Koffi, E. (2013). Confusion as a complement to intelligibility research. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference.* Aug. 2012. (pp. 54-65). Ames, IA: Iowa State University.

CONFUSION AS A COMPLEMENT TO INTELLIGIBILITY RESEARCH

Ettien Koffi, St. Cloud State University

Confusion research has been used for more than 50 years to test speech intelligibility in automatic speech recognition systems. In this paper, I apply its methodology and its findings to L2 English intelligibility research. Preliminary findings indicate that the Perceptual Distance Hypothesis (PDH) can help to predict vocalic substitutions that impinge on intelligibility and those that do not. Furthermore, PDH confirms findings by Derwing, Flege, Munro, and others, that the English vowels [æ, ʌ, ɑ] produced by many nonnative talkers are more likely to impinge on intelligibility than other vowels. I conclude, hesitantly and pending further research, that the available data confirms that confusion findings and intelligibility findings complement each other.

INTRODUCTION

The concept of intelligibility that Derwing and Munro have championed for about two decades has brought both enthusiasm and scrutiny among L2 pronunciation teachers and researchers.

Levis (2005, p. 370) notes that the "Intelligibility Principle" has begun to assert itself as a dominant force to be reckoned with. Dauer (2005, p. 548) agrees with this assessment, yet she contends that intelligibility is "difficult to define and measure." Indeed, Levis (2011) reports that a panel of experts did not come to a consensus on the factors that contributed to the unintelligibility of two L2 speech samples. It is, therefore, not an accident that the main theme of the 2012 PSLLT conference deals with ways of assessing intelligibility reliably. In this paper, I review the methodology employed in intelligibility research and suggest that confusion research can be used in tandem with intelligibility research to assess the latter more reliably. Confusion research has contributed significantly to both speech science and speech recognition research. However, it is only now that Miller and Nicely's (1955) groundbreaking work is trickling down into mainstream phonetic and phonology textbooks. Ladefoged and Disner (2012, pp. 99-113) and Johnson (2012, pp. 112-127) have included a chapter on confusion in the latest editions of their popular phonetic textbooks. I contend in this paper that confusion research can complement intelligibility research because it provides principled answers to many of the issues that intelligibility researchers are confronted with. For instance, in standard intelligibility methodology, native speaker judges are asked to render judgments as to whether or not a non-native production of a specific word is intelligible or not. If the word is deemed unintelligible, the researcher catalogues all the unintelligible segments and looks for a pedagogical solution to the problem. However, confusion researchers are also interested in identifying unintelligible sounds, but their quest does not stop there. They probe further to find out which other sounds in English the unintelligible sound is confused with. Their findings help establish confusion rates. They also help predict the likelihood of confusion between a given phone and all other phones in the language.

Intelligibility Research: Its Goals, Scope, and Blind Spots

Derwing and Munro's research agenda centers around a lexical trio: Intelligibility, comprehensibility, and accentedness. On paper, the distinction between these three concepts seems straightforward. However, during active listening, it is quite difficult to tease out the causality relationship between them, as noted by Jun and Li (2010, p. 58):

> The more interesting finding seems to be that when the raters were rating for accentedness, they mentioned comprehensibility, and in the same way, when the raters were rating for comprehensibility, they mentioned accentedness. This could point to the possible interrelationship, or "quasi-dependency" between comprehensibility and accentedness.

This echoes what Dauer (2005, p. 548) wrote almost a decade ago, namely that, "Although everyone states that the goal of pronunciation is intelligibility, it is difficult to define and to measure." It is indeed difficult to diagnose, let alone predict, how, when, and why an L2 pronunciation may be accented, incomprehensible, or unintelligible. During the 2010 PSLLT conference, Levis, the conference organizer, put together a panel of nine intelligibility researchers and asked them to listen and assess the speech of two international students. The panelists found it extremely hard to pinpoint what exactly caused intelligibility problems. Levis (2011, pp. 56-57) summarized the results of the panel discussions as follows:

> All participants listened to the free speech followed by the read speech of each student. None of the listeners had heard the recordings before. Everyone was provided with an unmarked copy of the reading passage for each student and with the questions the students were asked in the interview. Each recording was played once (for time's sake), and the panelists and audience members were asked to identify the features of the speech that most influenced the speaker's intelligibility or lack thereof. The task was quite difficult (even for the experts). Listening once was not felt to be enough, and everyone thought that another opportunity to listen would have been helpful.

It is actually good that the experts did not have another opportunity to listen to the speech sample, because in real life, more often than not, we have only one chance to listen to a speaker and make sense of what he/she is saying. For fear of appearing rude or causing L2 talkers to lose face about their language abilities, L1 hearers refrain from asking them to keep repeating themselves even if they do not understand a fair amount of what is being said. The very fact that the experts were not able to offer an accurate diagnosis upon hearing the speech sample only once speaks volumes about the fact that it is easier to describe intelligibility academically than to use it as a yardstick to gauge and predict how and why an L2 speaker may be unintelligible to native speakers of General American English (GAE). One can sense a tinge of disappointment in Levis' report as he wrote:

> They [the experts] were supposed to be able to judge which pieces of the student's performance were causing the biggest problems, and they found themselves not able to get a handle on the speech patterns. (Levis, 2011, p. 57).

