AN INTEGRATED APPROACH TO PRONUNCIATION: LISTENING COMPREHENSION AND INTELLIGIBILITY IN THEORY AND PRACTICE

Marnie Reed, Boston University

Christina Michaud, Boston University

This paper introduces a theoretical basis for understanding the relationship between speaking and listening as an auditory feedback loop, in which speakers use their own mental model of a sound as input. We argue that speaking and listening are related: production facilitates perception, promoting more intelligible spontaneous speech and enhanced listening comprehension. Since learners use their own speech production as a filter, we suggest a model based on convergent production. The model is compared to Flege's Speech-Learning Model and also applied to the practical problem of pronunciation instruction.

We identify four major components of pronunciation instruction: connected speech features, suprasegmental features, inflectional morphology, and segmentals. Within this integrated model of pronunciation, the route to successful listening comprehension is through auditory feedback wherein the learner's own increasingly target-like speech production facilitates and reinforces perception. We introduce specifics for promoting learners' convergent output: spoken models, visual aides, and oral or written descriptions. A teacher-student partnership in which teachers offer a principled approach to corrective feedback in the form of production prompts is advocated. In our model, pronunciation instruction accompanies and reinforces core language instruction, and integrated pronunciation instruction is viewed as a highly focused, metacognitive approach to the entire language classroom.

INTRODUCTION

Research findings and anecdotal observations alike indicate that adult second language learners often have intractable pronunciation problems that persist in the face of explicit instruction or correction. Mere target language input is insufficient to create changes (Flege & Hillenbrand, 1984; Flege, 1993; Strange, 1995), or learners would never need instruction. The long-established failure of adult learners to incorporate syntactic structures to which they are exposed (Pavesi, 1986; Long, Inagaki, & Ortega, 1998) has a parallel in pronunciation. Learners frequently seem not to "hear" the target pronunciation even when it is modeled by their teachers, instead continuing with their original, incorrect pronunciation.

One way of making sense of this problem is to conceive of the relationship between a learner's speaking (production) and listening (perception) as an auditory feedback loop, in which speakers use their own output—their own mental model of a sound—as input for their production. The role of auditory feedback, attested in research on speakers with normal hearing and profound hearing loss (Perkell, Matthies, Lane, Guenther, Wilhelms-Tricarico, Wozniak, & Guiod, 1997; Perkell, Guenther, Lane, Matthies, Perrier, Vick, Wilhelms-Tricarico, & Zandipour, 2000) also

helps account for the persistence of both spoken error and a lack of aural discrimination between the target sound and the error.

The question then arises of how to break this feedback loop; naturally enough, it can be broken when a speaker actually forms the target sound in his or her mouth and then combines this new and different motor-memory with a new acoustic image of the sound. In practical terms, we tell learners that when they finally "get it right"—produce a target sound—it will "feel wrong" to them, but they should nonetheless perform a mental freeze-frame at that moment.

For our purposes, it doesn't matter at all how to get speakers to produce the sound: teachers may use any means at their disposal, including aural models, mouth diagrams and mirrors, metalinguistic feedback, and pronunciation aids, such as rubber bands, Chinese yo-yos, and kazoos (Gilbert, 2005; Grant, 2004; Hewings, 2004). We believe that aural models will be less helpful, by and large, than other means, but teachers should on no account rule out any particular way of getting learners to produce the target sound (Couper, 2006; Gilbert, 2010). We agree with Flege (1993, 1995) that, when applicable, the contrasting nature of L1 and L2 features may help learners in their acquisition, but we argue that teachers must simply be more creative and proactive to help learners acquire pronunciation features that may not be as saliently different from their L1 analogs.

Often speakers *can,* in fact, produce the target sound (Fraser, 1999, 2000), but in limited phonological environments. Word-initial /w/ in "want or "what" is not problematic for Japanese speakers; however, since /ʊ/ does not occur in Japanese, word-initial /w/ in "would" or "woman" is difficult. By helping the speaker to isolate the target sound in a licensed consonant-vowel combination, we can start to help the speaker to form a new mental model in expanded CV syllable contexts.

