IOWA STATE UNIVERSITY

Using Acoustic Measurement as an Estimate of Vowel Intelligibility

PSLLT 2021
June 18-19

Mahdi Duris

Introduction

Background

- Intelligibility measurements of L2 speech involve 2 factors: the speaker and the listener
  - Listener's judgments of intelligibility by way of ratings
- In English segmentals: vowels are the nucleus (Fogerty & Humes, 2012)
  - Furthermore, vowel height (F1) carries 80% of the energy in a vowel (Ladefoged & Johnson, 2015)
- Estimating vowel intelligibility using acoustic phonetic measurements has been successfully done by Koffi (2021) for L2 intelligibility (12,000 tokens across 7 different L2s)
  - Acoustic thresholds that consider the Critical Band Theory (CBT), Just Noticeable Differences (JND) and Relative Functional Load (RFL) to posit the Acoustic Masking and Intelligibility (AMI) theory
  - AMI theory states that " **segments that are acoustically close may mask each other with only minimal risk to intelligibility, unless their relative functional loads indicate otherwise**." (p. 55)
  - Phoneticians can measure vowel intelligibility instrumentally ≠ listener ratings
  - L1 phonological background of L2 speakers is taken into account

Current Gap

- Focus on native listener ratings for intelligibility = lack of L2 speaker independent learning
- Most technology assisted tools designed to improve intelligibility do not give direct feedback to speakers in consideration of their L1 phonological background
  - Feedback by an ideal-IPA for the segment, coded with ARPABET. No threshold considerations or pedagogical solutions
- Koffi's AMI theory enables phoneticians to assess vowel intelligibility while considering the speakers L1 inventory of vowels and not requiring the judgements of native listeners.

To demonstrate the use of acoustic measurements as an estimate of vowel intelligibility, 32 advanced L1 Arabic speakers of English provided the data to respond to the following:

1. **What are the L2 vowel characteristics of the participants?**
   - This considers their L1 vowel inventory and provides a clear picture of potential problematic vowels
2. **How similar or different are these vowels compared with General American English vowels?**
   - Some vowels may be less problematic than others when considering intelligibility
3. **Do the differences in F1 interfere with intelligibility?**

Methodology

Participants

- 32 Saudi EFL adult teachers (23 females, 9 males)
- Mean age: 33 years old (ranging from 19 years to 53 years old)

Stimuli

- Modified read speech from the George Mason Speech Accent Archive: *Please call Stella…*

Data Analysis

- Each vowel analyzed was manually extracted from 3 different words using PRAAT
- 7392 tokens (11 vowels x 3 repetitions x 32 participants x 7 correlates) in total were measured with only F1 and F2 used for this study (2,112 tokens)

Table 1
Vowel sound names from the read speech.

Vowel sound and name										
fleece	**kiss**	**face**	**dress**	**trap**	**lot**	**thought**	**goat**	**foot**	**goose**	**strut**
[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Text equivalent										
please peas meet	with thick is	maybe faked paper	yellow edge red	ask pad mat	Bob dog frog	for bought corner	old go zone	good book cookie	blue scoop zoo	rubber duck must

Thresholds

- CBT: 60 Hz (F1) is a robust criterion for distinguishing between perceptually similar vowels (Labov et al., 2013).
- JND: If the distance between 2 vowels for F1 is ≤60 Hz = masking is likely. Complete masking occurs when the acoustic distance is ≤ 20 Hz (Koffi, 2021:75)
- RFL: The higher the RFL = the greater the likelihood of unintelligibility.
- Internal masking: how a speaker distinguishes between the frequencies of their own vowels
- External masking: how a speaker distinguishes between the frequencies of their vowel compared to the vowels of another speaker (speaker vs listener)

Table 2
Intelligibility Assessment Matrix (Koffi, 2021:75)

#	F1 Distance	Masking Levels	RFL	Intelligibility Rating
1.	> 60 Hz	No masking	0-24%	Good intelligibility
2.	41 Hz – 60 Hz	Slight masking	25-49%	Fair intelligibility
3.	21 Hz – 40 Hz	Moderate masking	50–74%	Average intelligibility
4.	0 Hz – 20 Hz	Complete masking	75–100%	Poor intelligibility

Results

L2 Vowel Characteristics for female participants

Figure 2
Vowel characteristics of female participants (internal vs external masking)