Two pages later, Levis writes, "What was clear was that even the experts do not completely agree on what most impacted intelligibility." I contend in this paper that insights from confusion research can help assess intelligibility more effectively.

Goals and Scope Confusion Research

In 1955 Miller and Nicely published an article entitled "An analysis of perceptual confusion among some English consonants." Since the publication of this article, confusion research has emerged as a fruitful area of scholarship in automatic speech recognition (ASR) and in speech perception studies. Confusion research depends crucially on the notion of perceptual distance between segments. This has led to the postulation of the Perceptual Distance Hypothesis (PDH), which I summarize and formulate for GAE as follows:

Perceptual Distance Hypothesis

> GAE segments that are in close perceptual proximity, *and* have a marginal relative functional load may be confused at no detriment to intelligibility, except as carriers of accentedness.

This postulate can be very useful in assessing the intelligibility of L2 English. In applying this concept, it is worth remembering that the perception of vowels is trickier than that of consonants. Ferrand (2007, pp. 263, 265, 287) suggests that this may be due to the fact that formant frequencies are the only cues that hearers have to distinguish between vowels. Johnson (2012, p. 108) describes an experiment in which the auditory perception of vowels was shown to be gradient, that is, the boundary lines between vowels often overlapped. Since vowels are perceptually harder to account for, if confusion findings can deal satisfactorily with vowels, then explaining the unintelligibility of consonants will be relatively easier. This is the reason why this paper is devoted exclusively to vowels.

Peterson and Barney (1952, p. 182) and Hillenbrand, Getty, Clark, and Wheeler (1995, p. 3108) provide the following confusion data about GAE vowels. Let's use the vowel [ɪ] to illustrate how the information in Tables 2 and 3 is interpreted by confusion researchers. Many GAE speakers produced the vowel [ɪ] (Spoken Stimuli). Many other GAE hearers were asked to indicate which vowel they heard (Perceived Stimuli). Overall, 92.8% of the hearers perceived [ɪ] accurately, but 0.5% of the hearers perceived it incorrectly as [ɪ], another group of hearers misperceived it as [ɛ] 6.7% of the time, some confused it with [æ] 0.1% of the time, and other hearers thought they heard [ɚ] 0.25% of the time.

Table 2
Peterson and Barney's Vowel Confusion Matrix

		Perceived Stimuli									
		[i]	[ɪ]	[ɛ]	[æ]	[ɑ]	[ɔ]	[ʊ]	[u]	[ʌ]	[ɚ]
Spoken Stimuli	[i]	99.9	.03	.05			.02				
	[ɪ]	.05	92.8	6.7	.01	>.01	>.01				.25
	[ɛ]		2.5	87.7	9.2	>.01	.02			.01	.49
	[æ]		>.01	2.9	96.5	.01	.01			.14	.37
	[ɑ]		>.01		.18	87.0	9.8	.67		2.2	.06
	[ɔ]			>.01	.01	5.7	92.8	.69	.04	.60	.13
	[ʊ]			>.01	>.01	.15	.49	96.5	.93	1.6	.18
	[u]			>.01		.01		.75	99.1		.01
	[ʌ]		>.01	>.01	.07	5.2	1.2	1		92.2	.2
	[ɚ]			.22	.05	.01	.02			.01	99.6

Note: Adapted from Peterson and Barney (1952, p. 182).

Table 3

Hillenbrand et al. (1995) Vowel Confusion Matrix

		Perceived stimuli												
		[i]	[ɪ]	[e]	[ɛ]	[æ]	[ɑ]	[ɔ]	[o]	[ʊ]	[u]	[ʌ]	[ɚ]	Total
Spoken Stimuli	[i]	99.6	0.1	0.1									0.1	99.9
	[ɪ]		98.8	0.2	0.9									99.9
	[e]	0.6	0.3	98.3	0.3		0.2	0.1		0.1				99.9
	[ɛ]		0.5		95.1	3.7	0.2	0.1		0.1		0.2	0.1	100
	[æ]				5.6	94.1	0.2							99.9
	[ɑ]		0.1		0.1	0.3	92.3	3.5	0.1	0.2	0.1	3.3		100
	[ɔ]						13.8	82.0		0.1	0.1	3.8	0.2	100
	[o]							0.1	99.2	0.5			0.2	99.5
	[ʊ]				0.1					97.5	1.3	1.0		99.8
	[u]								0.4	1.9	97.2		0.1	99.6
	[ʌ]						3.7	1.8	0.3	3.2	0.2	90.8	0.2	100.2
	[ɚ]	0.1				0.2				0.2			99.5	100

Note: Adapted from Hillenbrand et al. (1995, p. 3108).

