Towards an auditory reference system for primary vowel types: interviewing experts at ICPhS'99

Doris Mücke
Department of Phonetics, University of Cologne, Germany
E-mail: doris.muecke@uni-koeln.de

ABSTRACT

The power of traditional vowel classification is known to be moderate, leading especially in phonetic transcription to ambiguous classifications. A common auditory reference system - shared by all transcribers - might improve inter- and intra-speaker agreements. In order to suggest a first auditory subsystem, a series of auditory experiments was carried out with 35 of the world's leading phoneticians. They were asked to define their auditory concepts of primary vowel types by adjusting a vowel synthesiser. Mean values and standard deviations for the f1/f2-combinations are presented for three expert groups differing in their present IPA-usage. The criterion of auditory equidistance is examined statistically. Bias-effects due to listeners' transcription traditions and mother tongues are discussed and the assumption of a continuous perception mode for vowels is challenged.

1. INTRODUCTION

Vowels are articulated by an open approximation of the articulators. Except for closed vowels such as [i], [y] and [u] and some vowels of the front series such as [e] and [ɛ], there is no contact between active and passive articulators. In the case of open vowels such as [a] and [ɐ] as well as vowels of the back series such as [ɔ] and [ʌ], even the well-trained transcriber is incapable of making qualified statements about vocal tract configurations singularly on the basis of kinaesthetic awareness. In reality vowel descriptions are mainly based on assumptions of ideal central tongue positions: from a hypothetical highest-point of the tongue in the oral cavity two of the three-term labels, namely place and stricture of articulation, are derived. The scientific power of this method is moderate and leads commonly to ambiguous vowel description in phonetic transcription [1, 2]. We assume that some of the problems concerning the inter- and intra-speaker agreement on vowel transcriptions result from the lack of a common auditory reference system shared by all transcribers. Auditory reference, whether stored on sound storage media or generated by real-time speech synthesis, could be used in addition to traditional vowel parameters to support listeners' orientation in the continuous vowel space [3, 4]. The most famous sound scale for vowel qualities was developed by Daniel Jones one hundred years ago. Pursuing Jones' basic idea, this study aims to suggest a set of auditory reference sounds covering the primary IPA-symbols for vowels. With the generous support of the congress committee of the 14th ICPhS (San Francisco) a series of auditory experiments was carried out with 35 of the world's leading phoneticians. They defined their concepts of the primary vowel types by calibrating a formant synthesiser devised specifically for this purpose. While the complete interviews shall be presented within the scope of a dissertation project, this paper intends to focus on bias-effects on sound-symbol-association due to experts' transcription traditions and mother tongues [5].

2. PARTICIPANTS AND STUDY

The individual interviews lasted twenty minutes and took place in a quiet room at the congress area. Seven of 35 experts marked in the following list by an asterisk were interviewed proceeding the congress at several universities under equal conditions: *Robert Bannert, Angelika Braun, Ton Broeders, Gösta Bruce, Nick Cambpell, John Coleman, Olle Engstrand, John H. Esling, Gunnar Fant, Nina Grønnum, Carlos Gussenhoven, Bruce Hayes*, Jill House, Antti Iivonen, Sun-Ah Jun*, Pat Keating*, Klaus Kohler, Jens-Peter Köster, Florien Koopmans-van Beinum, Jody Kreiman*, Robert D. Ladd, Peter Ladefoged*, Raphaela Lauf, Ilse Lehiste, Ian Maddieson, Paroo Nihalani, Francis Nolan, Richard Ogden, Bernd Pompino-Marschall*, Jørgen Rischel, Walter F. Sendlmeier*, Wim van Dommelen, Vincent van Heuven, Wilhelm Vieregge, John Wells*. Each participant portrayed his ability to work with synthetic stimuli by callibrating corner vowels of their mother tongues. Most of the partcipants' native languages belonged to the Indo-European class of languages: British English (8 speakers), American English (5), German (8), Dutch (5), Danish (2), Swedish (3), Estonian (1), Finnish (1), Sindhi (1), Korean (1).

