Proceedings of the 2018 ASME International Conference and Exposition on Noise Control Engineering
NCAD2018
August 26-29, 2018, Chicago, IL, USA

NCAD2018-6128

SOUND QUALITY OF AIRCRAFT CABIN FOR VIP AND BUSINESS JETS

Nurkan Turkdogru Gurun
Aearo Technologies LLC, a 3M company
Indianapolis, Indiana, USA

Hemang N. Sheth
Aearo Technologies LLC, a 3M company
Indianapolis, Indiana, USA

ABSTRACT

This paper aims to identify the attributes that describe aircraft interior noise, determine most important psychoacoustic models that characterize cabin sounds, and construct a prediction model that can be utilized for VIP and business jets to evaluate subjective perception. In the first part, paired comparison listening tests and free verbalization are conducted with expert subjects who experienced VIP and business aircraft flight. The study generated a list of adjective pairs that describe perception of cabin sounds to be used for semantic differential listening tests. Multi-dimensional scaling is performed on paired comparison data. Results showed that subjects' decisions can be categorized in loudness and annoyance dimensions which are not necessarily linearly associated. The second part of the study is the development of a sound quality prediction model for aircraft cabin. Semantic differential tests are conducted with potential customers. Objective sound quality metrics are correlated to subjective test responses using principal components regression. This model is found to be most effective explaining pleasantness, comfort, and loudness perception. It is intended to be utilized to modify/redesign noise control treatments and sound signature of an aircraft. All listening tests were conducted inside an aircraft cabin simulator considering the influence of visual content.

1. INTRODUCTION

Significant reduction in the aircraft sound levels has been achieved over the last three decades. As current technical advances continue to be implemented, further reduction of noise levels becomes more challenging and expensive than it has ever been [1,2]. Advanced material technologies make it possible to reduce sound levels inside a fixed-wing aircraft as low as 50 dBSIL. Therefore, to accomplish higher customer acceptability, it is insufficient to solely focus on the level of noise but rather more emphasis needs to be placed on human factors regarding time structure and frequency content of sound.

Perception of sound is a complex, multi-sensory process evident by the responses under common perceptual conditions. Sound design must take the qualitative references between the senses into account. Human brain stores physical stimuli as subjective representations, which are collections of auditory, visual, tactile, and such data and their interactions. When the individual comes across a single, known stimulus (a vision, an audition, a smell, etc.), brain recalls it as the multi-sensory model it previously created [3]. Sound quality approach elicits the responses to multi-sensory experiences generated by a certain sound source and redesigns the source to satisfy the expectations of the individual, which are shaped by these experiences.

This paper is organized in two main parts. In the first part, authors conducted a series of paired comparison tests combined with free verbalization sessions and a questionnaire to determine semantic differentials that describe aircraft cabin sounds. The semantic differentials identified were used in the second part to conduct another set of listening tests with potential customers of the product. Then, a sound quality prediction model is developed by correlating subjective metrics obtained from listening tests with objective metrics computed with psychoacoustic and mathematical models representing different attributes of sound.