Once teachers can get a learner to produce the target sound, in any context, and on the road to creating a new mental model of it, both the learner's speaking and listening are expected improve. We believe that in a model of convergent production, production precedes and facilitates perception, promoting more intelligible spontaneous speech and enhanced listening comprehension as the speaker's production begins to converge on the target: as we tell learners, speaking helps listening. We posit that our production-centered approach to intelligibility is actually consistent with Flege's Speech-Learning Model (1993, 1995). Flege's SLM focuses on the mental and metacognitive work of perceiving a new target sound; our model does not contradict the SLM, but instead offers a specific and practical pedagogy to allow learners to create acoustic images of new sounds.

AN INTEGRATED MODEL

With what we see as this key relationship between speaking and listening in mind, therefore, we argue for an expanded model of pronunciation which includes both speaking and listening and addresses both suprasegmantals and segmentals. Figure 1 represents our Integrated Model of Pronunciation:

Pronunciation

Listening   Speaking

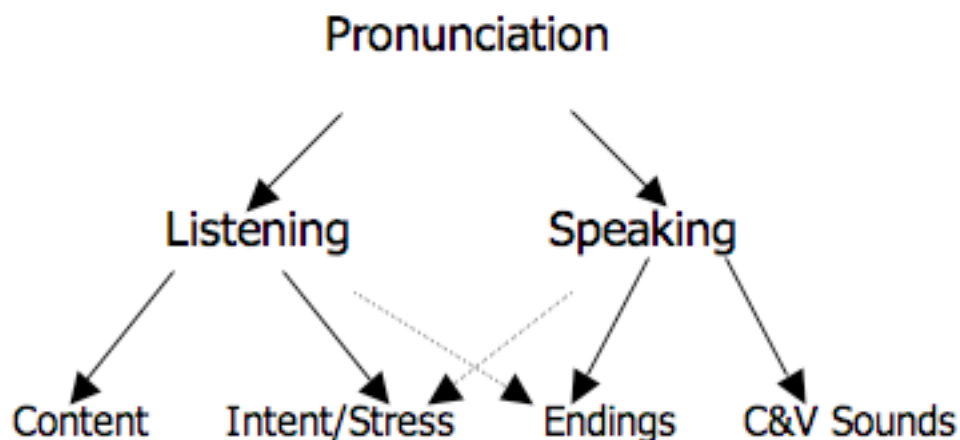Content   Intent/Stress   Endings   C&V Sounds

Figure 1. Integrated model of pronunciation

We identify four major components of pronunciation. Working from left to right in the figure above, we identify the first strand as listening for *content,* because learners who struggle with this strand of pronunciation tend to miss or misunderstand the actual content of what is being said. This strand contains the connected speech features—linked, reduced, deleted, altered, and contracted sounds. Though learners often think that the speed of English accounts for their listening problems, in fact these connected speech features are contributing sources. One of these features, reductions, has been frequently cited as a major contributing factor to misinterpretation of the homophones "to" and "two" in civil aviation (Waldock, 1994), resulting in a 1989 crash when an air traffic controller cleared an aircraft to descend to "two four zero zero" but the pilot interpreted the number "two" as the preposition "to" and descended accordingly (Cushing, 1995a, 1995b).

Next, we identify the strand of *intent*—the suprasegmental features of English (syllables, stress, intonation, rhythm, and timing) that convey speaker meaning. Learners who struggle with this aspect of pronunciation may understand every word a speaker says but may still miss the meaning; conversely, they may be unable to get their own content or intent across. This is why we show a dotted arrow coming down to this strand from speaking, as well; we argue that suprasegmentals are essential both for speaking and listening (Hahn, 2004). In a study currently being undertaken, advanced-level learners [Level D in an A-E level intensive English program, consisting of 14 students representing five native language groups: Arabic ($n = 4$), Korean ($n = 4$), Mandarin Chinese ($n = 3$), Spanish ($n = 2$), and French ($n = 1$)] were questioned prior to suprasegmental instruction on their opinions regarding the role of one component of suprasegmentals: intonation. Using a digital response system (clickers), students voted for their choice of response to the following question:

> Which is more important in a sentence?
>
> > a. the words in the sentence
> >
> > b. the intonation in the sentence

For this question, 73% (*n* =11) of the respondents voted for choice a: words, while 27% (*n* =3) of the respondents said that intonation was more important. Following completion of the first lesson on intonation and contrastive stress in their pronunciation text (Grant, 2009), students were asked their interpretation of a sentence with non-standard sentence-level stress. Using the digital response system, learners responded to a yes or no question after hearing an audio prompt:

> Audio Prompt: *The teacher* didn't grade your papers.

> Question: Were the papers graded?

> > Yes

> > No

The votes split evenly (*n* = 7 'yes' votes, 7 'no' votes), may reflect some effects of instruction and generating much discussion. When the correct response, (a), was displayed, a number of learners asked to hear the sentence again. Several could be seen and heard repeating the sentence under their breath, particularly the verb phrase "didn't grade." These respondents demonstrated grasp of the lexical, morphological, and syntactic content of the utterance, but missed the intended meaning, suggesting that learners are unaware of the importance of—in this particular case—contrastive stress and emphatic intonation, and their role in altering the meaning of an utterance. Attention to this suprasegmental strand is therefore essential for learner success in listening as well as in speaking

Returning to Figure 1, the third strand from the left is *endings*—in other words, the sounds of regular verb and noun inflectional morphology, which cause learners serious and intractable problems when they omit or mispronounce these endings. The fact that learners fail to pronounce these endings on regular nouns and verbs when reading aloud convinces us that the problem is not always a local grammar error, but is based in their internalization of the sound system. Teachers, based on learners' ability to state the grammar rules and produce these endings in controlled drills and tests, appear to succumb to what Gass and Varonis (1985) refer to as accommodation, or not noticing the errors in learners' spontaneous speech. Whether teachers do in fact unconsciously accommodate—i.e., no longer notice learners' errors in this area—or instead make a conscious decision to avoid addressing a seemingly intractable problem, these errors are nevertheless serious ones. Studies have shown that seemingly small errors with these endings nevertheless stigmatize learners and are responsible for communication breakdowns (Jiang, 2007; Lardiere, 1998; Long, 2003; Major, 1995, 1998). The misinterpretation of "He looked it up" as "He looked up" by a speaker who pronounces "look" plus past tense as two syllables attests to the role of the acoustic image and the interrelation of production and perception. For these reasons, we believe that this strand, too, is key for both learners' speaking and listening skills, as represented in Figure 1.

Finally, our fourth strand of pronunciation is that of individual *consonant and vowel sounds* (segmentals) which learners typically report to be their biggest pronunciation errors and main source of communication breakdowns (Derwing & Rossiter, 2002). Our experience in a pronunciation elective class with a Japanese neuroscientist who did not differentiate between liquids /l/ and /r/ or fricative /ʒ/ and affricate /dʒ/, rendering "brain regions" indistinguishable

from "brain lesions," attests to the need to address segmentals. Nevertheless, we weight this strand as relatively less important for intelligibility than the two strands toward the middle of the figure. Thus our model advocates a combined focus on segmental and suprasegmental features and an integrated approach to instruction.

THE MODEL IN CLASSROOM PRACTICE

We believe that this integrated model of pronunciation is particularly helpful to learners for making sense of their errors: teachers are able to break the large topic of pronunciation down into meaningful, yet discrete chunks, and then build it back up as learners master various patterns and sounds. Teacher tools that can help learners have success with these different parts of pronunciation abound: we suggest a teacher-learner partnership, where teachers guide learners through both the production of these different pronunciation topics and the metacognition necessary for "noticing" these patterns (Schmidt, 1990) and building up new acoustic images of the patterns (Neufeld, 1977).