On the basis of the information in Tables 2 and 3, it is possible to group GAE vowels into three perceptual categories: Perceptually close vowels, perceptually overlapping vowels, and perceptually distant vowels. This three-way division provides the basis for the discussions in the remaining sections of the paper.

Perceptually Close Vowels

The vowels [ɑ] and [ɔ] are perceptually very close. When these two vowels are produced, GAE hearers have a hard time telling them apart from each other. In Peterson and Barney (1952), the vowel [ɑ] was accurately perceived as [ɑ] 87% of the time. The vowel [ɔ] was also perceived accurately 92.8% of the time. The rest of the time, they were confused with each other. The vowel [ɑ] was confused with [ɔ] 9.8% of the time, while [ɔ] was confused with [ɑ] 5.7% of the time. The confusion rate of these two vowels is much higher in the Midwest according to Hillenbrand et al. (1995). In their data [ɔ] was confused with [ɑ] 13.8% of the time, whereas [ɑ] was confused with [ɔ] only 3.5% of the time. Many commentators, including Small (2005, pp. 74-75) have noted that the distinction between [ɔ] and [ɑ] is not made in the Midwest, the western United States, and in parts ofNew England. Consequently, pairs of words such as *cot/caught, tot/taught,* and *Don/Dawn* sound the same to the speakers of these dialects.

Formant frequency data from 22 Central Minnesota female speakers shows clearly that [ɑ] and [ɔ] have merged, as shown in Figure 1.

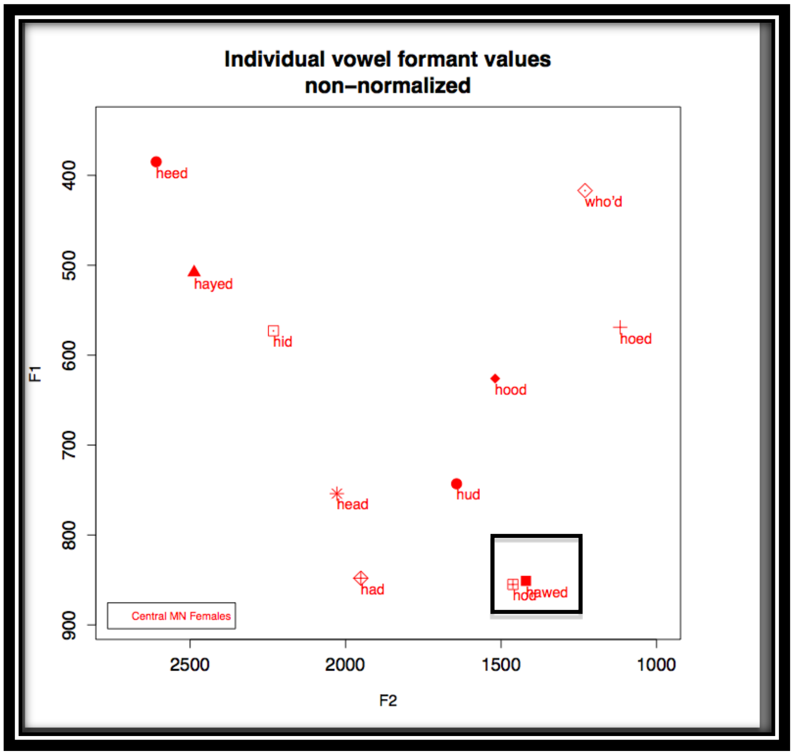

Figure 1. Acoustic Vowel Space of Central MN Females

The F1 formant of [ɑ] is 855 Hz, while that of [ɔ] is 851 Hz. The F2 formant of [ɑ] is 1420 Hz, and that of [ɔ] is 1462Hz. According to Baart (2010, p. 67), for two sounds to be perceived as acoustically different, there must be at least 200 Hz difference between the two F1s, and about 400 Hz difference between the two F2s. So, the difference of 4 Hz is insignificant for the two F1s, and so is the difference of 42 Hz for the two F2s. Therefore, acoustically speaking, [ɑ] and [ɔ] have merged in Central Minnesota. The merger of these two sounds was also reported in Ladefoged (2006, p. 89) who wrote that, "Many Midwestern and Californian speakers do not distinguish between [ɑ] and [ɔ]." Though Catford (1987, p. 90) rates the relative functional load of [ɑ] and [ɔ] at 26%, he also notes that in some dialects, the distinction is not made. Further evidence that [ɑ] and [ɔ] have merged in the speech of some people is found in Celce-Murcia, Brinton, Goodwin, & Griner (2010, p. 120). They instruct ESL teachers that /ɑ/ can be substituted for [ɔ] when teaching nonnative speakers. The Perceptual Distance Hypothesis proposed earlier works well for these two vowels; that is, since they have merged in many dialects of GAE, confusing one with the other, and vice versa, does not lead to unintelligibility.