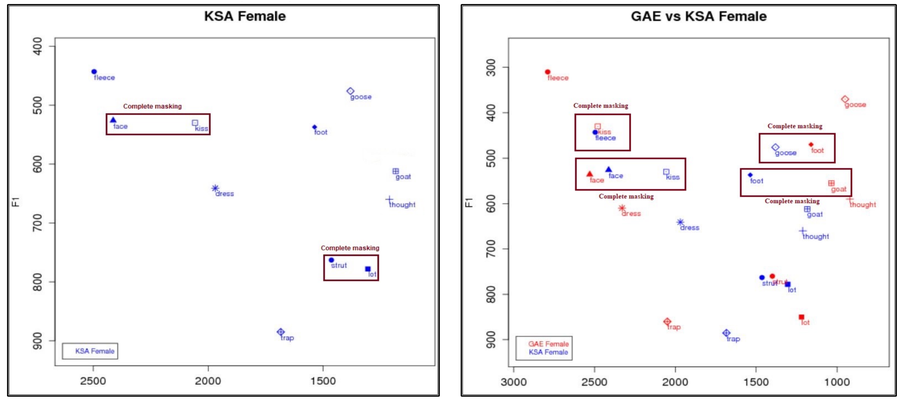

Table 3
Acoustic Masking and Intelligibility assessment for female participants

Vowel Pairs	F1 Distance	Internal Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	87 Hz	No masking	95%	Good intelligibility
[ɪ] vs. [e]	**4 Hz**	**Complete masking**	**80%**	**Poor intelligibility**
[ʌ] vs. [ɑ]	**14 Hz**	**Complete masking**	**65%**	**Average intelligibility**

Vowel Pairs	F1 Distance	External Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	**13 Hz**	**Complete masking**	**95%**	**Poor intelligibility**
[ɪ] vs. [e]	**6 Hz**	**Complete masking**	**80%**	**Poor intelligibility**
[e] vs. [ɛ]	84 Hz	No masking	53%	Good intelligibility
[u] vs. [ʊ]	**6 Hz**	**Complete masking**	**7%**	**Good intelligibility**
[ʊ] vs. [o]	**18 Hz**	**Complete masking**	**12%**	**Good intelligibility**

Summary: For female participants, only one pair of vowels are problematic when considering internal masking ([ɪ] vs. [e]). For external masking, two pairs of vowels are problematic ([i] vs. [ɪ] and [ɪ] vs. [e])

L2 Vowel Characteristics for male participants

Figure 3
Vowel characteristics of male participants (internal vs external masking)

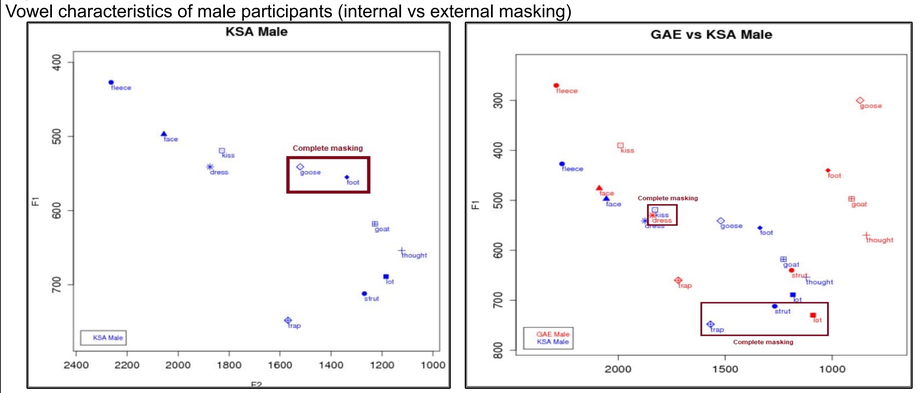

Table 4
Acoustic Masking and Intelligibility assessment for male participants

Vowel Pairs	F1 Distance	Internal Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	92 Hz	No masking	95%	Good intelligibility
[ɪ] vs. [e]	22 Hz	Moderate masking	80%	Poor intelligibility
[u] vs. [ʊ]	**14 Hz**	**Complete masking**	**7%**	**Good intelligibility**

