



Kent Academic Repository

Lengeris, Angelos (2016) *Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects*. *The Journal of the Acoustical Society of America*, 140 (4). EL314-EL319. ISSN 0001-4966.

Downloaded from

<https://kar.kent.ac.uk/59659/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1121/1.4964397>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects

[Angelos LengerisCC](#)

Citation: *J. Acoust. Soc. Am.* **140**, (2016); doi: 10.1121/1.4964397

View online: <http://dx.doi.org/10.1121/1.4964397>

View Table of Contents: <http://asa.scitation.org/toc/jas/140/4>

Published by the [Acoustical Society of America](#)

Comparison of perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects

Angelos Lengeris^{a)}

Department of English Studies and Linguistics, University of Kent, Canterbury,
Kent, CT2 7NF, United Kingdom
a.lengeris@kent.ac.uk

Abstract: This study compared the perception-production vowel spaces for speakers of Standard Modern Greek and two regional dialects. In experiment 1, participants produced the Greek vowels and chose vowel best exemplars (prototypes) in a natural sentence spoken in the participants' dialect. In experiment 2, the speakers who had made the recordings for experiment 1 chose themselves vowel prototypes. Cross-dialectal differences were found in both perception and production. Across dialects and experiments, participants' perceptual space was exaggerated compared to the acoustic one. Because participants' perceptual space in experiment 2 was calibrated to the participants' own voice, perception and production data are directly comparable.

© 2016 Acoustical Society of America

[CC]

Date Received: July 28, 2016 Date Accepted: September 20, 2016

1. Introduction

According to the “hyperspace effect,” when comparing perception and production data, listeners' preferred vowels are exaggerated compared to their productions. In a seminal paper, [Johnson *et al.* \(1993\)](#) asked American English participants to locate their vowel prototypes from a grid of synthetic vowels ranging in $F1/F2$ combinations using a method of adjustment and produce the same vowels. Two types of speech were elicited, normal and hyperarticulated speech. The perceptual vowel space was found to be more expanded than the acoustic one, especially when compared to the normal speech samples, with hyperarticulated vowels being closer to the vowel prototypes.

The robustness of the hyperspace effect has been demonstrated by studies adopting varying experimental protocols. For example, while the perceptual stimuli in the study of [Johnson *et al.* \(1993\)](#) consisted of isolated vowels from a synthetic voice unfamiliar to listeners, [Johnson \(2000\)](#) used hVd stimuli modeled after a speaker familiar to listeners and still found the hyperspace effect. In a detailed investigation of the hyperspace effect for the English vowel /i/, [Frieda *et al.* \(2000\)](#) used a more fine-grained stimulus set than [Johnson *et al.* \(1993\)](#) and report that mean results as well as individual performance confirmed that participants preferred more extreme /i/ tokens (i.e., lower $F1$ and higher $F2$) than those they produced.

[Whalen *et al.* \(2004\)](#) claimed that the hyperspace is an artifact of the methodology used by [Johnson *et al.* \(1993\)](#) and [Johnson \(2000\)](#). Two major objections were raised by the authors: First, because listeners were presented with a fixed 330-option grid and perception can be affected by the range of available responses, this could have resulted in forcing their choices toward the edge of the perceptual space. Second, perception and production results were not directly comparable since perception responses were mapped to a particular voice while production data were averaged across participants. One way to address the latter issue is to examine individual data, as done in [Frieda *et al.* \(2000\)](#). Even then, however, the argument raised by [Whalen *et al.* \(2004\)](#) seems reasonable since the vocal tract to which an individual's perceptual space is calibrated differs from her vocal tract (i.e., the vocal tract that produces the vowels).

