

MULTIDIMENSIONAL PERCEPTUAL UNFOLDING OF SPATIALLY PROCESSED SPEECH I: DERIVING STIMULUS SPACE USING INDSCAL

WILLIAM L MARTENS¹ AND NICK ZACHAROV^{1,2}

¹Multimedia Systems Laboratory, University of Aizu, Aizu, Japan.

wlm@u-aizu.ac.jp

²Nokia Research Center, Speech and Audio Systems Laboratory, Finland.

Nick.Zacharov@Nokia.com

Multidimensional perceptual unfolding was executed for a set of spatially-processed speech samples reproduced via headphones. This paper describes only the first stage of a two-part study employing the analytic technique termed external unfolding. In this first stage, global dissimilarity ratings were made for all pairwise comparisons of the experimental stimuli under each of four listening conditions, which included broadband or bandlimited speech samples presented either simultaneously or sequentially. These four datasets were analyzed independently using INDSCAL (INDividual Differences SCALing), a method for processing inter-stimulus dissimilarity data that has specific advantages over classical MultiDimensional Scaling (MDS) analysis. First, it is designed to characterize quantitatively the individual differences in responses obtained from a group of experimental subjects. Second, the spatial configuration of points derived for the experimental stimuli, termed the Stimulus Space, has an inherently unique orientation that has none of the ambiguity that makes the interpretation of classical MDS results problematic. The Stimulus Space derived in the first stage of this study is one of two inputs required for external unfolding. It is combined with discrete attribute ratings collected in the second stage of this study to reveal the principal perceptual attributes of the set of spatially-processed speech samples, and their relative salience under various conditions.

1. INTRODUCTION

Scientific exploration of spatial sound reproduction requires carefully constructed experiments for there to be progress in understanding and applying spatial sound processing technology. Because the primary goal of spatial sound reproduction is the creation and manipulation of auditory spatial imagery for the human listener, there is need for reliable techniques for evaluating the perceptual results of spatial sound processing under a variety of listening conditions. Psychophysical research has played an important role in the development of virtual acoustic rendering technology, addressing questions of virtual source spatial localization and auditory image quality [1]. But many psychophysical questions remain unanswered, perhaps due to the complexity of this interdisciplinary field, requiring expertise in acoustics, digital signal processing, computer-based acoustical modeling, and last but not

least, perceptual testing methodology.

The production of naturalistic auditory spatial imagery requires spatial sound processing that includes the simulation of discrete reflections with specified delay and direction [2]. Changes in the spatiotemporal distribution of these reflections produce changes in auditory spatial imagery on many perceptual dimensions, including apparent source width (ASW), apparent source distance (ASD), envelopment, spaciousness, etc. [3]. How to predict these changes is a significant problem that has only been addressed for variations under constrained stimulus conditions, such as selected musical sources presented via binaural reproductions of selected concert halls (for a representative treatment, see [4]). This paper reports an investigation that is no more broad than its predecessors, focusing as it does upon perceptual responses for a small selection of speech sources under a highly-constrained selection of room acoustic simulations. Though the results themselves should be of interest to those engaged in research and development of spatial sound processing technology (especially those interested in binaural telecommunication applications), another goal of the paper is to present a refined methodology for psychoacoustic investigation that has been termed multidimensional perceptual unfolding. In contrast to *internal unfolding* [5], the *external unfolding* employed in the current study requires the collection, for a single set of stimuli, of two types of subjective response data from a group of listeners: direct attribute ratings for each stimulus in a set of experimental stimuli, and global dissimilarity ratings for all pairwise comparisons of the stimuli. Only the dissimilarity ratings are analyzed in the first stage of this study, and the second stage that combines these results with the analysis of the discrete attribute ratings will be reported in a subsequent paper [6].

The use of external unfolding analysis in multidimensional perceptual studies is not new but has not been employed frequently in psychoacoustic investigation, despite the ready availability of appropriate software such as Ramsay's MULTISCALE [7]. He taught the advantages of joint analysis of dissimilarity and other types of judgments, such as pairwise preferences or direct attribute ratings of the stimuli with regard to defined properties:

There is usually some reason for supposing that the processes which give rise to the various types of judgments ... share features in common. In fact, these shared features may be exactly what is being investigated, so that the experimenter is interested in how the [perceptions and] cognitions ... of the stimuli give rise to a particular subject's evaluations of these same stimuli ([7], p. 149).

One goal of collecting direct ratings is to determine whether certain perceptual attributes are correlated with subjective preferences. Dissimilarity judgments are often included in such investigations in order to indicate the involvement of stimulus parameters for which direct ratings were not collected. Conversely, the dissimilarity-based structures can reveal which stimulus parameters do not enter into the subject's global evaluative reactions. In the broad research scheme within which the current study forms only the first stage, direct ratings are to be related to such dissimilarity-based structures via multidimensional perceptual unfolding, which benefits from having a good estimate of the Stimulus Space into which direct ratings may be "unfolded." The goal in this first stage is simply to derive, from judgments of the dissimilarity of a set of stimuli, a single representation of their multidimensional perceptual complexity.

2. GENERAL METHODS

2.1. Regarding psychoacoustic investigation

In scientific explorations of spatial sound reproduction, several varieties of psychoacoustic investigation can be identified: There are studies of physical acoustics in which attempts are made to generate

psychoacoustically meaningful objective measures of the proximal sound stimuli that are received by listeners. Though the goal of these attempts is to provide measures that might be useful in predicting human perception, the studies themselves are strictly focused upon identifying important features of the physical signals, the proximal stimuli that can be observed perhaps at the eardrums of the listener. Then there are studies of a psychological nature, in which the perceptual responses to spatial sound stimuli *per se* are the focus of the investigation. The scientific goals of these attempts include characterizing the features of auditory spatial imagery upon which human listeners can reliably report, and providing measures of the perceptual distinctions that can be made for a given range of spatial sound reproduction experiences. Finally, there are studies that are truly psychophysical in their design, that is, focused upon determining the relation between physical measures and psychological measures.

The scientific goals of psychophysical studies include predictive validation of objective measures of proximal sound stimuli (*predictor* variables), and evaluation of the multivariate interdependence of these objective measures in prediction of perceptual features of auditory spatial imagery (*response* or *criterion* variables). Shepard [8] referred to such endeavors generically as *psychophysical scaling*, while studies that exclude the involvement of physical predictor variables, focused only upon relations between spatial sound perceptions *per se*, were termed *perceptual scaling* studies. This paper focuses on the quintessential example of a *perceptual scaling* study, in that it treats the problem of uncovering the perceptual structure underlying judgments of inter-stimulus similarity. The most common analytic tool used for such exploratory investigation is MultiDimensional Scaling (MDS), in one of its many implementations that have evolved over several decades. It is instructive to read an early explanation of the role of MDS in this context from Torgerson's 1952 book [9]:

The traditional methods of psychophysical scaling presuppose knowledge of the dimensions of the area being investigated. The methods require judgments along a particular defined dimension . . . In many stimulus domains, however, the dimensions themselves, or even the number of relevant dimensions, are not known. What might appear intuitively to be a single dimension may in fact be a complex of several . . . Other dimensions of importance may be completely overlooked. In such areas the traditional approach is inadequate . . . This model differs from the traditional scaling methods in two important respects. First, it does not require judgments along a given dimension, but utilizes, instead, judgments of similarity between the stimuli. Second, the dimensionality, as well as the scale values, of the stimuli is determined from the data themselves.

2.2. MultiDimensional Scaling (MDS)

This paper addresses the need for a clear explanation of the theoretical basis and practical application of this popular exploratory technique, generically termed MDS, in the context of perceptual evaluation of spatial sound reproduction. While MDS has seen many fruitful applications in this area (e.g., [10]), its use as an exploratory technique has some clear limitations, especially with regard to the problem of interpreting the dimensions of a given MDS solution. When studying the auditory perception of complex, naturalistic stimuli, it is often the case that an explicit psychoacoustic theory for predicting psychophysical relationships between those stimuli is not available, or only prematurely specified. Under these circumstances it can be beneficial to collect subjective data about how perceptually similar or dissimilar stimuli are without specifying the ways in which stimuli may differ from one another. In the presence of data on such global differences between stimuli, the use of exploratory techniques such as MDS can be quite valuable.

In its classical form, MDS is a data analytic method that can be used to derive a spatial representation for a set of stimuli based upon a single set of measured similarities (or dissimilarities) between those stimuli. Four primary purposes of such perceptual scaling have been identified as follows [11]:

1. to create a low-dimensional representation of otherwise complex data,
2. to test how distinguishable multidimensional stimuli are from each other,
3. to discover the stimulus dimensions that underlying judgments of (dis)similarity,
4. to model the psychological dissimilarity between stimuli in terms of a distance function.

Of course, individual subjects may differ in how they form judgments of global dissimilarity, and so a refined method for doing a *weighted* MDS analysis [5] that takes such individual differences into account is to be recommended. This paper teaches the use of INDSCAL (INDividual Differences SCALing) [12] analysis as a powerful means for deriving an interpretable representation of the dimensions underlying reported inter-stimulus dissimilarities obtained from a potentially inhomogeneous group of subjects, each of which may place different *weights* upon each of the perceptual dimensions. While sets of dissimilarity data can be averaged across subjects to obtain one aggregated dataset for submission to classical MDS analysis [13], this paper shows the advantages provided by the INDSCAL model for the analysis of multiple sets of dissimilarity data, without requiring the assumption of a homogeneous group of subjects who share an identical perceptual structure for the stimuli. Beyond this, the two primary advantages of INDSCAL are as follows:

1. INDSCAL provides a quantitative characterization of the individual differences that exist within a group of experimental subjects, based upon dissimilarity judgments obtained from each subject. The individual differences are captured in a set of *weights* placed upon each of the stimulus dimensions by each subject.
2. INDSCAL provides an inherently unique configuration solution that requires no further analysis to find a meaningful rotation, in contrast to the orientational ambiguity inherent to classical MDS¹.

Interpreting the results of classical MDS is problematic because the solution can be rotated through an arbitrary angle without violating the structure of the solution. Of course, inter-stimulus distances remain invariant under rotation of both classical MDS and INDSCAL solutions alike; but the orientation of the INDSCAL solution is determined by modeling agreement between subjects. INDSCAL is designed to separate those factors that are common to a group of subjects from the ways in which subjects differ. The mathematical basis for these advantages are well explained in the book by Borg and Groenen [11], and are beyond the scope of this paper. Only an intuitive overview is provided here by describing the form of the matrices of input and output data required and produced by conventional analysis software packages.

The upper portion of Figure 1 illustrates the operation of Classical MultiDimensional Scaling (CMDS), in which a square matrix \mathbf{D} of dissimilarity judgments $\{\delta_{ij}\}$ is collected for all pairwise comparisons of stimuli i and j from within the set of n stimuli. The goal of MDS analysis is to find the coordinates $\{x_{is}\}$ for each stimulus i on each perceptual dimension s of a Euclidean Stimulus Space spanned by matrix \mathbf{X} with dimensionality p . Whereas the rows of these two matrices correspond to the same items, (i.e., the stimuli), the columns of the two matrices differ: While the columns of dissimilarity matrix \mathbf{D} also correspond to stimuli, the columns of Stimulus Space matrix \mathbf{X} correspond to dimensions. The derived coordinates $\{x_{is}\}$ should configure the stimuli such that the Euclidean distances between the stimuli match well the dissimilarity judgments $\{\delta_{ij}\}$ for those stimuli.

