative than the Huynh-Feldt correction factor, but not as conservative as the Lower-bound one.

From this repeated measures ANOVA it appears that the only significant within-subjects effect was that of Vowel ($F(1.436, 34.469) = 41.923$; $p < 0.0001$). This effect is such that the F0 of the high vowels is higher than that of the low vowels (high = 179 Hz vs. low = 160 Hz). The effects of Context and Repetition were not significant (Context ($F(4.461, 107.059) = 1.433$; $p = 0.224$); Repetition ($F(1.000, 24.000) = 1.417$; $p = 0.246$)). Furthermore, the analysis revealed a significant interaction between the factors Vowel and Language ($F(2.842, 34.469) = 6.398$; $p = 0.002$). Taken together, this means that F0 in the high vowels is higher than in the low vowels in all three language varieties. However, the exact magnitude of the difference depends on the language variety. This is illustrated in figure 1.

Note to Publisher: Insert Figure 1 about here

From figure 1 it appears that in native Arabic the average F0 difference between the high and low vowels is 8 Hz, while this difference amounts to 7 Hz for Dutch as an L2 and 21 Hz for native Dutch. This indicates that all three language varieties show IF0 in the traditional sense of the term.

Finally, there is a significant interaction between the factors Repetition*Sex ($F(1.000, 24.000) = 4.984$; $p = 0.035$). This indicates that in the female speakers, F0 in the second repetition (194 Hz) was lower than in the first repetition (208 Hz). In the male speakers F0 in the second repetition (148 Hz) was higher than in the first repetition (146 Hz). However, this observation has no direct implications for IF0.

As far as the between-subject effects are concerned, the ANOVA revealed a significant effect of Sex ($F(1, 24) = 5.970$; $p = 0.022$) in that F0 in female speakers (204 Hz) is higher than in male speakers (147 Hz). This is precisely what can be expected on the basis of anatomical differences between male and female speakers. The effect of Language on F0 was not significant (F(1, 24)

= 0.283; p = 0.756), nor was the interaction between Language and Sex ($F(2, 24) = 2.433$; p = 0.109). This means that the observed differences in overall vowel F0 in the three language varieties (native Arabic: 175 Hz, Dutch as an L2: 160 Hz, native Dutch: 170 Hz) are not significant.
The mean F0 values for the different variables in the experiment are summarized in table 2.

Table 2 Mean F0 for the high and low vowels in male and female speakers of Arabic, Dutch as an L2 and Dutch.

Language	Male				ΔMale	Female				ΔFemale
Arabic	/i/	148 Hz	/a/	142 Hz	**6 Hz**	/i/	208 Hz	/a/	197 Hz	**11 Hz**
	/u/	147 Hz	/a/	142 Hz	**5 Hz**	/u/	205 Hz	/a/	197 Hz	**8 Hz**
Dutch L2	/i/	166 Hz	/a/	163 Hz	**3 Hz**	/i/	200 Hz	/a/	190 Hz	**10 Hz**
	/u/	167 Hz	/a/	163 Hz	**4 Hz**	/u/	202 Hz	/a/	190 Hz	**12 Hz**
Dutch	/i/	135 Hz	/a/	119 Hz	**16 Hz**	/i/	216 Hz	/a/	194 Hz	**22 Hz**
	/u/	138 Hz	/a/	119 Hz	**19 Hz**	/u/	219 Hz	/a/	194 Hz	**25 Hz**