This study examined whether speakers of Standard Modern Greek (SMG) and two regional Greek dialects, Crete and Kozani Greek (a Southern and a Northern dialect, respectively), demonstrate the hyperspace effect. All three systems consist of five vowels /i, e, a, o, u/ but there are cross-dialectal differences in the positioning of vowels

^{a)} Author to whom correspondence should be addressed.

and in the total vowel space area covered with the SMG vowel space being larger than the non-standard ones (Lengeris and Nicolaidis, 2015). Compared to the wealth of cross-dialectal production studies (e.g., Adank *et al.*, 2007; Clopper *et al.*, 2005; Recasens and Espinosa, 2006; Jacewicz *et al.*, 2006), research on cross-dialectal perception of vowels is more limited and mostly concerns the effects of linguistic experience on the classification of dialectal variation (e.g., Clopper and Pisoni, 2004; Evans and Iverson, 2004; Jacewicz and Fox, 2012). In Greek, previous work has only examined the perception of SMG vowels showing that vowels are well separated from one another (Botinis *et al.*, 1997; Haws and Fourakis, 1995).

In experiment 1, participants produced the five Greek vowels in isolation and in nonsense words embedded in a carrier sentence and chose their best exemplar locations (prototypes) for those vowels. The synthetic vowels were played in a natural sentence uttered by a speaker of their dialect. Best exemplars were located using a goodness optimization method that allowed participants search for vowel prototypes from a large stimulus set containing more than 100 000 vowels. The magnitude of the stimulus set and the way the searching algorithm works minimizes potential biases in participants' responses discussed in Whalen *et al.* (2004). In experiment 2, the three speakers who had made the recordings for experiment 1 (i.e., one SMG, one Cretan, and one Kozani speaker) made some additional recordings and chose themselves vowel prototypes. Because these three speakers' perceptual space was mapped to their own voice, a direct comparison of their perception and production data was feasible thus addressing the second objection raised by Whalen *et al.* (2004).

2. Experiment 1

2.1 Participants

Thirty speakers were tested, ten for each dialect (five male, five female). Two speakers were dropped from the analysis because they could not reliably perform the task. Data were analyzed for 10 speakers from Athens (SMG), 9 from Crete, and 9 from Kozani with a mean age of 60 yrs (range = 43–73 yrs). The mean age of SMG speakers was 58 yrs (range = 43–70 yrs). The mean age of Crete speakers was 61 yrs (range = 47–73 yrs) and that of Kozani speakers was 62 yrs (range = 46–72 yrs). None reported any hearing or language impairments.

2.2 Materials and procedure

Production and perception data were collected as part of the VOCALECT project (www.vocalect.eu). For the purposes of VOCALECT, in addition to perception data, speech was collected from talkers of six Greek dialects producing a range of speaking styles from isolated vowels to conversational speech.

Production of isolated vowels and nonsense words in sentences. The 5 Greek vowels were recorded in isolation and in nonsense words in sentences for a total of 560 vowel tokens (28 speakers \times 5 vowels \times 2 tasks \times 2 repetitions). Isolated vowels were elicited by having participants read them in randomized order off a PowerPoint presentation. Being unaffected by coarticulation, they were considered to approximate hyper-articulated forms compared to vowels in sentences. Vowels in sentences were elicited as answers to questions. Participants heard recordings (embedded in the presentation) of a native speaker of their dialect via DT 770 PRO (beyerdynamic, Germany) headphones. After hearing the speaker say $\Delta\epsilon\nu\ \theta\alpha\ \pi\omega\ pVpV\ \pi\omega\theta\epsilon\nu\acute{\alpha}$. *Τί δειν θά πεις πουθενά;* “I am not going to say pVpV anywhere. What aren't you going to say anywhere (lit. nowhere)?,” participants repeated the first sentence. Recordings were made directly onto a laptop hard disk via a Blue Yeti microphone at a sampling rate of 44.1 kHz. For the acoustic analysis of the first vowel in pVpV, the $F1$ and $F2$ formant frequencies of each token were computed automatically in Praat (Boersma and Weenink, 2006) from a linear predictive coding analysis with 12 coefficients below 5 kHz at 20% and 80% of the duration of the vowel.