Vowel Pairs	F1 Distance	External Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	37 Hz	Moderate masking	95%	Average intelligibility
[ɪ] vs. [ɛ]	**11 Hz**	**Complete masking**	**54%**	**Average intelligibility**
[ʌ] vs. [ɑ]	**18 Hz**	**Complete masking**	**65%**	**Average intelligibility**
[æ] vs. [ɑ]	**18 Hz**	**Complete masking**	**76%**	**Poor intelligibility**

Summary: For male participants, none of their vowels are problematic when considering internal masking. For external masking, one pair of vowels are problematic ([æ] vs. [a])

Discussion

Summary

- For the female participants in this study
  - Only 2 high vowels : [i] and [u]
    - The kiss vowel [ɪ] is produced as a mid vowel
  - These speakers do not distinguish the following pairs in their own production:
    - [ɪ] vs. [ɛ] (ex: din vs. den)
    - [ʌ] vs. [ɑ] (ex: gut vs. got)
  - Only ([ɪ] vs. [e]) causes intelligibility issues based on the AMI theory when considering JND and RFL:
    - The acoustic distance is below the 20 Hz threshold
    - The functional load for this pair is high at 80%
  - When comparing their vowels to GAE:
    - The following vowels are problematic: [i], [ɪ], [u] and [ʊ]
    - Only two pairs causes severe intelligibility issues based on the AMI theory:
      - [i] vs. [ɪ] (ex: Pete vs. pit)
      - [ɪ] vs. [e] (ex: pit vs. pet)
- For the male participants
  - The participants did not produce any high vowels
    - They do not present any difficulties in distinguishing between their L2 vowels
    - Only 1 instance of internal masking for [u] vs. [ʊ] but with a low RFL
    - When compared to their GAE counterparts:
      - 3 pairs of vowels present an external masking but only 1 pair ([æ] vs. [ɑ]) causes poor intelligibility (ex: rad vs. rod)

Implications

The study shows a clear difference in vowel production between female and male participants.

Implications for female L1 Arabic speakers of English

- The participants vowel production was highly intelligible (81%)
- Only two pairs of vowels cause poor intelligibility and should be a focus in instruction when appropriate
  - Focus should be given in raising the frequencies of the problematic vowels
  - Differentiating between [i], [] and [e] vowels by using minimal pair awareness

Implications for male L1 Arabic speakers of English

- The participants were highly intelligible with only one pair causing poor intelligibility
- Focus in instruction should be given to raising the *trap* vowel [æ] and clearly differentiating it from the *lot* vowel [a]
  - Minimal pair activities could be an effective and simple way of increasing intelligibility

Pedagogical applications

- Using acoustic measurements as an estimate of vowel intelligibility can help learners and teachers by way of:
  - The Noticing Hypothesis
  - Provide an independent learning tool
  - Educators can incorporate this measurement to their pronunciation lessons

References

Fogerty, D., & Humes, L. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America, 131*(2), 1490-1501.

Koffi, E. (2021). Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility. CRC Press, 2020. Boca Raton

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound changes in Philadelphia: Linear incrementation, reversal, and reanalysis. Language, 89(1). Linguistic Society of America.

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics (7th ed.)*. Stamford, CT, USA: Cengage Learning

IOWA STATE UNIVERSITY

Using Acoustic Measurement as an Estimate of Vowel Intelligibility

PSLLT 2021
June 18-19

Mahdi Duris

Introduction

Background

- Intelligibility measurements of L2 speech involve 2 factors: the speaker and the listener
  - Listener's judgments of intelligibility by way of ratings

- In English segmentals: vowels are the nucleus (Fogerty & Humes, 2012)
  - Furthermore, vowel height (F1) carries 80% of the energy in a vowel (Ladefoged & Johnson, 2015)

- Estimating vowel intelligibility using acoustic phonetic measurements has been successfully done by Koffi (2021) for L2 intelligibility (12,000 tokens across 7 different L2s)
  - Acoustic thresholds that consider the Critical Band Theory (CBT), Just Noticeable Differences (JND) and Relative Functional Load (RFL) to posit the Acoustic Masking and Intelligibility (AMI) theory
  - AMI theory states that " **segments that are acoustically close may mask each other with only minimal risk to intelligibility, unless their relative functional loads indicate otherwise**." (p. 55)
  - Phoneticians can measure vowel intelligibility instrumentally ≠ listener ratings
  - L1 phonological background of L2 speakers is taken into account

Current Gap

- Focus on native listener ratings for intelligibility = lack of L2 speaker independent learning
- Most technology assisted tools designed to improve intelligibility do not give direct feedback to speakers in consideration of their L1 phonological background
  - Feedback by an ideal-IPA for the segment, coded with ARPABET. No threshold considerations or pedagogical solutions

- Koffi's AMI theory enables phoneticians to assess vowel intelligibility while considering the speakers L1 inventory of vowels and not requiring the judgements of native listeners.