As a second stage of the analysis, the obtained F0 values were converted into semitone values. This was done by converting the F0 measurements for [i:] and [a:] (front dimension) in corresponding target words into a semitone distance value. For this purpose the following formula was used: $12/\log2 \times \log(X_2/X_1)$, in which X_2 represented the F0 of a target word with a high vowel (e.g. 'mies'), while X_1 represented the F0 of the corresponding target word with a low vowel (i.e. 'maas'). The same procedure was applied to obtain IF0 measurements for [u:] and [a:] (back dimension), but in this case X_2 was the F0 value of the target word with [u:] (e.g. 'moes'), while X_1 represented the F0 of the corresponding target word with [a:] (i.e. 'maas'). As a consequence, IF0 in this paper is defined operationally as the F0 difference between high and low vowels in corresponding target words. This approach differs slightly from other studies on intrinsic vowel F0, which generally analyse the obtained F0 values for the high and low vowels in order to investigate a statistical effect of vowel height on F0. Our method was preferred since it enables an F0 normalisation for sex and individual anatomical differences between speakers and focuses on perception rather than production.
These IF0 values were analysed by means of a three-way repeated measures ANOVA with Language (3 levels: native Arabic, native Dutch, Dutch as an L2) and Sex (2 levels: male vs. female) as between-subject variables and Dimension (2 levels: front vs. back) and Context (10 levels: k_f, m_s, d_k, z_t, n_s, s_f, b_n, l_m, h_k, f_t) as within-subjects variables. The analysis was carried out in SPSS (version 17.0). In this analysis, the assumption of sphericity was investigated by Mauchly's test, which turned out to be significant for the within-subjects factor Context ($\chi^2(44) = 141.835$; $p < 0.0001$) and the interaction between Dimension and Context ($\chi^2(44) = 89.749$; $p < 0.0001$). Therefore, the degrees of freedom for Context and Context*Dimension were adjusted by means of the Greenhouse-Geisser correction factor: i.e. Context ($\varepsilon = .35$), Context*Dimension ($\varepsilon = .51$). On the basis of the adjusted degrees of freedom, none of the within-subjects factors nor interactions were significant. This non-significance is particularly noteworthy for the factor Dimension ($F(1.00, 22.00) = 1.439$; p = 0.243): IF0 in the front dimension was 1.21 ST, while it amounted to 1.35 ST in the back dimension.
As far as the between-subject effects are concerned, the ANOVA revealed a statistically significant effect of the speakers' language background on IF0 ($F(2, 22) = 7.556$; p = 0.003). This effect was such that IF0 in Arabic (0.74 ST) was smallest, followed by IF0 in the Dutch speech of the L2 learners (0.83 ST). IF0 in native speaker Dutch was largest and amounted to 2.28 ST. Post-hoc Bonferroni analysis for the different levels of the language variable indicated that only IF0 in native Dutch differed significantly from IF0 in the two other language varieties. The IF0 difference between native Arabic and Dutch as an L2 was not significant.

It should furthermore be mentioned that the effect of Sex on IF0 was not significant ($F(1, 22) = 0.414$; p = 0.527). In male speakers IF0 amounted to 1.16 ST, while IF0 in female speakers was

1.40 ST. Furthermore, the interaction between Language and Sex was not significant either ($F(2, 22) = 1.672$; $p = 0.211$). The mean IF0 values for the different variables involved are summarized in figure 2.

Note to Publisher: Insert Figure 2 about here

From figure 2 it appears that there is differential IF0 in male and female learners of Dutch as an L2: in this speaker group, IF0 in women (1.36 ST) was bigger than in the male participants (0.31 ST). The IF0 difference between genders in the two other language varieties was substantially smaller (Arabic: male = 0.58 ST, female = 0.89 ST; Dutch: male = 2.59 ST, female = 1.96 ST). The fact that this rather big difference in mean F0 in the non-native speakers of Dutch does not come out significant can be explained by the fact that the IF0 variance in this group was particularly big.

Besides the IF0 analysis, the formant values of the vowels of the three speaker groups were examined. Since the experiment contained formant values for male and female speakers, it was necessary to normalize the formant values in Hz. In doing so, a normalization procedure was needed which eliminated anatomical or physiological variation from the acoustic measurements, while preserving variation attributable to pronunciation differences between the vowels in the different groups of speakers. On the basis of a comparative study of 12 different vowel normalization procedures in Adank (2003) it was decided to normalize the formant measurements in Hz by using the Lobanov z-score transformation (Lobanov, 1971), since this method was best at eliminating anatomical variation while giving rise to the smallest reduction of pronunciation variation.