Location of best exemplars. Subjects participated in the best exemplars experiment around 3 months after completing the production tasks. The stimuli were synthesized vowels in the context /pVta/ (stressed on the first syllable) embedded in the natural carrier sentence $\Pi\epsilon\varsigma\ ___ \xi\alpha\nu\acute{\alpha}$ “Say $___$ again.” The carrier sentence was recorded by a male native speaker of the respective dialect and included the initial release /p/ burst and the final /ta/ from the recording. For each speaker/dialect, the synthesized vowels were created in a Klatt synthesizer (Klatt and Klatt, 1990) in cascade/parallel configuration and matched the natural vowels in terms of $F0$ and amplitude. The rest of the synthesis parameters were held constant across vowels and dialects with $F4$ and $F5$ set at 3500 and 4500 Hz, respectively, the formant bandwidths set at $B1 = 100$, $B2 = 180$,

$B3 = 250$, $B4 = 300$, $B5 = 550$, the tilt set at 0 dB slope and the open quotient set at 60%. The $F1$ and $F2$ frequencies changed in a linear way from the beginning to the end of the vowel. $F1$ formant frequency ranged between 5 and 15 Equal Rectangular Bandwidth (ERB) (Glasberg and Moore, 1990). $F2$ formant frequency started from 10 ERB, was at least 1 ERB higher than $F1$ and reached a limit defined by the equation $F2 = 25 - (F1 - F5) / 2$. The synthetic vowels were 1 ERB apart from each other and their durations spanned logarithmically in 7 steps (54, 75, 104, 144, 200, 277, and 383 ms, duration results are not reported here). Overall, 109 375 vowels were synthesized for each speaker/dialect.

Participants were tested individually in quiet rooms. Stimuli were presented over DT 770 PRO headphones from a laptop. Participants heard a synthesized vowel embedded in the natural sentence and were asked to rate on a continuous scale whether it was close to being a good example of a target word shown on a computer screen. A goodness optimization method (Evans and Iverson, 2004, 2007; Iverson and Evans, 2009) searched through a multidimensional space and located the best exemplar based on the participant's responses. For each vowel, there were seven search vectors (straight-line paths cutting through the space) and five trials per vector. Participants were able to find the best exemplar for a vowel after 35 trials with the whole procedure lasting around 20 min (for a detailed description of the method, see Evans and Iverson, 2007).

2.3 Results

Figure 1 plots the average best exemplar locations of SMG, Crete, and Kozani vowels. Best exemplars are represented by arrows from the onset to the offset of the $F1$ and $F2$ formant frequencies. An initial repeated-measures analysis of variance (ANOVA) was carried out on the Euclidean distances between pairs of adjacent best exemplars to investigate the within-subject effect of pair ($/i/-/e/$, $/e/-/a/$, $/a/-/o/$, $/o/-/u/$, $/u/-/i/$) and the between-subject effects of dialect (SMG, Crete, Kozani) and gender (female, male). Before performing the ANOVA, the onset and offset $F1$ - $F2$ frequencies of each best exemplar were averaged thus removing the formant movement. The ANOVA showed no effect of gender or any interactions, indicating that irrespective of the gender of the participants their perceptual space was calibrated to the voice played to them. A subsequent ANOVA, after averaging the results across gender, showed significant main effects of pair, $F(4,88) = 182.91$, $p < 0.001$ and dialect $F(2,22) = 4.16$, $p < 0.05$ and a significant pair \times dialect interaction, $F(8,88) = 6.23$, $p < 0.001$. Pairwise comparisons showed that the Euclidean distances were larger in SMG ($M = 6.15$ ERB) than they were in Crete ($M = 5.54$ ERB) and Kozani ($M = 5.42$ ERB). Simple effect tests exploring the interaction between pair and dialect showed that for $/i/-/e/$ Crete $>$ SMG, for