Perceptually Distant Vowels

The confusion data found in Peterson and Barney (1952) and Hillenbrand et al. (1995) suggest that the features [± tense] and [± back] are used by GAE speakers to perceive perceptual distance. GAE hearers hardly ever confuse tense and lax vowels. Peterson and Barney (1952) did not study the vowels [e] and [o] in their data because they took them to be diphthongs. However, since in the Midwest, most people do not perceive these vowels as diphthongs, Hillenbrand et al. (1995, p. 3108) included them in their study of GAE simple vowels. So, the discussion of [+tense] vowels [i, e, o, u, ɑ] is based entirely on Hillenbrand et al. When the features of tenseness and backness are used in tandem, we see that all tense vowels are perceived accurately: [i] 99.6%, [e] 98.3, [u] 97.2%, and [o] 99.2%. The only tense vowel that is somewhat less accurately perceived is [ɑ] 92.3%, for reasons that were explained in the previous section. Various studies of the acquisition of English vowels by L2 speakers do not show any pronunciation difficulties with these vowels (Koffi, 2012; Munro & Derwing, 2006). The reason for this is that the tense vowels [i, e, u, o, a] are attested in many languages. In fact, Crothers (1978) notes that [i, u, a] are universal, and [e, o] are near universal.

Overlapping Front Lax Vowels

GAE vowels that have the feature [-tense], i.e., lax vowels, tend to overlap in perceptual space. The English vowels in this category are /ɪ, ɛ, æ, ʊ, ʌ, ɔ/. This group of vowels can be further sub-categorized between front vowels and back vowels. Front lax vowels overlap among themselves, and back lax vowels also overlap among themselves. Since the former overlap is different from the latter, the overlapping patterns must be dealt with separately. This section is devoted to overlapping front lax vowels, and the next will focus on overlapping back lax vowels.

The confusion data shows that [ɪ] never overlaps with [æ]. This means the vowels [ɪ] and [æ] are perceptually distant. This makes sense since [ɪ] is a high vowel, and [æ] is a low vowel, we expect them to be perceptually different. According to Peterson and Barney's (1952) confusion data, the vowel that straddles both [ɪ] and [æ] is the mid vowel [ɛ]. It is confused with [æ] 9.2% of the time for Perceived Stimuli, and with [æ] 2.9% of the time for Spoken Stimuli. The total confusion rate between these two segments is 12.1%. The vowel [ɛ] is also confused with [ɪ] 2.5% of the time, and [ɪ] with [ɛ] 6.7% of the time, for a total confusion rate of 9.2%. The perceptual distance between these three vowels is shown in Figure 2:

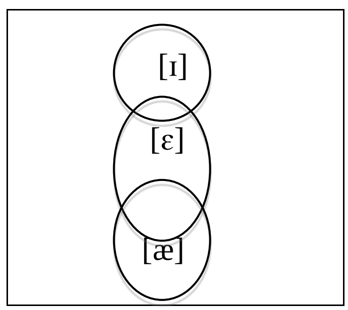

Figure 2. Overlapping between [ɪ, ɛ, æ]

The overlapping between [ɛ] and [ɪ], and [ɛ] and [æ] is phonologically conditioned. The phonological process is one of vowel raising. Some speakers raise [æ] to the height of [ɛ], and [ɛ] to the height of [ɪ] in predictable ways. The exact nature of the conditioning environment varies slightly from region to region, but the raising occurs immediately before voiced consonants. Ladefoged and Disner (2012, pp. 44-45) indicate <bad> [bæd] is pronounced [bɛd] in many of the northern metropolitan areas of the United States (Detroit, Rochester). In Central Minnesota, [æ] is raised to [ɛ] before [g]. In some regions, including the Ohio Valley, parts of Indiana, and in some southern states, [ɛ] is raised to the level of [ɪ] before [n]. So, words like *pen* and *Ben* are homophones with *pin* and *bin*. Lado and Fries (1958, p. 41) write in a note to instructors about teaching the distinction between [ɪ] and [ɛ] to nonnative speakers, saying that, "The distinction between [ɛ] and [ɪ] is likely to be more difficult than the distinction between [e] and [ɛ] and needs special attention."

The relative functional load between [ɪ] and [ɛ] is 54%, and that of [ɛ] and [æ] is 51%. This means that there are many lexical items in English in which these two vowels are contrastive. Consequently, L2 English talkers cannot confuse these pairs of vowels without impinging on intelligibility. GAE hearers expect [ɪ] and [ɛ], and [ɛ] and [æ] not to be confused except in cases where the confusion is predictable by phonological rules. In such environments, distinctions between the phonemes [ɪ] and [ɛ] are neutralized, as are distinctions between [ɛ] and [æ]. Even in such cases, if the hearer is not familiar with the dialect of the talker, intelligibility can occur. At a recent English Department meeting, the faculty representative of the department to the Interfaculty Organization reported that the "[ʃɛl] clauses" in the contract are being renegotiated. Many faculty members were confused and asked what the "[ʃɛl] clauses" were. The misunderstanding was not clarified until the faculty representative spelled the word. It turned out that the contract renegotiation deals with the clauses that have the modal verb "shall" in them. The talker is a native GAE speaker from a different area of the country. In her sociolect, [æ] has risen to [ɛ] before [l]. However, for most people in the audience, this phonological rule does not occur in their sociolect. This unexpected pronunciation of "shall" as [ʃɛl] led to an unintelligible exchange. This confusion confirms that Levis (2011, p. 64) is correct in stating that, "A lot of intelligibility comes down to our expectations. And any time you mess with expectations whether at the phonological level or at the lexical level or at the syntactic level or at the cultural level, you can impair intelligibility." In a more general sense, Rubin (2012, p. 12) echoes the same sentiment by noting that "Notwithstanding society's reliance on speech assessments, it should come as no surprise to discover that such perceptions are highly susceptible to the listener's own expectations of what she is about to hear."