Three sample entries in a pronunciation logbook, along with directions written for learners (Figure 2), show one way that metalinguistic feedback can help move a learner toward more target-like production—in this case, of the problem of initial consonant clusters, a syllable structure problem we classify in the second strand (suprasegmentals, or intent/rhythm) above:

Use a log to help you enter mistakes, label them by kind, and review a list of *your* biggest problems. If you keep making the same kinds of errors, use your log to help you correct them.

Everyone has different pronunciation problems. Look at the examples on the chart below to see the different difficulties three students had with the same word. To improve, students need to know what their individual mistakes are.

Word:	How should I say it?	How did I say it?	What was my mistake?	Other examples:
Student 1: speech	speech (1 syllable)	su-peech (2 syllables)	separating the first two consonants	su-trong/strong
Student 2: speech	speech (1 syllable)	es-peech (2 syllables)	adding a vowel sound at the front	es-port/sport
Student 3: speech	speech (1 syllable)	speech-ee (2 syllables)	adding a vowel sound at the end	each-ee/each

Figure 2. Sample pronunciation logbook

In Figure 3, we offer a checklist for learners who "know" the rule for the pronunciation of the regular past-tense verb ending, but nevertheless omit it entirely or incorrectly pronounce it as two syllables in verbs like *looked* and *used*. These errors, of course, relate to the third strand (verb and noun endings, or morphosyntax) in our integrated model of pronunciation:

Checklist:
How do you say the "-ed" ending on regular past tense verbs?

Look: Find the simple (root) form of a verb, without any endings.
Ask: What is the final sound (not letter)?
Is it: /t/ or /d/?

　　If yes . . . ✔
　　　　Add an extra syllable.
　　　　　　⇒ Say "-ed" as [Id].
　　If no . . .
　　　　Ask: Is the final sound unvoiced?
　　　　　　If yes . . . ✔
　　　　　　　　There is no extra syllable.
　　　　　　　　⇒ Say "-ed" as [t].
　　　　　　If no . . .
　　　　　　　　There is no extra syllable.
　　　　　　　　⇒ Say "-ed" as [d].

> *So, when saying the past tense ending:*
>
> 1. **Voiced** sounds use **voiced** endings, **[d]**.
>
> 2. **Unvoiced** sounds use **unvoiced** endings, **[t]**.
>
> 3. Sounds **/t/ or /d/** use an **extra syllable, [Id]**.

Figure 3. Checklist for pronouncing regular past tense endings

Teacher intervention is key to helping learners move from isolated production of a new pronunciation sound or pattern to full production of it in spontaneous speech, and in turn to full perception of it in other speakers' connected speech. Such intervention can be seen as the real work of the pronunciation classroom—it is what teachers are there to do, after all, and it is why we don't advise learners just to improve their pronunciation by merely reading a pronunciation book (even a good one). The model describes the intermediate stage of spoken proficiency when learners are able, sometimes, to produce a target sound or pattern on demand, but have not yet integrated it fully into their spontaneous speech as the stage of prompted production: through instruction in the sound or pattern, and careful prompting (again, the actual means of getting the learner to produce this sound may vary widely), learners will achieve their first success at producing the sound in a target-like fashion. Learners then move through a stage when teachers' corrective feedback is necessary to help them solidify the pattern in their own mouths and minds, before they fully form a new mental model of the sound or pattern. To better help learners, we suggest that the language of instruction match the language of correction; in other words, whatever metalinguistic feedback teachers wish to offer ("Say it as one syllable," "Don't use a voiced sound there," "Make the stressed syllable louder, longer, clearer, and higher," etc.) in the classroom, to individual learners, should be the same terms as those in which they teach the concept or pattern to begin with. That language should also be the same that teachers elicit from learners in the form of learner "tell-backs." Here we borrow the term from the literature on reading instruction (Vanderwood, 2007), where tell-backs are defined as classroom language that articulates, superficially for the teacher's benefit, but more significantly for the learner's own benefit, what the learner now knows about this concept (e.g., "I said it as two syllables, but I need to only say one; I need to omit the final syllable after the sibilant.") and what will help the