¹As noted by Shiffman et al [5], the non-rotatability of the INDSCAL solution assumes error free data. Some rotation may be justified in the presence of error.

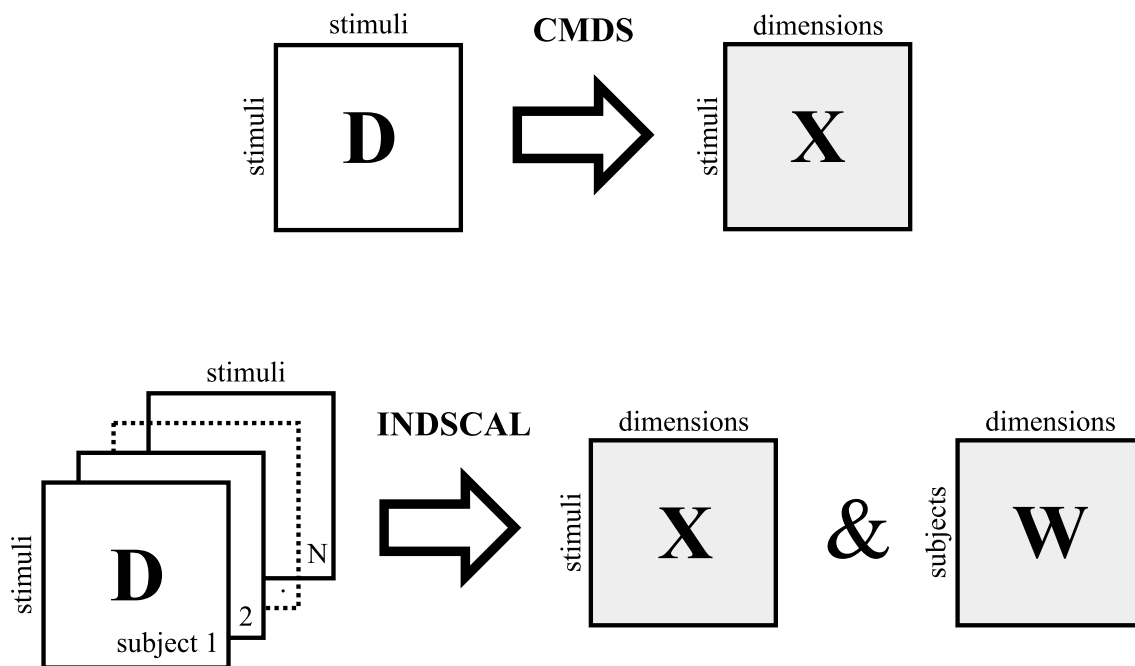


Figure 1: Comparison of Classical MultiDimensional Scaling (CMDS) and INDSCAL. Each box represents a matrix of data, the vertical word to the left of each identifying the nature of the row items, and the horizontal word above each identifying the nature of the column items. The large arrow identifies the type of analysis to which the input data matrices are submitted. Shading of a box indicates that the represented data matrix is the output of the respective types of analysis. See text for details.

If dissimilarity judgments are available for a number of listening subjects, a single group Stimulus Space can be derived using INDSCAL (INDividual Differences SCALing). For each individual k of m subjects, dissimilarity judgments $\{\delta_{ij,k}\}$ between stimuli i and j are collected for all pairwise comparisons of a given set of n stimuli. Just as in classical MDS, the primary goal of INDSCAL analysis is to find the stimulus coordinates $\{x_{is}\}$ within this Stimulus Space, but in contrast to CMDS, INDSCAL finds a group solution that provides coordinates for these stimuli on a set of perceptual dimensions that are common to all subjects. A Subject Space matrix **W** is also derived that reveals the weights $\{w_{ks}\}$ that each individual subject k placed on each dimension s of the p underlying dimensions in producing the inter-stimulus dissimilarity judgments. Note that while the columns of the two INDSCAL output matrices, **X** and **W**, correspond to the same items (i.e., the dimensions), the rows of the two matrices differ: While the rows of Stimulus Space matrix **X** correspond to stimuli, the rows of Subject Space matrix **W** correspond to subjects. Though such results might be interpreted directly according to observed relations between columns of **X** and a set of objectively measured stimulus parameters, more data is required for the best understanding of the set of perceptual dimensions. A subsequent paper will address the problem of interpreting the uniquely determined dimensions of the derived Stimulus Space via MultiDimensional Unfolding (MDU) of direct attribute ratings [6]. A brief overview of MDU is provided here to place the first stage of this study in context.

2.3. MultiDimensional Unfolding (MDU)

Direct attribute ratings (which can include preference ratings) may be submitted to classic MultiDimensional Unfolding (MDU) analysis to derive a Stimulus Space in which coordinates are provided for both stimuli and subjects. In addition to the matrix of coordinates for stimuli, **X**, a matrix **Y** of

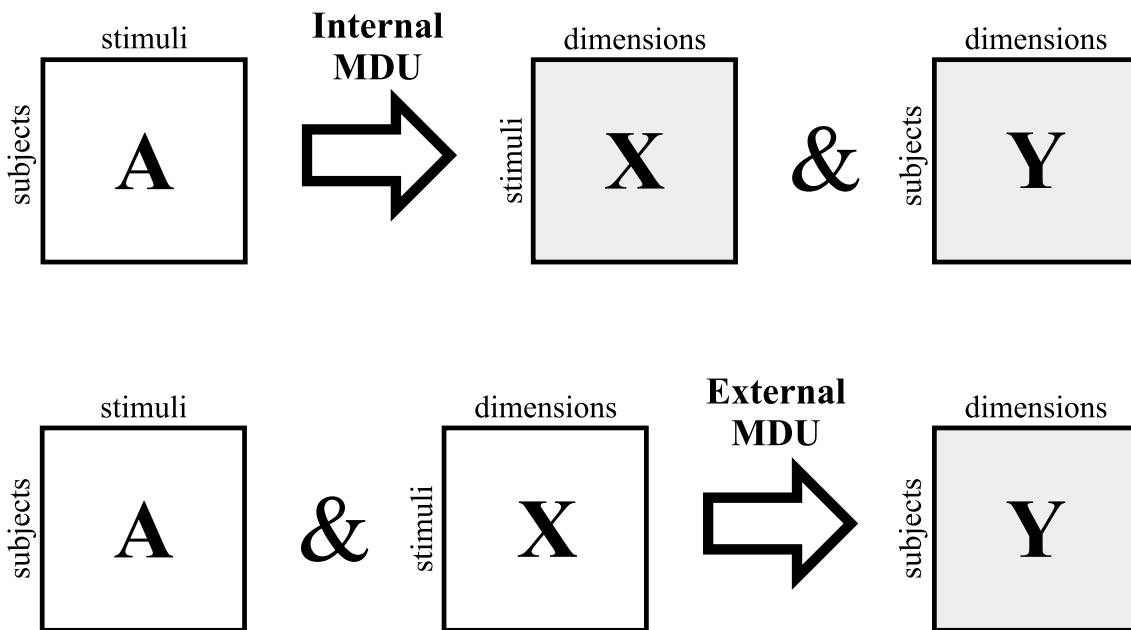


Figure 2: Comparison of internal and external MDU. As in Figure 1, each box represents a matrix of data, and the nature of the row and column items is identified by the words above and to the left of each box. Again, the large arrow identifies the type of analysis to which the input data matrices are submitted, and shaded boxes indicate data matrices that are the analysis outputs. See text for details.

ideal points $\{y_{ks}\}$ are found for each individual subject k on each dimension s in the p -dimensional Stimulus Space. The two principal forms of unfolding analysis, both illustrated in Figure 2, are termed *internal* and *external* MDU. When the only input to the analysis is a single matrix (**A**) of attribute ratings $\{\alpha_{ik}\}$ for each stimulus i of n stimuli and each individual k of m subjects, this type of analysis is termed *internal* MDU. If inter-stimulus dissimilarity judgments are also available for all pairwise comparisons those stimuli, then an alternative analysis is possible that has, in a sense, a more solid foundation in the data than *internal* MDU does, due to the uncertainty inherent in its joint derivation of both stimulus and subject coordinates. The alternative, illustrated in the lower portion of Figure 3, is *external* MDU. This analysis takes as an additional input the stimulus coordinate matrix **X** derived from a previous analysis (e.g., the Stimulus Space resulting from INDSCAL). Software packages providing *external* unfolding, such as PREFMAP-3 [14] and MULTISCALE [7], use prior estimates of **X** to aid in the determination, for each subject, of ideal stimulus coordinates in that Stimulus Space based upon the direct attribute ratings obtained from each subject. More confidence in the obtained matrix **Y** of the subject’s ideal points is developed when there is already some confidence in the validity of the Stimulus Space matrix **X** derived for the stimulus set. Gaining such confidence is precisely the motivation for the first stage of this study, and it is the focus of the remainder of this paper to document the stable derivation of Stimulus Space for the set of spatially-processed speech samples presented also for direct attribute rating.

3. EXPERIMENTAL DESIGN

This experiment was designed not to confirm a given hypothesis, but rather to explore the structure of the perceptual space associated with a set of spatially-processed speech samples, reproduced via headphones under various conditions. The central experimental hypotheses that motivated this study were not directly under test, but rather determined the stimulus variables to be manipulated in the

search for a multidimensional perceptual scaling result of value in subsequent experiments. These hypotheses regarding auditory spatial imagery were the following:

- Changes in the amount (i.e., duration) of indirect sound included in a virtual acoustic simulation will have a strong impact on perceived spatial qualities of the auditory image associated with speech sources.
- When the temporal distribution of simulated reflections is held constant, changes in the spatial distribution of simulated reflections will have an effect on additional spatial perceptual qualities of the auditory image that are relatively independent of those associated with variation in temporal distribution.
- Changes in stimulus bandwidth may potentially affect the contribution to auditory spatial quality made by various stimulus parameters.
- Listening simultaneously to two speech sources, lateralized to opposite sides of the listener, will potentially mask some of the spatial perceptual details that are more easily detected when listening to those same sources in an alternating sequence.

3.1. Stimulus generation

Dry speech samples were recorded by two male talkers in the University of Aizu's anechoic chamber (using an omnidirectional instrumentation microphone and Alesis ADAT at a sampling rate of 44.1 kHz). For the purposes of this experiment, four short, phonetically balanced and rich sentences were chosen for presentation via a headphone simulation of a small virtual acoustic space, typical of that that might be desired for a high quality binaural teleconferencing application. The choice to present complete sentences was based upon the assumption that speech samples shorter than these (around four seconds each) would not represent typical human telecommunications, and would therefore make the experimental task seem less natural. In each stimulus presentation, the subject heard all four of the following sentences:

Ford hit raw crime; No five leave court; Are raw moose lush; Inga buys ten.

In a typical teleconferencing situation, two talkers might be heard to be talking either simultaneously or in alternating succession, and therefore, conditions representing these two possible listening situations were included in this study. In the sequential condition, one talker located 40° to the listener's right would complete one sentence, after which the second talker would deliver a different sentence from a location 40° to the listener's left. All four sentences were heard in succession, alternating sides for each sentence. In the simultaneous condition, each talker would complete their two respective sentences, again located 40° to the listener's right and left, but the two would be heard at the same time rather than in succession. The opposing lateral positions of the two talkers in the simultaneous presentation condition created the possibility of masking early lateral reflections, those reflections being heard more clearly when they arrive from the walls of the model room opposite the location of the talker. In addition, the intelligibility should be reduced during simultaneous presentation.