2. DETERMINING SEMANTIC DIFFERENTIALS FOR AIRCRAFT CABIN SOUNDS

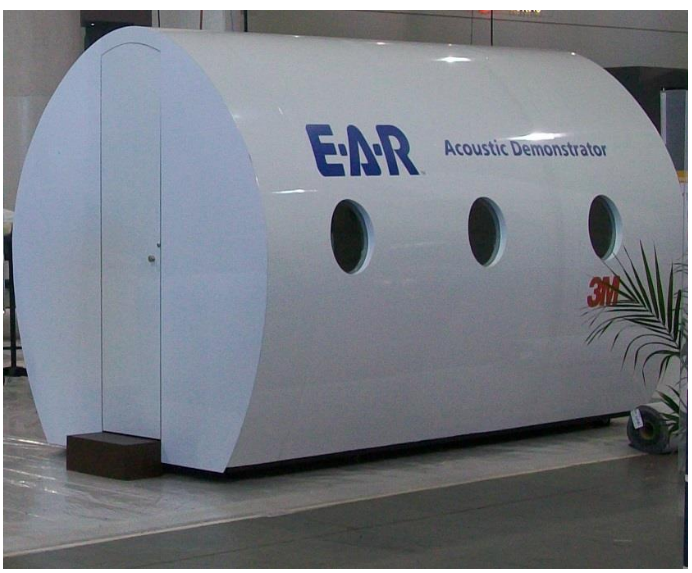
2.1. PREPARATION OF SOUND SAMPLES

Binaural sound recordings from two different aircraft - one VIP and one business jet- at cruise level and at several seat locations were used in the tests. The aircrafts were sound proofed with custom designed thermal and acoustic insulation. Data was acquired using Artemis binaural headsets BHS II and amplifier SQadriga II. Sound editing was done with Artemis Suite 6.1. Nine samples with 10.5 sec, one sample with 6.5 sec and one sample with 27 sec duration and all possible combinations of 11 samples, 55 pairs, were prepared.

2.2. LISTENING TEST ENVIRONMENT AND INSTRUMENTATION

Paired comparison listening tests are performed inside a 13′x8′x7′ room used to simulate an aircraft cabin. Figure 1(a) and 1(b) show the simulator cabin interior and exterior, respectively. Sound playbacks were executed with Sennheiser HD600 headphones and a Head Acoustics PEQ V equalizer. Subjects had full control of playing the sounds using Artemis Suite 6.1 on a laptop.

(a)



(b)

Figure 1 Test Environment (a) Interior (b) Exterior

2.3. PAIRED COMPARISON AND FREE VERBALIZATION TEST PROCEDURE

An expert group study composed of a survey on a list of 160 words and a series of paired comparison listening tests were conducted. Expert group was consisted of eight individuals with a diverse range of job functions from engineering to marketing, all of whom worked with VIP aircraft customers and experienced VIP aircraft flight. All individuals were required to take an audiometric test to ensure that they did not have any hearing loss.

In the first step, subjects were asked to complete a questionnaire. A set of 160 adjectives were gathered from several sound quality studies on aircraft, automobile, appliance sounds conducted in French, German, and English. None of the adjectives in English belonged to aircraft interior studies and to the authors' knowledge, there is no such study in the literature. The task was to choose 20-25 words from the given list of adjectives and categorize the chosen words into four dimensions of product sound -strength/magnitude, annoyance, amenity, and information content. A sample section of a filled-out questionnaire is shown in Table 1. Lyon states that combination of these four dimensions determines the acceptability, i.e. sound quality, of a product sound to the users [4]. The purpose was to find attributes, which differentiate aircraft sounds in American English. The subjects were not limited by this list. They were asked to add any adjective that is not included in the list but they think describes the aircraft sounds and, also, synonyms and antonyms of the words they picked from the list.

The second step was a combination of paired comparison tests and free verbalization. Paired comparison tests were carried out with 59 pairs of sounds - all combinations of 11 sound samples and repetitions of 4 pairs to check the consistency of the subject. Subjects were asked to rate the similarity of the two sounds for each pair on a scale from 1 to 7 as shown in Figure 2; 1 being "there is no difference between the two sounds" and 7 being "two sounds are extremely different". Pairs were represented in random order. Subjects did not have any information on sound samples rather than they were recorded inside an aircraft. They could listen each pair as many times as they wanted and turn back to a previous pair if they needed. Before starting the actual test, subjects were familiarized with the sounds and the evaluation process by a practice block composed of six pairs. All practice sound samples were different than the ones used for the actual test.

Table 1 Four dimensions of product sound questionnaire section

Strength/Magnitude			Annoyance			Amenity/Pleasantness			Information Content		
Adj.	**Syn.**	**Anton.**	**Adj.**	**Syn.**	**Anton.**	**Adj.**	**Syn.**	**Anton.**	**Adj.**	**Syn.**	**Anton.**
big	large	small	harsh	grating	enjoyable	clear	singular	muted	bright	crisp	low
booming	thunderous	muted	awful	bad	pleasant	humming	simple	shrill	buzzy	noise	clean

	No Difference						Extremely Different
Pair 1	1	2	3	4	5	6	7
Pair 2	1	2	3	4	5	6	7
Pair 3	1	2	3	4	5	6	7

Figure 2 Paired comparison test sheet example

In addition to rating the similarity of the sounds, the subjects had a free verbalization task. They were asked to interpret the sounds with their own words, describe unique attributes of each sample and common attributes of paired samples if they found the two similar. This technique allows the subjects to go beyond direct comparison of sounds and can be completed in a shorter time frame than other methods [5]. This task was completed every time they listened to a pair. Although author did not want to intervene the judgements of the subjects, some guidelines were provided.