Overlapping Back Lax Vowels

The situation with the back vowels [ʊ, ʌ, ɑ/ɔ] is similar to that of the overlapping front vowels discussed in the previous section. Here, the vowel that overlaps is [ʌ]. According to Peterson and Barney (1952), it is confused with [ɑ/ɔ] 5.2% of the time; [ɑ/ɔ] is confused with it 2.2% of the time. The vowel [ʌ] is confused with [ʊ] 1.6% of the time, and the latter is confused with the former 1.6% of the time. Among Midwesterners, [ʌ] is confused with [ɑ/ɔ] 3.7% of the time, and [ɑ/ɔ] is with [ʌ] 3.3% of the time. As for [ʌ] and [ʊ], the confusion rate in the Midwest is 3.2%, and [ʊ] with [ʌ] is 1%. Indeed, Small (2005, p. 79) notes that, "Students often confuse /ʌ/ with /ʊ/." The acoustic vowel space of Central Minnesota female speakers displayed in Figure 1 shows that [ʌ] is right in the middle of [ʊ] and [ɑ/ɔ]. It explains why [ʌ] is relatively easily confused with both [ʊ] and [ɑ/ɔ]. It has been reported in the L2 phonology literature that [u] and [ʊ] are merging (Celce-Murcia et al., 2010, p. 120). However, this claim is not supported by acoustic data or confusion data. From a standpoint of acoustic phonetics, what is happening is that [ʊ] is undergoing lowering. As a result, it is increasingly being confused with [ʌ]. I have done some informal confusion testing with Dragon Dictation, an automatic speech recognition application. The software does not sufficiently differentiate between <look> vs. <luck>, <book> vs. <buck>, <took> vs. <tuck>, <could> vs. <cud>, <put> vs. <putt>, <stud> vs. <stood> when these words are pronounced by some Central Minnesota speakers. More often than not, the software failed to recognize [ʌ] and perceived it as [ʊ]. I have done the same informal confusion testing with minimal pairs containing [ɑ/ɔ] vs. [ʌ] with the same results. The pairs of words tested were <duck> vs. <dock>, <dug> vs. <dog>, <hut> vs. <hot>, <cut> vs. <cot>/<caught>, <rub> vs. <rob>, <bust> vs. <bossed>, and <hug> vs. <hog> when these words are pronounced by Central Minnesotans. Occasional misunderstandings even among NAE talkers and hearers underscore the confusability of [ʌ]. Recently, the members of my daughter's middle school swim and dive team were completely confused when they were told that they should pick up their <mums> after practice. They all thought the sentence did not make sense because they did not know how to drive yet. Moreover, it is usually their moms who pick them up after practice, not the other way around.

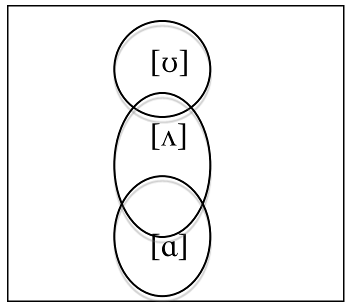

Figure 3. Overlapping between [ʊ, ʌ, ɑ]

Figure 3 provides a visual representation of the overlapping perceptual distance for these three back lax vowels. It underscores the fact that vowels [ɑ/ɔ] and [ʌ] are confused more often than [ʌ] and [ʊ]. Peterson and Barney's data gives a confusion rate of 7.2% for the former, and 3.2% for the latter. Hillenbrand et al. (1995) have a 7% confusion rate for [ɑ] and [ʌ], and a 4.2% rate for [ʊ] and [ʌ]. Catford (1987, p. 90) notes that the relative functional load of [ɑ/ɔ] and [ʌ] is 65%, while that of [ʊ] and [ʌ] is only 9%. Munro and Derwing (2006, p. 493) report that the Mandarin and Slavic participants in their experimental study did not improve their production of [ʊ]. Munro, Flege and MacKay (1996, p. 328) also write that the vowel [ʌ] was the most poorly identified by Canadian hearers who listened to speech samples produced by 240 Italian speakers. These studies indicate that the back vowels [ʊ, ʌ, ɑ/ɔ] are a challenge for L2 English talkers. This challenge is not likely to go away since their NAE (North American English) teachers themselves are increasingly failing to distinguish between them in their speech.