learner go about forming a new mental model. Over time, as learners demonstrate internalization and metacognitive awareness, external corrective feedback can take the form of increasingly precise, minimal, and unobtrusive 'production prompts' which allow the learner to self-correct and return to finish their thought. Figure 4 shows the path to intelligibility, and the connections among the different terms we have been using in this model:
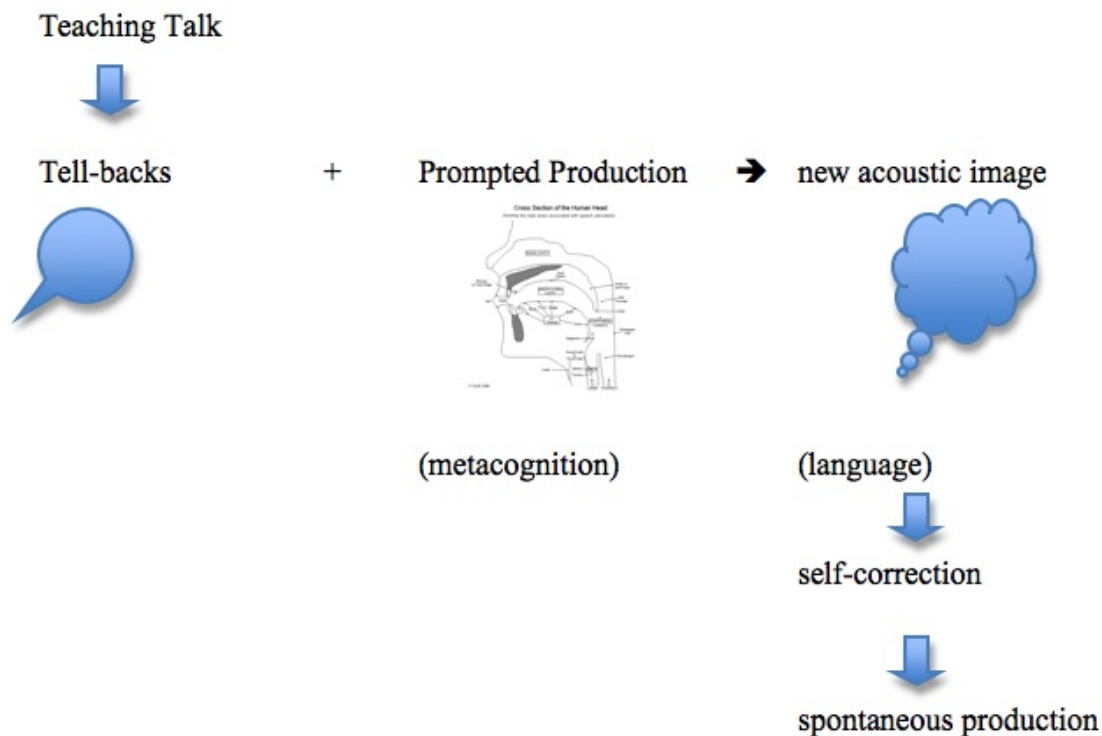

Figure 4. The path to intelligibility

In our model, pronunciation instruction accompanies and reinforces core language instruction. While Flege's SLM (1993, 1995) focuses primarily on segmentals, our model considers the acquisition of all aspects of pronunciation in a more integrated way; Flege's emphasis on the role of time, the nature of instruction, and the learner's internal representations of both L1 and L2 are also key to our understanding of the pronunciation acquisition process. Ultimately we view integrated pronunciation instruction as a highly focused, metacognitive approach to the entire language classroom, and a key step along the way to learners' increasing intelligibility.

