Mapping vowel categories at high fundamental frequencies using multidimensional scaling of cochlea-scaled spectra

3pSC36

Friedrichs, D.[1], Rosen, S.[1], Iverson, P.[1], Maurer, D.[2], & Dellwo, V.[3]

[1] Department of Speech, Hearing, and Phonetic Sciences, University College London, United Kingdom
[2] Institute of the Performing Arts and Film, Zurich University of the Arts, Switzerland
[3] Phonetics Laboratory, Institute of Comparative Linguistics, University of Zurich, Switzerland

BACKGROUND

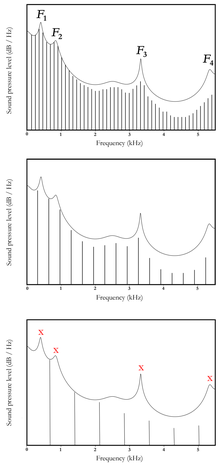
Vocalic identification in naturally produced steady-state vowels at f_os exceeding the F_1 they typically reveal in citation-form words has so far mainly been a concern of singing research, particularly in Western classical singing.

By now there is a large body of evidence indicating that the identifiability of vowels decreases with increasing f_o.

This seems plausible as the probably strongest cues to vowel category identification – formant frequencies (in terms of absolute or relative spectral maxima in the sound spectrum) – are largely undersampled at very high f_os.

→ Identification of single vowels has been shown to be compromised when f_o significantly exceeds the normal range of F_1 (e.g., Howie and Delattre, 1962).

In a recent study (Friedrichs et al., 2015), however, we found that the phonological function of the vowels /i y e ø ɛ a o u/ can be maintained at f_os up to 880 Hz.



Spectral undersampling of the vocal tract transfer function at high f_os (schematic illustration).

HYPOTHESIS AND OBJECTIVE

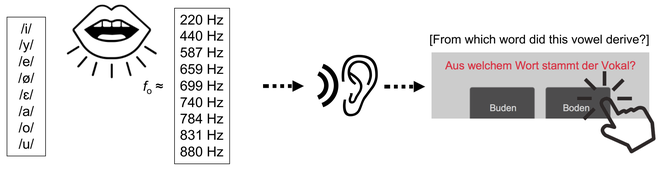
Vowel sounds may carry strong acoustic cues departing from common formant frequencies at high f_os.

Here, we carried out multidimensional scaling (MDS) analysis of cochlea-scaled spectra derived from the vowels used in Friedrichs et al. (2015).

METHODS

RECORDINGS AND IDENTIFICATION TEST

A female native speaker of German produced different disyllabic minimal pairs (e.g., /bu:dən/ vs /ba:dən/, /le:nə/ vs /lø:nə/) with contrasting vowels (/i y e ø ɛ a o u/) in the first syllable at nine f_os between 220 and 880 Hz. We asked her to focus exclusively on the intelligibility of speech. Forty native German listeners without reported hearing impairments (20 male, 20 female; mean age = 26.78 were divided into two subgroups (N=20 each) and identified the words in a binary choice identification task. Stimuli were presented as full words (stimulus condition 1) and quasi steady-state vowel centers (250 ms) (stimulus condition 2) in order to exclude formant-transitions and coarticulation phenomena.

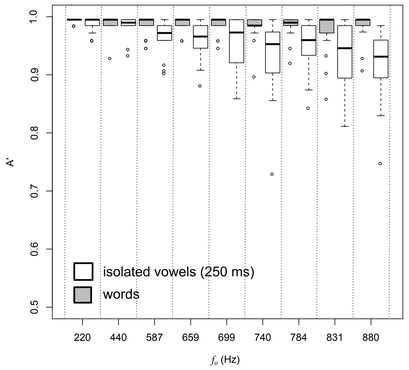


DATA ANALYSIS

Listeners' identification performance was calculated with the bias free non-parametric sensitivity measure A' from Signal Detection Theory (Stanislaw and Todorov, 1999).

Classical MDS analyses of cochlea-scaled spectra (Bark scale) was carried out by averaging the spectra across multiple files (only stimulus condition 2). Distances in the MDS space are linearly related to spectral distance.

RESULTS



Identification performance above chance (0.5).***

Words: $t^{17} = 83.43$, $p < .001$; isolated vowels: $t^{17} = 29.23$, $p < .001$

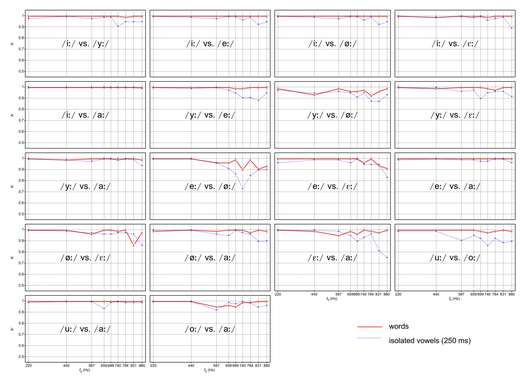
Poorer identification performance for isolated vowels than for words.***

$t[222.75] = 7.32$, $p < .001$

Poorer performance for isolated vowels (for each f_o level).***

T for 17 degrees of freedom ranged from 28.14 to 534.62. Each effect was highly significant ($p < .00028$).