The normalized values of F1 and F2 were analyzed statistically by means of a three-way repeated measures ANOVA with Language (three levels) and Sex (two levels) as between-subjects variables and Vowel, Context and Repetition as a within-subjects variables. In the analysis of F1, Mauchly's test of Sphericity was significant for Vowel ($\chi^2(2) = 23.945$; $p < 0.0001$), Vowel*Context ($\chi^2(170) = 268.684$; $p < 0.0001$), Vowel*Repetition ($\chi^2(2) = 6.319$; $p < 0.042$), Context*Repetition ($\chi^2(44) = 66.522$; $p < 0.019$) and Vowel*Context*Repetition ($\chi^2(170) = 254.934$; $p < 0.0001$). In order to provide protection against type 1 errors in these instances the degrees of freedom were adjusted by means of the Greenhouse-Geisser correction factor Epsilon: Vowel ($\varepsilon = .60$), Vowel*Context ($\varepsilon = .43$), Vowel*Repetition ($\varepsilon = .80$) Context*Repetition ($\varepsilon = .59$) and Vowel*Context*Repetition ($\varepsilon = .44$). The main aim of this analysis was to investigate whether there are any significant differences in vowel F1 in the three language varieties because F1 is generally assumed to correlate well with tongue height in vowel pronunciation (Ladefoged, 2001). The analysis revealed no significant effect of the language background of the speakers on F1 ($F(2, 23) = 0.780$; $p = 0.470$). This suggests that as far as this can be assessed acoustically there were no substantial differences in tongue height between the corresponding vowels in the three language varieties investigated in this study.

A similar analysis was carried out with F2 as a within-subjects variable. F2 is assumed to correlate rather well with the front-back dimension in vowel articulation (Ladefoged, 2001). In this analysis, Mauchly's test of Sphericity was significant for Vowel ($\chi^2(2) = 22.299$; $p < 0.0001$), Vowel*Context ($\chi^2(170) = 263.911$; $p < 0.0001$), Context*Repetition ($\chi^2(44) = 61.216$; $p = 0.050$) and Vowel*Context*Repetition ($\chi^2(170) = 314.805$; $p < 0.0001$). Accordingly, the degrees of freedom were adjusted by means of the Greenhouse-Geisser correction factor Epsilon: Vowel ($\varepsilon = .61$), Vowel*Context ($\varepsilon = .43$), Context*Repetition ($\varepsilon = .68$) and Vowel*Context*Repetition ($\varepsilon = .40$). This analysis showed no significant differences between the vowels of the three language varieties as far as the front-back distinction is concerned ($F(2,24) = 1.038$; $p = 0.370$). The mean values of F1 and F2 for the different language varieties are summarized in table 3.

Table 3 Mean formant values (in Hz) for the different vowels in the three language varieties (AR = native Arabic, L2 = Dutch as an L2, Du = native Dutch). The results have been separated out for male and female speakers.

	Male speakers						**Female speakers**					
	/i/		**/u/**		**/a/**		**/i/**		**/u/**		**/a/**	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
AR	316	2117	382	926	568	1426	350	2551	412	871	680	1564
L2	264	2213	337	1023	607	1177	364	2515	373	1016	759	1459
DU	252	2169	319	1069	619	1273	284	2616	316	879	820	1554

Formant 3 was not included in the analysis since it "has very little function in distinguishing the vowels" (Ladefoged, 2001, p. 46).

4.0 Discussion

This study is a carefully controlled investigation of intrinsic vowel F0 in three language varieties, viz. Arabic, Dutch and Dutch spoken as an L2 by native Arabic-speaking learners. It aimed to examine the hypothesis that whereas IF0 is caused by an automatic, physiologically determined mechanism, the effect can be enhanced to maximize vowel height contrasts, through an interacting mechanism under speakers' control. More specifically, it intended to investigate, first, to what extent the size of IF0 may be language specific; secondly, whether the size of IF0 may be related to the size of the vowel inventory; and finally, whether IF0 may be a feature that can be acquired in language learning.

A preliminary concern was to establish the presence or absence of an effect of vowel height on F0 in all three language varieties, i.e. to determine whether indeed they all exhibit IF0. Our results indicate that this is the case: in all three varieties, the mean F0 in high vowels is significantly higher than the mean F0 in low vowels. Thus, our findings are consistent with most previous research on IF0, confirming the widely attested universality of IF0 and corroborating the idea that IF0 is fundamentally physiologically determined. As to the underlying mechanism of this physiological determination, this paper does not have an innovative contribution to make. The idea that IF0 may be related to the biomechanical coupling between the articulatory and phonatory systems still seems to be very appealing but the precise nature of this mechanism does require further exploration.