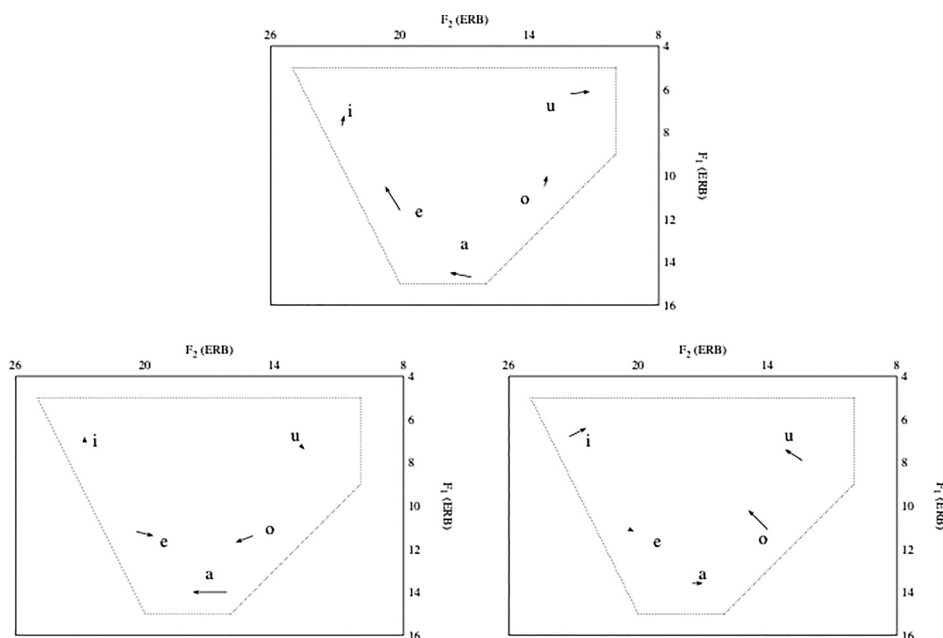


Fig. 1. Locations of vowel best exemplars for SMG (upper panel), Cretan (lower left panel), and Kozani Greek speakers (lower right panel). Best exemplars are represented by arrows from the starting to the ending $F1$ and $F2$ frequencies. Dotted lines indicate the limits of the synthesized vowels available to listeners.

/a/-o/ SMG > Kozani > Crete and for /o/-u/ Crete > Kozani, $p < 0.05$. The results thus confirmed the existence of dialect-induced differences in the positioning of vowels in the perceptual space.

To examine whether listeners' perceptual vowel space was larger than their production space, separate ANOVAs were run for male and female speakers with task (isolated vowel production, nonsense word production, vowel perception) and dialect (SMG, Crete, Kozani) as factors. Vowel space areas were calculated by dividing the vowel space into three triangles, calculating the area of each triangle using Heron's formula and summing the triangles. For male speakers, the ANOVA showed significant main effects of task, $F(2,36) = 108.39$, $p < 0.001$ and dialect, $F(2,36) = 18.06$, $p < 0.001$ and a significant task \times dialect interaction $F(4,36) = 6.39$, $p = 0.001$. Pairwise comparisons showed that, across tasks, the SMG vowel space was the largest (35.33 ERB²) with no difference between the Cretan (26.95 ERB²) and the Kozani (28.27 ERB²) spaces and that, across dialects, the perceptual vowel space was the largest (42.81 ERB²), followed by the acoustic space of isolated vowels (25.4 ERB²), which was in turn followed by the acoustic space of words in sentences (22.34 ERB²). Simple effect tests showed that the task \times dialect interaction occurred because the Cretan perceptual vowel space was larger than the acoustic spaces of both isolated vowels and words in sentences but the latter two did not differ from one another. For female speakers, the ANOVA showed a significant main effect of task, $F(2,30) = 108.39$, $p < 0.001$. Pairwise comparisons showed that, across dialects, the perceptual vowel space was the largest (41.78 ERB²), followed by the acoustic space of isolated vowels (32.64 ERB²), which was in turn followed by the acoustic space of words in sentences (24.58 ERB²). Numerically, across tasks the SMG vowel space was the largest (36.88 ERB²), followed by the Cretan (31.11 ERB²) and the Kozani (30.99 ERB²) space but the differences were not significant. Overall, the results showed that male and female speakers did not differ in terms of their perceptual vowel space (42.81 ERB² vs 41.78 ERB²) and that while female speakers' acoustic vowel space was, naturally, more expanded than males speakers' space, their perceptual space was still larger than their acoustic space (for an overview of the results, see Table 1).