Every session included a questionnaire concerning transcription usage and experience as well as the auditory definition of reference sounds by method of adjustment (MOA). The MOA relates to the adjustment of parameters of a speech synthesis until it generates the intended sounds, in our case reference sounds for primary vowel types [i], [e], [ɛ], [a], [ɑ], [ɔ], [o], [u]. Therefore we developed a graphical user interface for a KLATT-synthesiser ('Vowel Hunter') which permits the speaker to manipulate the variable control parameters f1/f2 in real time by moving a cursor through a virtual vowel space [6]. A zoom-function guarantees that individual segments of the vowel space are

ISBN 1-876346-48-5

graphically displayed. The segments are based on acoustic measurements of spoken vowels combined with an inverse filtered source of John Wells' voice (500 msec length produced with a falling tone) [7]. We implemented two optional faders for variations of f3/length. These were however seldom adjusted by the participants and were thus excluded from the analysis. It should be mentioned that reported f1/f2-combinations for listeners' perceptual targets in such MOA tasks are – compared to the well-known f1/f2-measurements of spoken vowels – affected by an expanded vowel space, recently confirmed as being robust [8].

3. BIAS DUE TO TRADITIONS

The participants were assigned to five groups according to educational traditions, present IPA usage and present occupation. The questionnaire demonstrates that even today's European scientists see a close relationship between the traditional Daniel Jones' theory and the current IPA vowel chart; only British phoneticians still practise an intensive training internalising cardinal vowel qualities, while others proceed to re-import the American IPA reception. The latter tradition creates terms like "IPA-vowels" instead of "cardinal vowels" using handy language-specific keywords (imitation techniques) and leaves the discussion of underlying concepts to Europeans. The auditory reference systems adjusted by the three expert groups A - C are presented in Figure 1 - 3, showing mean values (MV) and standard deviations (SD) for the f1/f2-combinations in Bark. The perceptual-acoustic vowel space illustrates mean values and related SD in form of dots and marks through the dots (y-axis: f1; x-axis: f2). Additionally, the criterion of auditory equidistance is examined statistically by calculating the distances between the adjusted vowels for the series of front and back vowels as well as the overall equidistance from [i] to [u]. For evaluation we used the rotation-dependent city-block-distance instead of the rotation-neutral Euklidean distance:

$$\text{distance} = |\ f1_{CV2} - f1_{CV1}\ |\ +\ |\ f2_{CV2} - f2_{CV1}\ |$$

While Group A and B intended to adjust language-neutral cardinal reference sounds, Group C defined language-specific IPA-vowels sounding appropriate for the MOA task. We left out two further expert groups, whose participants received some kind of cardinal vowel education, e.g. through private study. They adjusted a mix of cardinal and IPA-vowels comparable to the results of Group B.

Group A: 5 speakers (4 male, 1 female; same mother tongue), strictly trained in cardinal vowels in Great Britain (London school) and still practising there.

Group B: 7 speakers (7 male, different mother tongues) strictly trained in cardinal vowels originally in GB (mostly Edinburgh school), however practising abroad for some time.

Group C: 6 speakers (3 male, 3 female, different mother tongues) rejecting cardinal vowels and referring to language-specific sounds (new IPA-Tradition).

The comparison of Figures 1 – 3 reveals the highest agreement on adjusting auditory reference sounds for Group A (current British community) with a mean SDf1 of 0,31 Bark and SDf2 of 0,29 Bark for all adjusted vowels. In contrast, the mean standard deviation for Group B (ex-British community) is twice as high with 0,62 Bark for SDf1 and 0,75 Bark for SDf2 and reflects a low agreement on the formerly internalised cardinal reference sounds. Without regular practice the internal cardinal sound scale does not appear to be very stable. Surprisingly, the mean standard derivations for Group C (americanised IPA community) show a higher agreement than Group B with SDf1 of 0,45 and SDf2 of 0,48 Bark despite usage of pure imitation techniques with different keywords from varying languages.