The merger of [æ] and [ɑ] in L2 English

Peterson and Barney found that [æ] was never confused with [ɑ/ɔ], and that [ɑ/ɔ] was confused with [æ] only .18% of the time. Hillebrand et al. also reported that [æ] was confused with [ɑ/ɔ] only .2% of the time, and [ɑ/ɔ] was confused with [æ] .3% of the time. These confusion rates are so small that they are insignificant. Since GAE hearers hardly ever confuse these two low vowels, it suggests that they are perceptually distant. The Perceptual Distance Hypothesis predicts that if L2 English talkers confuse them, intelligibility issues are to be expected. Is this prediction borne out? To answer this question, let's examine my pronunciations of <i-pad> and <i-pod>. I went to an electronic retail store to take a look at <i-pads>[1] when they first came on the market. I told the retail merchant that I was looking for <i-pads> and wanted to take a look at what he had. The retail clerk was a native GAE speaker. He disappeared for a moment and came back loaded with all types of <i-pods>. Why did he misunderstand me? My <i-pad> was misunderstood for <i-pod> because I do not distinguish between [æ] and [ɑ/ɔ] sufficiently well. As a result, there is practically no distance between these two vowels in my acoustic vowel space:

[1] The following phonetic/phonological conventions are used throughout the paper: the symbols < > stand for graphemes, / / for phonemes, and [] for phones and allophones.