Required subject consistency rate was determined as 75% and was tested by the four repeating pairs. If the difference between the ratings of first and second play of the pair was not more than two for at least three pairs, the subject was considered consistent. All the subjects were consistent.

2.4. TEST RESULTS

2.4.1. DETERMINATION OF SEMANTIC DIFFERENTIALS

Listening test responses were reduced to a list of adjectives. Another list was constructed from the questionnaire choices. 66 adjectives were identified from questionnaire and 166 adjectives from listening tests, 31 of which were common. Number of repetitions per each adjective is counted for both questionnaire and listening tests. Then, the weighted average of the two lists was taken. The weights were 0.3 and 0.7 for questionnaire and listening test, respectively. It was observed that some words in these sets have similar meanings and most words can be categorized under a few titles. Ten categories were identified as pitch, intensity, regularity, pleasantness, nature, spatial, image, price, technical, and others. This categorization was important to create a diverse list and to keep it in the scope of study that is to redesign sound with the tools available. In this context, pitch, intensity, regularity, pleasantness, nature, spatial, and image categories were found suitable.

Table 2 shows the final set of adjective pairs. All the words with only one exception, atonal, were picked from the list. It should be noted that although most of the words chosen are among the highest ranked ones, to represent a more diverse range of attributes, some low ranked ones are included in the final semantic differential list. Additionally, there are three onomatopoeic adjectives - buzzing, droning, and hissing - without antonyms.

Table 2 Adjective Pairs Describing Aircraft Interior Noise

Category	Adjective Pair	
Pleasantness	Pleasant	Harsh
	Comforting	Discomforting
	Tonal	Atonal
Pitch	High Tone	Low Tone
Intensity	Loud	Soft
Regularity	Steady	Fluctuating
	Oscillating	Random
Spatial	Distant	Close
Image	Bright	Dark
Nature	Buzzing	Not Buzzing
	Droning	Not Droning
	Hissing	Not Hissing

2.4.2. ANALYSIS OF PAIRED COMPARISON TESTS

Interval scaled responses to 55 pairs of sound samples resulted in an 11x11 symmetric matrix for each subject. Data is analyzed using Multi-Dimensional Scaling (MDS). MDS is a commonly used method to reduce dimensionality and complexity of data [6]. The data is mapped on an n-perceptual dimensions plot where sound samples with similar characteristics cluster together. The number and meaning of perceptual dimensions are determined by the researcher. Usually, two or three dimensions produce meaningful results [7].

The input of MDS function is the average of symmetric matrices from eight subjects. MDS was performed in R software [8]. The results are shown in Figure 3. Most mentioned adjectives during free verbalization study were under pleasantness and intensity categories. For this reason, perceptual dimensions are chosen to be annoyance and loudness. Samples 3,4,5,7,9, and 11 were recorded at the same seat of an aircraft at different times. The cluster of these samples on the lower right corner of the plot indicates the reliability of the listening tests. Note that sample 3 is 27 sec and sample 7 is the first 10.5 sec extracted from sample 3. This was done to investigate the duration effect. Although these two samples appear very close to each other and have almost the same annoyance level, the one with longer duration was perceived louder. Further investigation is necessary to capture the duration effects.

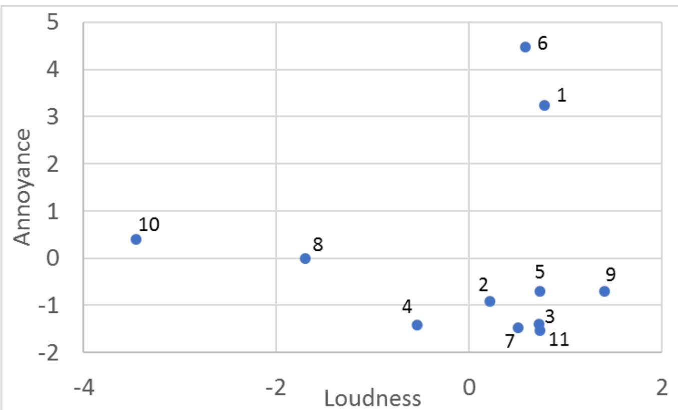


Figure 3 MDS results for the average of eight subjects.