To demonstrate the use of acoustic measurements as an estimate of vowel intelligibility, 32 advanced L1 Arabic speakers of English provided the data to respond to the following:

1. **What are the L2 vowel characteristics of the participants?**
   - This considers their L1 vowel inventory and provides a clear picture of potential problematic vowels
2. **How similar or different are these vowels compared with General American English vowels?**
   - Some vowels may be less problematic than others when considering intelligibility
3. **Do the differences in F1 interfere with intelligibility?**

IOWA STATE UNIVERSITY

Using Acoustic Measurement as an Estimate of Vowel Intelligibility

PSLLT 2021
June 18-19

Mahdi Duris

Methodology

Participants

- 32 Saudi EFL adult teachers (23 females, 9 males)
- Mean age: 33 years old (ranging from 19 years to 53 years old)

Stimuli

- Modified read speech from the George Mason Speech Accent Archive: *Please call Stella…*

Data Analysis

- Each vowel analyzed was manually extracted from 3 different words using PRAAT
- 7392 tokens (11 vowels x 3 repetitions x 32 participants x 7 correlates) in total were measured with only F1 and F2 used for this study (2,112 tokens)

Table 1

Vowel sound names from the read speech.

Vowel sound and name										
fleece	**kiss**	**face**	**dress**	**trap**	**lot**	**thought**	**goat**	**foot**	**goose**	**strut**
[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]
Text equivalent										
please	with	maybe	yellow	ask	Bob	for	old	good	blue	rubber
peas	thick	faked	edge	pad	dog	bought	go	book	scoop	duck
meet	is	paper	red	mat	frog	corner	zone	cookie	zoo	must

Thresholds

- CBT: 60 Hz (F1) is a robust criterion for distinguishing between perceptually similar vowels (Labov et al., 2013).
- JND: If the distance between 2 vowels for F1 is ≤60 Hz = masking is likely. Complete masking occurs when the acoustic distance is ≤ 20 Hz (Koffi, 2021:75)
- RFL: The higher the RFL = the greater the likelihood of unintelligibility.
- Internal masking: how a speaker distinguishes between the frequencies of their own vowels
- External masking: how a speaker distinguishes between the frequencies of their vowel compared to the vowels of another speaker (speaker vs listener)

Table 2

Intelligibility Assessment Matrix (Koffi, 2021:75)

#	F1 Distance	Masking Levels	RFL	Intelligibility Rating
1.	> 60 Hz	No masking	0-24%	Good intelligibility
2.	41 Hz – 60 Hz	Slight masking	25-49%	Fair intelligibility
3.	21 Hz – 40 Hz	Moderate masking	50–74%	Average intelligibility
4.	0 Hz – 20 Hz	Complete masking	75–100%	Poor intelligibility

IOWA STATE UNIVERSITY

Using Acoustic Measurement as an Estimate of Vowel Intelligibility

PSLLT 2021
June 18-19

Mahdi Duris

Results (1/2)

L2 Vowel Characteristics for female participants

Figure 2
Vowel characteristics of female participants (internal vs external masking)

