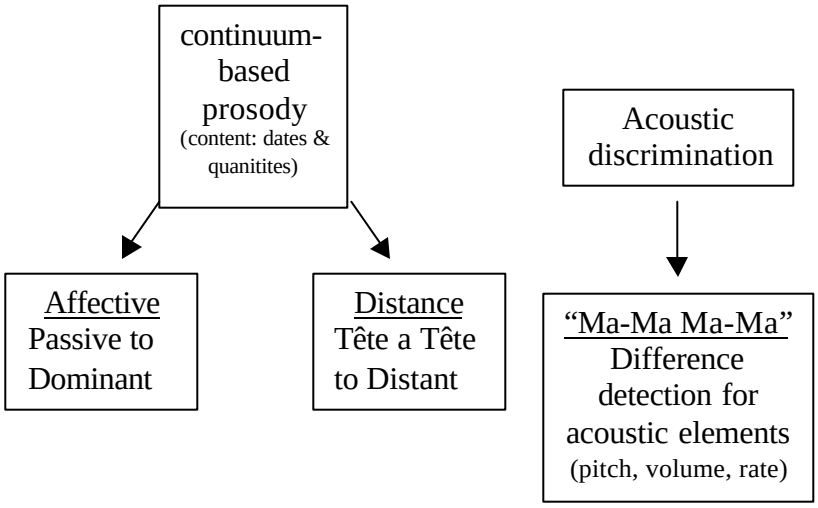
Prosody Comprehension Measures

The comprehension of prosody will be evaluated by using the Penn Continuum Evaluation of Prosody (PCEP-4, Figure 2.1). The PCEP-4 was developed for the proposed project. All stimuli for the PCEP-4 consist of 4 syllable utterances using various manipulations of prosody. The content includes dates, quantities, and single syllables repeated four times (e.g., "November fourth", "One thousand one", "ma-ma ma-ma"). The content of dates and quantities was selected to minimize semantic loading, as there are few semantic attributes and categorical kinds that are associated with content of this nature. Details regarding the development, architecture, normative inter-rater judgement data, and procedures for administering the PCEP-4 are described in subsequent sections.

Figure 2.1: Overview of the components of the Penn Continuum Evaluation of Prosody–4.



Continuum-based prosody. The continuum-based prosody stimuli include two conditions: affective and non-affective. The affective condition requires participants to listen to each stimulus, and make a continuum-based judgment of the emotion the speaker expressed. The affective continuum is phylogenically-based, and ranges from passive to dominant (Figure 2.3). The non-affective continuum uses interloculor distance to manipulate prosodic parameters in a non-emotional manner. This variant of prosody manipulates the distance between the speaker and the intended listener, ranging from a personal exchange at close proximity (referred to as "tête a tête") to a substantial distance between speakers (referred to as "distant"; Figure 2.2).

Participants are introduced to the distance continuum, and then asked to make judgments on a 5-point Likert-type scale for each stimulus of this type. The scale ranges from 1- tête a tête to 3- conversational distance to 5- distant.

Figure 2.2: The phylogenically-based continuum of non-affective prosody used in the PCEP-4.

1	2	3	4	5
Tête á tête Under breath		Conversational Distance		Distant In presence of loud ambient noise

Similarly, participants are asked to make judgments on a 5-point Likert-type scale for each stimulus that is affectively intoned. The scale ranges from 1- passive to 3- emotionally neutral to 5- dominant. Presentation of the affective and non-affective conditions of the PCEP-4 will be counterbalanced between participants.

Figure 2.3: The phylogenically-based continuum of affective prosody used in the PCEP-4.

1	2	3	4	5
Tentative Cautious Uncertain Passive		Affectively Neutral		Authoritarian Confident Forthright Dominant

Perception of Acoustic Elements. The perception of acoustic elements screening was devised as a background measure to identify the extent to which perception of discrete acoustic elements in naturalistic speech contributes to comprehension of prosody. The same professional actors that produced the continuum stimuli repeated the syllable "ma" 4 times, using regular cadence, and consistent pitch and volume. An exemplar from each actor was manipulated to create template utterances with consistent fundamental frequency, duration of syllables, and amplitude, while still maintaining natural sounding speech. The final syllables of these templates were manipulated to create different intensity levels of the three fundamental acoustic elements. Only one acoustic element from the neutral template will be manipulated for each token (see summary in Table 2.2).