3. Experiment 2

3.1 Participants

The participants were the three male speakers (one from each dialect) who had made the recordings used in experiment 1. Their mean age was 51 yrs (range = 35–65 yrs).

3.2 Materials and procedure

Production of isolated vowels and nonsense words in sentences. The five Greek vowels were recorded in isolation and in nonsense words in sentences. Instead of eliciting vowels in the sentence "I am not going to say pVpV anywhere" (as was done in experiment 1), it was decided to ask participants produce the vowels in the same sentence used in the best exemplar experiment ("Say /pVta/ again"). This way production and perception data were elicited in the exact same context. A total of 60 vowel tokens were recorded (3 speakers \times 5 vowels \times 2 tasks \times 2 repetitions), using the same recording procedure and equipment as in experiment 1.

Location of best exemplars. Same materials and procedures as in experiment 1, the only difference being that participants would hear their own voice during testing.

3.3 Results

Figure 2 plots the best exemplars (black arrows and symbols) and the vowels produced in nonsense words in sentences (gray arrows and symbols) by the SMG, the Cretan, and the Kozani speaker. As mentioned above, perception and production data have been elicited in the same context. All speakers showed expanded perceptual spaces

Table 1. Size of vowel space (ERB²) in three tasks for male and female speakers of SMG, Cretan, and Kozani Greek.

Task	Athens		Crete		Kozani	
	Male	Female	Male	Female	Male	Female
Production of isolated vowels	28.9	34.9	21.4	29.3	26	33.6
Production of nonsense words	25.8	27.5	21.9	25.2	19.4	20.1
Location of best exemplars	51	48.1	37.2	38.7	40.2	38.5