In addition to a broad-bandwidth presentation, the complete set of sequential and simultaneous speech samples also were presented in a bandlimited condition intended to match the wideband telephony bandwidth standard [15]² The concern here was, of course, whether the increased bandwidth

²True wideband telephony bandwidth as defined in [15], is specified as ranging from 100 Hz – 7 kHz; however the stimuli in this study were only lowpass filtered (using an order 9 Chebyshev type II filter with cutoff frequency 7 kHz,

proposed as an improvement over “telephone grade” audio is enough to provide little degradation of the spatial image in comparison to the broadband case [16].

The factorial combination of the two bandwidth levels and the double-talk conditions yielded the following four experimental conditions, referred to by the following numbers:

- Condition 1: Broadband (i.e. 22 kHz), sequential sentences,
- Condition 2: Broadband (i.e. 22 kHz), simultaneous sentences,
- Condition 3: Bandlimited (i.e. 7 kHz), sequential sentences,
- Condition 4: Bandlimited (i.e. 7 kHz), simultaneous sentences.

3.2. Stimulus generation

A virtual acoustic space can be created for the human listener through audio signal processing that is based upon a geometric model of that space. When the goal is the realistic simulation of the spatial sound stimulation listeners would hear were they located in such a space, the endeavor is termed “auralization,” a term introduced by Kleiner [17] as the auditory analog to visualization. The audio signal processing technology that makes auralization possible is termed “3D audio rendering,” bearing some resemblance to the 3D graphic rendering technology that makes possible the synthesis of realistic visual imagery. From the anechoic samples, it was desired to create a set of test stimuli that covered a wide range of auditory spatial image qualities. A range of spatial acoustic renderings were performed that were all characteristic of a small room.

Direct Sound Processing

Speech sources were placed at a range of 1 m and at angles of $\pm 40^\circ$ from the receiver position. In all cases except one, the direct sound was binaurally processed via convolution with measured head related transfer functions (HRTFs)³. In the one exceptional case, the two headphone signals were both convolved with the ipsilateral HRTF, and a spectrally-flat difference in level between the two ear signals was introduced by a 6 dB attenuation of signal presented to the contralateral ear. This condition was termed the **Level Difference (LD)** condition to contrast it with conventional binaural HRTF processing that presents the proper frequency-dependent **Interaural Level Difference (ILD)**. In both direct sound processing conditions, the identical **Interaural Time Difference (ITD)** was present, and so the two conditions differed only in that the interaural differences in level could be spectrally-flat, or could be **Head Related (HR)**. When sounds processed in these two ways (i.e., with LD and HR interaural cues) were compared under dry listening conditions (when no indirect sound simulation was included), the lateralization extent of the auditory image was nearly matched, though the difference in the quality of the auditory image was substantial.

through which the speech signal was passed first forward then backward to achieve 80 dB attenuation in the stopband with zero phase distortion in the passband).

³The employed head related transfer functions (HRTFs) were blocked-meatus, probe-tube measurements which had not been reduced to minimum-phase reconstruction. Though an adequate treatment of controversies surrounding the perceptual adequacy of minimum-phase HRTF reconstruction is beyond the scope of this paper, the authors would like to propose that the issue has not been convincingly resolved, excellent research efforts notwithstanding [18, 19, 20, 21]. The spatial sound processing for this study’s stimulus generation, therefore, did not use minimum-phase HRTF reconstruction, but rather used responses with measured phase intact for those spatial regions where the HRTF exhibited non-minimum-phase behavior. The HRTF dataset used was that collected by the first author at the behest of Intel Corp., for use in their 3D-RSX software for sound spatialization. These data (from subject MES) have been made available for non-commercial use via the following website: <http://www.u-aizu.ac.jp/~wlm/data/intel/>

Indirect Sound Processing

A three-dimensional image-model solution [22] was used to calculate the direction and effective range for discrete reflections arriving within the first 100 ms of the direct sound⁴. From this description of the temporal distribution of the indirect sound within a model room, discrete early reflections were simulated with the appropriate propagation delay and attenuation for each. The spatial distribution of these reflections was introduced into the simulation through convolution with the same measured HRTFs as were used to control direct sound incidence angle. The angles from the image model were rounded to match the nearest tabled values, which were spaced at 10 degree resolution in both azimuth and elevation. The goal here was to maintain the integrity of the original HRTF data. It was regarded as higher priority to avoid HRTF interpolation errors than to avoid deviations from the image-model-determined azimuth and elevation angles.

The sources were treated effectively as omni-directional in nature (no directional variation in source radiation was included for adjusting the spectral energy of discrete reflections). The reflective properties of the walls of the model room were specified as a match to plasterboard, using a filter with a broad peak between 500 and 1000 Hz and approximately 5 dB attenuation above 7 kHz. 2nd order reflections were passed through this “wall-material-simulation” filter twice, 3rd order reflections three times, and so on, leaving higher-order reflections quite dark.

A small virtual acoustic space was designed to provide a relatively dense temporal distribution of early reflections. In order to enable a comparison of auditory images with similar reverberance, but with varying spatial distribution of reflections, the modeled room was wide but not long (15 x 6 m), with a relatively tall ceiling (5 m). The source and receiver were displaced from the midline of the room to create a somewhat decorrelated interaural distribution of reflections (relative to the room’s lower-left-rear corner, source coordinates, in meters, were {9.8, 5.0, 1.8} and receiver coordinates were {7.2, 2.2, 1.2}). Figure 3 shows the “hedgehog” plot for a subset of the simulated early reflections for this configuration of room, source, and receiver. The source and receiver are located in the center panel of the figure, surrounded by eight panels representing the eight nearest image-model rooms located on the same, ear-level plane as the model room. The location of the mirror-image source in each room is indicated by a line segment connecting it to the receiver position in the model room. The temporal and spatial distribution of reflections are shown in Figures 4 and 5 respectively.

Whilst the room geometry creates a defined reverberation, as illustrated in Figure 4, a very specific reflection pattern is created by the room aspect ratio, as can be seen from the image source Figure 3. The azimuth and elevation angles of the reflections are shown in Figure 5. To further increase the perceptual dimensionality of the stimulus set, two room orientations were defined: wide and long. Whilst for both sets the temporal characteristic of the reverberation remained constant, the distribution of early lateral versus central (front/rear) reflection change significantly. For example in the wide case, the earliest 1st order reflections arrive from the front and rear walls within 10 ms. In the long room case these reflection were more lateralized. This should affect the perceived spaciousness of the room. Lastly, the length of the room impulse response convolved with the test samples was varied over the range 0, 10, 30, 45, 60 and 100 ms. The 0 ms condition represents anechoic HRTF processing (no indirect sound), whilst the others provide an increasingly reverberant rendering of the room. These variations in indirect sound duration were designed to modulate subjective responses

⁴The image-model solution is known to have poor performance for simulating the low-frequency response for small rooms, and more accurate results are potentially available from other methods such as the finite element method (FEM) and boundary element method (BEM) (see e.g., [23]). For the current simulation using speech samples with little energy below 100 Hz, the interest was not so much in accurate simulation of the resonance frequencies associated with specific room modes, or other features requiring extremely precise specifications of the model room boundaries; rather the interest was in simulations within reach of relatively low-cost implementation for realtime DSP applications.

such as externalization of the sound (i.e. out of head localization of the distal stimulus) [24], the sense of spaciousness [10], and other subjective attributes of the auditory image.

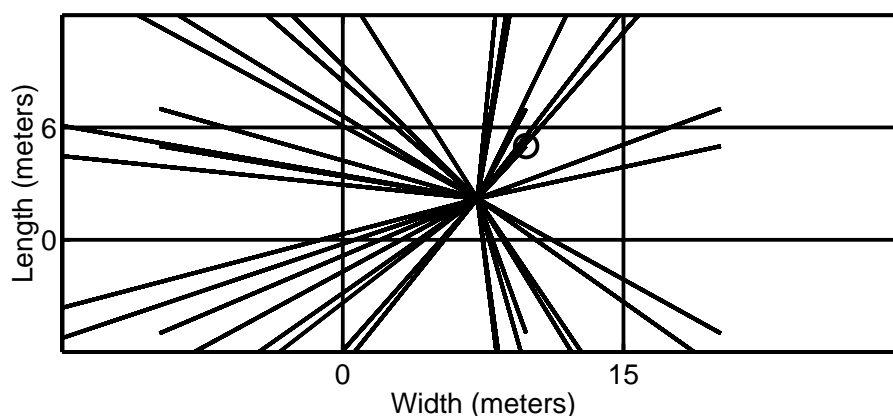


Figure 3: Graphical depiction of the image-model solution for a room of 15 x 6 x 5 m, showing reflection arriving only from the eight mirror-image rooms nearest the model room containing the source and receiver (all located on the ear-level plane). See text for details.

The range of indirect sound durations chosen were influenced by Barron's [25] study of the perceptual consequences of varying the relative delay and gain of a single reflection displaced laterally by 40° from the direct sound. For an otherwise anechoic stimulus, such variation in a single simulated reflection resulted in four distinct subjective effects: image shift, tone coloration, spatial impression, and disturbance (due to a clearly audible echo at longer delays). In anechoic simulations of multi-channel spatial sound reproduction incorporating simulated reflections delivered from loudspeakers located in many spatial directions (e.g., [26]), still more pertinent questions have been asked about the changes timbre associated with changes in the binaural sound field. For variations at the shortest reflection latencies (e.g., those under 10 ms) the variations in the auditory spatial image are not described as changes in tone color or timbre, but rather produce image broadening and displacement [27]. However, in the 10 to 45 ms range of reflection latency, strong tone coloration is the most frequently reported subject description.

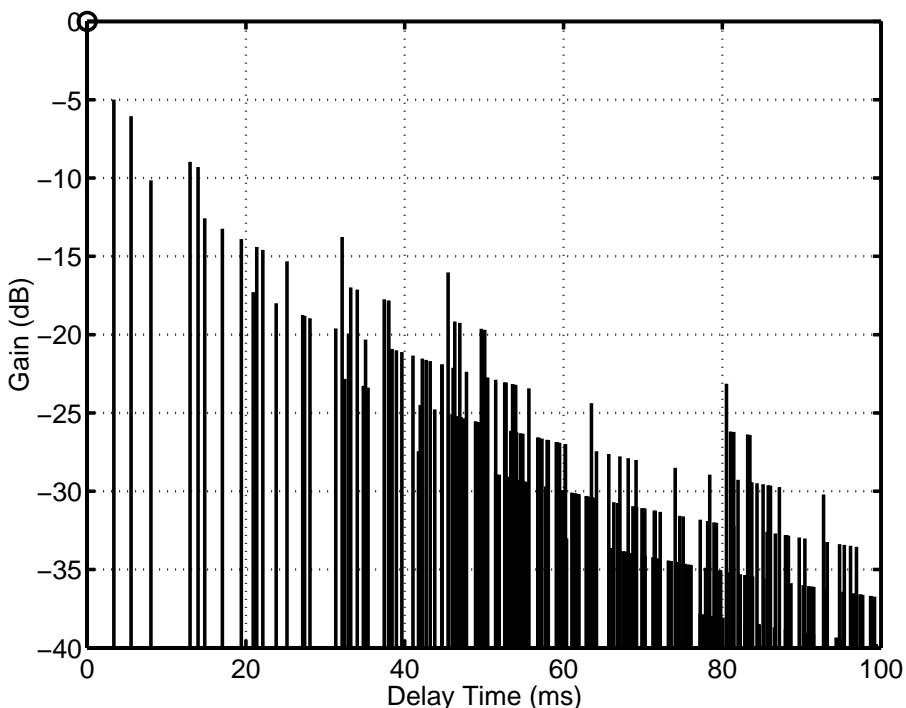


Figure 4: The echogram shared by the wide and long room, showing the gain on each discrete reflection as a function of time (prior to directional processing). Note that for this graphic, an extra 3 dB attention is applied to each reflection for each increase in reflection order.