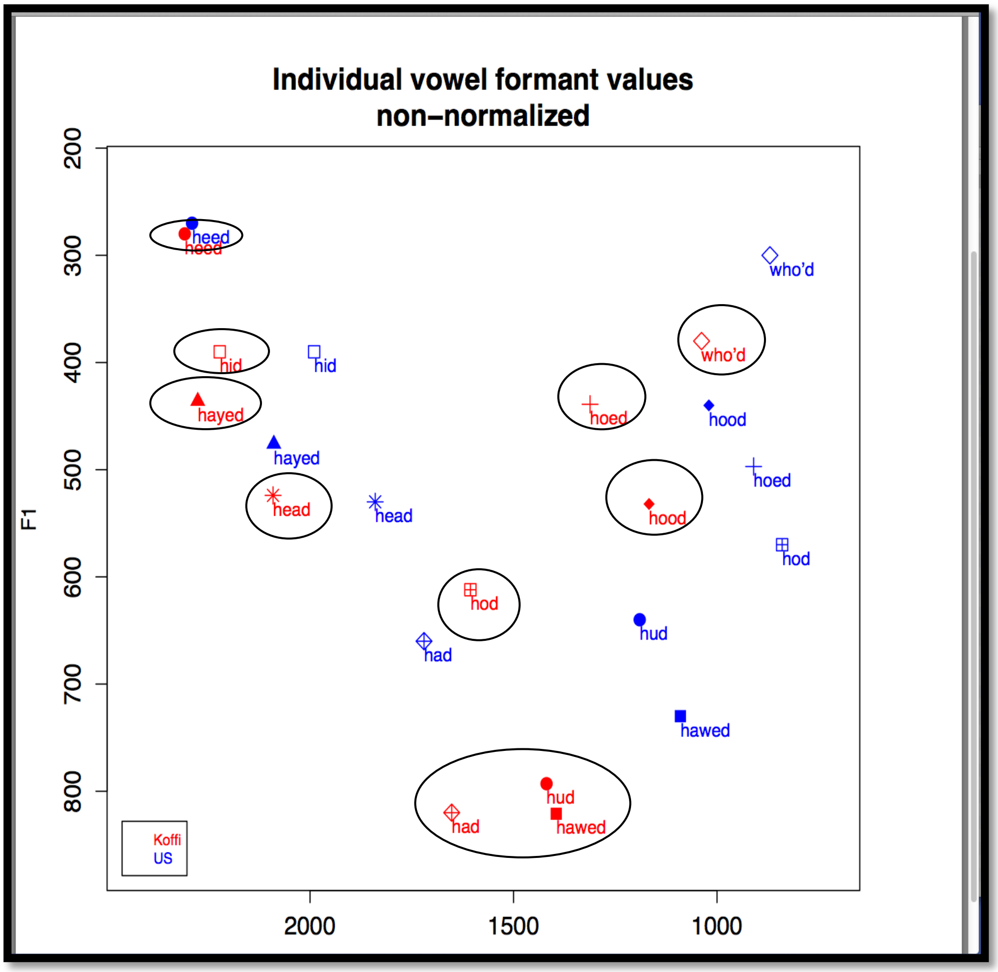

Figure 4. Koffi's English L2 acoustic vowel space

The F1 values for my [æ], [ʌ], and [ɔ/ɑ][2] are respectively 829 Hz, 793 Hz, and 821 Hz. The 7 Hz difference between my [æ] and [ɑ/ɔ] is below the threshold of human perception. Phoneticians say that humans cannot perceive frequencies that are below 20 Hz. Even the 36 Hz difference between my [æ] and [ʌ] is perceptually insignificant. My F2 values for the same vowels are respectively 1652 Hz, 1419 Hz, and 1606 Hz. As has been noted repeatedly in this paper, for two vowels not to be confused, an F2 frequency difference of 400 Hz is optimal. As for my [æ] and [ɑ], the acoustic vowel space shows clearly that they clustered together. This means that the vowels in my <pad>, <pod>, and <putt> are acoustically indistinguishable from one another. According to Munro et al. (1996, p. 328), I'm not the only L2 English talker who confuses these three vowels. They found that their NAE judges had trouble perceiving [ʌ] accurately when it was produced by Italian speakers, "Only one vowel could be said to be poorly identified overall: in the case of [ʌ], only 25% of the tokens were identified correctly. … The majority of

[2] My [ɑ] and [ɔ] are acoustically distinct. However, since many NAE speakers do not distinguish between the two, they mishear my [ɔ] as an [ɑ]. Circles are drawn around my vowels.

misidentified [ʌ] tokens were heard as [æ], though some were also heard as [ɑ]." Munro and Derwing (2006, pp. 491-2) also found that Mandarin and Slavic speakers had difficulty producing [æ]. They say that "although some /æ/ tokens were incorrectly produced as /ɛ/, a large number of incorrect productions from both speaker groups [i.e., Mandarin and Slavic speakers] fell into other categories. The majority of these cases were heard as /ɑ/ and /ʌ/." In other words, other non-native speakers find it challenging to produce [æ, ɑ, ʌ]. In this case, production difficulties lead to perception difficulties because these vowels have high frequency and moderately high functional load. Catford (1987, p. 89) ranks the relative functional load of [æ] vs. [ɑ]/[ɔ] at 76%, that of [æ] vs. [ʌ] at 68%, and [ʌ] vs. [ɑ]/[ɔ] at 65%. As for their frequency, Faircloth and Faircloth (1973, p. 57) list [æ] as the third most frequent vowel in English.

The combination of high frequency and high functional load makes the vowel /æ/ particularly troublesome for L2 English intelligibility. Munro and Derwing (2006, p. 491) write that /æ/ was the vowel that both Mandarin and Slavic speakers had the most trouble producing accurately. The worldwide distribution of [æ] may explain why speakers from various language backgrounds have a hard time producing it accurately. Cutler, Weber, Smits and Cooper (2004, p. 3675) found that many Dutch hearers could not distinguish [æ] from [ɑ] when they listened to American English. The confusion rate was 17.4%. Similarly, Lecumberri and Iragui (1997, p. 59) found that Spaniards confused [æ] with [ɑ] when they listened to British talkers. Data such as these confirm the claim that L2 English speakers cannot produce accurately phones that they cannot perceive accurately. Crothers (1978, p. 95) helps explain why [æ] is such a difficult phoneme to perceive and produce. He studied 209 languages in the Stanford Phonology Archiving Project. He found that only about 1% of the languages in the archive have both a low front [æ] and a low back [ɑ]. For the vast majority of these languages, there is only one low central vowel /a/. Acoustically, this low central /a/ is closer to /ɑ/ than it is to /æ/. Consequently, /æ/ is often produced in ways that it is perceived by NAE hearers either as [ʌ] or [ɔ/ɑ], but rarely as [æ].

This paper would be incomplete if I failed to make a passing remark about the intelligibility of [ɪ] and [ʊ]. These two vowels are often mentioned as being particularly difficult for L2 English speakers to produce. This may well be the case. However, exploratory L2 English acoustic vowel space data that I have collected over the years does not support this commonly held view. In fact, a comparison between Figures 1 and 4 shows that my [ɪ] is almost identical with the [ɪ] in Peterson and Barney (1952) whereas that of Central Minnesotans deviates significantly from NAE norms. Moreover, my [ʊ] is acoustically similar to the [ʊ] produced in Central Minnesota. My preliminary findings agree with what Munro and Derwing (2006, p. 488) found in their longitudinal study of vowel acquisition. They show that their participants produced /ɪ/ reasonably well over time. A comparison with Munro et al. (1996, p. 315) and Munro and Derwing (2006, p. 493) give conflicting results about [ʊ]. In an earlier study, Munro and his co-authors found that, "A larger number of Italian talkers produced [ʊ] tokens with native-like spectral and temporal properties." However, in a later study, he and Derwing found that their Mandarin and Slavic speakers did not show evidence of a noticeable improvement over time for /ʊ/. The acoustic vowel space data that I have of Mandarin and Slavic speakers do not show that they have any difficulty producing /ʊ/. Even if this were not the case, it is unlikely that their production of [ʊ] would result in unintelligibility. If they were to substitute [u] for [ʊ], that would not cause any intelligibility problem because the relative functional load between the two vowels is only 9%. Furthermore, Peterson and Barney (1952) and Hillenbrand et al. (1995) show that even NAE talkers and hearers often confuse the same two sounds. Consequently, even if nonnative speakers were to substitute [u] for [ʊ], NAE hearers would not be taken off guard.