The criterion of auditory equidistance propagated by Daniel Jones for all primary vowel types from [i] to [u] is contradicted by our data even for the adjustments of the trained British experts in Group A. First of all, distances are shorter between back vowels than between front ones reflecting the physical proportions of the oral cavity [1, 2]. Secondly, the distance between [a] and [ɑ] increases sharply to 5,33 Bark and does not fit in any of the series. Interestingly, the distances for the back series are almost equidistant (with a standard deviation of 0,22 Bark), while the distances of the front vowels show bias-effects due to British English speech styles. The relatively open stricture of English front vowels requires a shift towards a closer cardinal target and may lead to a perceptual overshoot, confirming contemporary popular opinion that cardinal vowels are not really language neutral. Apart from the British tradition, every participant of this MOA task intended to cover the continuous vowel space with a somehow useful, symmetrical grid. Consequently, all irregularities within a series of back or front vowels can be interpreted as being caused by bias-effects due to tradition and native language. Although Group C shows relatively high agreement in calibrating sounds, the perceptual-acoustic distances between the reference sounds are highly irregular. A sharp increase was observed between the stricture degrees "half-close" and "half-open" for both (front and back) vowel series, reflecting irregularities of native vowel systems such as American and German speech styles. The comparability of reference sounds to "IPA-vowels" is therefore strictly limited to languages with comparable vowel systems, thus challenging the assumption of a continuous perception mode for vowels.

4. THE ROLE OF NATIVE LANGUAGES

A comparison of the adjustments made by Groups A - C may be affected by the fact that speakers' mother tongues are mixed for Groups B and C but not for A (with the highest agreements). For this reason the experts have been grouped according to their mother tongues. However, none of the language groups reaches the high agreement of Group A. Even the group "British English" (8 listeners,

ISBN 1-876346-48-5

	MV f1	MV f2	SD f1	SD f2
i	1,88	14,99	0,35	0,28
e	3,67	14,38	0,53	0,13
ɛ	6,54	13,16	0,31	0,34
a	8,95	11,55	0,47	0,52
ɑ	7,09	8,08	0,13	0,34
ɔ	5,61	7,15	0,26	0,28
o	3,90	6,16	0,22	0,26
u	2,46	5,44	0,18	0,21
	Mean SD		0,31	0,29

Distance	
i e	2,40
e ɛ	4,09
ɛ a	4,03
a ɑ	5,33
ɑ ɔ	2,41
ɔ o	2,70
o u	2,16
front	0,78
back	0,22
overall	1,11

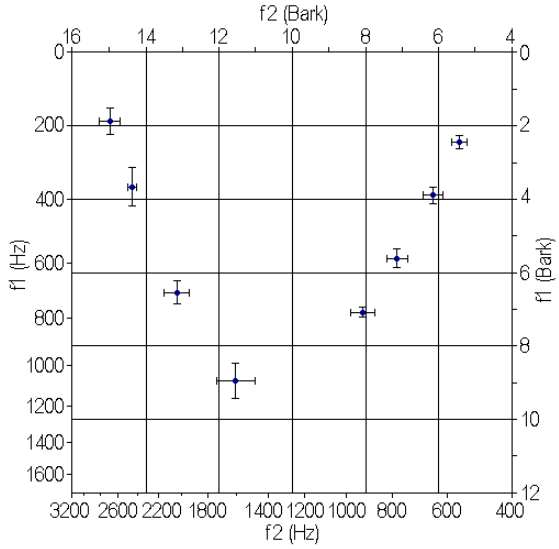

Figure 1: Group A

	MV f1	MV f2	SD f1	SD f2
i	2,09	15,10	0,64	0,28
e	3,73	14,32	0,45	0,61
ɛ	5,89	12,92	0,55	1,01
a	8,38	11,35	1,02	1,39
ɑ	7,24	7,92	1,08	0,78
ɔ	5,44	7,08	0,61	0,71
o	3,79	5,81	0,29	0,57
u	1,90	5,58	0,34	0,67
	Mean SD		0,62	0,75