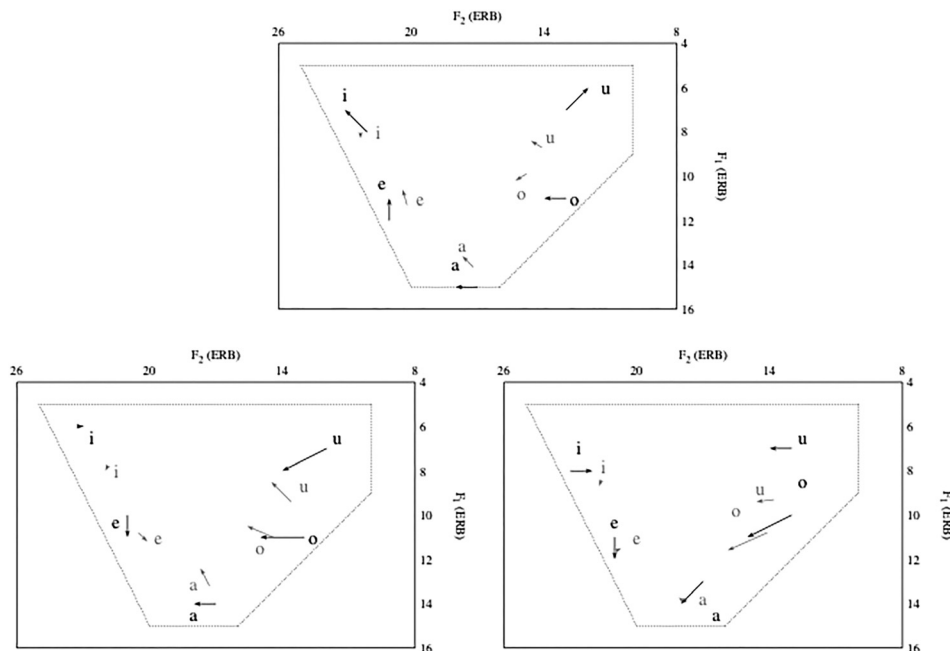


Fig. 2. Locations of vowel best exemplars (black arrows from the starting to the ending $F1$ and $F2$ frequencies and black symbols) and of vowels produced in nonsense words in sentences (gray arrows from the starting to the ending $F1$ and $F2$ frequencies and gray symbols) for the SMG (upper panel), the Cretan (lower left panel), and the Kozani Greek speaker (lower right panel). Dotted lines indicate the limits of the synthesized vowels available to listeners.

compared to the acoustic ones. This can be better illustrated in Table 2 showing the size of the perceptual space and the size of the acoustic spaces of isolated vowels and words in sentences (ERB^2) after removing the $F1-F2$ formant movement. The perceptual space of speakers who had mapped vowel prototypes to their own voice had approximately double the size of their acoustic space, even when compared with that of vowels spoken in isolation (with the exception of the Kozani Greek speaker where the difference was slightly smaller).

4. Discussion

The results of experiment 1 showed dialect-induced differences in the perceptual organization of vowel spaces with mean Euclidean distances between adjacent best exemplars being larger in the standard variety (SMG) than in the two non-standard varieties (Crete and Kozani Greek) and, at the same time, with the effect of dialect not being consistent across adjacent best exemplars. Similar dialectal effects in the acoustic vowel space are found in Greek and other languages (e.g., Adank *et al.*, 2007; Clopper *et al.*, 2005; Lengieris and Nicolaidis, 2015; Recasens and Espinosa, 2006; Trudgill, 2009; Jacewicz *et al.*, 2006). When comparing perception and production data, it was found that both male and female participants had perceptual spaces that were more expanded than their acoustic spaces for vowels spoken in isolation, which were in turn more expanded than their acoustic spaces for vowels spoken in a nonsense word in sentences. Female speakers' results are noteworthy because despite having larger (compared to male speakers) acoustic spaces their perceptual spaces were still larger than the acoustic ones. Group results obtained in experiment 1 were confirmed by a direct comparison of individual perception and production data in experiment 2; the perceptual space of all three speakers who had mapped vowel prototypes to their own voice was more expanded than their acoustic space.

The prediction that isolated vowels would be closer to perceptual targets than to vowels spoken in sentences was not confirmed. Two, not mutually exclusive, explanations could account for this finding. First, because participants were not instructed to

Table 2. Size of vowel space (ERB^2) in three tasks for the SMG, the Cretan, and the Kozani speaker.

Task	Athens	Crete	Kozani
Isolated vowels	26.1	22.5	24.2
Nonsense words	21.2	18.4	21.5
Best exemplars	51.4	44.9	37.3

speak clearly as done in previous studies (see, e.g., Frieda *et al.*, 2000), it is possible that the produced isolated vowels were not particularly hyperarticulated. Second, the availability of a large multidimensional space may have allowed participants to reach perceptual targets that were too extreme to approximate even in isolated vowel production.

In conclusion, the results of the two experiments are in support of the validity of the hyperspace effect. The fact they all individuals in experiment 2 had larger perceptual spaces than their acoustic spaces combined with the group results in experiment 1 collected from speakers of the standard variety in Greek and two regional dialects suggest that the hyperspace effect cannot be attributed to the methodology used. The effect is instead a robust phenomenon that reliably reveals listeners' internal structure of vowel representations, although, admittedly, since perceptual prototypes are more extreme than produced vowels, such representations cannot be drawn directly from the pool of vowel exemplars listeners have experienced.