One of the aims of this study was to investigate the hypothesis that the size of IF0 is language-specific and may differ across languages. Whalen and Levitt's (1995) survey, in which the IF0 values reported for the different languages ranged between 0.0496 ST and 4.72 ST, already offered an indication of the validity of this hypothesis. This is also confirmed by the results for the two native speaker groups in this study: the IF0 value of 0.74 ST for Arabic is considerably smaller than IF0 in Dutch, which amounts to 2.28 ST. It should be pointed out that the IF0 value of 0.74 ST for Arabic in this investigation is smaller than Whalen and Levitt's (1995) value of 1.17 ST for small vowel systems (4 or 5 vowels), while the value for Dutch found here (2.28 ST) is considerably bigger than the 1.70 ST associated with 12-vowel systems in Whalen & Levitt's (1995) survey. Our IF0 value for Belgian Dutch does agree remarkably well with the average IF0 value of 2.26 ST reported by Koopmans-Van Beinum (1980) in a study on the basis of four speakers of northern (Netherlandic) Dutch.

The existence of a significant difference in IF0 between Arabic and Dutch raises the question of how it should be accounted for. It should be quite clear that the cross-linguistic IF0 difference in this study is unlikely to result from methodological differences in the collection and analysis of the Arabic and Dutch speech materials. IF0 differences between languages have been observed before, but these have mostly been (maybe too easily) disregarded as experimental artifacts of methodological differences between studies in terms of the number of informants, the language materials used, recording and analysis procedures. In this study, the data collection and analysis, as well as the exact procedure of establishing IF0 were identical for Dutch and Arabic, so that the observed IF0 difference is unlikely to result from differences in methodology.

Furthermore, it should be pointed out that the IF0 difference in Arabic and Dutch is not related to vowel articulation differences between the two languages either, as far as can be assessed by measuring F1 and F2. Although we did not collect direct physiological information about vowel articulation in both language varieties, the spectrographic analysis (figure 2) of the native Dutch and Arabic vowels produced in this experiment did not reveal any significant differences

between corresponding vowels in both languages. The physiological hypothesis, in which IF0 is assumed to be entirely determined by vowel articulation, would have led us to expect that the distance between the high and low vowels in Dutch would be significantly bigger than in Arabic. In contrast, the absence of any significant differences in F1 between Arabic and Dutch indicates that the difference in the size of IF0 between Arabic and Dutch is not mirrored by differences in degree of opening between the corresponding Arabic and Dutch vowels. Consequently, a purely physiological explanation is unable to account for our findings.
A factor that may be relevant to account for the IF0 difference between Arabic and Dutch is the difference in size of the vowel inventories of both languages. With Arabic having a small 3-vowel inventory and only a small IF0 value (0.75 ST), and Dutch having a large 12-vowel inventory and an IF0 value three times as big as Arabic (2.28 ST), our findings are suggestive of a possible influence of vowel inventory size on IF0, and agree well with the hypothesis that IF0 can serve the purpose of perceptually enhancing vowel contrasts in languages with large vowel systems.
The third objective of this paper was to investigate whether IF0 might be a phonetic dimension that is 'learnable' by speakers acquiring a foreign language, and thus to find indications that IF0 might be actively controlled. This was examined by including an experimental group of native speakers of Arabic who were learning Dutch as an L2. The mean IF0 value of this group amounts to 0.83 ST, which does not differ significantly from IF0 in the native Arabic group, while being significantly smaller than the mean value of the native Dutch group. The spectrographic analysis indicates that the L2 learners' deviating IF0 value is not mirrored by articulatory characteristics deviating from those of the native Dutch speakers. This means that there are no height differences between the corresponding vowels of the native Dutch group and the learner group to account for the observed difference in IF0 between the two groups. In failing to show such a correlation, these results suggest that IF0 is not simply a by-product of vowel articulation. Rather, it appears that the articulatory characteristics of vowels and their IF0 value are two features that are at least in part acquired separately.
These findings thereby agree well with the mixed physiological-enhancement account which, as was mentioned above, would hold that a native-like pronunciation of vowels in L2 does not automatically leads to a native-like IF0 production: in addition to learning the articulatory characteristics of L2 vowels, L2 learners also have to learn the language-specific IF0 value. Interpreting our findings from this perspective, it can be assumed that there was no need for the learners of Dutch to make large adjustments to their Arabic articulatory patterns in order to reach the Dutch vowel targets, since the spectrographic data showed no notable differences between the Arabic and the Dutch articulatory norms, as exemplified by the native speaker groups. Not surprisingly then, the L2 learners' vowels did not exhibit pronunciation differences from the Dutch norms for [a:], [i:] and [u:]. On the other hand, they seem to have been unsuccessful at adapting their IF0 value to match the Dutch IF0 target, and apparently still produced an if0 value in Dutch comparable to their native IF0 value in Arabic.
Assuming that IF0 is fundamentally an automatic by-product of vowel articulation, but one that can be actively enhanced in order to maximize vowel contrasts, two final questions emerge. The first one relates to the exact size of IF0's physiological component. At this point, we can only speculate about the exact contribution of physiological determination by stating that it is likely to be relatively small: IF0 amounted to 0.74 ST in Arabic, a language with only three vowel contrasts and therefore presumably very little need for the perceptual enhancement. The fact that IF0 has generally been found to be a rather substantial effect may have to be accounted for by the fact that the largest proportion of IF0 research in the past has been carried out on languages with large vowel systems[2], like English and German. More carefully controlled cross-linguistic research on a typologically representative variety of languages with differently sized vowel inventories may shed more light on this matter.
The second question relates to whether the 2.28 ST IF0 of Dutch is big enough to be perceived by listeners as an effective shift in F1 in order to enhance vowel contrasts. Following the same