SUMMARY

Three important observations can be drawn from the analysis presented in this paper. First, the Perceptual Distance Hypothesis helps predict accurately which L2 English sounds are likely to be unintelligible to NAE hearers. Confusion data suggests that six of the 11 phonemic vowels have overlapping perceptual distance. The vowels in question are /ɪ, ɛ, æ, ʊ, ʌ, ɑ/ɔ/. Coincidentally, these vowels are lax. L2 English phonologists have known for a long time that lax vowels are harder for nonnative speakers to acquire than tense vowels. Secondly, formant measurements, segment frequency, and functional load help situate L2 English intelligibility within the wider context of speech recognition. Thirdly, confusion data that has been accumulated over half a century lends empirical support to L2 English intelligibility research. The Perceptual Distance Hypothesis explains why the vowels [æ, ʌ, ɑ/ɔ] are more prone to unintelligibility when uttered by nonnative speakers of English.

ABOUT THE AUTHOR

Ettien Koffi is Professor of Linguistics. He teaches the linguistic courses in the TESOL/Applied Linguistics M.A. program in the English Department at Saint Cloud State University, MN. He has written three linguistic books and is the author of many peer-reviewed articles on various topics in linguistics. His primary area of specialization is at the interface between acoustic phonetics and phonology. He also has extensive experience in emergent orthographies and in the acoustic phonetic and phonological description of dialect variation. He can be reached via email at: enkoffi@stcloudstate.edu.

REFERENCES

Baart, J. (2010). *A field manual of acoustic phonetics.* Dallas, Texas: SIL International.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory*. (pp. 87-100). Washington, D.C.: Teachers of English to Speakers of Other Languages.

Celce-Murcia, M., Brinton, D. & Goodwin, J., with Griner, B. (2010). *Teaching pronunciation: A reference for teachers of English to speakers of other languages, 2nd Ed.*, New York: Cambridge University Press.

Crothers, J. (1978). Typology and universals of vowel systems. In J. Greenberg (Ed.) *The universals of human language: Volume 2, phonology* (pp. 93-152). Stanford, CA: Stanford University Press.

Culter, A., Weber A, Smits, R., & Cooper N. (2004). Patterns of English phoneme confusions by native and non-native speakers. *Journal of the Acoustical Society of America, 116,* 3668-3678.

Dauer, R. M. (2005). The lingua franca core: A new model for pronunciation instruction? *TESOL Quarterly*, *39* (3), 543-550.

Faircloth, S. & Faircloth, M. (1973). *Phonetic science: A program of instruction*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Ferrand, C. (2007). *Speech science: An integrated approach to theory and clinical practice, 2nd Ed*. New York: Pearson Allyn And Bacon.

Hillenbrand, J., Getty, L.A, Clark, Michael J., & Wheeler K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97* (5), 3099-3111.

Johnson, K. (2012). *Acoustic and auditory phonetics, 3rd ed.,* Malden, MA: Wiley-Blackwell.

Jun, H. G., & Li, J. (2010). Factors in raters' perception of comprehnsibility and accentedness. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Iowa State, University (pp. 53-66). Ames, IA: Iowa State University.

Koffi, E. (2012). Acoustic phonetic self-discovery: The editor's prologue. *Linguistic Portfolios 1*, 1-4.

Ladefoged, P. (2006). *A course in phonetics*, 5th edition. New York: Thomson-Wadsworth.

Ladefoged, P. & Disner S. F. (2012). *Vowels and consonants, 3rd ed.,* Malden, MA: Wiley Blackwell.

Lado, R. & Fries, C.C. (1954). *English Pronunciation: Exercises in Sound Segments, Intonation, and Rhythm*. Ann Arbor, MI: University of Michigan.

Lecumberri, L. G., & Iragui, J. C. (1997). L2 perception of English vowels: Testing the validity of Kuhl's prototypes. *Revista Alicantina de Estudios Ingleses, 10,* 55-68.

Levis, J. M. (2005). Changing context and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(5), 369-377.

Levis, J. M. (2011). Assessing speech intelligibility: Experts listen to two students. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, Iowa State, University (pp. 56-69). Ames, IA: Iowa State University.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27* (2), 338-352.

Munro, M., & Derwing T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34,* 520-531.

Munro, M., Flege, J., & MacKay, I. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17, 313-334*.

Peterson, G., & Barney, L. (1952). Control methods in a study of the vowels. *The Journal of the Acoustical Society of America, 24*(2), 176-84.

Rubin, D. L. (2012). The power of prejudice in accent perception: Reverse linguistic stereotyping and its impact on listener judgments and decisions. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, Iowa State, University (pp. 11-17). Ames, IA: Iowa State University.

Small. L. H. (2005). *Fundamentals of phonetics: A practical guide for students, 2nd Edition*. New York: Pearson Education, Inc.