Since the output space of MDS reflects the rated similarity between each pair of items the annoyance values should be interpreted based on the relative relationships of pairs rather than the layout of the dimensions or the units provided. For this reason, on the annoyance dimension, samples 1 and 6 are found to be significantly more annoying than the rest. Sample 6 is from the front section of one of the aircrafts where acoustic insulation content is weaker compared to the main sections of the aircrafts where majority of the rest of the sounds were recorded. Sample 1 and sample 5 are acquired at the same row of the same aircraft separated by an aisle. Loudness of these two samples are perceived almost equal whereas sample 5 is significantly more annoying. It is authors' opinion that the noise sources at this location should be further investigated through the sound quality perspective. Although, sample 10 exhibits the lowest level in loudness, it has higher annoyance than all other samples except samples 1 and 6. Loudness of sample 6 is close to that of sample 9 whereas the subjects found sample 6 much more annoying than sample 9. These results suggest that a sound perceived as louder is not necessarily more annoying.

Zwicker Loudness vs time plots for five samples are shown in Figure 4. Calculated values of Zwicker Loudness rank similar to the loudness values obtained from subjective ratings. This result indicates that it is an appropriate decision to choose loudness as a perceptual dimension and Zwicker loudness is an effective tool to predict intensity perception.

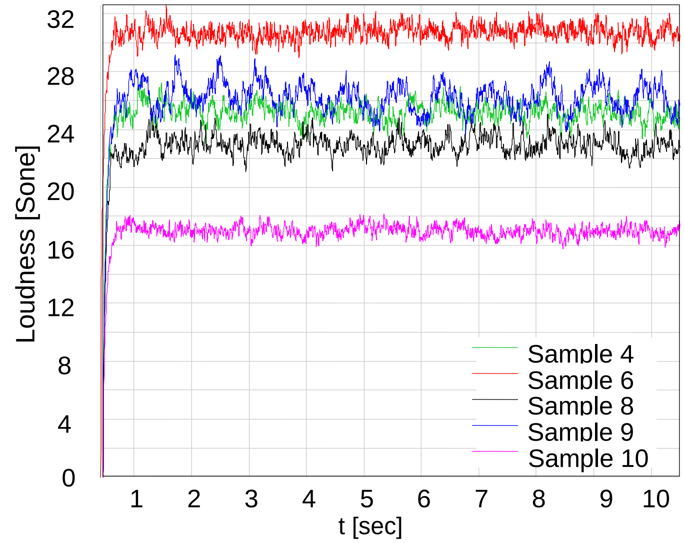


Figure 4 Metric MDS results for the average of eight subjects.

3. SOUND QUALITY PREDICTION MODEL

3.1. PREPARATION OF SOUND SAMPLES

Binaural sound recordings from four aircraft in either VIP or business jet category with custom thermal and acoustic insulation were used in the tests. 14 sound samples acquired at cruise level from several seat locations were all 10.5 seconds long.

3.2. SEMANTIC DIFFERENTIAL LISTENING TESTS

9 of 12 semantic differential pairs concluded in Section 2.4.1 were used to perform listening tests. Subjects were selected among potential VIP or business jet customers. Subjects were given a continuous scale to rate and were guided by adverbs at certain points of the scale. The experimenter then assigned a number from 0 to 100 based on the point subject marked.

Six subjects were trained with a practice block composed of three sound tracks before the actual test took place. All training sounds were different than the ones used for the actual test. Subjects evaluated 16 sound samples, two of which were repeated to evaluate subject consistency. All subjects achieved over 50% consistency on all sound tracks and were included in the analysis. Tests were performed inside the visual environment described in Section 2.2 with the instrumentation used for paired comparison tests.

3.3. OBJECTIVE DATA EVALUATION

21 sound quality metrics presented in Table 3 are calculated for 14 sound samples used during semantic differential listening tests. Artemis Suite software is used to

calculate most of the sound quality metrics. Among these, specific loudness is examined in three regions: low frequency region (below 500 Hz), mid frequency region (500-2000 Hz), and high frequency region (above 2000 Hz). Additionally, Low Frequency Sound Level (LFSL) is obtained by the summing up maximum noise level in each one-third octave band centered between 25 and 80 Hz [9].