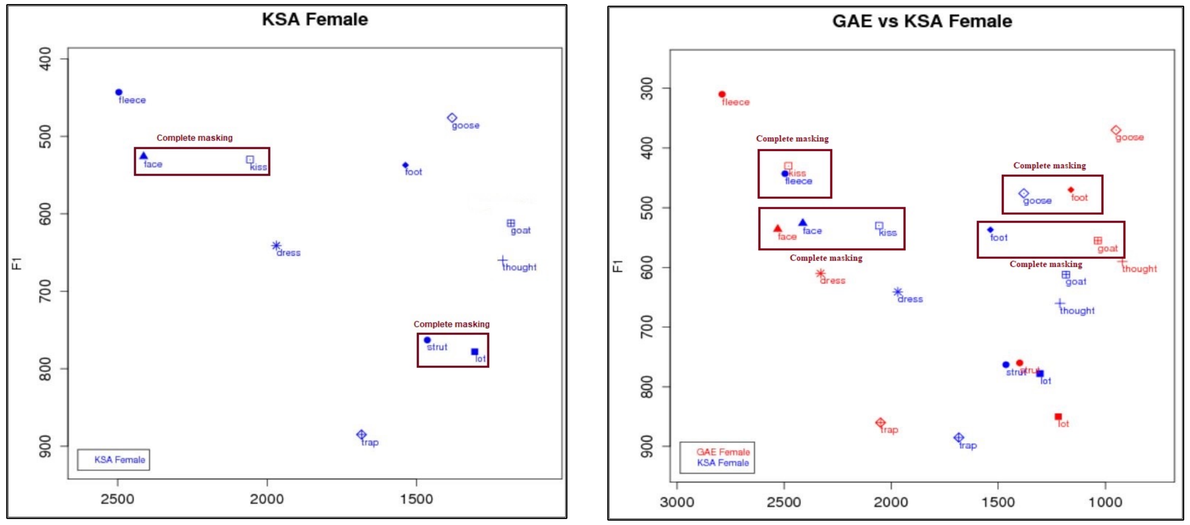

Table 3
Acoustic Masking and Intelligibility assessment for female participants

Vowel Pairs	F1 Distance	Internal Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	87 Hz	No masking	95%	Good intelligibility
[ɪ] vs. [e]	**4 Hz**	**Complete masking**	**80%**	**Poor intelligibility**
[ʌ] vs. [ɑ]	**14 Hz**	**Complete masking**	**65%**	**Average intelligibility**

Vowel Pairs	F1 Distance	External Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	**13 Hz**	**Complete masking**	**95%**	**Poor intelligibility**
[ɪ] vs. [e]	**6 Hz**	**Complete masking**	**80%**	**Poor intelligibility**
[e] vs. [ɛ]	84 Hz	No masking	53%	Good intelligibility
[u] vs. [ʊ]	**6 Hz**	**Complete masking**	**7%**	**Good intelligibility**
[ʊ] vs. [o]	**18 Hz**	**Complete masking**	**12%**	**Good intelligibility**

Summary: For female participants, only one pair of vowels are problematic when considering internal masking ([ɪ] vs. [e]). For external masking, two pairs of vowels are problematic ([i] vs. [ɪ] and [ɪ] vs. [e])

Results (2/2)

L2 Vowel Characteristics for male participants

Figure 3
Vowel characteristics of male participants (internal vs external masking)