Acknowledgments

This research has been co-financed by the European Union and Greek national funds (THALES: VOCALECT Grant No. MIS 379396). The author would like to thank Paul Iverson for offering the software he created for the purposes of locating vowel best exemplars, Mary Baltazani for her helpful comments on an earlier draft of this paper, and Anna Sfakianaki and Eleni Tsiartsioni for collecting dialectal data. Portions of this work were presented at the 18th International Congress of Phonetic Sciences (Glasgow, United Kingdom) and at the 14th Conference on Laboratory Phonology (Tokyo, Japan).

References and links

- Adank, P., Van Hout, R., and Smits, R. (2007). "An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties," *J. Acoust. Soc. Am.* **121**, 1130–1141.
- Boersman, P., and Weenink, D. (2006). "Praat: Doing phonetics by computer" [Computer software], University of Amsterdam, Amsterdam, The Netherlands.
- Botinis, A., Fourakis, M., and Hawks, J. W. (1997). "A perceptual study of the Greek vowel space using synthetic stimuli," in *Proceedings of the 5th Eurospeech*, pp. 1307–1310.
- Clopper, C., Pisoni, D., and de Jong, K. (2005). "Acoustic characteristics of the vowel systems of six regional varieties of American English," *J. Acoust. Soc. Am.* **118**, 1661–1676.
- Clopper, C. G., and Pisoni, D. B. (2004). "Some acoustic cues for the perceptual categorization of American English regional dialects," *J. Phon.* **32**, 111–140.
- Evans, B. G., and Iverson, P. (2004). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *J. Acoust. Soc. Am.* **115**, 352–361.
- Evans, B. G., and Iverson, P. (2007). "Plasticity in vowel perception and production: A study of accent in young adults," *J. Acoust. Soc. Am.* **121**, 3814–3826.
- Frieda, E., Walley, A., Flege, J., and Sloane, M. (2000). "Adults' perception and production of the English vowel /i/,", *J. Speech Hear. Res.* **43**, 129–143.
- Glasberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Hawks, J. W., and Fourakis, M. (1995). "The perceptual vowel spaces of American English and Modern Greek: A comparison," *Lang. Speech* **38**, 237–252.
- Iverson, P., and Evans, B. (2009). "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German Speakers," *J. Acoust. Soc. Am.* **126**(2), 866–877.
- Jacewicz, E., Fox, R., Holt, Y., and Salmons, J. (2006). "Acoustic characteristics of vowels in three regional dialects of American English," *J. Acoust. Soc. Am.* **120**(5), 3294.
- Jacewicz, E., and Fox, R. A. (2012). "The effects of cross-generational and cross-dialectal variation on vowel identification and classification," *J. Acoust. Soc. Am.* **131**(2), 1413–1433.
- Johnson, K. (2000). "Adaptive dispersion in vowel perception," *Phon.* **57**, 181–188.
- Johnson, K., Flemming, E., and Wright, R. (1993). "The hyperspace effect: Phonetic targets are hyperarticulated," *Lang.* **69**, 505–528.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *J. Acoust. Soc. Am.* **87**, 820–857.
- Lengieris, A., and Nicolaidis, K. (2015). "The vowels of Modern Greek dialects," in *36th Annual Meeting, AUTH Department of Linguistics*, April 24–25.
- Recasens, D., and Espinosa, A. (2006). "Dispersion and variability of Catalan vowels," *Speech Commun.* **48**, 645–666.
- Trudgill, P. (2009). "Greek dialect vowel systems, vowel dispersion theory, and sociolinguistic typology," *J. Greek Ling.* **9**, 165–182.
- Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., and Iskarous, K. (2004). "Vowel production and perception: Hyperarticulation without a hyperspace effect," *Lang. Speech* **47**(2), 155–174.
- www.vocalect.eu (2015). "Vowels of Greek dialects: phonetic and phonological analysis, dialect index construction, and diffusion of results through educational material." [Project website], University of Ioannina, Greece (Last viewed 3 May 2016).