Table 3 Principal Component Loadings for Sound Quality Metrics

Objective Metrics	PC 1	PC 2	PC 3	PC 4
Linear SPL [dB]	-0.15	-0.21	-0.42	-0.19
A-weighted SPL [dB]	-0.28	-0.16	0.08	-0.07
C-weighted SPL [dB]	-0.18	-0.34	-0.14	-0.15
G-weighted SPL [dB]	0.05	0.23	-0.46	-0.20
Low Frequency Sound Level [dB]	0.05	0.01	-0.54	-0.35
Loudness [Sone]	-0.30	-0.12	0.04	-0.04
Specific Loudness (Low) [Sone]	-0.17	-0.38	-0.03	-0.03
Specific Loudness (Mid) [Sone]	-0.30	-0.04	0.08	0.00
Specific Loudness (High) [Sone]	-0.23	0.29	0.05	-0.06
Sharpness [Acum]	-0.18	0.37	0.04	-0.04
Specific Roughness [Asper]	-0.24	-0.27	0.00	0.11
Specific Fluctuation Strength [Vacil]	-0.05	-0.40	0.12	-0.24
SIL3 vs. Time [dB]	-0.29	0.14	0.07	0.00
SIL4 vs. Time [dB]	-0.30	0.07	0.09	-0.02
Articulation Index [%]	0.30	-0.11	-0.07	0.02
Speech Intelligibility Index [%]	0.29	-0.13	-0.10	0.05
Specific Prominence Ratio [dB]	-0.11	0.18	0.13	-0.56
Tonality (max)	0.12	-0.02	0.36	-0.44
Tonality (average)	0.24	-0.18	0.17	-0.14
Impulsiveness [iu]	-0.25	0.10	-0.25	0.21
Kurtosis	-0.06	-0.14	-0.07	0.34
Proportion Variance (%)	49.51	21.75	12.69	7.64
Cumulative Variance (%)	49.51	71.27	83.96	91.60

Results for each metric is standardized with zero mean and unit standard deviation. Principal Components Analysis (PCA) was conducted to understand the dominant factors in the objective data by creating a new set of variables called Principal Components (PCs) through the linear combinations of these metrics. PCs, individually, being a single axis in data space, are orthogonal to each other. In practice, first few PC explain most of the total variance eliminating redundant information [10].

In this analysis, first four principal components (PCs) constituted 91.6% of the variability as shown in Table 3. Loadings determine how much each PC is influenced by an individual metric. Sound quality metrics that have the highest loadings for each PC is highlighted in Table 3. First PC is mostly influenced by the frequency region associated with speech interference and speech intelligibility and loudness in this region. For this reason, first PC is called loudness related to speech interference. Second PC has the highest loadings for specific fluctuation strength, loudness in low frequency region, and sharpness. This implies that this PC represents loudness in non-speech interference frequency region and slow fluctuations in loudness. Third and fourth have rather small proportional variance. Third PC reflects the sound levels at low frequency region and fourth PC represents the tonal content of the sound.

3.4. OBJECTIVE AND SUBJECTIVE DATA CORRELATION

A model utilizing sound quality metrics to predict subjective response is developed by correlating subjective and objective data. Principle Components Regression (PCR) is performed to correlate listening test responses for 9 semantic differential pairs with 4 principle components retained from PCA of metrics. Before correlation, responses to each semantic differential pair were also standardized.

Final model to predict is expressed as

$$Y=BX \qquad (1)$$

where Y is the 9x14 subjective response matrix that has 9 semantic differentials rated for 14 sound samples, X is the 21x14 objective response matrix describing 21 metrics calculated for 14 sound samples, and B is the 9x21 transformation matrix that will be used to predict subjective responses from given sound quality metric of new sound samples.