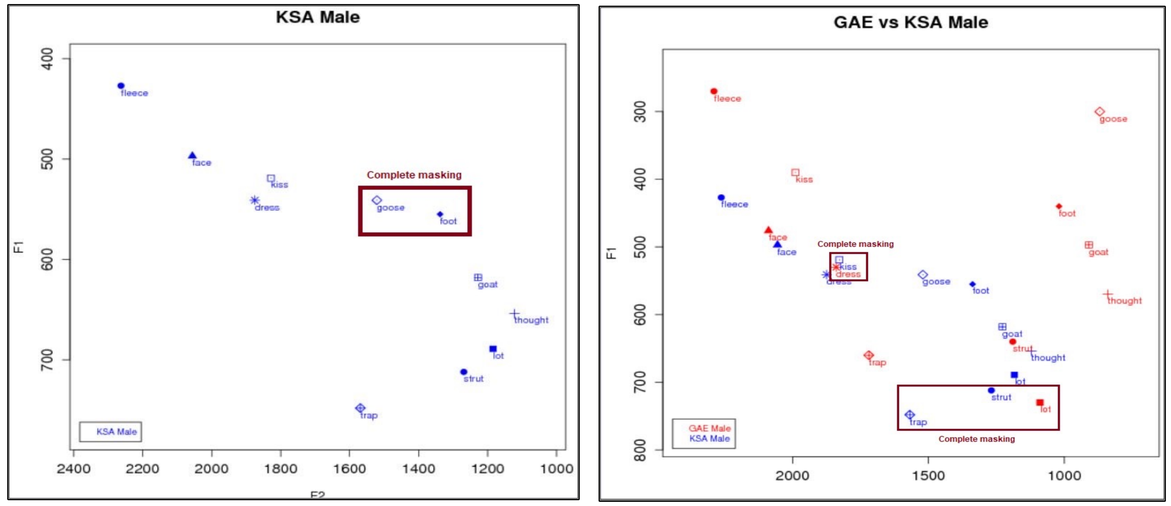

Table 4
Acoustic Masking and Intelligibility assessment for male participants

Vowel Pairs	F1 Distance	Internal Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	92 Hz	No masking	95%	Good intelligibility
[ɪ] vs. [e]	22 Hz	Moderate masking	80%	Poor intelligibility
[u] vs. [ʊ]	**14 Hz**	**Complete masking**	**7%**	**Good intelligibility**

Vowel Pairs	F1 Distance	External Masking Levels	RFL	Intelligibility Rating
[i] vs. [ɪ]	37 Hz	Moderate masking	95%	Average intelligibility
[ɪ] vs. [ɛ]	**11 Hz**	**Complete masking**	**54%**	**Average intelligibility**
[ʌ] vs. [ɑ]	**18 Hz**	**Complete masking**	**65%**	**Average intelligibility**
[æ] vs. [ɑ]	**18 Hz**	**Complete masking**	**76%**	**Poor intelligibility**

Summary: For male participants, none of their vowels are problematic when considering internal masking. For external masking, one pair of vowels are problematic ([æ] vs. [a])

IOWA STATE UNIVERSITY

Using Acoustic Measurement as an Estimate of Vowel Intelligibility

PSLLT 2021
June 18-19

Mahdi Duris

Discussion

Summary

- For the female participants in this study
  - Only 2 high vowels : [i] and [u]
    - The kiss vowel [ɪ] is produced as a mid vowel
  - These speakers do not distinguish the following pairs in their own production:
    - [ɪ] vs. [e] (ex: din vs. den)
    - [ʌ] vs. [ɑ] (ex: gut vs. got)
  - Only ([ɪ] vs. [e]) causes intelligibility issues based on the AMI theory when considering JND and RFL:
    - The acoustic distance is below the 20 Hz threshold
    - The functional load for this pair is high at 80%
  - When comparing their vowels to GAE:
    - The following vowels are problematic: [i], [ɪ], [u] and [ʊ]
    - Only two pairs causes severe intelligibility issues based on the AMI theory:
      - [i] vs. [ɪ] (ex: Pete vs. pit)
      - [ɪ] vs. [e] (ex: pit vs. pet)
- For the male participants
  - The participants did not produce any high vowels
    - They do not present any difficulties in distinguishing between their L2 vowels
    - Only 1 instance of internal masking for [u] vs. [ʊ] but with a low RFL
    - When compared to their GAE counterparts:
      - 3 pairs of vowels present an external masking but only 1 pair ([æ] vs. [ɑ]) causes poor intelligibility (ex: rad vs. rod)

Implications

The study shows a clear difference in vowel production between female and male participants. Yet, their vowels are highly intelligible.

Implications for female L1 Arabic speakers of English

- The participants vowel production was highly intelligible (81%)
- Only two pairs of vowels cause poor intelligibility and should be a focus in instruction when appropriate
  - Focus should be given in raising the frequencies of the problematic vowels
  - Differentiating between [i], [ɪ] and [e] vowels by using minimal pair awareness

Implications for male L1 Arabic speakers of English

- The participants were highly intelligible with only one pair causing poor intelligibility
- Focus should be given to raise the *trap* vowel [æ] and clearly differentiating it from the *lot* vowel [a] during instruction
  - Minimal pair activities could be an effective and simple way of increasing intelligibility

Pedagogical applications

- Using acoustic measurements as an estimate of vowel intelligibility can help learners and teachers by way of:
  - The Noticing Hypothesis
  - Provide an independent learning tool
  - Educators can incorporate this measurement to their pronunciation lessons

References

Fogerty, D., & Humes, L. (2012). The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *The Journal of the Acoustical Society of America, 131*(2), 1490-1501.

Koffi, E. (2021). Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility. CRC Press, 2020. Boca Raton

Labov, W., Rosenfelder, I., & Fruehwald, J. (2013). One hundred years of sound changes in Philadelphia: Linear incrementation, reversal, and reanalysis. Language, 89(1). Linguistic Society of America.

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics (7th ed.).* Stamford, CT, USA: Cengage Learning