ABOUT THE AUTHORS

Marnie Reed is an Associate Professor in the graduate TESOL Program and the Program in Applied Linguistics at Boston University. The focus of her research has been on second language acquisition, with a focus on the relationship between speech production and perception. She has

co-authored the books *Sound Concepts: An Integrated Pronunciation Course* (2005, McGraw-Hill) and its accompanying *Teacher's Manual* (2006), and *Goal-Driven Lesson Planning for Teaching English to Speakers of Other Languages* (2010, University of Michigan Press), which incorporates her work on pronunciation instruction.

Christina Michaud is a Lecturer in the College of Arts and Sciences Writing Program at Boston University and also a doctoral candidate in Language, Literacy, and Cultural Studies in the School of Education at BU. She teaches ESL composition as well as general first-year composition and research writing. She has co-authored two books and numerous conference presentations with Marnie Reed.

REFERENCES

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*(1), 46-66.

Cushing, S. (1995a). *Fatal words: Communication clashes and aircraft crashes*. Chicago, IL: University of Chicago Press.

Cushing, S. (1995b). Pilot-Air Traffic Control communications: It's not (only) what you say, it's how you say it. *Flight Safety Digest, 14*(7), 47-53

Derwing, T. M. & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30,* 155-166.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*, 1589-1608.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. & Hillenbrand, J. (1984). Limits on pronunciation accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America, 76*, 708-721.

Fraser, H. (1999). ESL pronunciation teaching: Could it be more effective? *Australian Language Matters, 7*(4), 7-8.

Fraser, H. (2000). Coordinating improvements in pronunciation teaching for adult learners of English as a second language. (Department of Education, Training and Youth Affairs, Australian National Training Authority ANTA Innovative Project) Canberra: Commonwealth of Australia.

Gass, S. M. & Varonis, E.M. (1985). Task variation and nonnative/nonnative negotiation of meaning. In S.M. Gass & C.G. Madden (Eds.), *Input in second language acquisition* (pp. 149-161). Rowley, MA: Newbury House Publishers.

Gilbert, J. B. (2005). *Clear speech: Pronunciation and listening comprehension in North American English. Teachers' resource book*. NY: Cambridge University Press.

Gilbert, J. B. (2010). Pronunciation as orphan: What can be done? *As We Speak: Newsletter of TESOL speaking, pronunciation, listening interest section, 7*(2).

Grant, L (2004). Demonstration presented at the 38th Annual TESOL Convention, Pre-Convention Institute: Essentials of Pronunciation, Long Beach, CA.

Grant, L. (2009). *Well said: Pronunciation for clear communication*. (3rd ed.). Boston, MA: Heinle & Heinle Publishers.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*, 201-223.

Hewings, M. (2004). *Pronunciation practice activities: A resource book for teaching pronunciatio*n. NY: Cambridge University Press.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*(1), 1-33.

Lardier, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14*(4), 359-379.

Long, M. (2003). Stabilization and fossilization in interlanguage development. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 4877-535). Malden, MA: Blackwell.

Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, *82*, 357-71.

Major, R. (1995). Native and nonnative phonological representations. *IRAL, 33*(2), 109-127.

Major, R. (1998). Interlanguage phonetics and phonology: An introduction. *Studies in Second Language Acquisition*. *20*(2), 131-137.

Neufeld, G. (1977). Language learning ability in adults: A study on the acquisition of prosodic and articulatory features. *Working Papers in Bilingualism*, *12*, 45-60.

Pavesi, M. (1986). Markedness, discoursal modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition*, *8*, 38-55.

Perkell, J. S., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication, 22*(2), 227-250.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. & Zandipour, M. (2000). A theory of speech motor control and supporting

data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28*(3), 233-272.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-158.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed), *Speech Perception and Linguistic Experience: Issues in cross-language research* (pp. 3-45). Timonium, MD: York Press.

Vanderwood, M. L., & Nam, J. (2007). Response to intervention for English Language Learners: Current developments and future directions. In S. R. Jimmerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *The Handbook of Response to Intervention: The Science and Practice of Assessment and Intervention* (pp. 408-417). NY: Springer.

Waldock, B. (1994, February 15). Major miscommunication mishaps. *Aviation Safety*, 3-5.