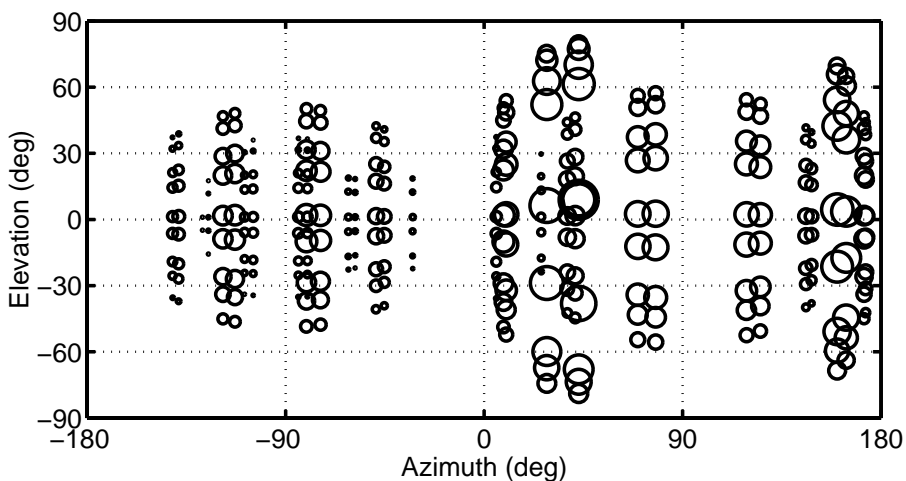


Figure 5: The spatial distribution of reflections calculated by the image model for the wide room configuration. Centered on the azimuth and elevation angle of each discrete reflection is a circle, the radius of which codes the gain in dB for each (and matches the gain values plotted on the vertical axis in the above echogram). The size of the emboldened circle at 40° establishes the 0 dB reference level for the direct sound.



Figure 6: GuineaPig [29] user interface for listener familiarization and training

Construction of the stimulus set⁵

The selected stimuli consisted of 12 spatializations for all sentences within each of the four main conditions and can be summarized as follows, with their associated symbols in brackets:

- Anechoic, Ipsilateral HRTF with spectrally-flat **L**evel **D**ifference (LD)
- Anechoic HRTF processed, ILD is **H**ead-**R**elated (HR)
- Anechoic HRTF + 10 ms wide room processing (● 10)
- Anechoic HRTF + 30 ms wide room processing (● 30)
- Anechoic HRTF + 45 ms wide room processing (● 45)
- Anechoic HRTF + 60 ms wide room processing (● 60)
- Anechoic HRTF + 100 ms wide room processing (● 100)
- Anechoic HRTF + 10 ms long room processing (○ 10)
- Anechoic HRTF + 30 ms long room processing (○ 30)
- Anechoic HRTF + 45 ms long room processing (○ 45)
- Anechoic HRTF + 60 ms long room processing (○ 60)
- Anechoic HRTF + 100 ms long room processing (○ 100)

3.3. Stimulus presentation

All tests were configured and run employing the Guinea Pig subjective testing software [28, 29] allowing for computerized representation of stimuli and automatic data collection in a monadic sequence⁶. Samples were stored on hard disk with a resolution of 16 bit at a sampling rate of 44.1 kHz. Sample playback was provided via the SGI 16 bit DAC line out, a Symmetrix SX204 headphone amplifier and Sennheiser HD580 headphones.

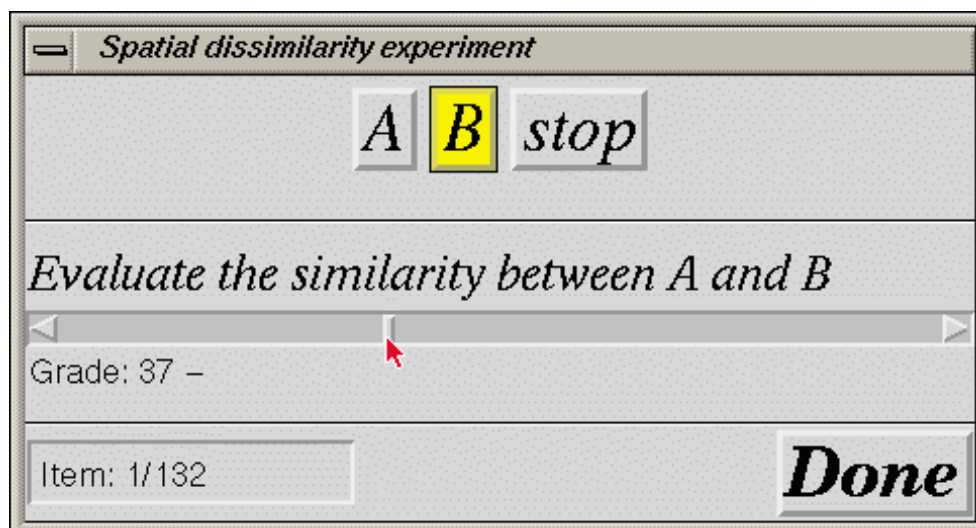


Figure 7: GuineaPig [29] user interface for the dissimilarity experiment

3.4. Listeners and training

Ten naive⁷ listeners were employed for this experiment to provide an estimate of how a population of typical subject would evaluate such samples. Little was known regarding the subjects experience, expertise, or the quality of their hearing.

Prior to commencing the experiment, all listeners were familiarized with the task with an oral instruction session and written instructions. They were further familiarized with the samples by listening to all samples at their leisure, employing the user interface illustrated in Figure 6 and allowed to perform a pilot session to become familiar with the user interface and grading scale. Only once these stages were completed were listeners allowed to continue onto the main dissimilarity experiment.

The user interface employed for the pilot and main experiment is illustrated in Figure 7. Subjects were asked to perform was to listen to each sample pair and rate their relative dissimilarity on a 100 point scale. A response of “0” implied that the two samples were perceived as “exactly the same” and a response of “100” implied that the two samples were perceived as “completely different”. Listeners were allowed to freely switch between samples as often as required until they were satisfied with their grading. Sample pairs were created for the full permutation set within each conditions, leading to 132 pairs per condition. A different order of presentation was created for each listener, in an attempt to minimize order effects.

4. ANALYSIS

The obtained dissimilarity data was submitted to INDSCAL analyses using the SPSS software package. The following two sections address the preliminary analyses and final Stimulus and Weight Space derivations, respectively. A third, complementary analysis was performed using the S-PLUS software package to find a hierarchical cluster representation for the stimuli in each of the four main conditions.

⁵These test stimuli and the original anechoic samples can be downloaded from the following web site: <http://www.nic.fi/~nickz/work/mdusamples.html>

⁶Once a sample pair is presented and graded on screen, it is removed prior to the presentation of the next sample pair, enhancing independence of judgments.

⁷A discussion of listen expertise and the meaning of naivete in this context is presented in [30]

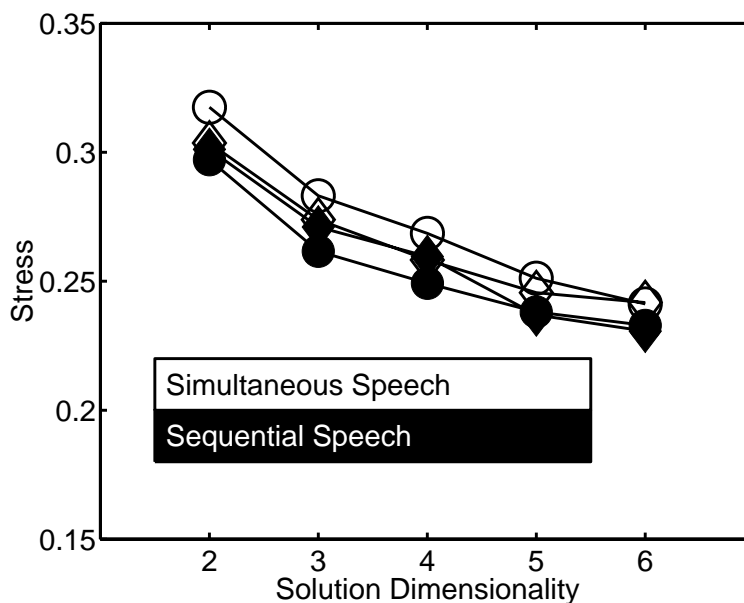


Figure 8: Stress plotted against dimensionality for conditions 1–4. \diamond : broadband, \circ : 7 kHz

4.1. Validation and dimensionality

One of the most important aspects of any multidimensional scaling analysis is the correct selection of the dimensionality prior to analysis, as discussed in ([31], pp.145–146), ([5], pp.10–14), ([11], pp.50–54). As noted by Wish and Carroll ([32], pp.321–322) the selection of dimensionality is “a balance between goodness of fit, interpretability and parsimony of data representation”. The normal manner in which one decides upon the suitable dimensionality of the solution, is to consider the stress vs. dimensionality plot. The appropriate dimensionality is often associated with a knee in the curve. However, as noted by Borg and Groenen ([11], p.52), this knee is often ill defined. This is the case with all four conditions in our experiments, as illustrated in Figure 8. It can be noted that there is no very clearly defined knee in the stress curve for any of the conditions. However, there is a slight jog in the curve around a dimensionality of 3. Another means of assessing the dimensionality would be to study the explained variance as a function of dimension for different dimensionality solutions. This is illustrated for the four conditions in Figure 9. Whilst this method is not widely discussed for application in multidimensional scaling, bar a short discussion by Borg and Groenen ([11], pp.50–54), such methods have been used in Generalized Procrustes Analysis ([33], pp.200–203). It is of interest to maximize the overall explained variance of the solution, though there is a need to limit the maximum number of dimensions considered, in particular if the increase in explained variance per dimension is less than ~ 0.05 (i.e. 5%). The use of an excessive number of dimensions will lead to “over fitting noise components” ([11], p.37), which is highly undesirable. Dimensions with a low contribution to the explained variance are a) tricky to explain b) may be associated with noisy data. Thus we have chosen to select a dimensionality of solution that meets both the stress and explained variance criteria. A three dimensional solution appears to be most suitable for all four conditions⁸.

Lastly, a matter to consider is whether our data is pure random noise. Such matters are of critical importance, as low stress values can be achieved from random noise datasets, as discussed by Borg

⁸It should be noted that multidimensional scaling is intended as a tool for dimensional reduction. In this respect, the aim is not to include as many dimensions as possible but as few as possible to meaningfully explain the data in a macroscopic sense.