Distance	
i e	2,43
e ɛ	3,55
ɛ a	4,06
a ɑ	4,58
ɑ ɔ	2,64
ɔ o	2,92
o u	2,12
front	0,68
back	0,33
overall	0,84

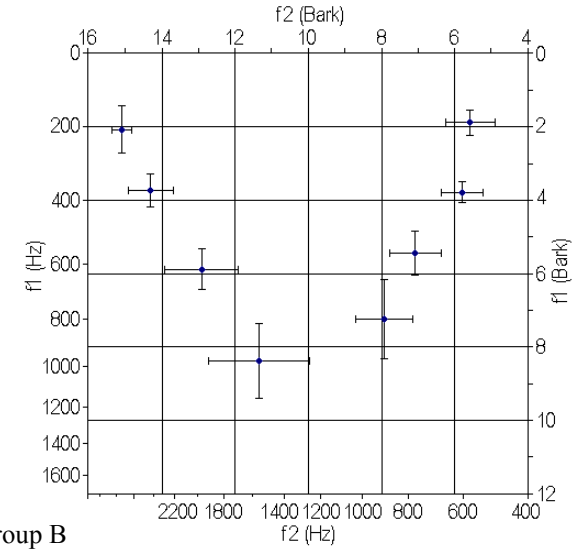

Figure 2: Group B

	MV f1	MV f2	SD f1	SD f2
i	1,99	14,58	0,44	0,36
e	3,38	14,37	0,37	0,24
ɛ	6,08	13,15	0,45	0,45
a	8,27	10,96	0,66	0,88
ɑ	7,11	8,68	0,57	0,60
ɔ	5,82	7,37	0,43	0,35
o	3,81	6,14	0,28	0,50
u	2,34	5,70	0,36	0,49
	Mean SD		0,45	0,48

Distance	
i e	1,61
e ɛ	3,92
ɛ a	4,38
a ɑ	3,44
ɑ ɔ	2,59
ɔ o	3,24
o u	1,91
front	1,21
back	0,54
overall	0,95

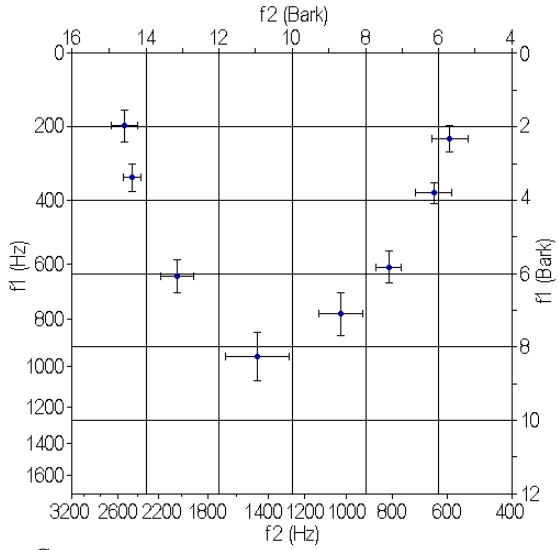

Figure 3: Group C

ISBN 1-876346-48-5 © 2003 UAB

	MV f1	MV f2	SD f1	SD f2
i	1,67	15,05	0,39	0,48
e	3,31	14,46	0,27	0,36
ɛ	5,85	13,39	0,41	0,28
a	8,18	11,07	0,60	0,99
ɑ	7,03	8,10	0,82	1,07
ɔ	5,54	7,23	0,42	0,59
o	3,70	5,72	0,23	0,47
u	1,95	5,41	0,27	0,48
	Mean SD		0,43	0,59

Distance	
i e	2,23
e ɛ	3,60
ɛ a	4,66
a ɑ	4,12
ɑ ɔ	2,37
ɔ o	3,35
o u	2,07
Front	0,99
Back	0,55
Overall	0,93