[2] In the Whalen and Levitt (1995) survey, of the 63 IF0 studies, 34 studies are on languages with more than 10 vowels, i.e. 54 %.

calculation of the magnitude of the shift in vowel identity as in Whalen & Levitt (1995, p. 360), an IF0 of 2.28 ST would give rise to a shift of 3.4 % in effective F1, which is just above the difference limen for changes in single formants proposed in Flanagan (1955). More importantly though, it should be pointed out that a pitch difference of 2.28 ST between high and low vowels is actually quite big as far as pitch differences go and should be readily perceivable ('t Hart & Collier, 1975).

5.0 Conclusions

This article set out to shed new light on the mechanism behind IF0, through a cross-linguistic investigation of the influence of vowel inventory size on IF0 and a study of IF0 in second language (L2) acquisition. An experimental investigation of IF0 was conducted in two native speaker varieties and one L2 variety. All three varieties exhibited IF0, whereas the size of the effect differed between the two native speaker varieties. This finding contributes to the view that IF0 is language-universal, while the size of the effect is language-specific and differs cross-linguistically. The larger IF0 value which was found for the language with the largest vowel inventory suggests that languages with large vowel systems may exploit IF0 to enhance vowel contrasts. This view presupposes that IF0 is not purely an automatic result of speech physiology. The latter hypothesis is given fuel by our finding that the difference in the size of IF0 between the two native speaker varieties was not mirrored by any significant differences in vowel articulation between the two languages. Also in line with that hypothesis is the finding that IF0 in Dutch produced by native Arabic-speaking learners resembled the Arabic IF0 value more than the Dutch value, while their vowels showed no notable differences in articulatory characteristics from both Dutch and Arabic. Overall, the results presented here are supportive of a mixed model of IF0 determination, which assumes a physiological basis for IF0, in combination with an interacting, controllable mechanism, available for language-specific manipulation.

Acknowledgements

The authors are grateful to all of the participants and to Helge Daniels for the Arabic translation of the reading materials. Many thanks also to the Antwerp centres for adult education LBC and Sité and the Antwerp Moroccan organization Ahlan, for their help in recruiting participants. The first author is a Ph.D. fellow of the Research Foundation – Flanders (FWO).

Reference List

Adank, P. (2003). Vowel Normalisation. A Perceptual-Acoustic Study of Dutch Vowels. Wageningen: Ponsen & Looijen.
Boersma, P., Weenink, D. (2006). Praat: doing phonetics by computer [Computer program]. Version 4.4.28, retrieved 16 August 2006 from http://www.praat.org/.
Caubet, D. (1993). *L'Arabe marocain. Vol. 1: phonologie et morphosyntaxe*. Leuven: Peeters.
Connel, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. Journal of Phonetics, 30, 101-129.
Diehl, R. (1991). The role of phonetics within the study of language. Phonetica, 48, 120-134.
Diehl, R. L., & Kluender, K. R. (1989a). On the objects of speech perception. Ecological Psychology, 1, 121-144.
Diehl, R. L., & Kluender, K. R. (1989b). Reply to commentators. Ecological Psychology, 1, 195-225.
Dyhr, N. (1990). The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels. Phonetica, 47, 141-154.
Field, A. (2009). Discovering Statistics Using SPSS. London: Sage.
Fischer-Jørgensen, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. Phonetica, 47, 99-140.
Flanagan, J. L. (1955). A difference limen for vowel formant frequency. Journal of the Acoustical Society of America, 27, 613-617.
Flege, J.E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. Studies in Second Language Acquisition, 26, 1-34.