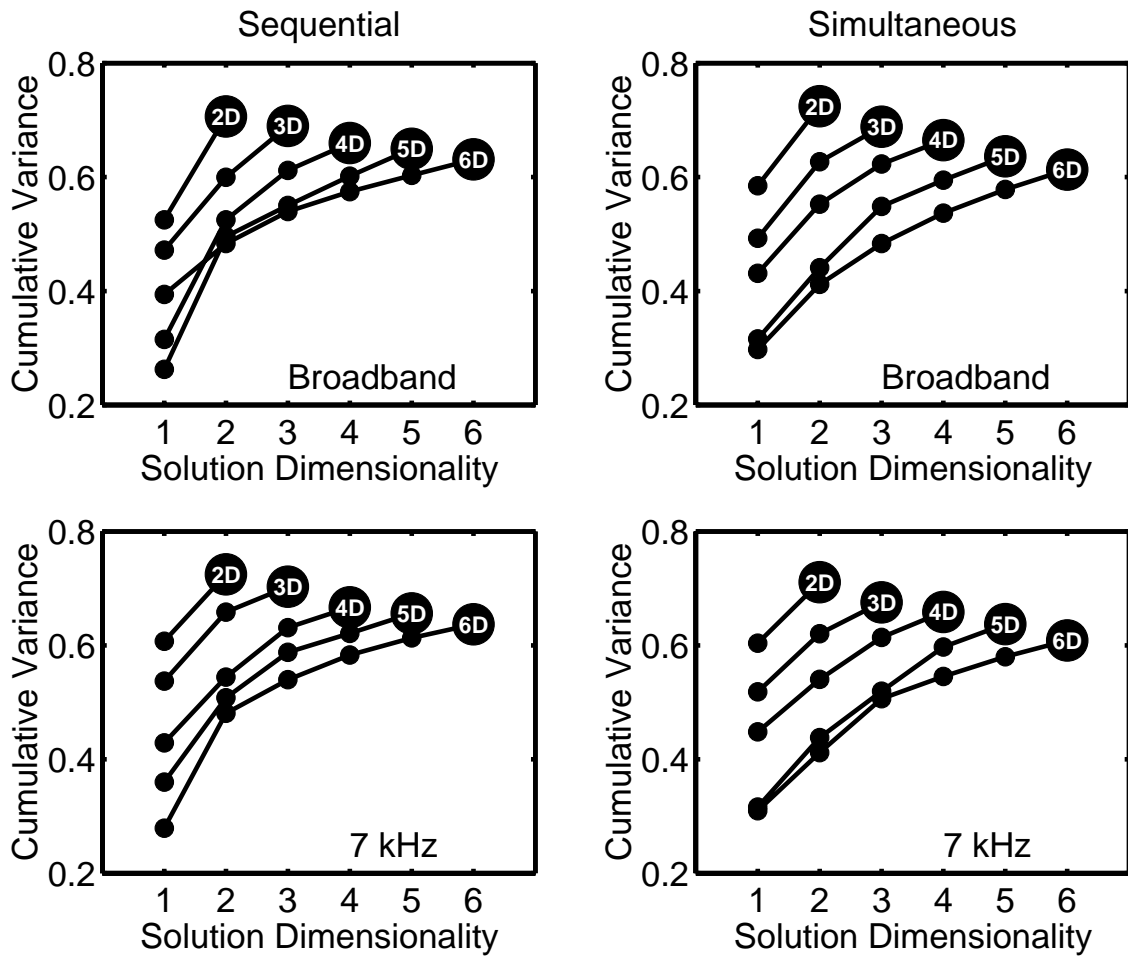


Figure 9: Explained variance as a function of dimensions for different dimensionality solutions.

et Groenen ([11], pp.38–45), for example. Furthermore, it was desirable to collect data from naive listeners to establish a more generalized opinion. However, as a result we have little awareness of the subjects reliability and repeatability as discussed in ([5], p.167), which might lead to noisy data. To test whether or not our data is pure noise, a dummy random dataset was created with the same matrix dimensionality. This was analysed in the same manner as the real dataset. For a solution dimensionality of 3, the stress of the random dataset was 0.354, compare to 0.26–0.28 for the real datasets. When checking the explained variance of the random data as a function of dimension, we find that ~ 0.06 is explained by each of the dimensions equally. This compares favorably with our data in which for the first three dimensions the contribution is 0.47, 0.13, 0.09 respectively, i.e. a total of 69% for condition 1. A similar level of explained variance occurs for all four conditions.

4.2. Stimulus and Weight Spaces

Having chosen the solution dimensionality of 3, both the Stimulus and Weight Spaces for each condition can be presented. It should be noted briefly at this point, that as each condition is independent of the other, so is the analysis. Whilst the INDSCAL results are normalised, no absolute comparison can be *assumed* across conditions neither with respect to the similarity of dimensions nor their scaling. However, a primary advantage of INDSCAL over other MDS methods is that the obtained dimensionality is often directly comparable to either a perceptual or physical dimension ([5], p.67).

The Weight Spaces (**W**) for each condition are illustrated in the lower portion of Figures 10–13. In all cases the three dimensional solution we can see a clear clustering of all subjects. Both the magnitude and direction of each weight vector are generally similar in all dimensions and for all conditions. This is a very encouraging result, as it suggests that subjects have similar options regarding all dimensions. Furthermore, this implies that we can analyse the Stimulus Space just once for all subjects. Another encouraging aspect is that, despite the similarities in magnitude and direction of the individual weights, there is enough variance within the cluster to aid in obtaining a unique orientation for the configuration solution. This variance is important for the INDSCAL procedure, as without it the solution would reduce to that of a Classical MDS solution provided with a single dissimilarity matrix, averaged over subjects, with all the associated limitations.

The Stimulus Spaces results (**X**) for each condition are illustrated in the upper portion of Figures 10–13. Whilst we have derived a Stimulus Space by this analysis, our aim here is not to describe these dimensions, as this will be addressed in a report of subsequent studies [6]. However, some observations can be made for each of the dimensions derived.

Dimension 1 contributes 47–54% of the explained variance in the INDSCAL solution for the four conditions, as illustrated in Figure 9. Considering the top-left plots of Figures 10–13, we can make the following observations regarding this first dimension of the solution. Firstly, we note that in all cases there is no differentiation between anechoic head-related (HR) processing and Level Difference (LD) stimuli. We can also observe that there is limited differentiation between the long and the wide room configurations, suggesting that this dimension maybe independent of those spatial distribution differences or the timbral qualities of the virtual acoustic space. It was also found that there is a strong association between dimension 1 and the “indirect sound duration”. This association is presented for the four conditions in Figure 14. A similarity can be found across all four conditions and we note that the HR and LD stimuli are associated with the lowest values of dimension 1, whilst the 100 ms stimuli are associated with the highest values of dimension 1. Whilst the experiment included no truly psychophysical investigation, it is of interest to observe that the curves illustrated in Figure 14 suggest a strong psychophysical relationship between indirect sound duration and dimension 1. The similarity across all four conditions and the associated high explained variance due to this dimension suggests that it is the dominant factor irrespective of the of the double-talk condition (sequential

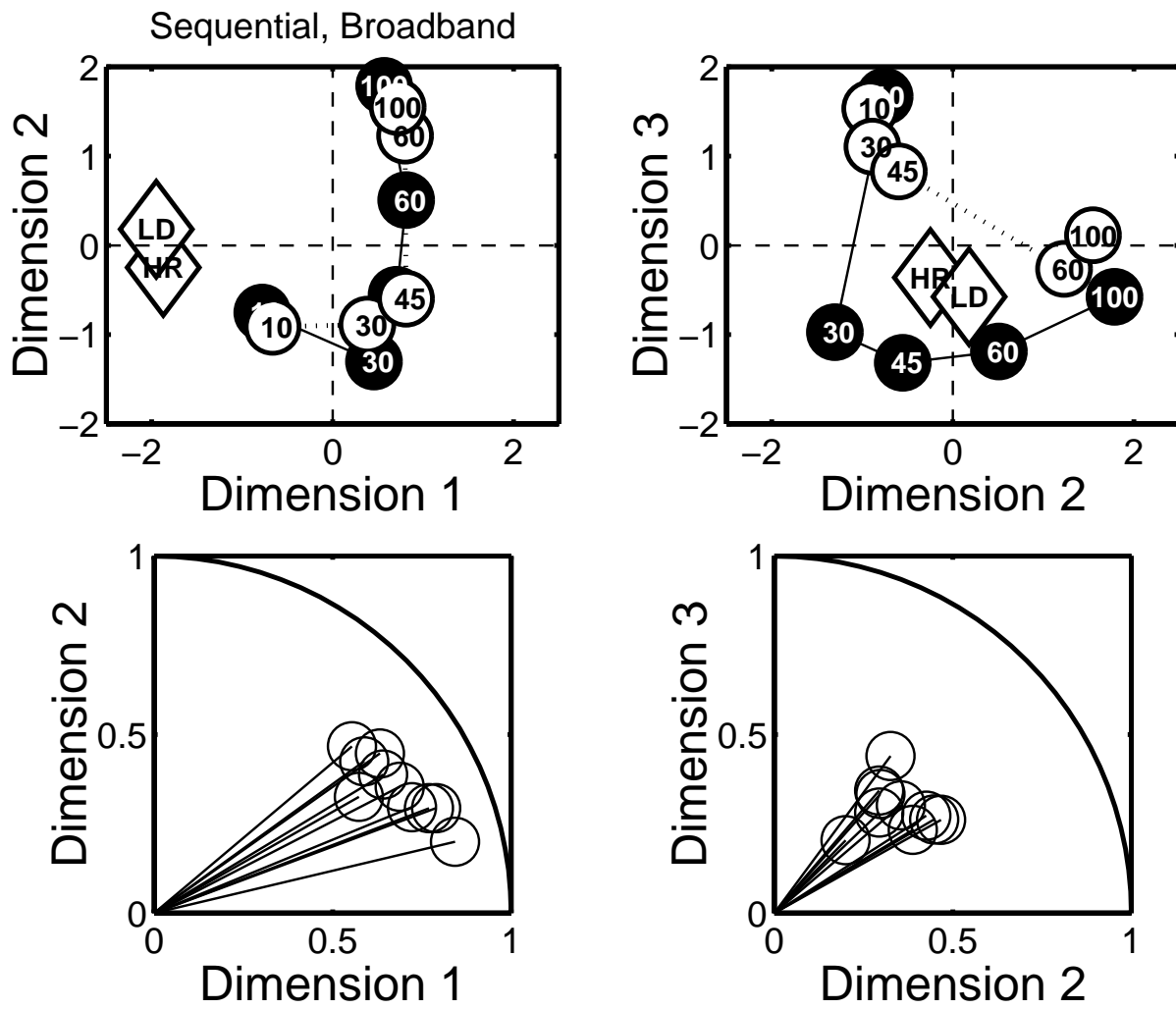


Figure 10: Condition 1 (broadband, sequential sentences) three-dimensional INDSCAL solution. Upper plots illustrate the Stimulus Space result (\mathbf{X}) and lower plots show the Weight Space result (\mathbf{W}) for dimensions 1–3. ●: wide room configurations, ○: long room configurations.