Figure 5 shows the variance percentage that the four PC model explains for each subjective variable. Variables defining pleasantness, comfort, and loudness are explained with a variance above 75%. Bright-dark pair that links sound to a visual perception, tonal-atonal and high tone-low tone pairs that are associated with the tonal characteristics of the sound, and buzzing-not buzzing pair implying the randomness in the signal rate are above 60%. Distant-Close pair stays below 50% which indicates that distance perception associated with sound is not well represented by this model. Test conditions can be improved for better visual-auditory connections for subjects to be able to extract spatial information with spatial perception better. The reason Steady-Fluctuating pair is left at 22% might be that there is not much significant difference between sound samples in terms of steadiness since they are all acquired during cruise.

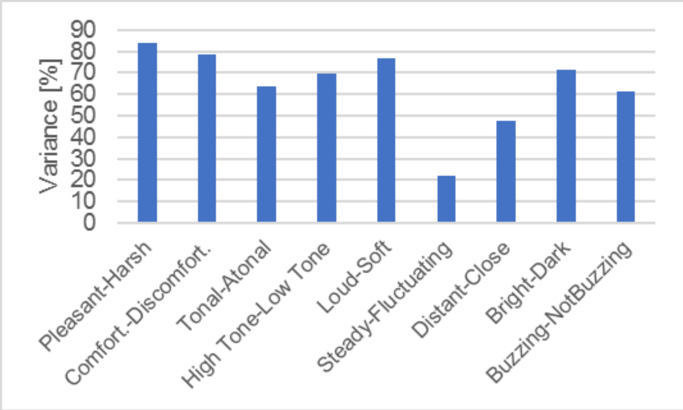


Figure 5 Subjective Variable Variances.

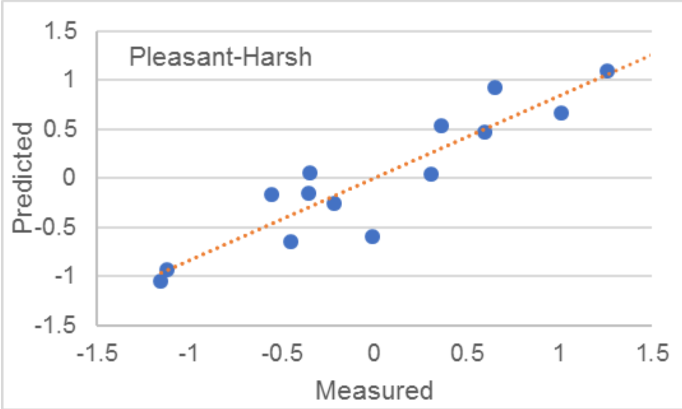
Figure 6 shows how well the model can predict listening test ratings for pleasantness, comfort, and loudness attributes. R^2, i.e. goodness of fit, values for Pleasant-Harsh, Comforting-Discomforting, and Loud-Soft pairs are 0.78, 0.68, and 0.68, respectively.

R software is used to perform statistical analysis.

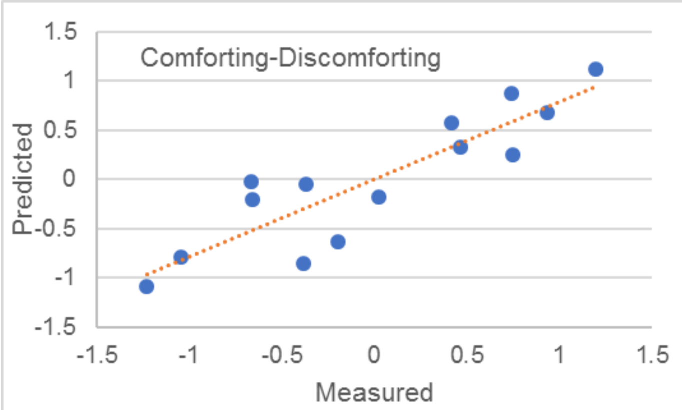
4. CONCLUSIONS AND FUTURE WORK

This paper explored the prominent effects of aircraft cabin sounds on customer preferences using sound quality approach focused on the VIP and business jet market. The purpose of the first part of the effort was to create a list of semantic differentials that describe human perception of aircraft interior noise to be used in listening tests. A subjective study was conducted with binaural data acquired inside two different aircrafts during cruise. In addition to the test procedure developed and adjective pair list generated, a series of conclusions were reached on the data collected from paired comparison tests. MDS results suggested that subjects did not only decide based on the intensity of the sounds. While their ratings showed consistency with the Zwicker loudness metric, there was another effect making samples with lower loudness less favorable than those with higher loudness. Associated with the jury responses, this effect was described as annoyance.