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. Applied Psycholinguistics, 23, 567-598.
Gandour, J. & Weinberg, B. (1980). On the relationship between vowel height and fundamental frequency: evidence from esophageal speech. Phonetica, 37, 344-354.
Katz, W. F., & Assmann, P. F. (2001). Identification of children’s and adults‘ vowels: intrinsic fundamental frequency, fundamental frequency dynamics, and presence of voicing. Journal of Phonetics, 29, 23-51.
Kingston, J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. Language and Speech, 35, 99-110.
Kingston, J. (2007). Segmental influences on F0: Automatic or controlled? In C. Gussenhoven & T. Riad (Eds.), Tones and Tunes. Vol. 2: Experimental Studies in Word and Sentence Prosody (pp. 171-210). Berlin: Mouton de Gruyter.
Kingston, J., & Diehl, R. (1994). Phonetic knowledge. Language, 70, 419-454.
Koopmans-Van Beinum, F. J. (1980). Vowel Contrast Reduction. An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions. Amsterdam: Academische Pers.
Ladefoged, P. (2001). Vowels and Consonants. An Introduction to the Sounds of Languages. Oxford: Blackwell.
Lehiste, I. (1977). Suprasegmentals. Cambridge: MIT Press.
Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. Journal of the Acoustical Society of America, 49, 606-608.
Max, L. & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. Journal of Speech, Language & Hearing Research, 42, 261–270.
Mitchell, T. (1990). Pronouncing Arabic. Oxford: Clarendon.
Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: acoustic measurements and accentedness ratings. Language and Speech, 36, 39-66.
Pettorino, M. (1987). Intrinsic pitch of vowels: an experimental study on Italian. In Proceedings XIth International Congress of Phonetic Sciences, 1, 138-141. Tallinn: Academy of Sciences of the Estonian SSR.
Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: a review. Journal of Phonetics, 29, 191-215.
Reinholt Petersen, N. (1978). Intrinsic fundamental frequency of Danish vowels. Journal of Phonetics, 6, 177-189.
Reinholt Petersen, N. (1986). Perceptual compensation for segmentally conditioned fundamental frequency perturbation. Phonetica, 43, 31-42.
Sapir, S. (1989). The intrinsic pitch of vowels: theoretical, physiological and clinical considerations. Journal of Voice, 3, 44-51.
Silverman, K. (1984). What causes vowels to have intrinsic fundamental frequency? Cambridge Papers in Phonetics and Experimental Linguistics, 3, 1-15.
Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America, 79, 1086-1100.
‘t Hart, J., & Collier, R. (1975). Integrating different levels of intonation analysis. Journal of Phonetics, 3, 235-255.
Traunmüller, H. (1981). Perceptual dimension of openness in vowels. Journal of the Acoustical Society of America, 69, 1465-1475.
Verhoeven, J. (2005). Belgian Standard Dutch. Journal of the International Phonetic Association, 35, 243-247.
Verhoeven, J., Connell, B. & M. Swerts (forthcoming). Intrinsic vowel pitch in Hamont Dutch.
Verhoeven, J., & Van Bael, C. (2002). Akoestische kenmerken van de Nederlandse klinkers in drie Vlaamse regio’s. Taal en Tongval, 54, 1-23.
Whalen, D. H., Gick, B., Kumada, M. & Honda, K. (1998). Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. Journal of Phonetics, 27, 125-142.
Whalen, D., & Levitt, A. (1995). The universality of intrinsic F0 of vowels. Journal of Phonetics, 23, 349-366.

Whalen, D. H., Levitt, A. G., Hsiao, P. & Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month-old French- and English-learning infants. Journal of the Acoustical Society of America, 97, 2533-2539.

Figure 1
Mean F0 for the high and low vowels in the three language samples. Error bars show 95.0% confidence intervals of mean.

Figure 2
Summary of the IF0 differences in terms of language variety and sex. The top graph summarizes the results in the Back dimension of the vowel space, while the bottom graph illustrates the front dimension. Error bars show 95.0% confidence intervals of mean.