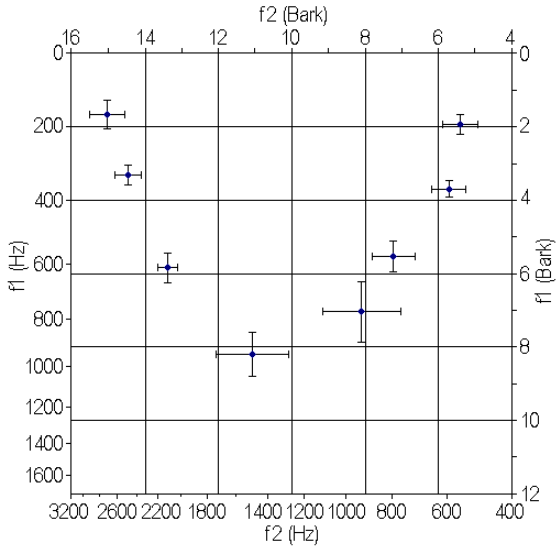

Figure 4: German group

cardinal reference sounds targeted) with a mean SD of 0,53 Bark for f1 and 0,56 Bark for f2 shows higher deviations than the language-mixed Group C. The results of the German group (7 speakers) in Figure 4 demonstrate an even stronger bias effect of the mother tongue on auditory spacing. There is high agreement on all vowel qualities occurring in the mother tongue such as [i], [e], [ɛ], [ɔ], [o], [u], but very low agreement on targeted A-qualities (deviations of 0,6 Bark for SDf1[a]; 0,99 Bark for SDf2[a]; 0,82 Bark for SDf1[ɑ]; 1,07 Bark for SDf2[ɑ]). The uncertainties on defining reference sounds for open corner vowels are due to the non-occurrence of a front and back A-quality in most German speech styles, being merged into a central vowel quality. Although some of the listeners of the German language group intended to define reference sounds for cardinal vowels and others for IPA-vowels, all adjustments can be interpreted as a simple extension of the German vowel system. Familiar phonetic targets were used wherever possible. The same was true of the language groups "American English", "Dutch" and "Swedish" in the MOA task [5].

5. CONCLUSION

The highest agreement on defining reference sounds for cardinal vowels has been found for listeners of the British tradition. High levels of agreement were also found for language-specific "IPA" reference sounds adjusted by listeners of native languages with similar vowel systems. The lowest levels of agreement were observed for listeners of the ex-British tradition; the lack of practice leads to a high degree of inconsistency by switching from native to cardinal sounds. Nevertheless, all investigated reference systems have proved to be affected by bias-effects due to listeners' mother tongues, that challenges the assumption of a continuous perception mode for vowels. Due to the observed bias-effects, an auditory reference system common to all transcribers cannot be put forward. We take the view that investigations of transcribers' auditory reference systems are an appropriate way of measuring and systematising bias-effects and should improve inter-speaker agreement. This remains a task for further studies.

ACKNOWLEDGEMENTS

We were fortunate in getting support from all participants and from the congress committee of the 14th ICPhS. The project owes much to a DAAD grant and to Georg Heike's supervision.

REFERENCES

[1] P. Ladefoged, "What are linguistic sounds made of?", *Language,* vol. 56/2, pp. 485-502, 1980.

[2] J. Laver, 1994. *Principles of phonetics*, Cambridge: Cambridge Univ. Pr., 1994.

[3] W. Meyer-Eppler, *Grundlagen und Anwendungen der Informationstheorie*, Berlin: Springer, 1959.

[4] D. Jones, *An Outline of English Phonetics*, Cambridge: Heffer & Sons, [9]1962 [[1]1918].

[5] D. Mücke, *Vokale: Synthese und Transkription*, Köln, Univ., Diss., (in progress).

[6] D. H. Klatt, "Software for a cascade/parallel formant synthesizer", *JASA*, vol. 67/3, pp. 971-995, 1980.

[7] J. Wells and J. House, *The Sounds of the International Phonetic Alphabet*, London: Phonetics & Linguistics U.C.L. [Audio CD; supplement], 1995.

[8] K. Johnson, "Adaptive Dispersion in Vowel Perception", *Phonetica*, vol. 57, pp. 181-188, 2000.