Table 2.2: Summary of type of acoustic manipulations for the acoustic perception screening. The factor of manipulation varied from actor to actor. For actual acoustic parameters for each actor, refer to Appendix D.

Acoustic Element Manipulated (all other acoustic elements held constant)	Intensity and Direction of Manipulation				
Amplitude	Lower	Low	Normalized to fundamental frequency, Duration of utterance, and	High	Higher

Fundamental Frequency	Lower	Low		High	Higher
Duration	Longer	Long		Short	Shorter

Acoustic manipulations of these recordings produced 48 utterances with various acoustic manipulations of the final syllable. Twelve utterances will be presented in blocks by acoustic element with an equal number of copies of the normalized template utterance. Participants will be asked to identify if the pitch, volume, or rate of last syllable sounds different than the rest of the utterance using a 5 point-Likert scale (Figure 2.4). This scaling will afford the comparison of the acoustic screening with other PCEP-4 conditions.

Figure 2.4: The difference judgement continuum of acoustic perception used in the PCEP-4. Participants will be asked to: "Rate how different the volume is on the last syllable compared to the rest of the phrase." Similar instructions will orient the participant to changes in duration and pitch.

1	2	3	4	5
Much less Intense	Slightly Less Intense	Similar	Slightly More Intense	Much more Intense

Development of the PCEP-4

Four young adult actors (2 males, 2 females) were trained using quazi-standardized procedures to elicit various manipulations of prosody (see Actors' Manual, Appendix B). Briefly, actors were given a written manual that included an overview of the project, described the phylogenically-valid affective and non-affective continua, and written samples of the 4-syllable scripts (i.e., dates and quantities). Actors were asked to state dates and quantities using exemplars from the phylogenically-valid affective and non-affective continua proposed in the preceding literature review (Figures 2.2 and 2.3) and detailed in a written manual. These actors were instructed to produce extremes of these exemplars ("1" and "5" on both the affective and non-affective continua) as well as affectively neutral utterances spoken at a conversational interlocutor distance (i.e., distance and emotion neutral, or "3" on the affective and non-affective continuum, Figures 2.2 and 2.3). The experimenter also provided one-on-one training to actors, directing them

to each mental/emotional set, and providing the 4-syllable scripts. Actors were also coached to produce several exemplars of the baseline acoustic measure (i.e., the "ma-ma ma-ma" series). Each stimulus is less than 1.5s in duration, and the corpus is appropriate for both behavioral and *f*MRI paradigms.

Recordings were made in a sound recording studio with the direction from the experimenter. Recordings were made on two channels. Two microphones were used: 1) a stand-mounted boom Shure SM94, and 2) a headset Sennheiser HMD 410. The utterances of actors were recorded directly into WAVES+ datafiles at a sampling rate of 16K on aUnix-based machine via a DAT-link. Simultaneous recordings were also made onto DAT tapes (a Sony PCM-R300 deck at 44.1 KHZ) through a Mackie 1202-VLZ mixer. The recording sessions were segmented into individual 4-syllable tokens of no more than 1.5 seconds in duration (i.e., appropriate for use in *f*MRI paradigms). Single channel clips of these tokens were then demultiplexed. A routine that maximizes gain was run to normalize the tokens with respect to amplitude. The nature of the 4-syllable utterances used in this corpus are rich in fricatives (e.g, dates). Because most of the energy in the pronunciation of fricatives is carried at high frequencies, this caused some distortion of the original tokens. However, the sampling rate was dense enough to create tokens where the content and attitude of the speaker are easily understandable. These gain-maximized, single-channel demultiplexed tokens were converted into 11.025 kHz .wav files. These .wav files were converted to SoundEdit16 proprietary format so that they could interface with the PsyScope software (Cohen, 1993), which controls the presentation of the PCEP to participants.

The acoustic discrimination stimuli were further manipulated using the following techniques. Duration of the syllables was measured and only those exemplars with consistent duration were selected for template tokens. A template token for each actor was created by adjusting the amplitude to approximately half of the maximum gain of each actor. These templates were further equalized for fundamental frequency between the syllables. The pitch contour within each syllable was maintained to make the template sound more like naturalistic speech. Details on the manipulations of pitch, duration, and volume for the "ma-ma" stimuli can be found in Appendix D.