Box plots showing the distributions of A' (y-axis) for all vowel pairs that were tested at nine f_os (x-axis). Cond. 1, words: white; cond. 2, isolated vowels (250 ms): gray. A' reaches from 0.5 (chance) to 1 (max. performance).



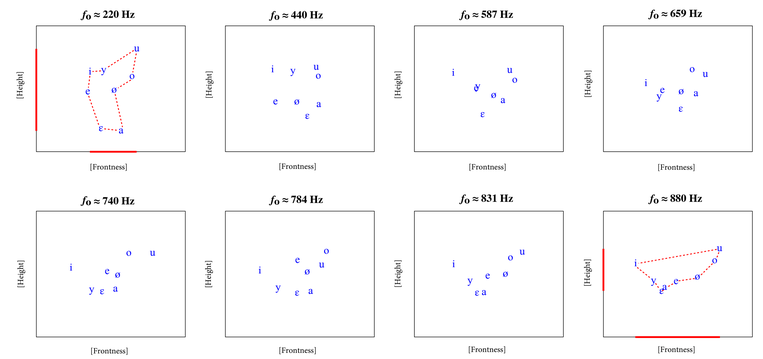
Equally high performance in word condition over entire f_o range; significantly decrease with f_o in the isolated vowel condition.

Words: $F^{8,153} = 1.01$, $p = .39$; isolated vowels: $F^{8,153} = 5.14$, $p < .001$

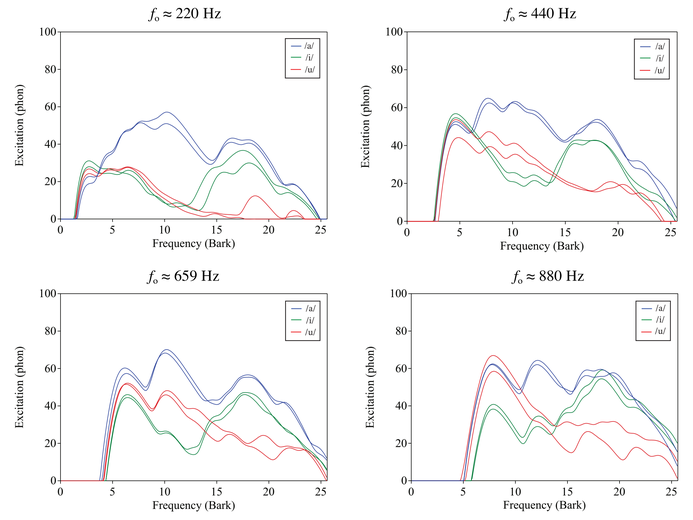
No /a/-bias, except for the pair /a/ vs. /ɛ/ at f_o = 831 and 880 Hz under the isolated vowel condition.

A' = 0.81 (831 Hz); A' = 0.75 (880 Hz); $B''_D = 0.8$ (831 Hz); $B''_D = 0.89$ (880 Hz)

A' (y-axis) for words (red solid lines) and isolated vowels (blue dotted lines) for each of the minimal pair contrasts at the nine investigated f_os (x-axis). A' reaches from 0.5 (chance) to 1 (max. performance).



MDS plots showing spectral distances of the investigated vowels at each f_o calculated by averaging cochlea-scaled spectra (Bark) across multiple files (N=324).



Averaged cochlea-scaled spectra of the point vowels /i a u/ at moderate and high f_os derived from the stimuli used in condition 1 and 2 (N=14, for each condition and at each f_o).

DISCUSSION

The phonological function of vowels can be maintained at f_os up to at least 880 Hz, that is, in cases when the normal range of F_1 of some vowels (in particular high vowels) is substantially exceeded.

Listeners do not rely on consonantal context phenomena for their identification performance.

MDS reveals that with increasing f_o, the vowel height dimension partially collapses, but the front/back distinction expands, thus allowing the vowels to be distinguished.

This indicates that the perceptual space is reorganized and vowel height and frontness are being combined in a correlated way at higher f_os.

FUTURE RESEARCH

As our vowels did not show typical spectral dynamic phenomena of continuous speech, it is questionable to what degree phenomena such as *vowel inherent spectral change* (Nearey and Assmann, 1986) might explain listeners' identification performance. It seems however reasonable to carry out identification tests with **resynthesized steady-state vowels.** This way the influence of sweeping harmonics and breath noise (sampling the vocal tract transfer function) could be examined.

Furthermore, the effects of **talker variability** and **multiple response options** should be investigated.

REFERENCES

1. Friedrichs, D., Maurer, D., and Dellwo, V. (2015) "The phonological function of vowels is maintained up to fundamental frequencies of 880 Hz," *J. Acoust. Soc. Am.*, 138(1), EL36-EL42.
2. Howie, J., and Delattre, P. (1962). "An experimental study of the effect of pitch on the intelligibility of vowels," *Natl. Assoc. Teachers Singing Bull.*, 18, (4), 6–9.
3. Nearey, T., and Assmann, P. (1986). "Modeling the role of inherent spectral change in vowel identification," *J. Acoust. Soc. Am.*, 80, 1297-1308.
4. Stanislaw, H., and Todorov, N. (1999). "Calculation of signal detection theory measures," *Behav. Res. Methods Instrum. Comput.*, 31, 137–149.