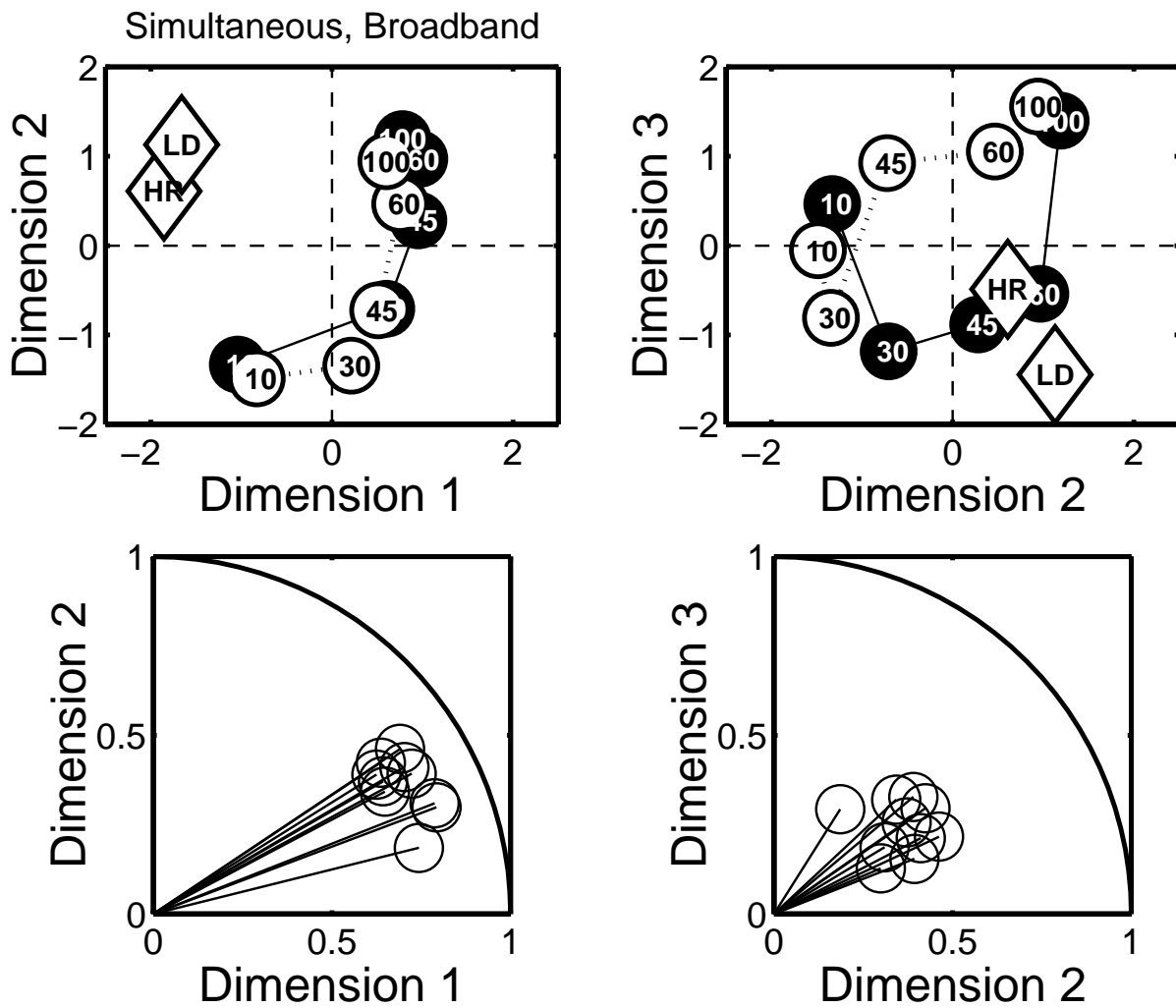


Figure 11: Condition 2 (broadband, simultaneous sentences) three-dimensional INDSCAL solution. Upper plots illustrate the Stimulus Space result (\mathbf{X}) and lower plots show the Weight Space result (\mathbf{W}) for dimensions 1–3. ●: wide room configurations, ○: long room configurations.

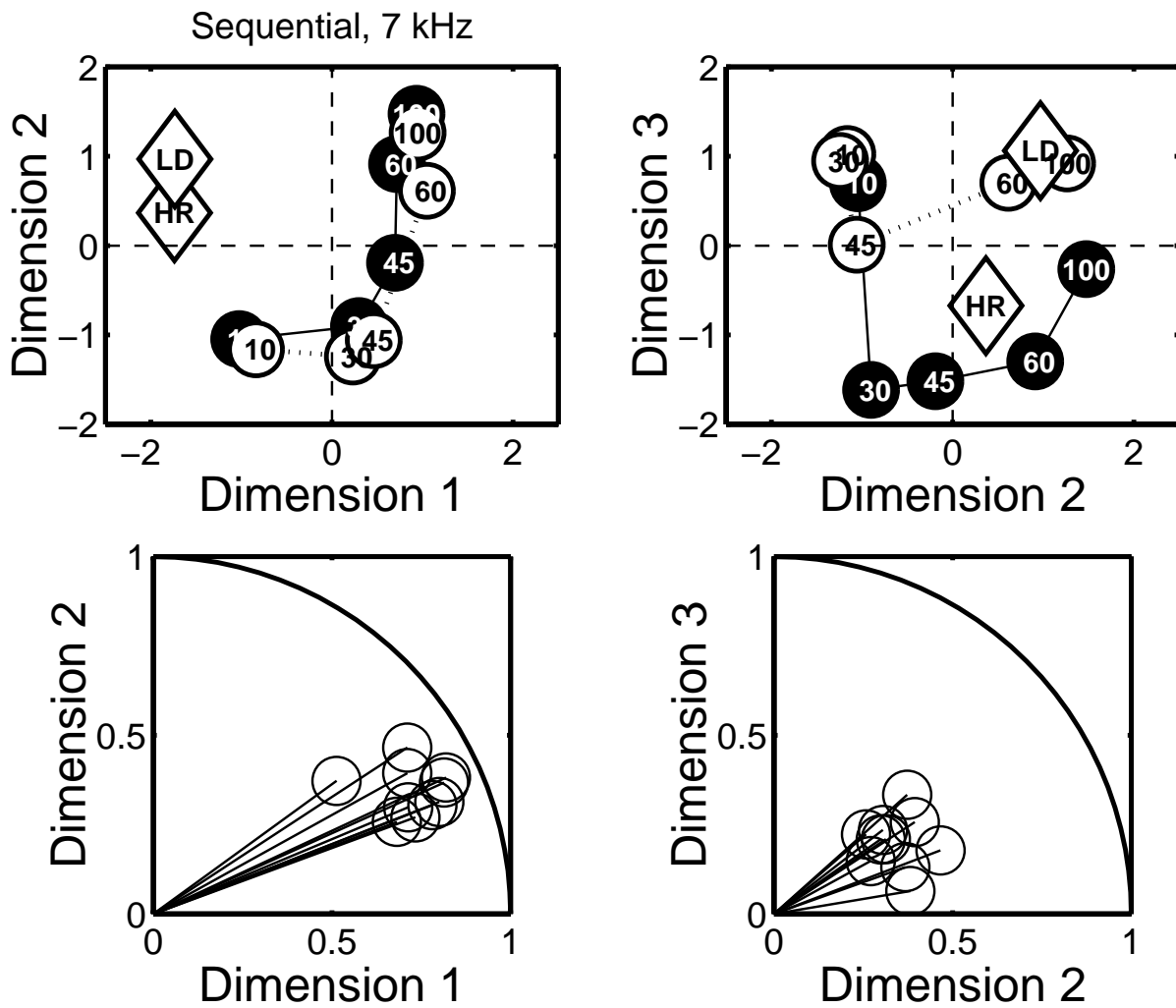


Figure 12: Condition 3 (7 kHz, sequential sentences) three-dimensional INDSCAL solution. Upper plots illustrate the Stimulus Space result (**X**) and lower plots show the Weight Space result (**W**) for dimensions 1–3. ●: wide room configurations, ○: long room configurations.

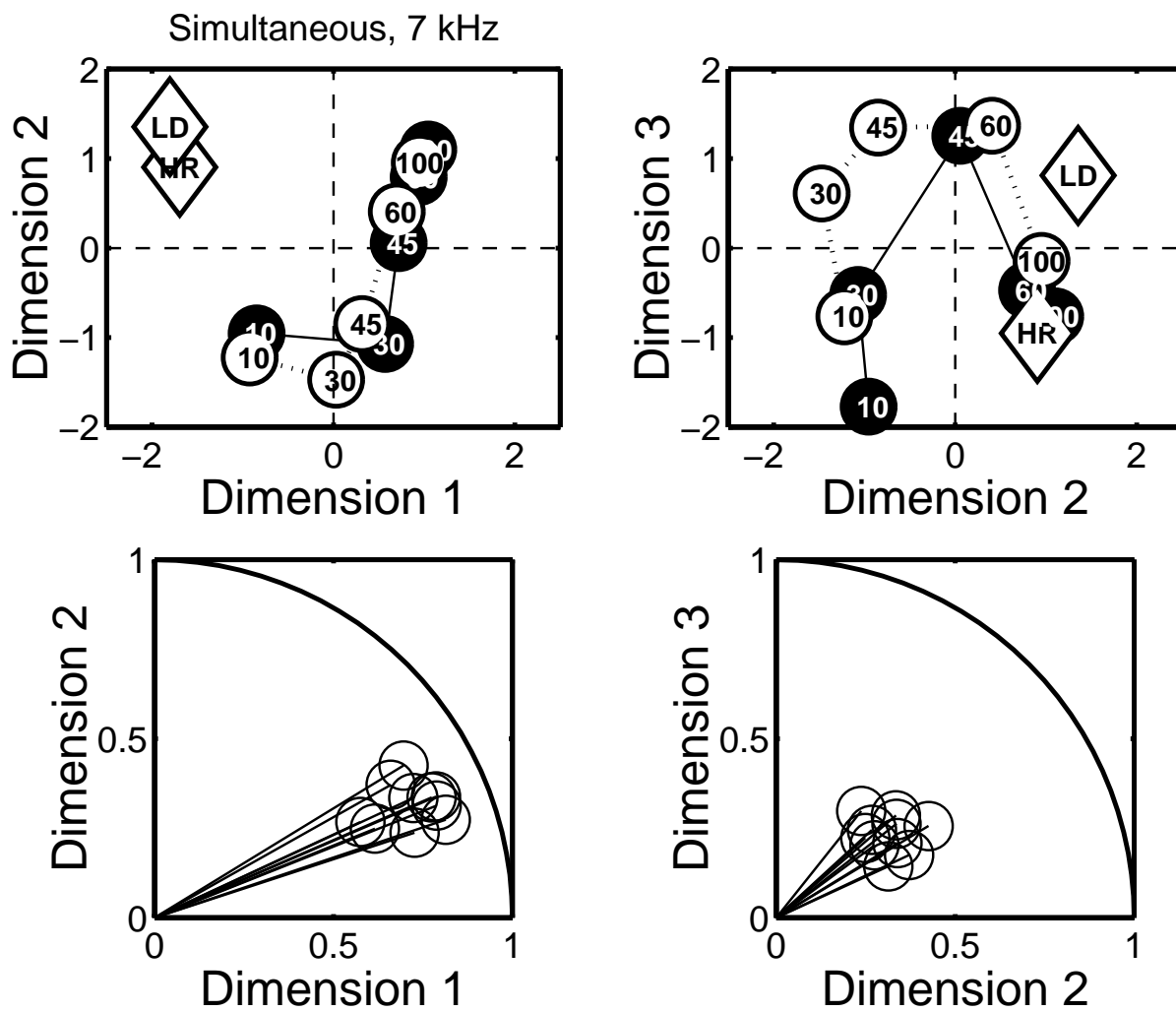


Figure 13: Condition 4 (7 kHz, simultaneous sentences) three-dimensional INDSCAL solution. Upper plots illustrate the Stimulus Space result (**X**) and lower plots show the Weight Space result (**W**) for dimensions 1–3. ●: wide room configurations, ○: long room configurations.