Normative data on the PCEP-4. The PCEP-4 includes 2 classes of prosodic manipulations: continuum-based prosody, and acoustic discrimination (Figure 2.1). Several hundred stimuli comprise the entire PCEP-4 corpus. A pilot study was conducted to shorten the continuum-based portion of the corpus, and to maximize the internal validity of this investigation. The goal of the pilot study was to select the best exemplars of the affective and distance continuum from the existing corpus, and to evaluate the psychometric properties of the PCEP-4 (including inter-judge agreement). This was achieved by asking healthy young participants to make affective and distance continuum judgments for the corpus of stimuli. Twenty healthy young adult participants judged 4-syllable exemplars using the same Likert-type continua as the actors used to render the tokens. These ratings were summarized and ranked by standard deviation, and the tokens that received the most consistent ratings were chosen to be included in the corpus. Consistency ratings were operationally defined as having a low standard deviation of ratings across healthy young participants. Yet each exemplar with a high consistency rating also needed to have a mean rating close to the intended continuum intensity. For example, if an actor was asked to produce an exemplar of a "5" on the affective continuum, and all healthy normal subjects also rated this stimulus as a "5", the item received a standard deviation of zero, and was included in the final corpus of stimuli. The tokens that were in the top fifth of consistency ratings were chosen to be included in the final version of the PCEP-4. No exemplar had a rating standard deviation of more than 0.60. In the case of "ties" (i.e., there were more than exemplars at a given consistency rating than needed for a particular stimulus type), the experimenter judged the better exemplars by using criteria such as subjective ratings of representativeness of the continuum, and sound quality. A summary of these pilot study data can be found in Appendix E.

Normative data has been gathered on the acoustic perception screening task of the PCEP-4 (i.e., the "ma-ma" stimuli) to ensure its psychometric properties are acceptable for use as a background task. Fifteen healthy young individuals were asked to complete the task. Two variables were used to assess accuracy: difference detection and absolute summed differences. Difference detection accuracy was determined by summing the items that were correctly judged as different on the last syllable with those correctly judged as being the same on the last syllable and dividing this sum by the total number of items. Absolute summed differences were determined by subtracting the observed from expected values for each item, then summing those absolute differences. The items in the task were manipulated until comparable accuracy scores were

obtained from at least 5 healthy young adults, and results were not at ceiling. Results are summarized in Table 2.3.

Table 2.3: Summary scores for initial normative data on the PCEP-4 acoustic perception task.

	Duration n=15		Pitch n=5		Volume n=10	
	difference accuracy	sum of absolute (exp-obs) differences	difference accuracy	sum of absolute (exp-obs) differences	difference accuracy	sum of absolute (exp-obs) differences
mean	88.33	11.13	85	11	85.94	11
StDev	9.41	5.19	10.94	5.56	6.79	4.22
min	71.878	3	71.88	4.85	71.88	6
max	100	24	96.88	19	93.75	20

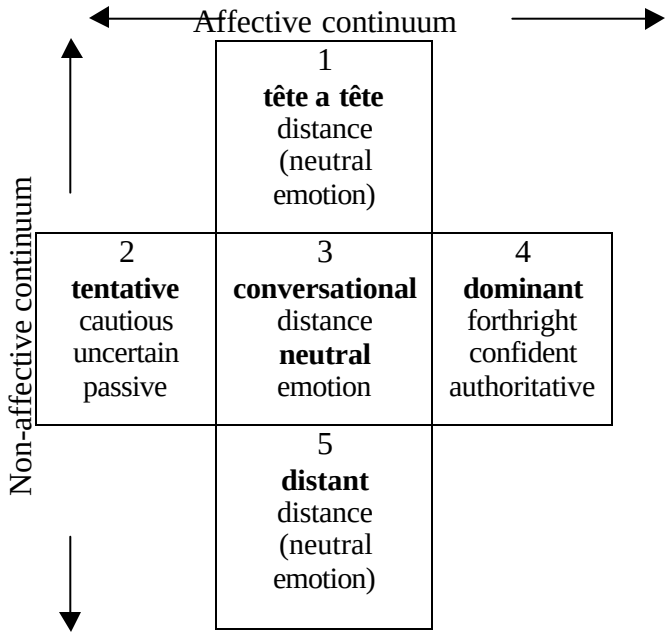
Architecture and Procedure of PCEP-4. The PCEP-4 will be presented on a G3 series PowerBook computer (Apple Computers, Cupertino, CA) using PsyScope software (Cohen et al., 1993) and professional quality audio headphones (Optimus Pro 50 MX). Presentation of the affective and non-affective condition will be counterbalanced, and an interactive training will occur prior to each condition to introduce participants to each continuum. During the interactive training session, participants will hear 30 of the exemplars, (5 each of intensity 1, 3, and 5 for each condition (i.e., affective and non-affective) with the best consistency ratings from the pilot study). Following the training, the experiment program will present 90 unique auditory stimuli at 4500ms intervals. Participants will respond to either the non-affective or the affective continuum judgments using an array of buttons representing a Likert-type scale (1 through 5) for each stimulus. A customized foam-core keyboard overlay will be used to simplify the visual response choices for participants. This overlay is sturdy enough for participants to rest their hands close to the response buttons, thus decreasing the latency to respond and minimizing the motor components of the task. Affective and non-affective conditions will be counterbalanced, and an interactive training will occur prior to each condition.