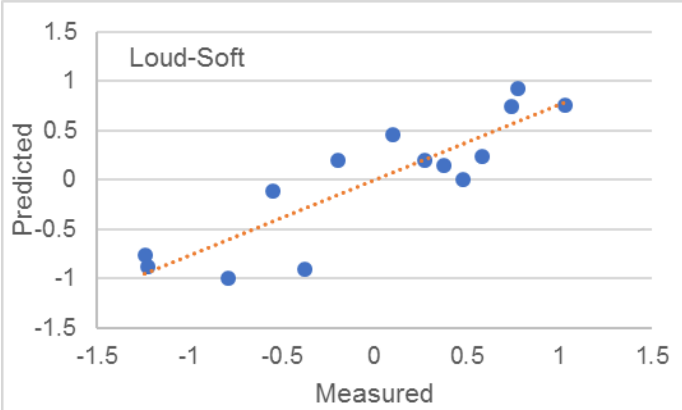
In the next section, 21 sound quality metrics were identified and calculated for 14 sound samples acquired at several seat locations of four different aircraft. PCA performed on the data set provided four PCs that account for 91.6% of the variability. Most influence is found to be coming from speech interference and speech intelligibility frequency regions and loudness in these regions. This was followed by sharpness and slow fluctuations in loudness. Low frequency region and tonal content of the sound had small contributions on the overall. In addition, semantic differential listening tests were conducted with potential customers and the responses were correlated to objective sound quality metrics using PCR. Sound quality prediction model developed, best presented pleasantness, comfort and loudness aspects of the cabin acoustics.



(a)



(b)



(c)

Figure 6 Listening test results (Measured) versus sound quality model predictions (Predicted) values of (a) Pleasant-Harsh, (b) Comforting-Discomforting, and (c) Loud-Soft attributes for 14 sound samples. Dotted lines represent the line fit for the data.

Authors' future focus is directed on improving the model functionality by supporting it with more subjective testing. Additionally, spatial variations of the acoustic signature inside an aircraft are as important as the temporal variations.

Sound signature differs between locations as dominant noise sources change along the length of the aircraft. How to handle multiple noise sources in the sound design process and how sound quality analysis should be integrated in the product design are the upcoming challenges for the researchers.

5. REFERENCES

[1] Barbot, B., Lavandier C., Cheminee, P., “Perceptual representation of aircraft sounds”, Applied Acoustics Vol. 69 No. 11 (2008): pp. 1003-1016.

[2] Schutte, M., Muller, U., Sandrock, S., Griefahn, B., Lavandier, C., Barbot, B., “Perceived quality features of aircraft sounds: An analysis of the measurement charactreristics of a newly created semantic differential”, Applied Acoustics Vol. 70 No. 7 (2009): pp. 903-914.

[3] Haverkamp, M., “Look at that sound! Visual aspects of auditory perception” 3. Congreso Intyernacional de Sinestesia, Ciencia y Arte, Granada, 2009.

[4] Lyon, R., “Product Sound Quality – from Perception to Design”, The Journal of Acoustical Society of America Vol. 108 No. 5 (2000): pp. 2471.

[5] Altinsoy, E., Jekosch, U., “The Semantic Space of Vehicle Sounds: Developing a Semantic Differential with Regard to Customer Perception”, Journal of Audio Engineering Society Vol. 60 No. 1/2 (2012): pp. 13-20.

[6] Otto, N., Amman, S., Eaton, C., Lake, S., Guidelines for Evaluations of Automotive Sounds”, SAE Technical Paper 1999-01-1822, 1999.

[7] Hout, M., Papesh, M., Goldinger., “Multidimensional Scaling”, WIREs Cogn. Sci. Vol. 4 (2013): pp. 93-103.

[8] Mevic, B., Wehrens, R., “Introduction to the pls Package”, December 2016. URL: mevik.net/work/software/pls.html

[9] More**,** S. R., **“**Aircraft Noise Characteristics and Metrics.**”** PhD Thesis. Purdue University**,** West Lafayette**,** IN**.**

[10] Bowen, D., “Correlating Sound Quality Metrics and Jury Ratings”, Sound & Vibration Magazine, September 2008.