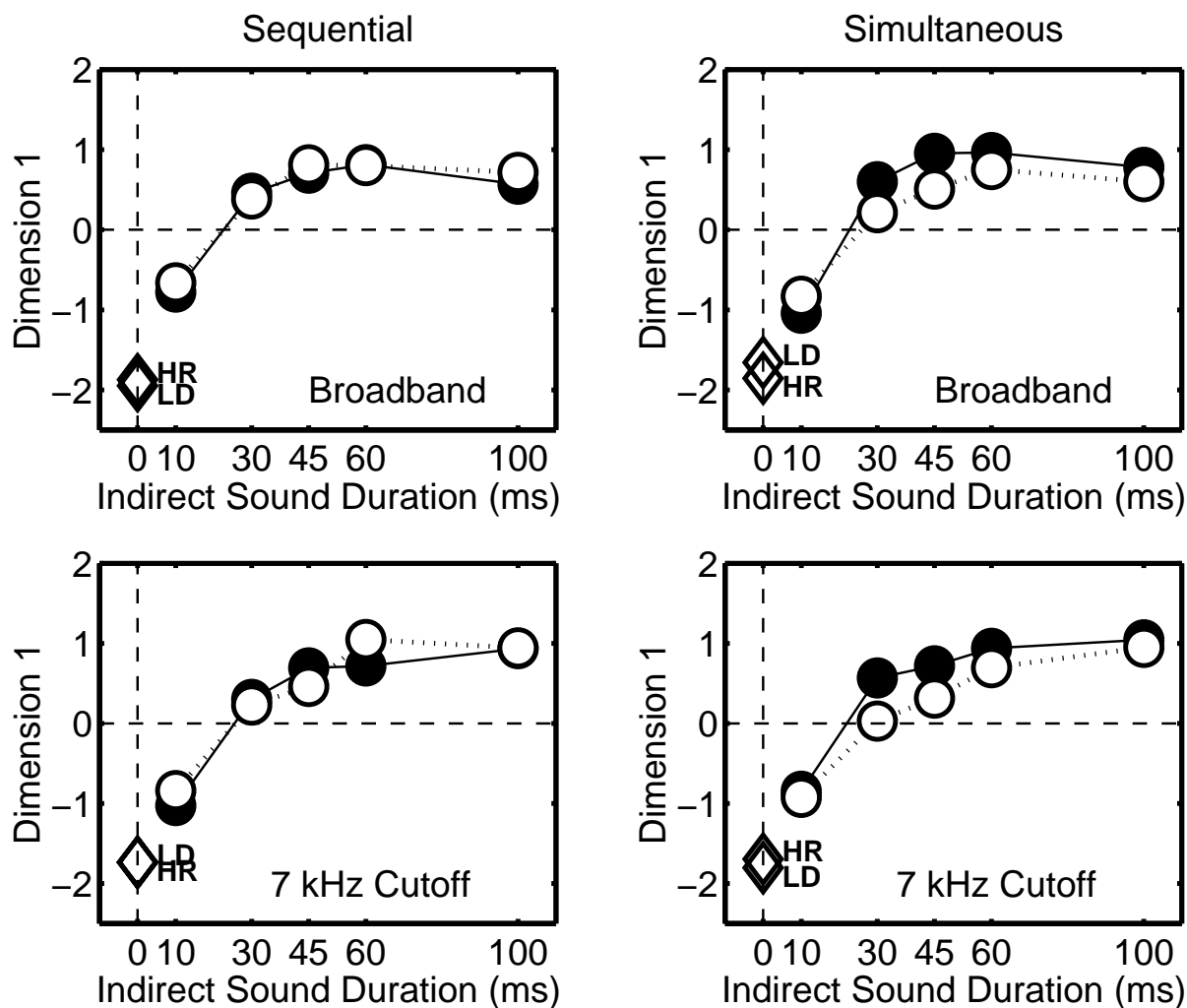


Figure 14: Stimulus (X) space for dimension 1 against indirect sound duration for the three-dimensional INDSCAL solution for all condition. ●: wide room configurations, ○: long room configurations.

versus simultaneous) or the bandwidth of the reproduction.

The second dimension is presented in the upper half of Figures 10–13, which describes 10–13% of the variance for the four conditions. Once again we can see that there is little differentiation between the wide and the long room configurations in all conditions. Also we see that there is little to differentiate the LD and HR stimuli. There can also be found a strong temporal grouping of the stimuli. However, this is where the similarity with dimension 1 ceases. The main feature to note for dimensions 2 is the association of the LD, HR and 100 ms stimuli, which have the highest values in dimension 2. For most of the conditions we find that the stimuli in the temporal range 10–~45 ms lie in the negative range of dimension 2, whilst the 0 ms (i.e. LD and HR) and ~45–100 ms stimuli lie in the positive region of the dimension. Barron [25] showed that reflection latency in the range 10–30 ms is associated with subjective reports of strong tonal coloration. Such coloration is less strong when reflections latency is in less than 10 ms and more than 45 ms. This timbral interpretation of dimension 2 is speculative and requires further substantiation.

The third and last dimension under consideration in this analysis describes 6–9% of the variance in the broadband cases and 4-5.5% of the variance in the 7 kHz bandwidth conditions. This suggests

that in the 7 kHz presentation, the perceptual variation on the third dimension of the Stimulus Space is masked to a greater extent than in the broadband presentation. The results for dimension 3 can be found for all four conditions in the upper-right of Figures 10–13. For this dimension we can observe some major differences compared to dimensions 1 and 2. Firstly, some significant differences can be found between conditions. We also note that there are often significant differences between the perception of the LD and HR stimuli, and also there are often large differences between the long and the wide room stimuli. In the generation of the stimuli, as discussed in section 3, it has been considered that the use of the long and the wide room cases might provide some variation in the breadth of the spatial image. Whilst it is difficult to define the meaning of dimension 3, there is a suggestion from these observations that it might be related to auditory source width (ASW) or envelopment.

As a final note, whilst a three dimensional solution has been selected for analysis, one of the key factors for this selection is interpretability of the dimensions ([11], pp.50–53; [5], p.50). The interpretation of the dimensions of the Stimulus Space will be presented in part II of this series [6], to be published in the near future, and will show the results of *external unfolding* of the direct attribute ratings combined with the dissimilarity data collected here. At that point it may be of interest to study higher order dimensions than proposed here, though their contribution of explained variance is limited to less than 7.8% per dimension for all conditions.

4.3. Cluster analysis

Alternative methods for exploring the results of the obtained inter-stimulus dissimilarity data (proximity estimates) provide additional opportunities for insight into the underlying perceptual relationships between the stimuli. The three complementary types of generic analysis of proximity estimates, MDS, tree-fitting, and clustering, are contrasted in an excellent review by Shepard [8]. In order to provide a different perspective from that provided by the INDSCAL analysis, the data obtained under the four conditions in this study were submitted to agglomerative hierarchical clustering (generated using S-PLUS routine `agnes` using a Euclidean distance metric; divisive hierarchical clustering gave virtually the same result). Figures 10 and 11 show how the stimuli are grouped together into clusters, and how those clusters in turn can be grouped together to form broad categories based upon gradually higher perceptual dissimilarity criteria (these criteria are specified as the *Height* values on the vertical axis of the plots). The clustering results show a non-spatial representation of the stimulus similarity structure that complements the spatial representation derived using INDSCAL. Note that the two anechoic stimuli are always clustered at a relatively low height under all conditions, and that they are joined with the cluster containing the stimuli processed with 10 ms of indirect sound. These clusters do not join the rest of the stimuli until a very high threshold is reached, indicating that they belong together, perhaps differing from the other stimuli in a categorical manner. Note the relationship to coordinates for these stimuli on dimension 1 of the INDSCAL results: Without exception, and under all four main conditions, these stimuli (**L_10**, **L_10**, **HR**, and **LD**) exhibit negative values while the remainder of the stimuli exhibit positive values on dimension 1.

An additional detail that is worth noting is the difference in the cluster results due to sequential versus simultaneous presentation. In the two simultaneous presentation conditions, the hierarchical structure is such that the **L_45** stimulus always groups closely with the two stimuli processed with 60 ms of indirect sound **L_60** and **W_60**. In the two sequential presentation conditions, however, the **L_45** stimulus does not group so closely with the stimuli processed with greater durations of indirect sound, joining the cluster only after **L_60** and **W_60** have already joined with **L100** and **W100**.

5. SUMMARY AND CONCLUSIONS

This study illustrates that spatially processed speech samples, both in a narrow and broadband sense, can create significant multidimensional perceptual variations. It is apparent that even naive subjects perceive such stimuli a) in a similar manner b) in a multidimensional manner. This suggests that the design of any system employing such spatial sound processing technology should approach the perceptual optimisation from a multidimensional standpoint.

The results of the INDSCAL analysis provide 3 salient dimensions, for which we have good confidence. In this study we have made some observations about the possible physical association of these three dimensions. These observations suggest that dimension 1 might be associated with the indirect sound duration, dimension 2 with tonal coloration associate with certain latency reflections and dimension 3 with aspects of spatial impression associated with reflections latency and direction of arrival.

However, it should be noted that these observations still need to be substantiated through the multidimensional perceptual unfolding, to be presented in the second paper of this series [6].

Dimensions 1 & 2 of the solution are noted to be quite stable across all four test conditions, representing different reproduction bandwidths and sequential/simultaneous sentence presentation. Dimension 3 is found to be more complex in nature and varies across the four conditions. However, due to the low explained variance of this latter dimensions, further analysis with this dimension should consider whether or not this data is noisy.

Lastly, a cluster analysis provided a complementary, non-dimensional perspective on the perceptual structure of the stimuli. The groupings of stimuli that were based upon the same inter-stimulus dissimilarity data as that submitted to INDSCAL. These groupings provided independent verification of the dimensional distinctions made for the stimuli within the derived Stimulus Space.

The benefits of the INDSCAL procedure can be seen from the uniformity of the output for dimensions 1 and 2 across all conditions. These dimensions are very similar in nature, and the INDSCAL output allows the experimenter to quickly ascertain the associated perceptual or physical association of dimensions due to the non-rotatability and uniqueness of the solution.

6. FUTURE WORK

The second stage of this work is to perform a direct attribute rating experiment employing the sample stimuli. This data will be subsequently analyses employing the PREFMAP implementation of external MDU, in order to study the ideal point model for this data set. This work will be presented as the second paper in this study.