For the behavioral judgments, stimuli will be presented in a random order within each non-affective and affective condition. In other words, a block of 30 exemplars from affective continuum value 1 and 30 exemplars from affective continuum value 5 will be presented in

randomized fashion with 30 affectively neutral stimuli (90 items total). Participants will be asked to respond to the prompt: "Is the emotion expressed more tentative, neutral, or more dominant?" by selecting a response from a 5-point Likert-type scale. Similarly, 30 exemplars of non-affective continuum value 1 and 30 exemplars non-affective continuum value 5 will be will be presented in randomized fashion with 30 distance-neutral stimuli. For this portion, participants will be asked to respond to the prompt: "Is the speaker very close, neutral or far away from the intended listener?" by selecting a response from a 5-point Likert-type scale. Stimuli are presented every 4500 milliseconds, and participants are instructed to respond as soon as they can make an accurate judgement (even if they have made a decision before the item is finished playing, see instructions in Appendix A).

The design and stimulus presentation and dependent variables are summarized in Tables 2.4 and 2.5.

Table 2.4: Affective and non-affective stimuli types presented in the PCEP-4. (Numbers refer to stimuli codes used in Table 2.4.)



Acoustic Discrimination. The acoustic discrimination screening was devised as a baseline measure to identify the extent to which perception of pitch, rate and volume of speech contributes to

comprehension of prosody. The same professional actors that produced the continuum stimuli repeated the syllable "ma" four times, using regular cadence, volume and consistent pitch. Acoustic manipulations of these recordings produced 48 utterances with various acoustic manipulations of the final syllable. Twelve utterances will be presented in blocks by acoustic element with an equal number of copies of the normalized template utterance. Participants will be asked to identify if the pitch, volume, or rate of last syllable sounds different than the rest of the utterance using a 5 point-Likert scale (Figure 2.4).

Participants will listen to each stimulus, and judge how different the final syllable is from the rest of the utterance using a 5-point Likert-scale (see Appendix A for administration instructions). Similar to the other PCEP-4 measures, the acoustic discrimination task will have an interactive training session and be presented using PsyScope software on a PowerBook G3 series computer. Participants will be instructed to respond by using the number keys 1 through 5 with the customized computer keyboard overlay.

Table 2.5: Outline of stimuli presentation for the behavioral version of the PCEP-4. Behavioral responses will be recorded on the keyboard using keys 1, 2, 3, 4, 5. See legend in Table 2.3 for further description of stimulus types.

	Stimuli presentation	Cognitive decision	Behavioral response	Dependent Measures
Prosody judgments	1. Tête a tête 3. Neutral 5. Distant *n=90, random presentation*	Is the speaker very close, neutral or far away from the intended listener?	1= very near 2= close 3= conversation 4= across room 5= far	Accuracy Sum of absolute difference scores (difference score= observed minus expected) Latency to respond
	2. Tentative 3. Neutral 4. Dominant *n=90, random presentation*	Is the emotion expressed more tentative, neutral, or more dominant?	1= very passive 2= passive 3= neutral 4= dominant 5= very dominant	
Acoustic discrimination	4 intensity levels of fundamental frequency, duration, amplitude	Rate how different the pitch is on the last syllable compared to the rest of the phrase. Rate how different the duration is on	1= much less intense 2= slightly less intense 3= similar 4=slightly more	Accuracy Sum of absolute difference scores (difference score= observed minus expected)

	from 4 actors; and neutral tokens *n=32, blocked presentation*	the last syllable compared to the rest of the phrase. Rate how different the volume is on the last syllable compared to the rest of the phrase	intense 5=much more intense	Latency to respond