ACKNOWLEDGMENTS

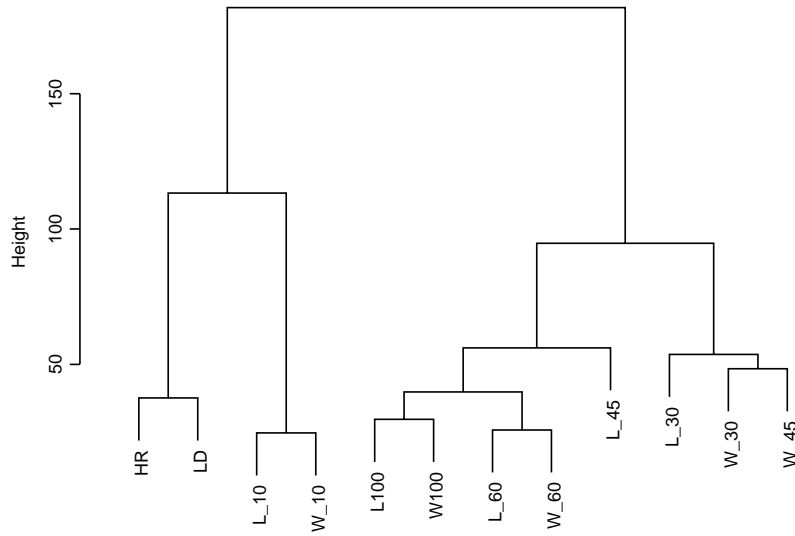
We would like to thanks Jyri Huopaniemi for his creative ideas regarding the development of test stimuli. Akiko Yoshida is thanked for her assistance in finalizing the subjective tests. The listeners are gratefully thanked for their patience with this experiment. The second author performed this work during a Foreign Research Fellowship at the University of Aizu. This author would like to thank the Nokia Foundation for funding support.

REFERENCES

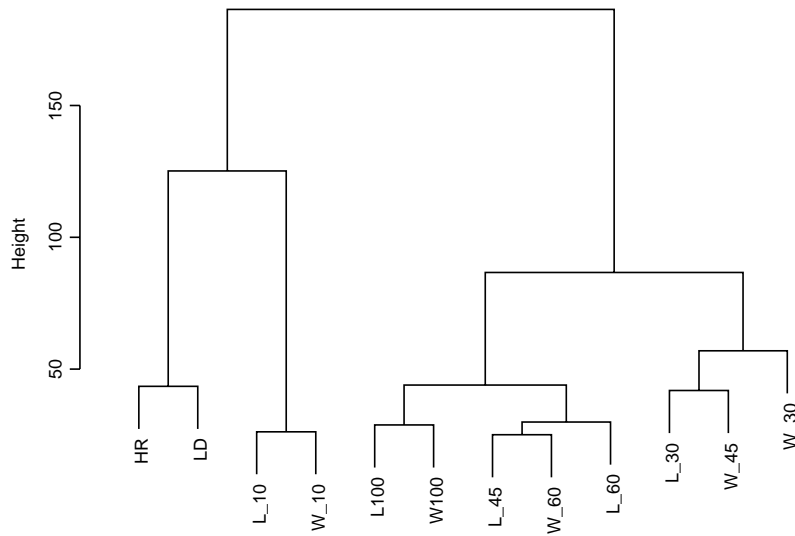
- [1] W. L. Martens, "The role of psychophysical research in the development of virtual acoustic rendering technology," in *Proceedings of the 7th Western Pacific Regional Acoustics Conference*, (Kumamoto, Japan), 2000.
- [2] G. S. Kendall and W. L. Martens, "Simulating the cues of spatial hearing in natural environments," in *Proceedings of the International Computer Music Conference*, (Paris), pp. 111–126, Computer Music Association, 1984.
- [3] D. Griesinger, "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," *Acta Acustica*, vol. 83, pp. 721–731, 1997.
- [4] Y. Ando and D. Noson, eds., *Music and Concert Hall Acoustics*. London: Academic Press, 1996.
- [5] S. S. Shiffman, M. L. Reynolds, and F. W. Young, *Introduction of multidimensional scaling*. Academic press, 1981.
- [6] N. Zacharov and W. L. Martens, "Multidimensional perceptual unfolding of spatially processed speech ii: Interpreting stimulus space using direct ratings." in preparation, 2001.
- [7] J. O. Ramsay, "MULTISCALE: Four programs for multidimensional scaling analysis," tech. rep., National Education Resources, Inc., Chicago, 1978.
- [8] R. N. Shepard, "Multidimensional scaling, tree-fitting, and clustering," *Science*, vol. 210, pp. 390–398, 1980.
- [9] W. S. Torgerson, *Theory and methods of scaling*. New York: Wiley, 1958.
- [10] J. Blauert and W. Lindemann, "Auditory spaciousness: Some further psychoacoustic analyses.," *Journal of the Acoustical Society of America*, vol. 80, no. 2, part 2, pp. 533–542, 1986.
- [11] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer-Verlag, 1997.
- [12] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.
- [13] L. R. Tucker and S. J. Messick, "An individual differences model for multidimensional scaling," *Psychometrika*, vol. 38, pp. 333–368, 1963.
- [14] J. Meulman, W. J. Heiser, and J. D. Carroll, "PREFMAP-3 user's guide," tech. rep., Bell Telephone Laboratories, Murray Hill, NJ, USA, 1986.
- [15] ITU-T, *Recommendation P.830, Subjective performance assessment of telephone-band and wideband digital codecs*. International Telecommunications Union, Telecommunications Standardization Sector, 1996.

- [16] D. R. Begault, "Virtual acoustic displays for teleconferencing: Intelligibility advantage for "telephone grade" audio," in *Proceedings of the Audio Engineering Society 98th Int. Conv.*, Audio Eng. Soc., 1995.
- [17] M. Kleiner, "Auralization an overview," *Journal of the Audio Engineering Society*, vol. 41, no. 11, pp. 861–874, 1993.
- [18] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1637–1647, 1992.
- [19] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "On the minimum-phase approximation of head-related transfer functions," *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995.
- [20] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer function phase spectra," *Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2821–2840, 1999.
- [21] J. Plogsties, P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Audibility of all-pass components in head-related transfer functions," in *Proceedings of the Audio Engineering Society 108th International Conference*, (Paris), 2000.
- [22] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, April 1979.
- [23] A. Pietrzyk, "Computer modeling of the sound field in small rooms," in *Proceedings of the Audio Engineering Society 15th Int. Conf.*, pp. 24–31, Audio Eng. Soc., 1998.
- [24] F. E. Toole, "In-head localization of acoustic images," *Journal of the Acoustical Society of America*, vol. 48, pp. 943–949, 1969.
- [25] M. F. E. Barron, "The subjective effects of first reflections in concert halls - the need for lateral reflections," *Journal of Sound and Vibration*, vol. 15, pp. 475–494, 1971.
- [26] S. Bech, "Perception of timbre of reproduced sound in small rooms: Influence of room and loudspeaker position," *Journal of the Audio Engineering Society*, vol. 42, no. 12, pp. 999–1007, 1994.
- [27] M. B. Gardner, "Image fusion, broadening, and displacement in sound location," *Journal of the Acoustical Society of America*, vol. 56, pp. 939–943, 1974.
- [28] J. Hynninen, "GuineaPig - Overview," <http://www.acoustics.hut.fi/~hynde/>, 1998.
- [29] J. Hynninen and N. Zacharov, "GuineaPig - A generic subjective test system for multichannel audio," in *Proceedings of the Audio Engineering Society 106th Int. Conv.*, Audio Eng. Soc., 1999.
- [30] N. Zacharov and V. V. Mattila, "GLS - A generalised listener selection procedure." unpublished, 1998.

- [31] W. R. Dillon and M. Goldstein, *Multivariate analysis: methods and applications*. John Wiley & Sons, 1984.
- [32] M. Wish and D. Carroll, *Handbook of statistics: Classification pattern recognition and reduction of dimensionality*, vol. 2, ch. 14: Multidimensional scaling and application, pp. 317–345. North-Holland Publishing Company, 1982.
- [33] G. Dijksterhuis, *Multivariate analysis of data in sensory science*, vol. 16 of *Data handling in science and technology*, ch. 7: Procrustes analysis in sensory research, pp. 185–217. Amsterdam, The Netherlands: Elsevier, 1996.

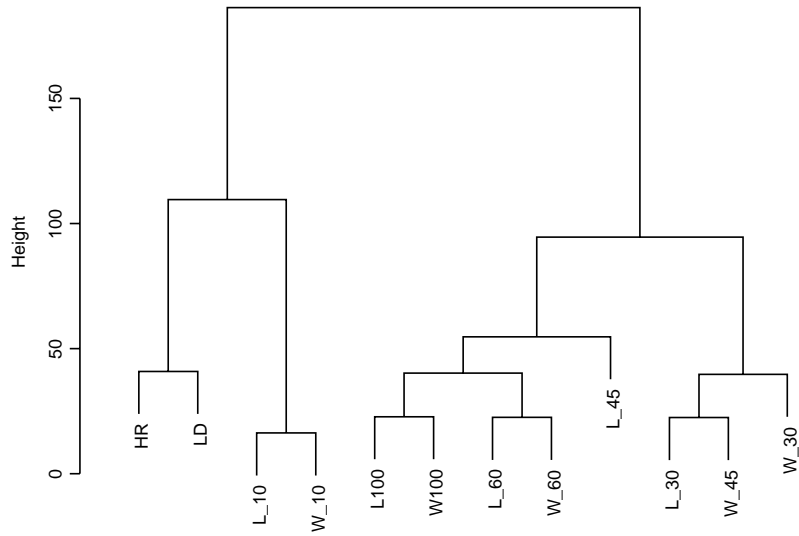


(a) Condition 1, sequential

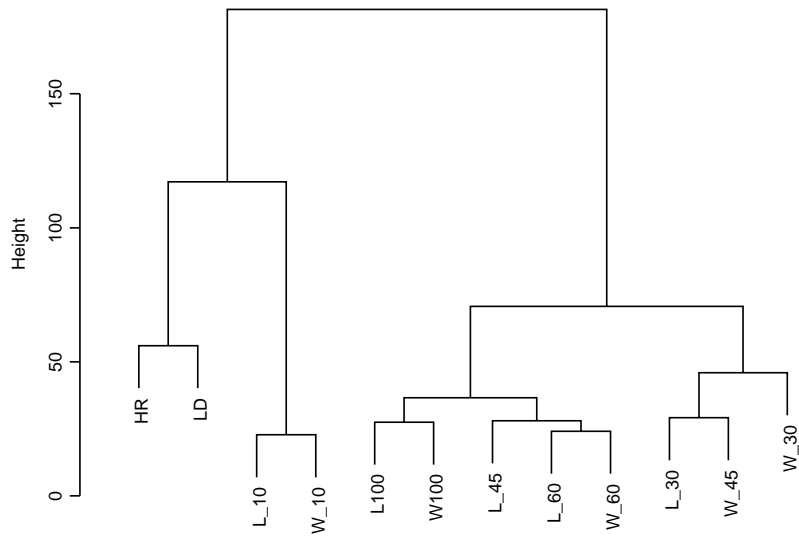


(b) Condition 2, simultaneous

Figure 15: Agglomerative hierarchical cluster analysis, for broad bandwidth.



(a) Condition 3, sequential



(b) Condition 4, simultaneous

Figure 16: Agglomerative hierarchical cluster analysis, for 7 kHz bandwidth.